1st International Conference on Multidisciplinary Engagement

Modeling Stunting Prevalence in Districts and Cities on Java Island

Afifah Nuraini Majid[1]*, Dewi Retno Sari Saputro[2], Purnami Widyaningsih[3]
[1,2,3]Department of Mathematics, Sebelas Maret University, Ir. Sutami No. 36A, Surakarta 57126, Indonesia

Corresponding Author.
**Email*: afifahnmajid@student.uns.ac.id

Abstract: *Stunting is a chronic nutritional problem that remains a major concern in Indonesia, particularly on the island of Java, which accounts for a significant proportion of cases nationwide. Poverty and inadequate food intake are known to be associated with the prevalence of stunting, but the relationship between these variables is not always linear. This research seeks to apply a smoothing spline approach within a nonparametric regression framework to examine the association between stunting prevalence, poverty levels, and the proportion of households experiencing inadequate food consumption across regencies and cities on Java in 2024. The data used consisted of 119 observations, divided into 80% training data and 20% testing data. Model estimation was carried out through a penalized least squares framework, and the smoothing parameter was obtained using the GCV approach. The results show that the optimal parameters were obtained at a combination of* $\lambda_1 = 16681$ *and* $\lambda_2 = 464.15$ *with a minimum GCV value of 25.2396. An MSE value of 28.75 indicates that the model is capable of modeling the relationship between predictor variables and stunting prevalence. The R-squared coefficient of 0.2446 indicates that 24.46% of the variation in stunting prevalence can be explained by the model. Thus, spline smoothing regression is effective in capturing nonlinear relationship patterns in stunting data on the island of Java.*
Keywords: *Stunting, Smoothing Spline, Nonparametric Regression, Generalized Cross Validation*

1. INTRODUCTION

Stunting is a condition of growth failure in young children resulting from chronic malnutrition during the first 1,000 days of life, characterized by height below the standard for their age [1, 2]. Its effects extend beyond physical growth including cognitive development, disease risk, future productivity, and economic well-being. The rate of stunting in Indonesia is still considered substantial and has become a key concern for the government in its efforts to reduce child malnutrition [3]. Java Island is home to more than half of Indonesia's population and is the primary contributor to national stunting cases. Four provinces on Java Island are among the six provinces with the highest number of stunting cases in Indonesia, collectively accounting for nearly 50% of the national total [4, 5].

In general, poverty and malnutrition are factors closely associated with stunting. Poverty limits access to nutritious food and health services, while malnutrition is reflected in the high prevalence of underweight toddlers [6]. Based on [7], underweight status, real per capita expenditure, education, and waste management have a significant effect on stunting on the island of Java, although the approach is still within the framework of parametric regression. Meanwhile, Fadlirhohim et al. [8] used truncated spline semiparametric regression, but still assumed a partially linear relationship and required the explicit determination of knot points.

In practice, the relationship between stunting prevalence and factors such as poverty and malnutrition does not always follow a clear linear pattern [9, 10]. Visualization using scatter plots often reveals curved trends or irregular fluctuations. This indicates that parametric and semiparametric approaches may not be fully capable of representing nonlinear relationships. Therefore, a nonparametric regression approach is needed because it does not assume a specific functional form from the outset.

One nonparametric regression method that can be used is smoothing spline regression, which flexibly estimates the regression function using the penalized least squares (PLS) approach [11, 12]. This method allows the curve to adapt to the data pattern without requiring a predefined functional form. The degree of curvature of the curve is controlled through the smoothing parameter, thereby balancing model accuracy and the smoothness of the function [13]. The selection of the smoothing parameter is performed using the Generalized Cross Validation (GCV) method to obtain an optimal model without overfitting [14]. Therefore, this method is capable of producing stable estimates that effectively capture complex relationships within the data.

The novelty of this study lies in the application of fully nonparametric spline to stunting data in districts and cities on the island of Java, without incorporating parametric components and without explicitly specifying knot points, as is done in the truncated spline approach. This approach enables the estimation of smoother, continuous, and adaptive curves in capturing the nonlinear relationship patterns between stunting prevalence, poverty, and undernutrition. Therefore, this study focuses on nonparametric smoothing spline regression to stunting data on the island of Java.

2. METHOD

This applied study uses a nonparametric smoothing spline regression approach to investigate how stunting prevalence is related to poverty and food insecurity across Java Island. The research process was conducted in several stages, namely data collection and analysis, bibliometric analysis using VOSviewer, smoothing spline regression modeling, and the selection of smoothing spline parameters using GCV.

2.1 Data and Variables

The data sources used for this study are derived from secondary data published by the Indonesian Nutrition Status Survey (SSGI) and the Central Statistics Agency (BPS). The data used are from 2024 and consist of 119 data points covering the prevalence of stunting among toddlers, the population living in poverty, and the proportion of households with inadequate food consumption across all regencies and cities on the island of Java. For modeling purposes, the data was then divided into two groups: 80% training data for model development and 20% testing data for model validation. The three research variables are shown in Table 1.

Table 1. Response and predictor variables of the research data

Variables	Research Data
Y	Prevalence of child stunting on the island of Java
X_1	Percentage of the population living in poverty
X_2	Prevalence of inadequate food intake

2.2 Research Methodology

The steps carried out in this study to achieve the research objectives are as follows.

1. Splitting the data into training and testing sets.
2. Conducting initial data exploration using scatter plots.
3. Normalizing the predictor variables using the Min–Max scaling method.
4. Constructing a nonparametric smoothing spline regression model.
5. Determining the optimal smoothing parameter (λ) using the GCV method.
6. Deriving the smoothing spline model equation.
7. Validating the model.
8. Interpreting the model results.

2.3 Bibliometric Analysis with VOSviewer

VOSviewer is a software tool designed to support users in creating, analyzing, and visualizing bibliometric maps using scientific publication data [15, 16]. Using VOSviewer, users can display large-scale bibliometric maps with clear and easy-to-understand visualizations, thereby simplifying the process of interpreting relationships between research elements, such as connections among authors, keywords, journals, and scientific documents [17]. Additionally, this software is widely utilized in bibliometric studies due to its ability to present the patterns and structural development of a scientific field in a systematic and informative manner. Through features such as network and overlay visualizations, VOSviewer also enables researchers to identify emerging research trends and evolving topics over time.

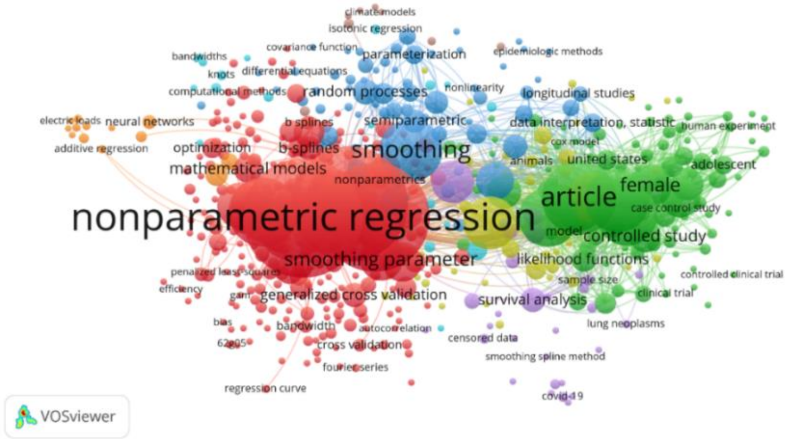


Figure 1. Bibliometric network of nonparametric regression and smoothing spline

Figure 1 shows a bibliometric map generated by VOSviewer, featuring three clusters distinguished by different colors. The red cluster covers nonparametric regression methods such as splines and GCV, the blue cluster addresses parameters, while the green cluster discusses the application of statistical models across various fields. Circle size reflects how often a term appears, with bigger circles representing higher frequency. The circles for “smoothing” are relatively small, indicating that research utilizing the term “smoothing spline” remains limited, thus requiring further study.

2.4 Smoothing Spline Regression

Spline smoothing regression is a nonparametric regression approach applied to model the relationship between independent variables and response variables where the form of the function is not explicitly known [18]. This method is intended to generate a regression function that is both smooth and flexible, allowing it to adapt to data patterns without requiring a predetermined functional form [19]. In this study, smoothing spline regression is used to model complex relationships that cannot be represented by a simple linear function. Unlike parametric regression, this method allows the data to determine the shape of the curve. As a result, it is more flexible in capturing nonlinear patterns. In simple terms, smoothing spline attempts to find a balance between fitting the data closely and maintaining a smooth curve. A very flexible curve may follow the data too closely (overfitting), while a very smooth curve may ignore important patterns.

The first step in smoothing spline analysis is collecting data consisting of pairs of predictor x_i and response y_i variables for $i = 1,2, \dots, n$. The basic form of the nonparametric regression model is written as

$$y_i = f(x_i) + \varepsilon_i, i = 1,2, \dots, n.$$

In this model, $f(x_i)$ is the regression function to be estimated from the data, while ε_i is the random error component, assuming that the expected value of the error is zero $E(\varepsilon_i) = 0$ and that the variance of the error is constant $\mathrm{Var}(\varepsilon_i) = \sigma^2$ [20].

Furthermore, in the application of spline smoothing, we use a spline function, which is a piecewise function consisting of several polynomials smoothly connected at specific points, called knots [21, 22]. A commonly used spline function is the cubic spline, which uses third-degree polynomials for each segment and has continuous first and second derivatives at these knot points [23]. Mathematically, the general form of an mth-order spline function with k knot points is written as

$$f(x) = \beta_0 + \sum_{j=1}^{m} \beta_j x^j + \sum_{j=1}^{k} \theta_j \left(x - \kappa_j\right)_+^m , \quad (1)$$

where β_0 is the model parameter, while β_{j} is the parameter for the variable $x^j \cdot \theta_j$ refers to the coefficient associated with the j, and K^j is the location of the j knot. The truncated power function, defined as $(x-\kappa_j)_+^m$, describes the relationship between x and the j knot. This function is written as

$$(x-\kappa_j)_+^m = \begin{cases} (x-\kappa_j)^m, & \text{if } x > \kappa_{j,} \\ 0, & \text{if } x \leq \kappa_{j.} \end{cases}$$

Spline functions are piecewise but remain continuous up to the $m-1$. derivative. In the context of cubic splines ($m = 3$), this approach is widely used because it provides smooth and stable function estimates [24].

In the next step, the smoothing spline $f(x_i)$ can be obtained from observational data consisting of pairs of response and independent variables that can accurately model the data while having a small error variance [25, 26]. Therefore, by applying data of size n, the function $f(x_i)$ is obtained by minimizing the PLS function, written as

$$PLS = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \int [f''(x)]^2 dx. \quad (2)$$

The PLS function consists of two main components. The first component measures how well the model fits the data, while the second component controls the smoothness of the curve. The parameter λ determines the trade-off between these two components. In this equation, the first term $\sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2$ represents the total squared error between the observed value y_i and the predicted value $f(x_i)$ also known as the goodness-of-fit. Meanwhile, the second term $\lambda \int [f''(x)]^2 dx$ is a penalty for the roughness of the function f, where $f''(x)$ is the second derivative of the function.

For the PLS function in Equation (2), the matrix form is written as

$$\boldsymbol{PLS} = (\boldsymbol{Y}-\boldsymbol{f})^{\boldsymbol{T}}(\boldsymbol{Y}-\boldsymbol{f}) + \boldsymbol{\lambda} \boldsymbol{f}^{\boldsymbol{T}} \boldsymbol{K} \boldsymbol{f}, \quad (3)$$

where the observed data vector $\boldsymbol{Y} = [y_1, y_2, \dots, y_n]^T$ contains the measured values, while $\boldsymbol{f} = [f(x_1), f(x_2), \dots, f(x_n)]^T$ is the vector of estimated functions based on that data. The matrix K is a curvature penalty matrix formed based on the second-order derivative (finite difference) approach. The parameter λ is a scalar smoothing parameter that controls how smooth the estimated curve is.

The estimate of the function $\hat{f}(x)$ is obtained by minimizing the PLS function. This minimization process is carried out by taking the first derivative of the PLS function in Equation (3) with respect to f and setting it equal to zero. From this optimality condition, a equation is obtained, which can be written as

$$(\boldsymbol{I} + \boldsymbol{\lambda K})\boldsymbol{f} = \boldsymbol{Y}, \quad (4)$$

where matrix I is an $n \times n$ identity matrix, while λ is the smoothing parameter that controls the smoothness of the function. Based on Equation (4), the function estimate can be written as

$$\hat{\boldsymbol{f}}(\boldsymbol{x}) = (\boldsymbol{I} + \boldsymbol{\lambda K})^{-\boldsymbol{1}} \boldsymbol{Y},$$

where the matrix $(I + \lambda L)^{-1}$ is an $n \times n$ matrix also known as the smoothing matrix S_λ, which enables efficient estimation of the spline function.

2.5 Selecting the Optimal Parameter with GCV

In the PLS function, λ serves as a smoothing parameter used to regulate the trade-off between how well the model fits the data and the smoothness of the function. If λ is too small, it can cause overfitting, whereas if it is too large, it can eliminate important information from the data. The optimal value of λ is determined using methods such as GCV, which is based on the balance between prediction error and model complexity. The GCV formula can be written as

$$GCV(\lambda) = \frac{MSE}{\left(1 - \frac{1}{n}\text{tr}(S_\lambda)\right)^2}, \quad (5)$$

where the Mean Square Error (MSE) is written as

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(y_i - \hat{f}(x_i)\right)^2, \quad (6)$$

where $\text{tr}(S_\lambda)$ is the sum of the diagonal elements of the smoothing matrix S_λ, which reflects the model's effective degrees of freedom. The value $\hat{f}(x_i)$ is the estimated regression function at point xi obtained from the PLS minimization process.

3. RESULTS AND DISCUSSION

This section presents the results of the analysis and discussion of the application of nonparametric smoothing spline regression to stunting data on the island of Java. The analysis begins with a description of the research data to outline the characteristics of the variables used, followed by an exploration of the patterns of relationships among variables, modeling using smoothing splines, and model validation.

3.1 Description of Research Data

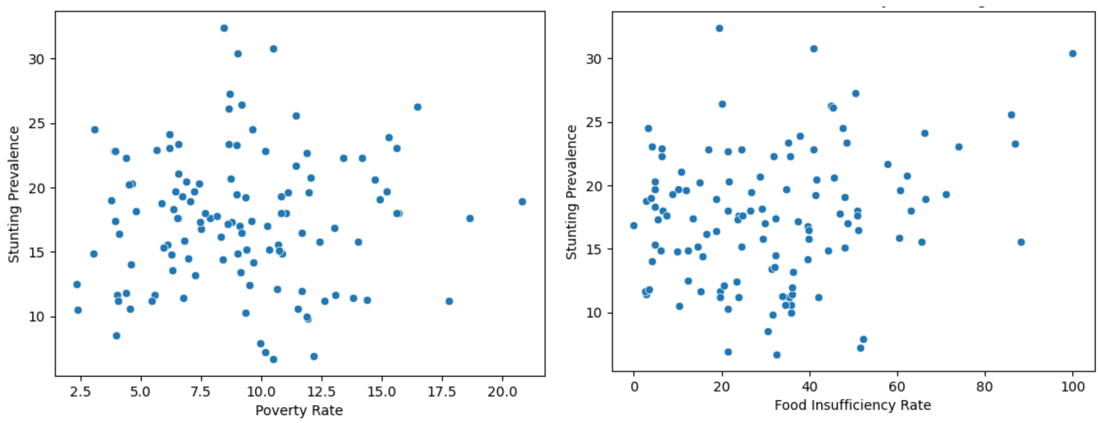
According to the 2024 SSGI results, stunting prevalence among toddlers in Indonesia was recorded at 19.8%. This figure represents a decrease from the 2023 rate of 21.5%, indicating an improvement in children's nutritional status nationwide. Nevertheless, this prevalence still reflects that stunting remains a significant public health issue requiring sustained mitigation efforts. This condition highlights the need for continuous monitoring and evaluation of factors associated with stunting to support more effective policy interventions. In addition, understanding the distribution and variability of the data is important to provide a clearer picture of the underlying patterns. The data overview in this study is presented through descriptive statistical analysis shown in Table 2.

Table 2. Summary of Research Variable Statistics

Variable	Mean	Minimum	Maximum	Variance
Y	17.48	6.70	32.40	2.67
X_1	9.07	2.34	20.83	1.38
X_2	79936.19	208	254566	2.85

3.2 Scatter Plot

A scatter plot is a diagram that displays the distribution of data points on a Cartesian plane to explore the pattern of interaction between variables. In this research, scatter plots were generated for each predictor variable against the response variable.



(a) (b)

Figure 2. Scatter plots of relationships: (a) prevalence of stunting versus prevalence of the poor; (b) prevalence of stunting versus the number of people with inadequate food consumption.

Figure 2 shows that the distribution of data points between the prevalence of stunting and the prevalence of the poor, as well as with inadequate food consumption, does not form a linear pattern or any clear specific pattern. The data points are randomly scattered with considerable variation across nearly the entire range of predictor values, and exhibit fluctuating trends and changes in slope in some sections, making them unsuitable for representation by a parametric linear regression model. These results indicate that the relationships between variables are nonlinear, rendering parametric regression models with specific functional forms less appropriate for use. Therefore, this study employs a nonparametric regression approach using smoothing splines.

3.3 Modeling with Nonparametric Spline Regression

Data from 119 districts and cities were divided into two groups 80% as training data for model development and 20% as testing data for validation purposes. The division was performed randomly to ensure proportional data representation in both groups. Prior to the modeling process, predictor variables were normalized using the Min-Max Scaling method so that all variables fell within the range of 0 to 1. This normalization process aims to avoid scale differences between variables that could affect the stability of parameter estimates in the smoothing spline model.

The smoothing spline model is estimated using the PLS approach as described in Equation (2), with the smoothing parameter λ controlling the balance between model fit and the smoothness of the function. The search for λ values is conducted within the range of 10^{-2} to 10^{5} to cover various levels of smoothing, ranging from highly flexible curves to very smooth ones. Several candidate λ values were tested, and for each candidate, the GCV value was calculated as formulated in Equation (5). The optimal λ value was determined by selecting the smallest GCV value, as it reflects a balance between model fit and complexity. Based on the calculation results for various combinations of smoothing parameters, the GCV values obtained are shown in Table 3.

Table 3. Values of λ and GCV

λ_1	λ_2	GCV
16681	464.15	25.23
2782.55	464.15	25.35
100000	464.15	25.40
16681	2782.55	25.55
16681	100000	25.56
16681	16681	25.59
2782.55	2782.55	25.60
2782.55	100000	25.62
2782.55	16681	25.63
100000	2782.55	25.75

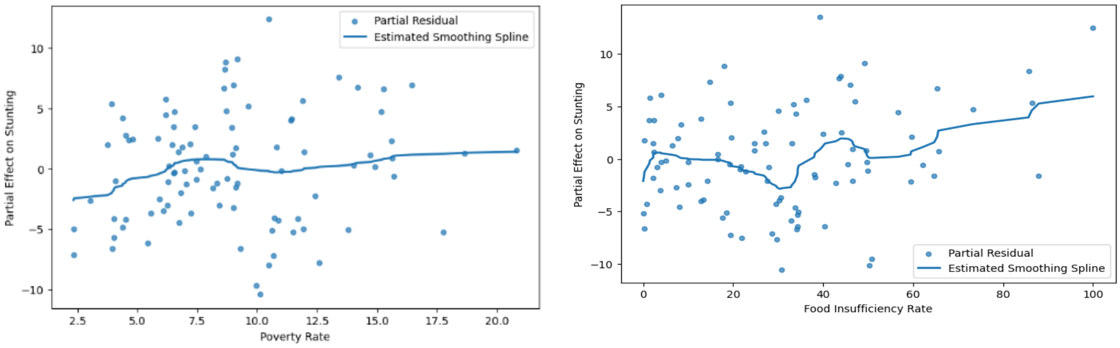
The first row of Table 3 shows that the smallest GCV value is 25.23, with the parameter combination λ_1 = 16681 and λ_2 = 464.15. This value is smaller than other λ combinations, so it was selected as the optimal smoothing parameter. The differences in GCV values between parameter combinations are relatively small, but this combination of λ_1 and λ_2 consistently yields the minimum value. This indicates that the resulting smoothing level provides the best balance between model fit to the data and function complexity. The relatively large value of λ_1 indicates that smoothing on the first variable tends to be stronger, while the smaller value of λ_2 suggests that the second variable requires higher curve flexibility to capture the relationship patterns in the data.

Theoretically, a smoothing spline has a control point at every observation point, resulting in a highly flexible function. However, having the same number of control points as data points leads to an extremely complex equation that is difficult to write out explicitly. To obtain a more concise yet representative model, a cubic spline basis approach with a limited number of control points is used. The control points are selected based on data quantiles in the range of 5% to 95%. The selection of quantiles aims to ensure that the control points are evenly distributed following the data distribution and are not overly influenced by outliers. After determining the optimal smoothing parameters using the GCV method, these values are used to construct the spline function based on Equation (1), which can be written as

$$\hat{Y} = 11{,}87 + 0{,}43x_1 + 1{,}30x_2 + \sum \theta_{1j}\left(x_1 - \kappa_{1j}\right)_+^3 + \sum \theta_{2j}\left(x_2 - \kappa_{2j}\right)_+^3$$

$$\hat{Y} = 11{,}87 + 0{,}43x_1 + 1{,}30x_2 + 9{,}24(x_1 - 3{,}94)_+^3 - 13{,}40(x_1 - 4{,}07)_+^3 + \cdots + 0{,}20(x_2 - 49{,}42)_+^3 - 0{,}16(x_2 - 50{,}10)_+^3. \quad (7)$$

To provide a visual representation of the nonlinear relationships generated by the model, the following graphs show the smoothing spline curves between each predictor variable and the prevalence of stunting.



(a) (b)

Figure 3. Smoothed spline regression curves: (a) prevalence of stunting versus prevalence of the poor; (b) prevalence of stunting versus the number of people with inadequate food intake.

Based on the graph of the partial effect of poverty on stunting prevalence, it is evident that the relationship formed is nonlinear. At low to moderate poverty levels, the curve tends to rise gradually, then undergoes a change in slope within a certain range and rises again at higher poverty levels. This pattern indicates that the effect of poverty on stunting is not constant across the entire range of values.

Meanwhile, the partial effect of food consumption inadequacy also exhibits a more fluctuating nonlinear pattern. At low to moderate levels, the curve temporarily decreases before rising again, and at higher values, a more consistent upward trend is observed. This indicates that as food consumption inadequacy increases, the prevalence of stunting tends to rise, particularly at higher levels. Overall, both graphs confirm that the relationship between the predictor variables and the prevalence of stunting is nonlinear, making the spline smoothing approach more appropriate than a linear model.

3.4 Model Validation

Based on the model validation results, the MSE value for the test data was 28.75. The model's prediction error was calculated using the MSE as expressed in Equation (6). The MSE represents the mean of the squared differences between the observed values and the predicted values generated by the model.The obtained MSE value of 28.75 indicates the average prediction error of the model. Although the value is not very small, the model is still able to capture the general pattern of the data. This suggests that the smoothing spline model provides a reasonably good fit for describing the relationship between the variables. Therefore, the resulting model can be used to describe the relationship between poverty, inadequate food consumption, and the prevalence of stunting on the island of Java.

Meanwhile, the coefficient of determination (R^2) value of 0.2446 indicates that approximately 24.46% of the variation in stunting prevalence can be explained by the poverty and food insecurity variables in the model. The remaining 75.54% is influenced by other factors outside the model. This moderate R^2 value suggests that stunting is a complex phenomenon influenced by multiple determinants beyond the variables included in this study. However, the prediction accuracy could be improved by incorporating additional relevant variables in future research.

4. CONCLUSION

Based on the results, this study concludes that the nonparametric smoothing spline regression model is able to capture the relationship between poverty, inadequate food consumption, and stunting prevalence in Java. The optimal smoothing parameters were obtained at $\lambda_1 = 16681$ and $\lambda_2 = 464.15$, which yielded the minimum GCV value. The resulting model has an MSE value of 28.75, indicating that the model has adequate predictive performance for new data. The spline smoothing regression model equation is presented in Equation (7) with a coefficient of determination R^2 of 0.2446, indicating that approximately 24.46% of the variation in stunting prevalence can be explained by the model, while the remainder is influenced by other factors outside the scope of this study. These findings indicate that the smoothing spline approach is able to capture nonlinear patterns in stunting data. Future research is recommended to include additional variables and spatial factors to better explain the variation in stunting prevalence. In addition, the results of this study have practical implications. Since poverty and inadequate food consumption are associated with stunting, policies should focus on improving economic conditions and access to nutritious food. This can help reduce stunting prevalence more effectively.

ACKNOWLEDGEMENTS

The authors express their sincere appreciation to all parties who contributed to this study, especially in formulating the mathematical model, selecting appropriate statistical approaches, and carrying out the estimation and validation processes. Insightful scientific discussions concerning model assumptions, inferential validity, and predictive performance evaluation have significantly strengthened the methodological rigor of this study. The authors further appreciate the support in providing data and computational facilities, which enabled the effective implementation of algorithms and numerical analysis. These contributions have helped ensure that the results obtained are reliable and reproducible.

REFERENCES

[1] M. de Onis and F. Branca, "Childhood stunting: A global perspective," *Matern. Child Nutr.*, May 2016, doi: 10.1111/mcn.12231.

[2] R. Dwi, T. Setyawan, T. Setyawan, and E. Sucipto, "The impact of multisectoral nutrition interventions on stunting rates among children under five years old: A systematic review," *Int. J. Med. Sci. Health Res.*, vol. 10, 2024.

[3] S. R. Arifin, I. Tanziha, F. Zuhra, N. N. Hadi, and A. Ahmad, "Association between nutrition intervention program indicators and stunting prevalence among toddlers in Indonesia," *AcTion Aceh Nutr. J.*, vol. 10, no. 2, p. 227, Jun. 2025, doi: 10.30867/action.v10i2.1852.

[4] A. N. Widya, B. R. Samudro, and E. Gravitiani, "Stunting in Java Island: Spatial and risk factor analysis," *Econ. Dev. Anal. J.*, vol. 13, no. 2, 2024.

[5] Y. P. Devi et al., "Spatial analysis of stunting prevalence according to family data collection indicators in Indonesia," Media Publ. Promosi Kesehat. Indones., vol. 8, no. 3, pp. 210–220, Mar. 2025, doi: 10.56338/mppki.v8i3.6931.

[6] N. Petry, O. Obeid, J. Wirth, et al., "The impact of poverty on child malnutrition and health in Lebanon: The need for multisectoral interventions," *Int. J. Equity Health*, vol. 24, p. 267, 2025, doi: 10.1186/s12939-025-02652-7.

[7] V. A. D. Djara, Y. Andriyana, and L. Noviyanti, "Modelling the prevalence of stunting toddlers using spatial autoregressive with instrument variable and S-estimator," *Commun. Math. Biol. Neurosci.*, vol. 2022, 2022, doi: 10.28919/cmbn/7234.

[8] R. D. Fadlirhohim, A. T. R. Dani, and A. T. R. Dani, "Modeling stunting prevalence in Indonesia using spline truncated semiparametric regression," *BAREKENG J. Math. Appl.*, vol. 18, no. 3, pp. 2015–2028, 2024, doi: 10.30598/barekengvol18iss3pp2015-2026.

[9] T. Handayani, S. Sifriyani, and A. T. Rian Dani, "Stunting prevalence modeling using nonparametric regression of quadratic splines," *J. Varian*, vol. 7, no. 2, pp. 149–160, Jun. 2024, doi: 10.30812/varian.v7i2.2916.

[10] G. Moonga, S. Böse-O'Reilly, U. Berger, K. Harttgen, C. Michelo, D. Nowak, U. Siebert, J. Yabe, and J. Seiler, "Modelling chronic malnutrition in Zambia: A Bayesian distributional regression approach," PLoS ONE, vol. 16, no. 8, p. e0255073, Aug. 2021, doi: 10.1371/journal.pone.0255073.

[11] A. K. Kuchibhotla and R. K. Patra, "Efficient estimation in single index models through smoothing splines," *Bernoulli*, vol. 26, no. 2, pp. 1587–1618, 2020, doi: 10.3150/19-BEJ1183.

[12] R. Du and H. Yamada, "Principle of duality in cubic smoothing spline," *Mathematics*, vol. 8, no. 10, pp. 1–19, Oct. 2020, doi: 10.3390/math8101839.

[13] B. Lestari, Fatmawati, I. N. Budiantara, and N. Chamidah, "Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing spline and kernel estimators approaches," *J. Phys. Conf. Ser.*, Dec. 2019, doi: 10.1088/1742-6596/1397/1/012064.

[14] T. Ampa, I. N. Budiantara, and I. Zain, “Selection of optimal smoothing parameters in mixed estimator of kernel and Fourier series in semiparametric regression,” *J. Phys. Conf. Ser.*, Dec. 2021, doi: 10.1088/1742-6596/2123/1/012035.

[15] N. J. van Eck and L. Waltman, “Citation-based clustering of publications using CitNetExplorer and VOSviewer,” *Scientometrics*, vol. 111, no. 2, pp. 1053–1070, May 2017, doi: 10.1007/s11192-017-2300-7.

[16] D. Guleria and G. Kaur, “Bibliometric analysis of ecopreneurship using VOSviewer and RStudio Bibliometrix, 1989–2019,” *Library Hi Tech*, vol. 39, no. 4, pp. 1001–1024, 2021, doi: 10.1108/LHT-09-2020-0218.

[17] N. Ersen, İ. Akyüz, and K. C. Akyüz, “Bibliometric analysis of the International Journal in Wood Science using visualization mapping method,” *Eurasian J. Forest Sci.*, vol. 12, no. 2, pp. 47–65, Aug. 2024, doi: 10.31195/ejejfs.1467759.

[18] H. F. F. Mahmoud, “Parametric versus semi and nonparametric regression models,” *Int. J. Stat. Probab.*, vol. 10, no. 2, p. 90, Feb. 2021, doi: 10.5539/ijsp.v10n2p90.

[19] N. P. A. M. Mariati, I. N. Budiantara, and V. Ratnasari, “The application of mixed smoothing spline and Fourier series model in nonparametric regression,” *Symmetry*, vol. 13, no. 11, Nov. 2021, doi: 10.3390/sym13112094.

[20] Y. Jiao, G. Shen, Y. Lin, and J. Huang, “Deep nonparametric regression on approximate manifolds: Nonasymptotic error bounds with polynomial prefactors,” Ann. Stat., vol. 51, no. 2, pp. 691–716, 2023, doi: 10.1214/23-AOS2266.

[21] D. Kitahara, L. Condat, and A. Hirabayashi, “One-dimensional edge-preserving spline smoothing for estimation of piecewise smooth functions,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Brighton, U.K., 2019, pp. 5611–5615, doi: 10.1109/ICASSP.2019.8683260.

[22] S. O. Adams and M. A. Zubair, “On some techniques of selecting spline smoothing parameters for a correlated dataset with autocorrelation structure in the residual,” *World J. Adv. Res. Rev.*, vol. 17, no. 2, pp. 68–78, Feb. 2023, doi: 10.30574/wjarr.2023.17.2.0216.

[23] M. Bica and D. Curilă, “The Akima’s fitting method for quartic splines,” *J. Numer. Anal. Approx. Theory*, vol. 51, no. 2, pp. 155–166, Dec. 2022.

[24] J. Yu, Y. Chen, W. Zhong, and P. Ma, “Smoothing splines approximation using Hilbert curve basis selection,” J. Comput. Graph. Stat., vol. 31, no. 3, pp. 802–812, 2022, doi: 10.1080/10618600.2021.2002161.

[25] S. O. Adams, D. A. Obaromi, and A. A. Irinews, “Goodness of fit test of an autocorrelated time series cubic smoothing spline model,” *J. Niger. Soc. Phys. Sci.*, vol. 3, no. 3, pp. 191–200, Aug. 2021, doi: 10.46481/jnsps.2021.265.

[26] R. Risnawati, A. A. R. Fernandes, and Nurjannah, “The estimation function approach smoothing spline regression analysis for longitudinal data,” in IOP Conf. Ser.: Mater. Sci. Eng., vol. 546, no. 1, p. 052064, 2019, doi: 10.1088/1757-899X/546/5/052064.