International Journal of Innovation Scientific Research and Review
Vol. 07, Issue, 05, pp.8350-8353, May 2025
Available online at http://www.journalijisr.com
SJIF Impact Factor 6.599

IJISRR

Research Article **ISSN: 2582-6131**

MODELING ENVIRONMENTAL QUALITY INDEX IN INDONESIA WITH MULTIVARIATE ADAPTIVE REGRESSION SPLINES

*** Marfa Audilla Fitri, Suliyanto, M. Fariz Fadillah Mardianto, Elly Ana**

Statistics Department of Mathemathics, Airlangga University, Surabaya, Indonesia.

Received 02nd March 2025; Accepted 03rd April 2025; Published online 16th May 2025

ABSTRACT

Aims: This study aims to analyze the factors influencing Indonesia’s Environmental Quality Index (IKLH) using nonparametric regression methods, specifically Multivariate Adaptive Regression Spline (MARS) and multipredictor truncated spline. Study Design: A quantitative study applying nonparametric regression analysis. **Place and Duration of Study:** Data were collected from the Central Bureau of Statistics and the Ministry of Environment for 34 provinces in Indonesia in the year 2022. **Methodology:** The study used secondary data comprising predictor variables such as Human Development Index (HDI), population density, proper sanitation, poverty percentage, and Gross Regional Domestic Product (GRDP). Two nonparametric regression methods were applied: MARS and multipredictor truncated spline. Model performance was assessed using Mean Squared Error (MSE), Generalized Cross-Validation (GCV), and R^2 values. **Results:** The truncated spline method demonstrated superior performance with MSE = 5.63308, GCV = 10.42, and R^2 = 82.63%, compared to MARS with MSE = 7.685, GCV = 16.014, and R^2 = 79.3%. The analysis confirmed that HDI, population density, proper sanitation, poverty percentage, and GRDP significantly influence IKLH. **Conclusion:** The findings highlight the key determinants of environmental quality in Indonesia and emphasize the effectiveness of nonparametric regression approaches in environmental studies. These insights provide a valuable reference for policymakers in designing strategies to improve environmental sustainability.

Keywords: IKLH, Spline Truncated, MARS, Nonparametric Regression.

INTRODUCTION

Indonesia, with its abundant natural resources, faces significant challenges in environmental conservation. The rapid growth of Gross Regional Domestic Product (GRDP) and high urbanization rates have increased environmental pressure, including the consumption of natural resources, greenhouse gas emissions, and environmental degradation due to economic activities and infrastructure development. Global assessments indicate that Indonesia’s environmental preservation efforts remain relatively low compared to other countries in the Asia-Pacific region. The 2022 Environmental Performance Index recorded Indonesia’s score at 28.2 out of 100, ranking 164th out of 180 countries globally and 22nd out of 25 in the Asia-Pacific region [1]. Additionally, Indonesia’s Environmental Quality Index (IKLH) in 2022 was 72.42, still categorized as moderate [2]. IKLH plays a crucial role in achieving the Sustainable Development Goals (SDGs), particularly in six key environmental aspects covered in SDG points 6, 11, 12, 13, 14, and 15 [3]. The government has implemented various environmental management programs, including air pollution control, water pollution mitigation, land degradation prevention, peatland ecosystem protection, and coastal and marine pollution control. Although IKLH increased by 0.97 points from 2021, seven provinces have yet to meet the target [4], indicating that several factors influence environmental quality and require further analysis.

To understand the significant factors affecting IKLH, regression analysis is employed. Regression models are categorized into three types: parametric, nonparametric, and semiparametric [5][6]. Parametric regression is used when the curve shape is known [7], whereas nonparametric regression is more suitable when the curve shape is unknown [8]. Therefore, this study applies the Multivariate Adaptive Regression Spline (MARS) method, which excels in identifying interactions among predictor variables and handling non-linear relationships [9][10].

Based on previous studies, factors suspected to influence IKLH include the Human Development Index (HDI), population density, proper sanitation, poverty percentage, and GRDP [11]. Given the possible interactions among these factors and the non-linear relationship patterns [12], MARS is chosen to model IKLH. The findings of this study are expected to serve as a reference for policymaking to improve environmental quality in Indonesia.

METHODOLOGY

Data Research

This study utilizes secondary data obtained from the official websites of the Central Bureau of Statistics (BPS) and the Ministry of Environment and Forestry, accessible at https://www.bps.go.id/. The dataset includes the Environmental Quality Index (IKLH), Human Development Index (HDI), Population Density, Proper Sanitation, Poverty Percentage, and Gross Regional Domestic Product (GRDP) for 34 provinces in Indonesia in 2022.

Research Variables

The research variables consist of IKLH as the response variable, while HDI, Population Density, Proper Sanitation, Poverty Percentage, and GRDP serve as predictor variables, all measured on a ratio scale with continuous data.

Flowchart

The following is a picture of the research flow presented in Figure 1.

***Corresponding Author: Marfa Audilla Fitri,**
Statistics Department of Mathemathics, Airlangga University, Surabaya, Indonesia.

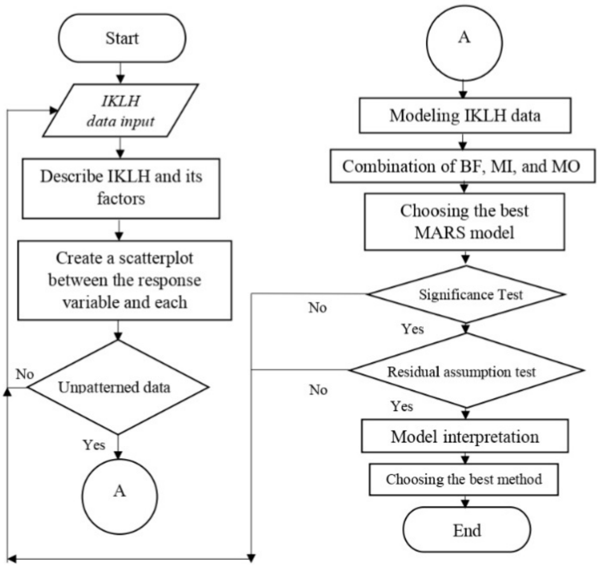

Fig 1. Flowchart

RESULTS AND DISCUSSION

A. Descriptive Statistics

This research presents descriptive statistical analysis using bar charts and scatterplots. The bar chart is used to provide an overview of environmental quality index in Indonesia. On the other hand, scatterplots are used to observe the distribution pattern of the research data and serve as an initial step in identifying the potential for applying nonparametric approach methods. The following is a bar chart of environmental quality index in Indonesia.

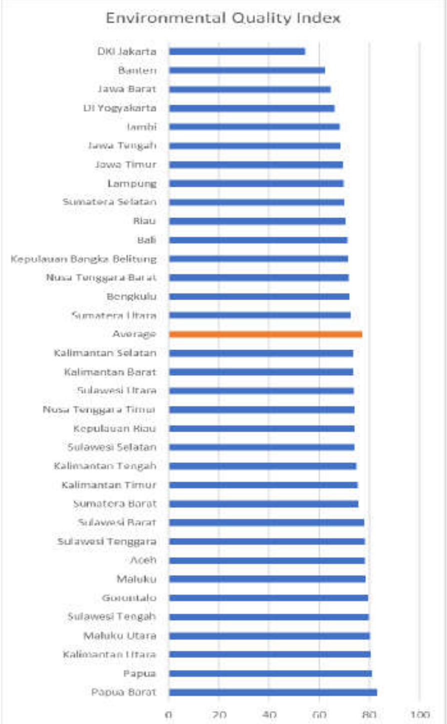

Fig 2. Environmental Quality Index Diagram

As show in Fig. 1 based on the Environmental Quality Index chart, Papua Barat has the highest environmental quality index at 83,31, while DKI Jakarta has the lowest at 54,57. Additionally, the average index is 77,33. The distribution of the index values does not follow a clear trend, suggesting that a nonparametric regression approach may be suitable for analyzing the relationship between environmental quality and its influencing factors.

B. MARS Model Parameter Estimation

The calculation is done using MARS by combining BF, MI, and MO. The best model was derived from a combination of BF 10, MI 2, and MO 0. The GCV result is 16,014; $R^2$is 0,79; and MSE is 7,685. The R^2 indicates that 82% of the variation in the response variable can be explained by the predictor variables.

The estimated value of the basis function can be calculated after determining best model. The estimation results for the best model in modeling environmental quality index in Indonesia are as follows.

$BF1 = \max\,(0, X_1 - 264)$

$BF2 = \max(0, 264 - X_1)$

$BF3 = \max\,(0, X_4 - 4{,}53) \times BF2$

$BF4 = \max\,(0, X_2 - 13092{,}81) \times BF1$

MARS model is as follows:

$$\hat{Y} = 73{,}187 - 0{,}009BF_1 + 0{,}002BF_3 + 0{,}472 \times 10^{-7}BF_4$$

#

From equation (3), the complete MARS model estimation is obtained as follows:

$$\hat{Y} = 73{,}187 - 0{,}009(X_1 - 264) + 0{,}002(X_4 - 4{,}53)(264 - X_1) + 0{,}472 \times 10^{-7}(X_5 - 13092{,}81)(X_1 - 264)$$

C. Significance Test of MARS Model

Significance test of MARS model is conducted using two approaches: simultaneous and partial. Simultaneous testing aims to evaluate if all the basis function coefficients in the MARS model have a combined effect on the response variable. The results of the simultaneous test for the model's basis function coefficients are as follows.

TABLE I SIMULTANEOUS TEST OF MARS

Test Statistics	Value

According to Table 1, the p-value is $0{,}22329 \times 10^{-9}$, smaller than significance level of $\alpha = 0.05$. So the decision rejects H_0 and the conclusion is that there is at least one $\alpha_m \neq 0$.

Then, partial testing is carried out to determine whether any of the basis function coefficient in the MARS model significantly impacts the response variable. The results of the partial test for the model's basis function coefficients are as follows.

TABLE 2 PARTIAL TEST OF MARS

Parameters	$P - Value$	Decision
Constant	0.999×10^{-15}	Reject H_0
$BF1$	0.199×10^{-4}	Reject H_0
$BF3$	0.798×10^{-4}	Reject H_0
$BF4$	0.121×10^{-3}	Reject H_0

According to Table 2, the p-value for each basis function is smaller than the significance level $\alpha = 0.05$. So the decision taken is reject H_0, so it is concluded that α_m is not equal to zero, with the value of $m = 1, 3, 4$.

D. Interpretation of the Best MARS Model

Once the best model is identified, the significance of the variables in the model is tested, and the assumptions regarding the residuals are checked. Additionally, the response variable and its predicted values can be visualized through a plot to compare the two. The resulting plot is presented below.

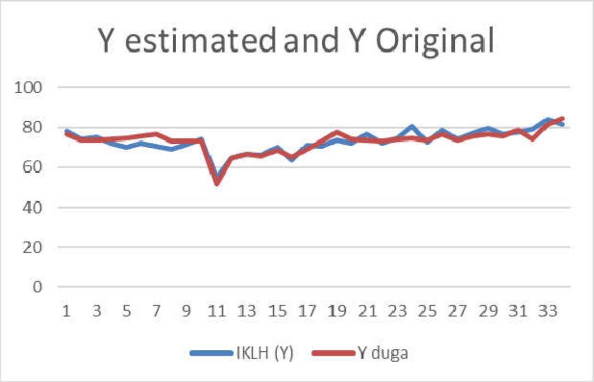

Fig 3. Plot of Y with $\widehat{\mathbf{Y}}$

As shown in Fig. 2, the estimated value $(\hat{Y})$ is closely aligned with the actual value (Y). The results of the MARS model interpretation that has been obtained in equation (4) can be described as follows:

1. Basis Two Function $(BF1)$

$$BF2 = \begin{cases} (X_1 - 264) & ; for\ X_1 > 264 \\ 0 & ; for\ the\ other\ X_1 \end{cases}$$

The interpretation of the value of base function one $BF1$ with a coefficient of -0.009 means that every one unit increase in $BF1$ will reduce the IKLH by 0.009, with base functions $BF3$ and $BF4$ considered constant. In addition, another meaning is that provinces twith a population density value of more than 264 people/km^2 will significantly contributeto the decline in IKLH. This is in line with the conditions in the field, namely that the higher the population density, the chance of IKLH also decreases.

2. Basis Three Function $(BF3)$

$$BF3 = \begin{cases} (X_4 - 4{,}53)(X_1 - 264)\ ; \\ for\ X_1 > 4{,}53\ \ and\ X_1 < 264 \\ 0 \qquad ; for\ the\ other\ X_1\ and\ X_4 \end{cases}$$

The interpretation of the value of base function three $BF3$ with a coefficient of 0.002 means that every one unit increase in $BF3$ will increase the IKLH by 0.002, with base functions $BF1$ and $BF4$considered constant. In addition, another meaning is that provinces that have a percentage value of the poor population of more than 4.53% and a population density of less than 264 inhabitants/km^2 will contribute significantly to the increase in IKLH. However, conditions in the field show that the higher the population density, the opportunity to increase IKLH actually tends to decrease, indicating that the quality of human resource management and the environment is an important factor in maintaining this balance.

3. Basis Four Function $(BF4)$

$$BF4 = \begin{cases} (X_5 - 13092{,}81)(X_1 - 264); \\ for\ X_4 > 13092{,}81\ and\ X_1 > 264 \\ 0 \quad ; for\ the\ other\ X_1\ and\ X_5 \end{cases}$$

The interpretation of the value of base function four $BF4$ with a coefficient of $0{,}472 \times 10^{-7}$ means that every one unit increase in $BF4$ will increase the IKLH by $0{,}472 \times 10^{-7}$, with base function $BF3$considered constant. In addition, another meaning is that provinces that have a GRDP percentage value of more than 13092,81 and a population density of more than 264 inhabitants/km^2 will contribute significantly to improving the IKLH. However, conditions in the field show that the higher the population density, the negative impact on the environment tends to increase, so the opportunity to improve IKLH becomes more difficult. This indicates that the quality of human resource management and sustainable development planning are essential to maintain the balance between economic growth and environmental quality.

CONCLUSION

Based on the findings of this research, it can be inferred that the lowest environmental quality index in Indonesia is 54,57 (DKI Jakarta Province), while the highest is 83,31 (Papua Barat Province). Then, the best model was derived from a combination of BF10, MI 2, and MO 0. There are 3 significant Basis Functions, namely $BF1, BF3, BF4$. The GCV result is 16,014; is 0,79; and MSE is 7,685. The R^2 value of 0,79 indicates that 79% of the variation in the response variable can be explained by the predictor variables.

Acknowledgements

The authors extend their sincere appreciation to the Statistics Study Program of Universitas Airlangga, the Central Bureau of Statistics for providing the data, and all individuals who contributed to this research.

Competing interests

Authors have declared that no competing interests exist.

Authors' Contributions

Marfa Audilla Fitri designed the study, performed the statistical analysis, wrote the protocol, and drafted the first version of the manuscript. Suliyanto and M. Fariz Fadillah Mardianto managed the data analysis and interpretation. Elly Ana conducted the literature review and provided critical revisions. All authors read and approved the final manuscript.

REFERENCES

[1] Yale College, "Yale Center for Environmental Law & Policy. 2022 EPI Results - Environmental Performance Index.," Yale Center for Environmental Law & Policy.

[2] Kementerian Lingkungan Hidup dan Kehutanan, "Profil Indeks Kualitas Lingkungan Hidup 2022.," Jakarta, 2023.

[3] Badan Perencanaan Pembangunan Nasional (Bappenas), Pilar Pembangunan Lingkungan. Badan Perencanaan Pembangunan Nasional, 2017.

[4] Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia, " Kualitas Lingkungan Hidup Indonesia Meningkat dalam Lima Tahun Terakhir," https://ppid.menlhk.go.id/berita/siaran-pers/6972/kualitas-lingkungan-hidup-indonesia-meningkat-dalam-lima-tahun-terakhir.

[5] B. W. Silverman, "Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting," Journal of the Royal Statistical Society. Series B (Methodological), vol. 47, no. 1, pp. 1–52, 1985.

[6] M. F. Qudratullah, Analisis Regresi Terapan Teori, Contoh Kasus dan Aplikasi dengan SPSS. Yogyakarta: Andi Offset, 2013.

[7] A. Islamiyati, “Model Spline Kubik Dengan Titik-Titik Knots Dalam Regresi Nonparametrik,” Nov. 06, 2019. Doi: 10.31227/Osf.Io/Z3dte.

[8] B. Lestari, Fatmawati, I. N. Budiantara, And N. Chamidah, “Estimation Of Regression Function In Multi-Response Nonparametric Regression Model Using Smoothing Spline And Kernel Estimators,” J Phys Conf Ser, Vol. 1097, P. 012091, Sep. 2018, Doi: 10.1088/1742-6596/1097/1/012091.

[9] A. Wibowo, “Multivariate Adaptive Regression Splines Modeling For Household Food Security In Central Borneo Province 2017,” Global Science Education Journal, Vol. 1, No. 1, Pp. 39–47, Apr. 2019, Doi: 10.35458/Gse.V1i1.5.

[10] J. H. Friedman, “Multivariate Adaptive Regression Splines,” The Annals of Statistics, vol. 19, no. 1, Mar. 1991, doi: 10.1214/aos/1176347963.

[11] S. Risambessy, S. N. Aulele, and F. K. Lembang, “Misclassification Analysis of Elementary School Accreditation Data in Ambon City Using Multivariate Adaptive Regression Spline,” Jurnal Matematika, Statistika dan Komputasi, vol. 18, no. 3, pp. 394–406, May 2022, doi: 10.20956/j.v18i3.19451.

[12] N. Chamidah et al., “Consistency and Asymptotic Normality of Estimator for Parameters in Multiresponse Multipredictor Semiparametric Regression Model,” Symmetry (Basel), vol. 14, no. 2, p. 336, Feb. 2022, doi: 10.3390/sym14020336.

****