NONPARAMETRIC REGRESSION ANALYSIS USING SMOOTHING SPLINES AND TRUNCATED SPLINES IN MODELING THE AIR QUALITY INDEX IN INDONESIA

Thesis

By

NADHIA AZ ZAHRA
2217031013

FACULTY OF MATHEMATICS AND NATURAL SCIENCES
LAMPUNG UNIVERSITY
BANDAR LAMPUNG
2026

ABSTRACK

NONPARAMETRIC REGRESSION ANALYSIS USING SMOOTHING SPLINES AND TRUNCATED SPLINES IN MODELING THE AIR QUALITY INDEX IN INDONESIA

By

Nadhia Az Zahra

The Air Quality Index (AQI) is a composite indicator that reflects regional air quality conditions and is influenced by multiple determinants with complex and nonlinear relationships. In such circumstances, parametric regression may be restrictive because it requires a predetermined functional form. This study applies spline based nonparametric regression using smoothing splines and truncated splines to model AQI in Indonesia and to compare the performance of both approaches. AQI is treated as the response variable, while population density, land cover area within and outside forest areas, and the number of motor vehicles are considered as predictor variables. For smoothing splines, the optimal smoothing parameter is selected using Generalized Cross Validation, whereas truncated splines are estimated using Ordinary Least Squares under various knot configurations and selected based on the minimum Generalized Cross Validation value. Model performance is evaluated using Generalized Cross Validation, Mean Squared Error, and Adjusted R squared. The study aims to identify the most appropriate model and to determine key factors influencing AQI variation in Indonesia, thereby providing empirical support for environmental policy making.

Keywords: Air Quality Index (AQI), nonparametric regression, smoothing splines, truncated splines, Generalized Cross Validation (GCV), Mean Squared Error (MSE), Adjusted R^2.

ABSTRAK

ANALISIS REGRESI NONPARAMETRIK MENGGUNAKAN SMOOTHING SPLINE DAN TRUNCATED SPLINE DALAM MEMODELKAN INDEKS KUALITAS UDARA DI INDONESIA

Oleh

Nadhia Az Zahra

Indeks Kualitas Udara (IKU) merupakan indikator komposit yang mencerminkan kondisi kualitas udara suatu wilayah dan dipengaruhi oleh berbagai determinan dengan hubungan yang kompleks dan nonlinier. Dalam kondisi seperti ini, regresi parametrik dapat bersifat terbatas karena memerlukan bentuk fungsi yang telah ditentukan sebelumnya. Penelitian ini menerapkan regresi nonparametrik berbasis spline menggunakan smoothing splines dan truncated splines untuk memodelkan AQI di Indonesia serta membandingkan kinerja kedua pendekatan tersebut. AQI diperlakukan sebagai variabel respon, sedangkan kepadatan penduduk, luas tutupan lahan di dalam dan di luar kawasan hutan, serta jumlah kendaraan bermotor dipertimbangkan sebagai variabel prediktor. Pada smoothing splines, parameter penghalus optimal dipilih menggunakan Generalized Cross Validation, sedangkan truncated splines diestimasi menggunakan Ordinary Least Squares dengan berbagai konfigurasi titik knot dan dipilih berdasarkan nilai Generalized Cross Validation minimum. Kinerja model dievaluasi menggunakan Generalized Cross Validation, Mean Squared Error, dan Adjusted R squared. Penelitian ini bertujuan untuk mengidentifikasi model yang paling sesuai serta menentukan faktor faktor utama yang memengaruhi variasi AQI di Indonesia, sehingga dapat memberikan dukungan empiris bagi perumusan kebijakan lingkungan.

Kata-kata kunci: Indeks Kualitas Udara (IKU), regresi nonparametrik, smoothing splines, truncated splines, Generalized Cross Validation (GCV), Mean Squared Error (MSE), koefisien Determinasi.

NONPARAMETRIC REGRESSION ANALYSIS USING SMOOTHING SPLINES AND TRUNCATED SPLINES IN MODELING THE AIR QUALITY INDEX IN INDONESIA

NADHIA AZ ZAHRA

Thesis

In a Partial Fulfillment of The Requirements for
BACHELOR OF MATHEMATICS

In the

Department of Mathematics

Faculty Of Mathematics And Natural Sciences

FACULTY OF MATHEMATICS AND NATURAL SCIENCES
LAMPUNG UNIVERSITY
BANDAR LAMPUNG
2026

Thesis Title	:	**NONPARAMETRIC REGRESSION ANALYSIS USING SMOOTHING SPLINES AND TRUNCATED SPLINES IN MODELING THE AIR QUALITY INDEX IN INDONESIA**
Student Name	:	**Nadhia Az Zahra**
ID Number Of Student	:	**2217031013**
Study Program	:	**Mathematics**
Faculty	:	**Mathematics and Natural Sciences**

APPROVAL

1. Supervisory Committee

Dr. Khoirin Nisa, S.Si, M.Si
NIP 197407262000032001

Misgiyati, S.Pd, M.Si
NIP 198509282023212032

2. Head of Mathematics Department Mathematics

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

VALIDATED BY

1. Examination Committee

Head	:	**Dr. Khoirin Nisa, S.Si, M.Si**	....
Secretary	:	**Misgiyati, S.Pd, M.Si**	....
Examiner Non-Supervisor	:	**Drs. Nusyirwan, M.Si**	....

2. Dean of the Faculty of Mathematics and Natural Sciences

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Thesis Defense Date: **29 January 2026**

STUDENT THESIS STATEMENT

The undersigned below:

Name	:	**Nadhia Az Zahra**
ID Number Of Student	:	**2217031013**
Major	:	**Mathematics**
Thesis Title	:	**Nonparametric Regression Analysis Using Smoothing Splines And Truncated Splines In Modeling The Air Quality Index In Indonesia**

Hereby declare that this thesis is the result of my own work and all writings contained in this thesis have followed the rules of scientific writing at the Lampung University.

Bandar Lampung, 29 January 2026

Author,

METERAI TEMPEL

1FB56ANX282658369

Nadhia Az Zahra

BIOGRAPHY

Nadhia Az Zahra was born in Bandar Lampung City, Lampung Province on January 9, 2005. She is the third of three children of Mr. Zainal Hasan and Mrs. Rohila Dahlan. She has two older sisters named Yulia Sari and Fani Mutia Sari respectively.

She began her early education at Pratama I Kindergarten in the academic year 2009-2010. She continued her elementary school at SDN 1 Tanjung Agung in 2010-2016. Then took junior high school education at SMPN 04 Bandar Lampung in 2017-2019 and continued her high school education at SMAN 10 Bandar Lampung in 2019-2022.

In 2022, Nadhia Az Zahra continued her undergraduate education at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Lampung University through the National Selection for State University Admission (SNMPTN).

During her university years, she participated as member of the Anniversary of the Mathematics Department (DINAMIKA) in 2023 and 2024. From January 2 to February 7, 2025, she completed a professional internship at the Perum Bulog in Bandar Lampung city. Furthermore, from July to August 2025 she conducted a Real Work Lecture (KKN) in Labuhan Ratu Raya, Bandar Lampung. She also completed an internship at BRI Insurance in Bandar Lampung from August 21 to October 10, 2025.

WORD OF INSPIRATION

"Life can be heavy, especially if you try to carry it all at once. Part of growing up and moving into new chapters of your life is about catch and release."
(Taylor Swift)

"And Allah is the best of planners."
(Q.S Ali Imran: 54)

"Altitude depends on your attitude"

DEDICATION

With heartfelt gratitude to Allah SWT, for His grace, guidance, and gifts that have enabled me to complete this thesis, I dedicated this work to:

My beloved parents, Zainal Hasan and Rohila Dahlan, and my dearest sisters, Yulia Sari and Fani Mutia Sari, as well as my beloved brothers in law, Vik Handra and Rafi Febrian, and my lovely nephews, Rakha and Rasya, who have given me endless moral and material support throughout my educational journey. Thank you for your endless love and sacrifies, which have shaped the person I am today.

My academic supervisors and examiners, whose invaluable guidance and feedback have played a crucial role in the completion of this thesis.

To my dear friends and companions, thank you for your unwavering support and shared journey, for always being there through every step along the way.

To my beloved almamater, Lampung University.

ACKNOWLEDGEMENT

Praise be to Allah SWT and blessings and greetings be upon the beloved propher Muhammad SAW. With His mercy and grace, the author was able to complete this thesis entitled “Nonparametric Regression Analysis Using Smoothing Splines and Truncated Splines In Modeling The Air Quality Index In Indonesia”.

The completion of this thesis cannot be separated from the support, guidance, and assistance of many individuals. Therefore, the author would like to express sincere thanks to:

1. Mrs. Dr. Khoirin Nisa, S.Si., M.Si. as rhe primary supervisor, for her time, guidance, and invaluable feedback throughout the process of writing this thesis.
2. Mrs. Misgiyati, S.Pd., M.Si. as the second supervisor, for her insightful guidance, support, and knowledge that have greatly benefited the author.dapat menyelesaikan skripsi ini.
3. Mr. Drs. Nusyirwan, M.Si. as a thesis examiner who has provided constructive criticism and suggestions during the thesis preparation process.
4. Mr. Dr. Aang Nuryaman, S.Si., M.Si., as the Head of the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Lampung University.
5. Mr. Dr. Ahmad Faisol, S.Si., M.Sc., as Academic Supervisor, who has provided guidance and support during the study period.
6. Mr. Dr. Eng. Heri Satria, S.Si., M.Si as the Dean of the Faculty of Mathematics and Natural Sciences, Lampung University.
7. All lecturers, staff, and employees of the Mathematics Department, Faculty of Mathematics and Natural Sciences, Lampung University, who have shared their knowledge and extended their support to the author.

8. My beloved Father Zainal Hasan and Mother Rohila Dahlan, who have given endless support and prayers. Thank you for the love, patience, and understanding that have always been given to me.

9. The two beloved sisters, Yulia Sari and Fani Mutia Sari, along with my two dear brothers in law, Vik Handra and Rafi Febrian, and my lovely nephews, Rakha and Rasya, have always given the author endless encouragement, prayers, and affection. May they always remain close, bring happiness to the family and continue to be a source of motivation in completing this thesis.

10. Ardila, Erisa, Hani, Kinasih, and Nazwa, whose constant motivation and support throughout the author's college years have been truly invaluable.

11. Makanbatu, my high school friend, whose friendship, encouragement, and cheerful spirit have always supported and entertained the author.

12. Afina, Baiq, and Sheilla who have always been there to listen, accompany, and bring comfort and laughter to the author during the process of completing this thesis.

13. All the people who have helped the author during the process of completing this thesis, supported the author in every situation, comforted the author, and always been by the author's side through all circumstances.

14. Lastly, thanks to me. Be proud of yourself for trying and struggling this far. Thank you for perseverance, for never giving up in any difficulty, and for your strong spirit of learning throughout the process of completing this thesis. This is the first step toward greater achievements to be proud of in the future.

The author acknowledges that this thesis is far from perfect and may still contain shortcomings, both in presentation and writing technique. Therefore, constructive criticism and suggestions are highly welcomed to improve the quality of this work. The author hopes that it will be useful for the readers.

Bandar Lampung, 29 January 2026

Nadhia Az Zahra

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURE

I. INTRODUCTION

1.1 Background and Issues

Regression analysis is a statistical method used for the investigation of interconnections between two or more variables. Practically, the variables are commonly divided into independent and dependent ones. The independent variables represent factors, which potentially may influence the outcomes, while the dependent ones mark the effects or results of influences (Ali, 2021). In addition, regression is also commonly used for estimating the robustness of the connection between a set of many variables and a specified response variable. Using the framework of randomized clinical trials, the examination is usually predefined and formulated for the purposes of clarity (Zapf, 2024). Typically, the regression analysis may also be undertaken utilizing three different methodologies namely parametric, nonparametric, and semiparametric (Dani, 2021).

Parametric regression applies when the connection between the variables is precisely specified and presumed to adopt a specified structure (Suparti, 2018). In a situation when the data pattern is not specified, a better representation is given by the nonparametric regression (Carolin Wongkar, 2023). The semiparametric regression incorporates the aspects of the parametric as well as the nonparametric approach into a unified model (Dani, 2021).

The three regression methods have distinctive uses. Parametric regression is commonly used when the functional form of the regression relationship is predescribed from theory or from previously established research outcomes. Nevertheless, not all data sets may be adequately described with this method, especially when the interaction between the responses and the predictor variables remains indeterminable or vague (Yavuz, 2022). This method may impose problematic or inaccurate understandings when used under these conditions. The opposite, however, would be nonparametric regression, which is commonly preferred over the parametric and semiparametric methods due to the simpler and more accommodative assumptions underlying this approach, leading to more accurate estimates that exhibit lower sensitivities to data pattern variation.

One of the well known parametric methods for estimating regression curves would be the splines regression. This method is especially effective because of the ability of the splines to accommodate fluctuations of the data behavior over particular sub-intervals (Handayani, 2024). Splines are also commonly used for the nonparametric regression due to the clear visualization interpretation, flexibility, and ability to handle smoothly changing functions. There also exist a lot of splines estimation methods including Truncated Splines, Smoothing Splines, and others (Mariati, 2021).

These methods used for the estimation of splines are the smoothing splines and the truncated splines. Smoothing splines happen to be a general nonparametric regression methodology since they can balance the exactness of model fitting with the smoothness of the estimated curve by means of a smoothing parameter. In spite of the computational intensity of the procedure, which could restrict the applicability of the procedure for large datasets (Xu, 2020), the procedure affords success with datasets with smooth characteristics and is reliable in the identification of the pattern changing over specified intervals (Mariati, 2021).

Truncated splines are polynomial forms defined in segmented elements (**? ?**). They do pose some sort of limitations, especially when the situation involves a high-order splines, a high number of knots, or tightly located knots, which may cause the matrix of the normal equations to approach singularity and thereby render the solving process challenging. This procedure does have limitations, especially if the order of the splines is high, the number of knots is high, or the knots close to each other. In these cases, the matrix of the normal equations may approach singularity and hence posing difficulty for the solving process (Ariesta, 2021). Despite these setbacks, the segmented form of truncated splines allows them to accommodate local attributes of the data more effectively than the ordinary polynomials, leading to regression curves fitting the dataset more precisely (Anggreni, 2018).

Several previous studies have applied nonparametric regression using smoothing and truncated splines. For example, (Mariati, 2021) used smoothing splines amid nonparametric regression analysis for social data, which demonstrates the applicability of the technique for depicting complex data trends. At the same time, (Nurcahayani, 2019) used truncated splines for modeling the average years of education for several districts in Java, which shows the capability of the methodology for effectively outlining tendencies for specific sub-intervals. In addition, comparative studies, like those from (Nyoman Budiantara, 2019), reported that the smoothing splines set a better standard than the truncated splines for modeling the human blood pressure based on the Mean Squared Error (MSE).

Nonparametric splines regression, including smoothing and truncated splines, is especially suited for the analysis of social and environmental data, for which the interplay between variables usually is vague, complicated, and nonlinear. Splines techniques provide smooth estimates but maintain responsiveness at local variation points, making them of special interest for the modeling of the Air Quality Index (AQI), a parameter depending on many determinants which change from region to region in Indonesia.

Air Quality Index (AQI) is a parameter formulated to indicate the condition of the air within a specified area, calculated from the composite value of many pollutant parameters as specified by the current regulations (Arini, 2023). The AQI acts as a numeric value, providing a determination of air cleanliness or level of pollution. The AQI is usually assigned a range of six classes, which are good, satisfactory, moderately polluted, poor, very poor, and hazardous (Pant, 2023). The ranges are set depending on the NO2 and SO2 parameter data, as increased levels of these emissions negatively impact air quality.

Other air quality factors used for this research involve population density, the use of the land, as well as the number of vehicles. The population density usually correlates with high human activities, thus potentially leading to a high volume of emissions (Pant, 2023). The kind of land cover is also directly correlated with the capacity of the environment for absorbing the pollutant. Thus, the region with less vegetation signifies a region with poor air quality (Kumar Singh, 2024). Conversely, the number of vehicles also serves as a prime air pollution source, and this mainly concerns the cities with heavy traffic.

Therefore, an appropriate modeling method should be utilized to capture the dynamic nature of the time series of Air Quality Index of Indonesia. These air quality observation data are appropriate for this form of data as the nonparametric splines regression model has also been referred to as particularly robust at revealing nonlinear trends (Gutiérrez, 2016).

Based on the description above, this research attempts to apply the smoothing splines and truncated splines models for estimating AQI in Indonesia and compare the performance of both methods in modeling the AQI in Indonesia.

1.2 Research Problems

The research problems can be formulated as follows:

1. How can nonparametric regression models using smoothing splines and truncated splines be developed to model the Air Quality Index (AQI) in Indonesia?
2. Which splines model provides the best performance based on selection criteria such as Generalized Cross Validation (GCV), coefficient of determination (R^2), and Mean Squared Error (MSE)?
3. What factors significantly influence the Air Quality Index in Indonesia according to the best performing splines model?

1.3 Research Objectives

The objectives of this study are as follows:

1. To develop nonparametric regression models using smoothing splines and truncated splines for modeling the Air Quality Index in Indonesia.
2. To determine the best model for each splines method based on selection criteria such as Generalized Cross Validation (GCV), the coefficient of determination (R^2), and Mean Squared Error (MSE).
3. To identify the factors that have a significant influence on the Air Quality Index in Indonesia based on the best performing splines model.

1.4 Research Benefits

The benefits of this study are:

1. Development of a nonparametric model for understanding the Air Quality Index in Indonesia.
2. Identification of the best model that enables more accurate air quality predictions.
3. Identification of the key factors influencing the Air Quality Index, providing support for environmental policy making.

II. LITERATURE REVIEW

2.1 Regression Analysis

According to (Kumari, 2018), regression is a statistical technique used to examine the relationship or influence between one variable and another. When the relationship is linear, the analysis is called linear regression. Linear regression involving only one dependent variable and one independent variable is referred to as simple linear regression, while if the dependent variable is influenced by two or more independent variables, the analysis is called multiple linear regression.

In addition to analyzing the relationship between paired data x and y, regression can also be applied for prediction purposes (Heinze, 2024). Let x denote the predictor variable and y the response variable for n observations $\{(x_i, y_i)\}_{i=1}^{n}$. The linear relationship between these variables can be expressed using the general regression model:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{2.1}$$

In this model, y_i represents the dependent variable for the i-th observation, ϵ_i is the error term assumed to follow a normal distribution with a mean of zero and constant variance σ^2, and $f(x_i)$ indicates the regression function or curve (Heinze, 2024).

In estimating a regression function or curve, there are two main approaches, namely parametric and nonparametric. In parametric regression, the form of the regression curve is assumed to be known in advance, which makes the method relatively rigid (Yavuz, 2022). In contrast, nonparametric regression does not require a predetermined curve shape, providing greater flexibility and making it suitable for various real-world applications. A widely used nonparametric approach is splines regression.

In estimating a regression function or curve, there are two main approaches, namely parametric and nonparametric. In parametric regression, the form of the regression curve is assumed to be known in advance, which makes the method relatively rigid (Yavuz, 2022). In contrast, nonparametric regression does not require a predetermined curve shape, providing greater flexibility and making it suitable for various real-world applications. A widely used nonparametric approach is splines regression.

2.2 Parametric Regression

The parametric regression approach is applied when the form of the regression curve or the pattern of the relationship between the response and predictor variables is already known. Some types of models included in parametric regression are linear, quadratic, cubic, and polynomial regressions of degree k, as well as other variations (Erlando, 2022). Among these forms, polynomial regression models are the most commonly used. In general, a polynomial regression equation of degree k can be expressed as follows:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \cdots + \epsilon_i \tag{2.2}$$

Where:

y_i : Response variable.

$x_i, \ldots, x_i^n$: Predictor variables.

$\beta_1, \ldots, \beta_n$: Unknown parameters.

ϵ_i: Independent error, $N \sim (0, \sigma^2)$.

2.3 Nonparametric Regression

With the advent of the latter part of the 1990s, the use of nonparametric regression as a data analysis approach witnessed a significant growth. This development finds impetus from the requirements for methods for the effective handling of heterogeneous data that often fail to meet the parametric assumptions. As a highly dynamic approach, this methodology assumes great importance in the study of longitudinal data (Sriliana, 2022).

Typical cases for the usage of the nonparametric regression involve when the shape of the regression curve is unspecified or when limited prior knowledge ascertains the shape of the data pattern (Chen, 2022). The general characterization of a nonparametric regression model may be declared as follows:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{2.3}$$

Where:

y_i : Response variable.
$f(x_i)$: A function f representing the regression curve whose form is unknown.
x_i : Predictor variable.
ϵ_i : Independent error, $N \sim (0, \sigma^2)$.

2.4 Smoothing Splines Regression

The nonparametric regression method of smoothing splines is essentially a function capable of accurately capturing data patterns while producing a relatively low variance error (Centofanti, 2023). In addition, smoothing splines regression is also applied in data smoothing by utilizing spline functions, such that the resulting estimates yield a small residual sum of squares. The smoothing splines estimator $\hat{f}_\lambda$ for f is defined as the one that minimizes the following Penalized Least Square (PLS) function (Nisa, 2018):

$$Q_\lambda(s) = \sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int_0^1 [f''(x)]^2 \, dx.$$

Where:

λ : The smoothing parameter controlling the smoothness of the curve.
$\sum_{i=1}^{n} [y_i - f(x_i)]^2$: The residual sum of squares.
$\lambda \int_0^1 [f''(x)]^2 \, dx$: The roughness penalty.

According to (Paulina, 2022), the value of the parameter λ lies within the range from 0 to infinity. When λ approaches zero, the resulting curve tends to be rougher but the model becomes highly accurate, since the distance between the observed data and the predicted values is very small. Conversely, as λ increases towards infinity, the curve becomes smoother with minimal curvature, but the model loses precision as the distance between the observed and predicted values becomes larger, consistent with the roughness penalty.

To determine the smoothing parameter in smoothing splines regression, one of the methods that can be employed is Generalized Cross Validation (GCV), as selecting an appropriate parameter is crucial for obtaining an optimal curve estimator (Li, 2023). The formula for the GCV method in nonparametric smoothing splines regression is given by:

$$GCV_\lambda = \frac{\text{MSE}}{\left(\frac{1}{n}\text{trace}[\mathbf{I} - \mathbf{S}_\lambda]\right)^2}$$

Where:

n : Number of observations.
I : Identity matrix.
MSE : $\frac{1}{n}\sum_{i=1}^{n}[y_i - f(x_i)]^2$.
$\mathbf{S}_\lambda$: $(\mathbf{I} + \lambda\mathbf{K})^{-1}$.
K : $WB^{-1}W^T$.
W : A matrix of dimension $n \times (n-2)$.
B : A matrix of dimension $(n-2) \times (n-2)$.

2.5 Truncated Splines Regression

According to (Putra, 2015), a splines refers to a piecewise polynomial structure with smoothness which may construct regression functions nearby the data. Truncated splines regression analysis only consists of a single response variable for univariate as well as multivariate nonparametric regression settings. For the univariate nonparametric truncated splines regression case, a single response variable exists with a single predictor variable (Kurniawati, 2019).

The merits of truncated splines include great suitability for data with variations arising within specified sub-intervals (Sriliana, 2022). Application of knot points also provides for the formation of relatively smooth regression curves as it can accommodate instances of data forms involving sudden rises as well as sudden falls (Goepp, 2018). In the formation of a truncated splines regression model, various factors come into consideration including the definition of the degree of the regression model, the number of knot points, as well as the location of the knot points (Yang, 2023). A truncated splines function of order p with knot points $K_1, K_2, \ldots, K_r$ forms a regression curve denoted as $f(x_i)$, which may be defined by the expression (Kurniawati, 2019):

$$f(x_i) = \sum_{k=0}^{m} \beta_k x_i^k + \sum_{k=1}^{r} \beta_{(m+l)}(x_i - K_1)_+^m$$

The truncated function includes truncated components as follows (Dani, 2021):

$$(x_i - K_l)_+^m = \begin{cases} (x_i - K_l)^m, & \text{if } x_i \geq K_l \\ 0, & \text{if } x_i < K_l \end{cases}$$

Where:

λ : The smoothing parameter controlling the smoothness of the curve.
$\sum_{i=1}^{n} [y_i - f(x_i)]^2$: The residual sum of squares.
$\lambda \int_0^1 [f''(x)]^2 \, dx$: The roughness penalty.

Based on equation (2.4), the order k represents the degree of the polynomial, while the knot points indicate where changes in the data pattern occur. When equation (2.4) is substituted into equation (2.3), the truncated splines nonparametric regression model takes the following form:

$$y_i = \sum_{k=0}^{m} \beta_k x_i^k + \sum_{k=1}^{r} \beta_{(m+l)}(x_i - K_1)_+^m + \epsilon_i \quad (2.6)$$

In equation (2.6), the truncated splines nonparametric regression model can be expressed in matrix form as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p & \vdots & (x_1-K_1)_+^m & \dots & (x_1-K_r)_+^m \\ 1 & x_2 & x_2^2 & \dots & x_2^p & \vdots & (x_2-K_1)_+^m & \dots & (x_2-K_r)_+^m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^p & \vdots & (x_n-K_1)_+^m & \dots & (x_n-K_r)_+^m \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p+r} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{2.7}$$

It can also be written using matrix notation as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

Let y denote the response vector of dimension $n \times 1$, X the design matrix of dimension $n \times (m + r + 1)$, $\boldsymbol{\beta}$ the parameter vector to be estimated with dimension $(m + r + 1) \times 1$, and ϵ the error vector of dimension $n \times 1$. In this study, the parameter vector $\boldsymbol{\beta}$ is estimated using the Ordinary Least Squares (OLS) approach, which minimizes the sum of squared residuals. Accordingly, the OLS estimator of $\boldsymbol{\beta}$ can be obtained by solving the following optimization problem:

$$\mathbf{min}\, \boldsymbol{Q}(\boldsymbol{\beta}) = \mathbf{min} \left\{ (\boldsymbol{y} - \boldsymbol{X\beta})^{\boldsymbol{T}} (\boldsymbol{y} - \boldsymbol{X\beta}) \right\} \tag{2.8}$$

Expanding Equation (2.8) yields:

$$\begin{aligned} \boldsymbol{Q}(\boldsymbol{\beta}) &= (\boldsymbol{y} - \boldsymbol{X\beta})^{\boldsymbol{T}} (\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= (\boldsymbol{y^T} - \boldsymbol{\beta^T X^T})(\boldsymbol{y} - \boldsymbol{X\beta}) \\ &= \boldsymbol{y^T y} - 2\boldsymbol{\beta^T X^T y} + \boldsymbol{\beta^T X^T X \beta} \end{aligned} \tag{2.9}$$

Taking the first derivative of $\boldsymbol{Q}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ gives:

$$\frac{\partial \boldsymbol{Q}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X^T y} + 2\boldsymbol{X^T X \beta} \tag{2.10}$$

By setting the first-order condition in Equation (2.10) equal to zero, we obtain:

$$-2\boldsymbol{X^T y} + 2\boldsymbol{X^T X \beta} = \mathbf{0} \tag{2.11}$$

which simplifies to:

$$\boldsymbol{X^T X \beta} = \boldsymbol{X^T y} \tag{2.12}$$

Thus, the OLS estimator $\boldsymbol{\beta}$ is expressed as:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T y} \tag{2.13}$$

Based on model (2.6), using one response variable and one predictor variable, a truncated splines nonparametric regression model can be constructed with degree p (where $p = 1, 2, 3$) at knot points K_1, K_2, K_3. The models that can be applied include:

a. Linear truncated splines ($p = 1$) with one knot (K_1):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - K_1)_+$$

b. Quadratic truncated splines ($p = 2$) with one knot (K_1):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 (x_1 - K_1)_+^2$$

c. Cubic truncated splines ($p = 3$) with one knot (K_1):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4 (x_1 - K_1)_+^3$$

d. Linear truncated splines ($p = 1$) with two knots (K_2):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - K_1)_+ + \hat{\beta}_3 (x_1 - K_2)_+$$

e. Quadratic truncated splines ($p = 2$) with two knots (K_2):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 (x_1 - K_1)_+^2 + \hat{\beta}_4 (x_1 - K_2)_+^2$$

f. Cubic truncated splines ($p = 3$) with two knots (K_2):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4 (x_1 - K_1)_+^3 + \hat{\beta}_5 (x_1 - K_2)_+^3$$

g. Linear truncated splines ($p = 1$) with three knots (K_3):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_1 - K_1)_+ + \hat{\beta}_3 (x_1 - K_2)_+ + \hat{\beta}_4 (x_1 - K_3)_+$$

h. Quadratic truncated splines ($p = 2$) with three knots (K_3):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 (x_1 - K_1)_+^2 + \hat{\beta}_4 (x_1 - K_2)_+^2 + \hat{\beta}_5 (x_1 - K_3)_+^2$$

i. Cubic truncated splines ($p = 3$) with three knots (K_3):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_1^2 + \hat{\beta}_3 x_1^3 + \hat{\beta}_4 (x_1 - K_1)_+^3 + \hat{\beta}_5 (x_1 - K_2)_+^3 + \hat{\beta}_6 (x_1 - K_3)_+^3$$

To obtain the optimal truncated splines regression model, special attention must be given when determining the knot points ($K_1, K_2, \ldots, K_n$). The presence of knot points divides the curve into several segments according to the specified locations. Therefore, the selection of knot points should not be done arbitrarily but must consider the data distribution pattern carefully.

The calculation for knot values can be broken down as follows:

$$\Delta_j = \frac{X_{j,\text{max}} - X_{j,\text{min}}}{L - 1} \tag{2.14}$$

The k-th grid point is calculated as:

$$g_{j,k} = X_{j,\text{min}} + (k - 1)\Delta_j \tag{2.15}$$

In choosing the best truncated splines regression model, the Generalized Cross Validation (GCV) value is used as a reference. Optimal knot points are determined based on the minimum GCV value. The smaller the GCV value obtained, the better the truncated splines regression model at the optimal knot points (Sifriyani, 2018). The formula for calculating the GCV value is expressed as:

$$GCV(K_1, K_2, \ldots, K_n) = \frac{\text{MSE}(K_1, K_2, \ldots, K_n)}{\left(\frac{1}{n}\text{trace}[\mathbf{I} - \mathbf{A}(K_1, K_2, \ldots, K_n)]\right)^2} \tag{2.16}$$

Where:

n = number of data points
$\mathbf{I}$ = identity matrix
$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
$(K_1, K_2, \ldots, K_n) =$ knot points

$\mathbf{A}(K_1, K_2, \ldots, K_n) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{m-1} & (x_1 - K_1)^{m-1} & \cdots & (x_1 - K_h)^{m-1} \\ 1 & x_2 & \cdots & x_2^{m-1} & (x_2 - K_1)^{m-1} & \cdots & (x_2 - K_h)^{m-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^{m-1} & (x_n - K_1)^{m-1} & \cdots & (x_n - K_h)^{m-1} \end{bmatrix}$$

2.6 Model Evaluation

2.6.1 Mean Squared Error(MSE

The selection of knot locations is crucial in obtaining the best splines regression model. One of the commonly used methods for selecting knot locations and determining the best regression model is by using the Mean Squared Error (MSE). MSE can also be applied to detect outliers in the data. This is because the computation of MSE involves squaring the residuals, which causes large errors to have a much greater influence compared to small errors. In other words, when a prediction deviates substantially, the squaring process amplifies its contribution to the overall MSE value (Chicco, 2021).

Mean Squared Error (MSE) represents an estimate of the residual variance. A model with the minimum MSE is considered the best model, as it indicates that the estimates are closer to the actual values (Tripena, 2022). The MSE function for the splines model is given by:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where:

y_i : The i-th actual value.
$\hat{y}_i$: The i-th predicted value.
n : Number of observations.

2.6.2 Coefficient Determination (R^2)

Adjusted R^2 is a way to measure how well a regression model fits the data and how accurate it is. Adjusted R^2 is the R^2 value that has been penalized for including independent variables that may not be relevant to the model. This means that it can stop the model from looking strong just because it has too many variables (Li, 2024). The formula for the coefficient of determination is as follows:

$$R^2_{\text{adj}} = 1 - \frac{\text{SSE}/(n - \text{df}_{\text{eff}})}{\text{SST}/(n-1)}$$

where SSE (Sum of Squared Errors) measures the remaining error between the observed values and the model predictions, and SST (Total Sum of Squares) measures the total variation in the observed data.

2.7 Air Quality Index (AQI)

As shown by (Arini, 2023), the Air Quality Index (AQI) is an air quality parameter, calculated from the cumulative value of the various air pollutant parameters of a region for a specified time. In general, air pollution has been well established as a big challenge in the majority of the cities of the world, including Indonesia. This conceptualization is supported by the rising tendencies of ambient air pollutant parameters such as carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and particulate matter (TSP, PM_{10}, and $PM_{2.5}$). Corresponding with the rising population, the demand for transport as well as energy also increases. The more frequent usage of transport facilities as well as the rising energy usage contribute significantly towards enhancing the level of air pollution, which further influences the environment as well as the general well-being of the population.

In general, the installation of the ambient air quality monitoring instrumentation shall meet the following requirements:

a. Sited in a clearing without obstruction which offers a minimum clearance angle of 120° from obstacles, including structures or raised vegetation,

b. With sampling inlet height of not less than two meters above the ground level for particulate and gaseous emissions,

c. At least 20 meters from the closest emission source, and

d. In the case of industrial activities, the sampling spot must comply with current standards for the prevention of air pollution emissions from stationary sources.

Air Quality Index (AQI) is classified into a number of classes for the interpretation of the condition of the atmosphere for a given area. As stated by the Regulation of the Minister of Environment and Forestry of the Republic of Indonesia Number 27 of the year 2021, the AQI consists of five classes. Precisely, an AQI value between 90 and 100 goes by the name very good, whereas a value between 70 and below 90 qualifies as good, a value between 50 and below 70 goes by the name moderate. Conversely, air quality with a value between 25 and below 50 qualifies as poor, and when the AQI value measures between 0 and below 25, this corresponds to a very poor air quality class.

In accordance with the Regulation from the Minister of Environment and Forestry of the Republic of Indonesia, Number 27 of the year 2021, regarding the Environmental Quality Index, the selection of stations for monitoring ambient air quality uses the Indonesian National Standard (SNI) specifying the sampling methods. The selection criteria for the sites include the areas which are typified by heavy traffic around main streets, industrial areas, highly populated residential centers, diversified land cover both with and without forests, and office buildings which experience little impact from transport activities.

Among the key factors leading to air quality concerns is the elevated population density, which also provides a growing demand for energy, transport, and industrial operations. This growth also precipitates a deteriorating state of local air quality. The evidence shows a probable deteriorating state of air quality by a population density boost since it amplifies the amount of anthropogenic activities adding up to the emissions of harmful gases. Consequently, population density serves as a significant parameter used when determining the Air Quality Index (AQI).

Population density and land cover are crucial factors affecting air quality. Open green spaces act as pollutant sinks, so a reduction in vegetation cover can worsen air pollution scenarios (Cintoro, 2023). Vegetation cover plays a significant role in reducing pollutants, especially in cities with high transportation and industrial activities. Thus, the area and extent of land cover, including forests and green spaces, are important when studying air quality. Another key factor is the number of motor vehicles, a major source of air pollution in Indonesia. Motor vehicles are significant emitters of pollutants, particularly particulates and exhaust gases (Fitriyah, 2022). As the number of vehicles increases, emissions also rise, contributing to a decline in air quality. Therefore, the quantity of motor vehicles is a critical parameter for assessing variations in the Air Quality Index (AQI).

III. RESEARCH METHODS

3.1 Time and Place of Research

The research was conducted in the odd semester of the 2025/2026 academic year at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung.

3.2 Research Data

The secondary data used in this research were obtained from the publications of Statistics Indonesia (BPS), available through the official website `https://www.bps.go.id`, for 32 provinces in Indonesia in 2023. The collected data include the response variable in the form of the air quality index, with predictor variables consisting of population density by province (people/km²) (X_1), land cover area within and outside forest areas (Thousand Hectares) (X_2), and the number of motor vehicles (X_3).

3.3 Research Methods

The analysis was carried out by employing two separate nonparametric regression techniques to compare smoothing splines and truncated splines regression by utilizing the RStudio software. The research methodology used here consists of the following:

1. Execute operations of pre-processing data, including verifying the completeness of data and scaling if required.
2. Conduct exploratory data analysis to determine the preliminary relationship

between the predictor variables X_1 (population density), X_2 (land cover), and X_3 (number of motor vehicles) and the outcome variable Y (air quality index).

3. Utilization of nonparametric regression modeling adopting two techniques, smoothing splines and truncated splines entails the following procedure:

    - For smoothing splines, estimating the best smoothing parameter (λ) using the Generalized Cross Validation (GCV) approach.
    - For truncated splines, selecting several candidate knot points, formulating splines models for each candidate knot using the Ordinary Least Squares (OLS) methodology, and subsequently measuring the GCV for each model.

4. Choice of optimum smoothing splines model from the minimum GCV value.

5. Selecting the truncated splines model corresponding with the minimum GCV value by the number and the location of the knot.

6. Comparison between the two models (smoothing splines and truncated splines) output using the evaluation criteria such as coefficient of determination (Adj R^2), Mean Squared Error (MSE).

7. Interpretation of model fit and the relative performances of the smoothing splines and the truncated splines.

8. Inference from the analysis and the comparison of the two splines regression models for the air quality index of Indonesia during the year 2023.

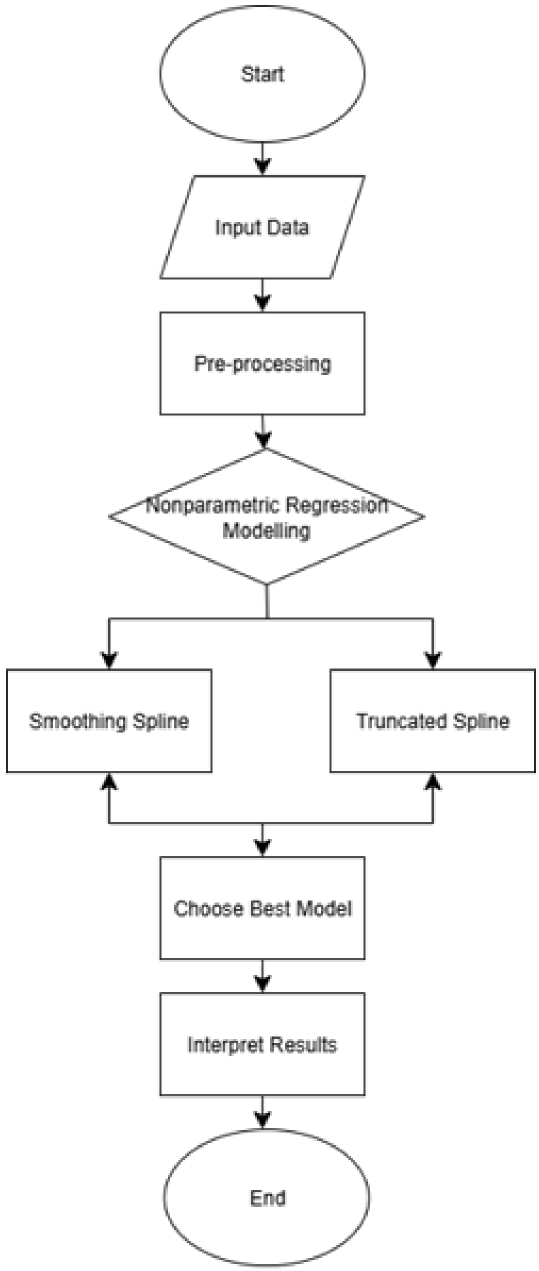

Figure 1 Smoothing Splines and Truncated Splines In Modeling The Air Quality Index In Indonesia

V. CONCLUSION

Based on the results of research conducted on nonparametric regression using smoothing splines and truncated splines to model the Air Quality Index (AQI) in Indonesia, the following conclusions were obtained:

1 Two nonparametric regression models, namely smoothing splines and truncated splines, were successfully developed to model the Air Quality Index (AQI) in Indonesia. Both models were chosen for their ability to capture the nonlinear relationship between AQI and factors such as population density, land cover, and number of motor vehicles. The development of these models involved selecting the optimal lambda (λ) value using Generalized Cross Validation (GCV) to ensure the best balance between model accuracy and the smoothness of the resulting curve.

2 The evaluation results showed that the smoothing splines model performed better than the truncated splines model. Based on the Adjusted R^2 value, the smoothing splines model explained 79.4% of the variation in AQI, while the truncated splines model explained only 72.9% of the variation. This indicates that smoothing splines is more effective at capturing the complex relationship between variables that affect air quality. Additionally, the Mean Squared Error (MSE) for the smoothing splines model was 0.0716, which is lower than the MSE of 0.2196 for the truncated splines model, indicating that the smoothing splines model is more accurate in predicting AQI.

3 The smoothing splines model also showed that the optimal lambda (λ) values were used to assess the impact of each factor on AQI changes. Based on the lambda analysis, it was found that population density and number of motor vehicles had smaller lambda values, indicating that these two factors have a

more significant influence on AQI changes in Indonesia. Although land cover also has an effect, the lambda value for this factor is larger, suggesting that its influence on air quality is not as significant as the other two factors. This shows that the smoothing splines model assigns more weight to the more dominant factors affecting AQI, such as population density and the number of motor vehicles.

LITERATURE

Ali, P., & Younas, A. (2021). Understanding and interpreting regression analysis. *Evidence-Based Nursing*, 24(4): 116–118.

Anggreni, N. P. R., Suciptawati, N. L. P., & Srinadi, I. G. A. M. (2018). Nonparametric truncated splines regression model for tuberculosis cases in Bali Province in 2016. *E-Jurnal Matematika*, 7(3): 211.

Ariesta, D., Gusriani, N., Parmikanti, K., & Padjadjaran University Team. (2021). Parameter estimation of nonparametric B-splines regression model on maternal mortality rate. *Jurnal Matematika UNAND*, 10(3): 342–354.

Arini, D. R., Purnawan, C., Rahayu, E. S., & Purnomo, N. A. (2023). Contribution of regional air quality index as part of achieving sustainable development goals (Case study: Magetan Regency). *ENVIRO: Journal of Tropical Environmental Research*, 24(2): 36.

Carolin Wongkar, D., & Fahmuddin, M. S. (2023). Nonparametric truncated splines regression analysis of factors influencing open unemployment rate in South Sulawesi Province. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 5(2): 55–63.

Centofanti, F., Lepore, A., Menafoglio, A., Palumbo, B., & Vantini, S. (2023). Adaptive smoothing splines estimator for the function-on-function linear regression model. *Computational Statistics*, 38(1): 191–216.

Chen, L., Smetanina, E., & Wu, W. B. (2022). Estimation of nonstationary nonparametric regression model with multiplicative structure. *Econometrics Journal*, 25(1): 176–214.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7: e623.

Cintoro, B. M., Ofrial, S. A. M. P., Sulistyorini, R., & Herianto, D. (2023). Analisis emisi polutan kendaraan akibat kegiatan transportasi di lingkungan Universitas Lampung. *REKAYASA: Jurnal Ilmiah Fakultas Teknik Universitas Lampung*, 27(3): 22–24.

Dani, A. T. R., & Adrianingsih, N. Y. (2021). Nonparametric regression modeling with truncated splines estimator vs Fourier series. *Jambura Journal of Mathematics*, 3(1): 26–36.

Dani, A. T. R., Adrianingsih, N. Y., Ainurrochmah, A., & Sriningsih, R. (2021). Flexibility of Nonparametric Regression Splines Truncated on Data without a Specific Pattern. *Jurnal Litbang Edusaintech*, 2(1): 37–43.

Dani, A. T. R., & Ni'matuzzahroh, L. (2022). Application of polynomial truncated splines family in nonparametric regression. *Inferensi*, 5(1), 37.

Erlando, Yundari, & Helmi. (2022). Semiparametric truncated splines regression modeling on open unemployment rate data in Indonesia. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, 11(3): 523–532.

Fitriyah, F., & Indriyani, Y. S. (2022). The effect of motor vehicles on air pollution in Ciruas Subdistrict, Serang, Banten. *Journal of Sustainable Civil Engineering (JOSCE)*, 4(2): 89–98.

Goepp, V., Bouaziz, O., Nuel, G., & Splines, G. N. (2018). Regression with Automatic Knot Selection. *TEST*, 27(4): 806–828.

Gutiérrez, L., Mena, R. H., & Ruggiero, M. (2016). A time dependent Bayesian nonparametric model for air quality analysis. *Annals of Applied Probability*, 26(5): 2889–2918.

Handayani, T., Sifriyani, S., & Rian Dani, A. T. (2024). Stunting Prevalence Modeling Using Nonparametric Regression of Quadratic Splines. *Jurnal Varian*, 7(2): 149–160.

Heinze, G., Baillie, M., Lusa, L., Sauerbrei, W., Schmidt, C. O., Harrell, F. E., & Huebner, M. (2024). Regression without regrets –initial data analysis is a prerequisite for multivariable regression. *BMC Medical Research Methodology*, 24(1).

Kumar Singh, H., Kumar Patel, P., Singh, S., & Er Hrishikesh Kumar Singh, C. (2024). Evaluation of Predictive Models for Air Quality Index Prediction in an Indian Urban Area. *Journal of Indian Association for Environmental Management*, 44(3): 31–40.

Kumari, K., & Yadav, S. (2018). Linear regression analysis study. *Journal of the Practice of Cardiovascular Sciences*, 4(1): 33.

Kurniawati, N. A., & Budiantara, I. N. (2019). Modeling open unemployment rate in West Java Province using nonparametric truncated splines regression. *Jurnal Sains dan Seni ITS*, 8(2): 2337–3520.

Li, Ken. (2024). Informativeness of Performance Measures: Coefficients or R-Squareds? *Journal of Risk and Financial Management*, 17: 481.

Li, Z., & Cao, J. (2023). Automatic search intervals for the smoothing parameter in penalized splines. *Statistics and Computing*, 33(1).

Mariati, N. P. A. M., Budiantara, I. N., & Ratnasari, V. (2021). Smoothing splines estimator in nonparametric regression (Application: Poverty in Papua Province), pp. 309–314. Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020). Advances in Social Science, Education and Humanities Research, Vol. 528. Atlantis Press, Paris.

Nisa, K., Herawati, N., & Setiawan, E. (2018). Analisis regresi nonparametrik dengan teknik smoothing, hlm. 1-20. *Prosiding Seminar dan Rapat Tahunan (Semirata 2018)*, Medan, Indonesia.

Nurcahayani, H., Budiantara, I. N., & Zain, I. (2019). Nonparametric truncated splines regression on modelling mean years schooling of regencies in Java. *AIP Conference Proceedings*, 2194(1): 020073.

Nurlaelah, Firmansyah, & Mualafah, U. (2024). The influence of economic value added (EVA), return on investment (ROI), current ratio (CR), and price-to-book value (PBV) on stock prices. *The Indonesian Journal of Management and Accounting*, 12(1): 45–54.

Nyoman Budiantara, I., & Lestari, B. (2019). Comparison of Smoothing and Truncated Splines Estimators in Estimating Blood Pressure Models. *International Journal of Innovation, Creativity and Change*, 5(3): 1177–1199.

Pant, A., Joshi, R. C., Sharma, S., & Pant, K. (2023). Predictive Modeling for Forecasting Air Quality Index (AQI) Using Time Series Analysis. *Avicenna Journal of Environmental Health Engineering*, 10(1): 38–43.

Paulina, E., & Sarah Sobariah, S. (2022). Smoothing splines regression on modeling the effect of rainfall on chili prices in Bandung City, pp. 1-5. Proceedings of the National Seminar on Economics and Accounting (Senakota).

Putra, I., Srinadi, I., & Surmajaya, I. (2015). Splines regression modeling (Case study: Herpindo Jaya Ngaliyan Branch). *Pemodelan Regresi Splines*.

Regulation of the Minister of Environment and Forestry of the Republic of Indonesia Number 27 of 2021 concerning the Environmental Quality Index, Pub. L. No. 27 of 2021, State Gazette of the Republic of Indonesia No. 1426 (2021).

Sifriyani, Budiantara, I. N., Kartiko, S. H., & Gunardi. (2018). Selection of optimal knot point and best geographic weighted on geographically weighted splines nonparametric regression model. *MethodsX*, 13: 102802.

Sriliana, I., Budiantara, I. N., & Ratnasari, V. (2022). A Truncated Splines and Local Linear Mixed Estimator in Nonparametric Regression for Longitudinal Data and Its Application. *Symmetry*, 14(12).

Suparti, Prahutama, A., & Santoso, R. (2018). Mix local polynomial and splines truncated: The development of nonparametric regression model. *Journal of Physics: Conference Series*, 1025(1).

Tripena, A. (2022). The use of mean square error method in second-order splines regression, pp. 269–284. Proceedings of the National Seminar on Mathematics and Its Applications (Senamantra), Universitas Jenderal Soedirman, Purwokerto.

Xu, D., & Wang, Y. (2020). Low Rank Approximation for Smoothing Splines via Eigensystem Truncation.

Yang, G., Zhang, B., & Zhang, M. (2023). Estimation of Knots in Linear Splines Models. *Journal of the American Statistical Association*, 118(541): 639–650.

Yavuz, E., & Şahin, M. (2022). Investigation of Parametric, Non-Parametric and Semiparametric Methods in Regression Analysis. *Sakarya University Journal of Science*, 26(6): 1111–1116.

Zapf, A., Wiessner, C., & König, I. R. (2024). Regression Analyses and Their Particularities in Observational Studies Part 32 of a Series on Evaluation of Scientific Publications. *Deutsches Arzteblatt International*, 121(4): 128–134.