CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (2) (2026), Pages 2181–2200
p-ISSN: 2086-0382; e-ISSN: 2477-3344

# Spatial Dependence and Co-Kriging Prediction of Stunting Prevalence Using the Linear Model of Coregionalization

Zakiyah Mar'ah*[1], Isma Muthahharah[1], and Mahendri Deayu Putri[2]

[1] *Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Indonesia*
[2] *Department of Nutrition, Faculty of Sports and Health Sciences, Universitas Negeri Makassar, Indonesia*

**Abstract**

Stunting remains a major public health challenge in Indonesia, marked by substantial regional disparities. This study analyzed the spatial dependence structure of stunting prevalence and its determinants using the Linear Model of Coregionalization (LMC) and co-kriging, which exploit spatial cross-correlation among variables rather than autocorrelation alone. Secondary data for 38 provinces in 2024 from the Indonesian Health Profile comprised stunting prevalence and nine explanatory variables, modeled jointly through empirical semivariograms and cross-semivariograms. The selected model combined a nugget effect with a spherical structure of 600 km effective range (LMC = Nug + Sph(600km)), the nugget was non-zero for every variable and reported explicitly. Stunting prevalence showed moderate spatial dependence (SDP = 68.78%). Low birth weight showed positive cross-spatial dependence with stunting, whereas exclusive breastfeeding and pediatric health service coverage showed negative dependence. The poverty pattern was weak and mixed in the full sample but turned clearly positive once the two provinces reporting 0% prevalence were excluded. These are co-spatial patterns among provincial aggregates, not causal effects. Co-kriging revealed distinct clustering, with East Nusa Tenggara, parts of Papua, and Sulawesi showing relatively high predicted prevalence. Leave-one-out cross-validation gave a mean error of 0.264 and an RMSE of 6.423 percentage points, close to the between-province standard deviation (7.68), with prediction standard errors of 5.6–7.7, indicating moderate accuracy. Excluding the 0% values left the dependence class, effective range, and cross-covariance signs unchanged. Multivariate geostatistics thus offers spatial evidence for geographically targeted, ecological-level stunting interventions.

**Keywords:** co-kriging; linear model of coregionalization; multivariate geostatistics; spatial dependence; stunting prevalence



## 1. Introduction

Stunting remains a major public health concern in Indonesia and a significant barrier to human development. As a manifestation of chronic early-childhood undernutrition, it is driven by a complex interplay of determinants, immediate factors such as inadequate dietary intake and poor health, and underlying factors related to environmental quality, socioeconomic status, education, and household food security [1]. Without sustained intervention, stunting carries

*Corresponding author. E-mail: zakiyahm@unm.ac.id

Received: June 22, 2026 Revised: August 30, 2026 Accepted: September 16, 2026
DOI: https://doi.org/10.18860/cauchy.v11i2.44312

substantial long-term consequences, including higher risk of low birth weight, greater vulnerability to infectious and non-communicable diseases (hypertension, cardiovascular disease, obesity), impaired cognitive development, and reduced adult productivity [2]. Against this challenge, Indonesia has made notable progress, the national stunting rate fell from 24.4% in 2021 to 21.5% in 2022, 21.1% in 2023, and 19.8% in 2024 (SSGI) [3], suggesting that coordinated interventions have improved child nutritional outcomes nationally.

This national improvement, however, masks considerable regional disparities. In 2024, provincial prevalence reached 37% in East Nusa Tenggara against a national mean of 21.09%, with a between-province variance of 59.05 (standard deviation $\approx$ 7.68 percentage points, coefficient of variation $\approx$ 36%). Progress has been faster in western Indonesia, whereas East Nusa Tenggara still exceeds 30%, underscoring the need for geographically targeted interventions [4]. Such disparities imply that the underlying determinants, poverty, access to safe drinking water, and healthcare accessibility, are distributed non-randomly across space [5], plausibly exhibiting spatial dependence and positive autocorrelation that produce clustered patterns of nutritional vulnerability and persistent pockets of disadvantage. This study does not assume such structure a priori but quantifies it formally through empirical semivariogram modeling and the spatially dependent proportion (SDP) index.

Univariate spatial models cannot fully represent the multifactorial nature of stunting because they ignore the spatial covariance among determinants. Co-kriging based on the Linear Model of Coregionalization (LMC) addresses this by characterizing spatial variability and cross-spatial dependence among multiple variables simultaneously [6]. By integrating the variance and covariance structures of stunting prevalence and its determinants into a valid positive semi-definite model, the LMC provides a rigorous basis for representing complex cross-variable spatial dependence [7], allowing co-kriging to draw on correlated secondary variables to improve prediction accuracy, reduce estimation uncertainty, and yield a more informative map of the geographic distribution of stunting [8].

The choice of an LMC-based co-kriging framework over more familiar spatial models is deliberate. Ordinary and universal kriging are univariate: the former uses only the autocovariance of stunting prevalence, and the latter adds a large-scale trend but still ignores correlated determinants [9, 10]. Regression kriging [11] and spatial regression models, SAR and CAR [12] and the Besag–York–Mollié (BYM) model [13], treat covariates as fixed terms and model only the residual spatial structure, so they do not represent the spatial cross-covariance between the response and each determinant. Geographically weighted regression (GWR) [14] and its multiscale extension (MGWR) [15], widely applied in stunting research [1, 8], let coefficients vary across space but are local regression techniques rather than geostatistical predictors, and they yield neither a valid multivariate covariance model nor optimal predictions with uncertainty at unsampled locations. By contrast, the LMC jointly models the direct and cross semivariograms of stunting prevalence and its determinants within a single positive semi-definite covariance structure [16, 17], and co-kriging exploits this structure to borrow strength from correlated secondary variables, an approach well suited to a setting where the response is observed at only 38 provincial locations while its determinants are spatially structured and mutually correlated.

Prior studies underscore the value of multivariate spatial modeling for stunting. One reported significant spatial heterogeneity in factors associated with three stunting-related indicators across Southeast Sulawesi, showing that multivariate approaches capture interdependencies among indicators and produce more comprehensive risk maps than univariate methods [18]. Another found substantial spatial variation in the determinants of nutritional outcomes and showed that modeling relationships among indicators improved the accuracy of spatial assessments [19]. These studies share a limitation, however: they describe how regression relationships vary across space, through geographically weighted or segmentation-based models, but do not explicitly model the spatial cross-covariance among indicators or generate geostatistical predictions with quantified uncertainty at unsampled locations. This gap, the absence of a valid multivariate covariance model linking stunting prevalence to its determinants, motivates the LMC-based co-kriging framework adopted here.

By integrating a validated LMC and co-kriging, this study moves beyond descriptive mapping to model-based predictions of stunting prevalence together with their co-kriging prediction variance, which the secondary variables can lower relative to univariate kriging [20]. The contribution is one of integrated application rather than methodology, no new mathematical procedure is introduced, since the LMC, the Goulard–Voltz algorithm, and the co-kriging equations are all standard. What the study adds is their joint application to stunting in Indonesia, a single valid LMC fitted across all ten variables rather than a series of pairwise models, coregionalization matrices checked for positive semi-definiteness and reported, a co-kriging predictor validated by leave-one-out cross-validation against a univariate benchmark, and the spatially structured share of variation for each variable quantified through the SDP index. Because the units of analysis are provinces, the resulting evidence is intended to help policymakers decide where to target stunting-reduction programs, not to draw conclusions about individual children.

# 2. Methods

This section describes the data, analytical procedures, model specification, and validation strategy used to characterize multivariate spatial dependence and predict stunting prevalence across Indonesian provinces.

## 2.1. Data

Secondary data used in this study were obtained from the *Indonesia Health Profile 2024* published by the Indonesian Ministry of Health [21]. The response variable ($Y$) is the 2024 stunting prevalence in the 38 provinces of Indonesia, accompanied by nine explanatory variables covering demographic, socioeconomic, and health aspects: $X_1$) percentage of Low Birth Weight (LBW) infants, $X_2$) proportion of households with access to safe drinking water, $X_3$) Human Development Index (HDI), $X_4$) mean years of schooling, $X_5$) percentage of the population living below the poverty line, $X_6$) percentage of infants younger than six months receiving exclusive breastfeeding, $X_7$) mean maternal age at first marriage, $X_8$) percentage of antenatal care coverage, and $X_9$) percentage of pediatric health service coverage. Every variable comes from this single publication, refers to the same reference year, and was compiled for the same 38 provinces. Relying on one source keeps definitions, denominators, and recording periods aligned across variables, therefore comparisons between regions are not disturbed by differences in the way the data were collected. Table 1 sets out the definition, denominator, and unit of each variable, and these are applied consistently throughout the manuscript.

**Table 1:** Definition, Denominator, and Unit of Each Variable

| Code | Variable name | Operational definition | Denominator | Unit |
|---|---|---|---|---|
| $Y$ | Stunting prevalence | Percentage of children aged 0–59 months classified as severely stunted or stunted on the height-for-age index (HAZ), that is, a $Z$-score below $-2$ SD | Children aged 0–59 months measured | % |
| $X_1$ | LBW infants | Newborns with a birth weight below 2,500 grams | Newborns weighed | % |
| $X_2$ | Households with access to safe drinking water | Households with access to an adequate drinking water source, namely a main source of piped water, protected water, or rainwater; for users of bottled or refilled water, the source counts as adequate if the water used for bathing and washing comes from a protected source | All households | % |

| Code | Variable name | Operational definition | Denominator | Unit |
|---|---|---|---|---|
| $X_3$ | Human Development Index (HDI) | Composite index of life expectancy, expected years of schooling, mean years of schooling, and adjusted per capita expenditure | Not applicable (composite index) | Index 0–100 |
| $X_4$ | Mean years of schooling | Mean years of schooling of the population aged 15 years and over | Population aged 15 years and over | Years |
| $X_5$ | The population living below the poverty line | Population whose average monthly per capita consumption expenditure falls below the poverty line | Total population | % |
| $X_6$ | Exclusive breastfeeding coverage | Coverage of infants given breast milk alone from birth until six months of age, with no other food or drink except medicine, vitamins, and minerals | Infants aged six months | % |
| $X_7$ | Mean maternal age at first marriage | Mean age at first marriage among women | Women who have ever been married | Years |
| $X_8$ | Antenatal care coverage | Coverage of K6 antenatal care, that is, pregnant women receiving antenatal services to standard with at least six contacts, of which at least two involve examination by a physician | Target number of women giving birth | % |
| $X_9$ | Pediatric health services coverage | Coverage of children under five served by Stimulation, Early Detection, and Intervention for Child Growth and Development (SDIDTK) at community health centres | Target number of children under five | % |

Note: All variables were obtained from the Indonesia Health Profile 2024 [21], the reference year for all variables is 2024.

The nine explanatory variables are factors repeatedly reported as related to stunting in Indonesian studies. Variable $X_8$ was included as a proxy for access to maternal health services. Its value exceeds 100% in several provinces because the denominator is a projected number of pregnancies rather than an actual count, therefore its relationship with stunting is interpreted with caution. A preliminary screening of the pairwise correlations between each candidate variable and stunting prevalence was also carried out to confirm their relevance before they were entered into the multivariate model.

## 2.2. Analysis Stages

All analyses were conducted in R (version 4.5.1) using the `gstat` package for empirical variogram estimation, linear model of coregionalization fitting, and co-kriging, together with the `sp` and `sf` packages for spatial-data handling and the `spdep` package for supporting diagnostics. The analytical procedure consisted of the following steps.

1. Exploring the spatial distribution of stunting prevalence across the provinces of Indonesia.
2. Estimating the empirical semivariogram in Eq. (1) [22],

$$\widehat{\gamma}_X(h) = \frac{1}{2N(\mathbf{h})} \sum_{(i,j)=1}^{N(\mathbf{h})} (x_i - x_j)^2 \,, \tag{1}$$

and the cross-semivariogram in Eq. (2) [22],

$$\widehat{\gamma}_{XY}(h) = \frac{1}{2N(\mathbf{h})} \sum_{(i,j)=1}^{N(\mathbf{h})} (x_i - x_j)(y_i - y_j) \,, \tag{2}$$

with $i, j = 1, 2, \ldots, k$; $N(h)$ is the number of observation pairs at lag distance $h$; $h$ denotes the separation distance vector; $x_i$ and $x_j$ represent the observations of variable $x$ at locations

$i$ and $j$, respectively; and $y_i$ and $y_j$ represent the observations of variable $y$ at locations $i$ and $j$, respectively.

3. Interpretation of spatial variability using the semivariogram [23] and characterization of inter-variable spatial relationships through the cross-semivariogram [24].
4. Construction of the Linear Model of Coregionalization (LMC) framework in Eq. (3) [16],

$$\gamma_{ij}(h) = c_{ij,0}\, g_0(h_0) + c_{ij,1}\, g_1(h_1), \tag{3}$$

where $c_{ij,0}$ is the semivariogram coefficient associated with the nugget effect, whose sill corresponds to the variance attained when the variogram stabilizes and is equivalent to the sample variance. The coefficient $c_{ij,1}$ represents the contribution of the basic semivariogram structure, whereas $g_0$ denotes the nugget effect, reflecting variability at very short distances. A nonzero semivariogram value near the origin indicates the presence of a nugget effect. Moreover, $g_1$ denotes the basic semivariogram model, and $h_0$ and $h_1$ represent the corresponding lag distances for the nugget and structural components, respectively. The present study adopted the nugget model in Eq. (4) and the spherical model in Eq. (5), both of which satisfy the positive semidefiniteness requirement. In these models, $h$ denotes the separation distance, $a$ the range parameter at which the variogram reaches the sill, and $c$ the sill parameter [25].

$$g_0(h_0) = \begin{cases} 0, & \text{if } h = 0, \\ 1, & \text{otherwise}, \end{cases} \tag{4}$$

$$g_1(h_1) = \begin{cases} \left(\dfrac{3h}{2a}\right) - \left(\dfrac{h}{2a}\right)^3, & \text{if } h \le a, \\ 1, & \text{otherwise}. \end{cases} \tag{5}$$

The nugget effect reflects variability attributable to several sources, including intrinsic small-scale variability, unobserved external influences, measurement errors, and information loss associated with very short separation distances. The relative contribution of the nugget component to the total variance is of particular importance in spatial analysis, as the difference between the total variance and the nugget variance corresponds to the spatially structured component that can be modeled through spatial dependence.

5. Estimation of the coregionalization matrix in Eq. (6) [26],

$$\mathbf{\Gamma}(h) = \begin{bmatrix} \gamma_{11}(h) & \cdots & \gamma_{1p}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{np}(h) & \cdots & \gamma_{pp}(h) \end{bmatrix} = \sum_{s=0}^{M} \mathbf{C}_s\, g_s(h_s) = \sum_{s=0}^{M} \begin{bmatrix} c_{11,s} & \cdots & c_{1p,s} \\ \vdots & \ddots & \vdots \\ c_{p1,s} & \cdots & c_{pp,s} \end{bmatrix} g_s(h_s), \tag{6}$$

where $\mathbf{C}_s$ denotes the positive semidefinite matrix of semivariogram coefficients (coregionalization matrix), $\mathbf{\Gamma}(h)$ is the semivariogram matrix at lag distance $h$, with diagonal entries representing direct semivariograms and off-diagonal entries representing cross-semivariograms, $g_s$ is the estimated positive semi-definite semivariogram function, and $s = 0, 1, \ldots, m$ indexes the spatial structures incorporated into the model.

In the present study, the LMC was constructed by fitting both direct semivariogram and cross-semivariogram models. Each estimated coregionalization matrix was then projected onto the nearest positive semidefinite matrix through a least-squares procedure in which negative eigenvalues were set to zero. This projection guarantees that the resulting LMC is positive semidefinite, which is the condition required for a mathematically valid multivariate spatial covariance structure. Positive definiteness is a stronger property and is not required [27].

6. Evaluation of multivariate spatial variability across different lag distances. The proportion of spatially structured variation among regions was measured using the Spatially Dependent

Proportion (SDP) index in Eq. (7) [28],

$$SDP(\%) = \left(\frac{C_1}{C_0 + C_1}\right) \times 100. \tag{7}$$

7. Estimation of stunting prevalence using multivariate co-kriging interpolation based on the fitted LMC, as given in Eq. (8) [6, 29],

$$\widehat{Y}(u) = \sum_{i=1}^{n} \lambda_i Y(u_i) + \sum_{i=1}^{n}\sum_{j=1}^{m} \beta_{ij} X_j(u_i), \tag{8}$$

with $\widehat{Y}(u)$ denoting the predicted value of the primary variable at location $u$, $Y(u_i)$ the observed value of the primary variable at the $i$-th sampling location, $X_j(u_i)$ the observed value of the $j$-th secondary variable at the same location, $\lambda_i$ the kriging weight assigned to the primary variable, $\beta_{ij}$ the co-kriging weight associated with the $j$-th secondary variable, $n$ the number of observations of the primary variable, and $m$ the number of secondary variables.

8. Generation of thematic maps to visualize the spatial distribution and interrelationships among variables based on the multivariate co-kriging results.
9. Evaluation of the predictive performance of the co-kriging model using Mean Error (ME) in Eq. (9) and Root Mean Square Error (RMSE) in Eq. (10), which quantify prediction bias and overall estimation accuracy, respectively [30],

$$ME = \frac{1}{N}\sum_{i=1}^{n}\left[Y(x_i) - \widehat{Y}(x_i)\right], \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left[Y(x_i) - \widehat{Y}(x_i)\right]^2}. \tag{10}$$

10. Development of evidence-based and geographically targeted policy recommendations for stunting reduction based on the results of the multivariate spatial analysis.

### 2.3. Data Preprocessing, Estimation, Model Selection, and Validation

Each province was represented by the geographic centroid (longitude and latitude) of its area, and pairwise distances were computed as great-circle distances in kilometers, representing an areal unit by a single point is an approximation, so the fitted model describes variation between provincial centroids and carries no information about within-province variation. The 0% stunting prevalence recorded for Highland Papua and Central Papua most likely reflects incomplete reporting in these newly established provinces rather than a true absence of cases, these values were retained as recorded in the main analysis, and their influence was assessed through a sensitivity analysis that treated both as unavailable and repeated the entire procedure (empirical semivariogram estimation, LMC fitting, SDP computation, co-kriging, and leave-one-out cross-validation) on the remaining 36 provinces (Subsection 3.3). All variables were analyzed on their original measurement scales, so the coregionalization matrices are reported in each variable's original units, no logarithmic or power transformation was applied, and because the direct and cross semivariograms were fitted jointly through the LMC, prior standardization was not required. Stunting prevalence and the nine determinants were assembled into a single multivariable model, each specified with a constant but unknown mean, from which the empirical direct and cross semivariograms were estimated jointly by the method of moments (Eq. (1) and Eq. (2)) using the default lag configuration in `gstat` (cutoff at one-third of the bounding-box diagonal, equal-width lag classes). The analysis assumed intrinsic stationarity, examined by inspecting the semivariograms for unbounded growth; as no strong global trend was evident at the provincial scale, the assumption was considered adequate.

The LMC was constructed by fitting a common set of basic structures, a nugget effect and a single bounded structure, to all direct and cross semivariograms simultaneously. Among the exponential, Gaussian, and spherical alternatives, the nugget-plus-spherical combination was retained for its smallest weighted residual sum of squares and best cross-validation performance, yielding LMC = Nug + Sph(600km) (Subsection 3.2); individual models were fitted by weighted least squares, with weights proportional to the number of pairs per lag divided by the squared lag distance. The matrices $\mathbf{C}_0$ and $\mathbf{C}_1$ were estimated with the iterative Goulard–Voltz algorithm [31] (`fit.lmc` in `gstat`), which updates all sills jointly under the constraint that each matrix remains positive semi-definite: at each iteration the matrix was eigen-decomposed, negative eigenvalues set to zero, and the matrix reconstructed from the truncated non-negative eigenvalues. The LMC requires only positive semi-definiteness (all eigenvalues $\geq 0$), not the stronger positive definiteness, so a fitted matrix may rest on the boundary of the positive semi-definite cone with one or more zero eigenvalues; the eigenvalues of both matrices are reported in Subsection 3.2 (Table 5).

Spatial prediction used ordinary co-kriging, with stunting prevalence as the primary variable and the nine determinants as secondary variables, their means treated as unknown but locally constant. At each location the co-kriging weights ($\lambda$ and $\beta$ in Eq. (8)) were obtained by solving the ordinary co-kriging system built from the direct and cross covariances implied by the fitted LMC, subject to the unbiasedness constraints that the primary weights sum to one and the secondary weights to zero. Predictive performance was assessed by leave-one-out cross-validation, each province predicted in turn from the remaining 37, using the Mean Error (ME, Eq. (9)) and Root Mean Square Error (RMSE, Eq. (10)); because RMSE is in the units of the response, it was interpreted relative to the national mean (21.09%) and the between-province standard deviation (7.68 percentage points). To quantify the added value of the multivariate approach, the LMC-based co-kriging was benchmarked against univariate ordinary kriging of stunting prevalence alone under the same procedure (Subsection 3.2, Table 7). All computations were carried out in R using the `sp` package for spatial referencing and the `gstat` package for variogram estimation, LMC fitting, co-kriging, and cross-validation.

## 3. Results and Discussion

This section presents the empirical findings, beginning with an exploratory assessment of stunting prevalence, followed by the fitted multivariate spatial structure, co-kriging predictions, sensitivity analysis, and interpretation of the results.

### 3.1. Exploratory Data Analysis

An exploratory analysis was first conducted to characterize the distribution of stunting prevalence across provinces in Indonesia. Descriptive statistics were used to summarize the main characteristics of the data, and thematic maps were employed to visualize the spatial distribution of stunting prevalence and identify potential geographic patterns. Table 2 reports the descriptive statistics of stunting prevalence.

Based on Table 2, stunting prevalence across Indonesia in 2024 ranged from a minimum of 0%, recorded in Highland Papua and Central Papua. These zeros most likely reflect data limitations, unavailable measurements, missing records stored as zero, or reporting challenges in these newly established provinces, rather than a true absence of cases. They were nonetheless retained as recorded in the main analysis, and their effect was quantified rather than assumed: Subsection 3.3 reports a sensitivity analysis treating both as unavailable and repeating the full modelling sequence on the remaining 36 provinces. The highest prevalence was in East Nusa Tenggara (37%). The mean across provinces was 21.09%, an unweighted average in which each province contributes equally, so it exceeds the population-weighted national estimate of 19.8%

**Table 2:** Descriptive statistics of stunting prevalence

| Statistic | Value | Location |
|---|---|---|
| Minimum | 0.00 | Highland Papua and Central Papua |
| Maximum | 37.00 | East Nusa Tenggara |
| Mean | 21.09 | – |
| Variance | 59.05 | – |
| Median | 22.05 | – |
| Std. deviation | 7.68 | – |
| IQR | 8.38 | – |

(SSGI 2024 [3]), where low-prevalence, populous Java provinces carry greater weight. A variance of 59.05 indicates substantial heterogeneity in prevalence among provinces.

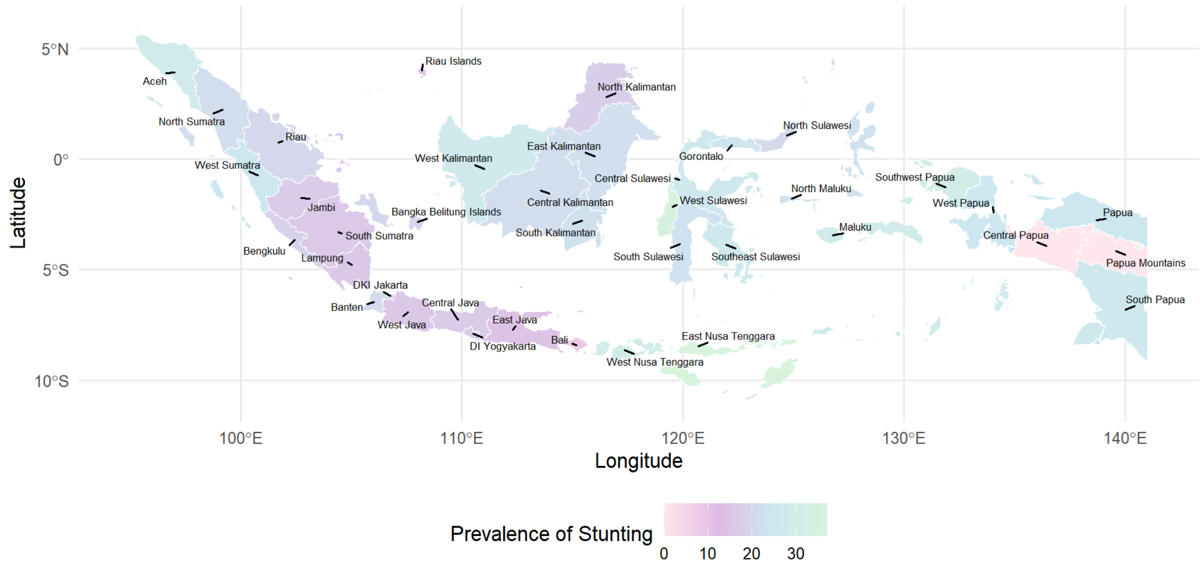



**Fig. 1:** Spatial distribution of stunting prevalence in Indonesia [21]

Based on the thematic map in Fig. 1, the spatial distribution of stunting prevalence in Indonesia exhibits considerable regional variation. The eastern part of the country, particularly East Nusa Tenggara and several provinces in Papua, is characterized by relatively higher stunting prevalence, as indicated by the darker color intensity. In contrast, several provinces in Sumatra, Kalimantan, and Sulawesi display low to moderate prevalence levels. This pattern suggests the existence of pronounced spatial disparities in nutritional and health outcomes across regions, which may be associated with differences in socioeconomic conditions, access to healthcare services, sanitation infrastructure, and food availability.

## 3.2. Linear Model of Coregionalization (LMC)

Construction of the LMC began with estimation of the direct and cross-semivariograms for the study variables. The direct semivariograms characterise the spatial variability of each variable individually, whereas the cross-semivariograms quantify the spatial cross-correlation between variable pairs. Jointly, these functions specify the multivariate spatial dependence structure required for a valid LMC and for the subsequent co-kriging predictions. The estimated direct semivariograms are reported in Table 3.

Based on Table 3, the variables exhibited distinct patterns of spatial variability across Indonesian provinces. Stunting prevalence ($Y$) showed semivariance increasing with lag distance, indicating spatial dependence, geographically neighboring provinces tend to have more similar prevalence than distant ones. Most determinants also showed spatial dependence, with $X_5$, $X_6$,

**Table 3:** Summary of the estimated direct semivariograms

| Variable | Minimum semivariance | Maximum semivariance | Spatial dependence class |
|---|---|---|---|
| $Y$ | 3.84 | 52.32 | Moderate |
| $X_1$ | 0.72 | 4.48 | Strong |
| $X_2$ | 9.40 | 184.30 | Moderate |
| $X_3$ | 9.99 | 30.23 | Moderate |
| $X_4$ | 0.66 | 2.18 | Moderate |
| $X_5$ | 0.78 | 27.79 | Strong |
| $X_6$ | 8.05 | 378.70 | Strong |
| $X_7$ | 0.31 | 1.38 | Moderate |
| $X_8$ | 108.18 | 535.24 | Moderate |
| $X_9$ | 38.56 | 343.51 | Strong |

and $X_9$ rising markedly with separation distance before levelling off, the shape expected under spatial dependence. The absolute size of the semivariance values, however, is not comparable across variables, because semivariance is in the squared units of each variable, percentage-scale variables with large variances ($X_6$, $X_8$, $X_9$) necessarily yield large values, while those varying over narrow ranges ($X_1$, $X_4$, $X_7$) yield small ones. Cross-variable comparisons of spatial-dependence strength therefore rely on the dimensionless SDP index reported later in this subsection. The corresponding semivariogram plots are displayed in Fig. 2.

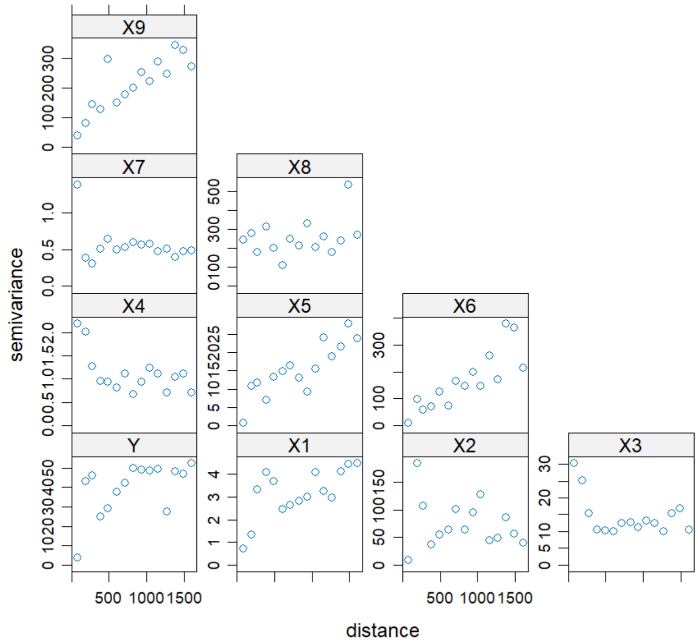



**Fig. 2:** Semivariogram plot of the response and the nine explanatory variables

Fig. 2 shows that the empirical semivariograms of $X_1$ (LBW infants), $X_4$ (mean years of schooling), and $X_7$ (mean maternal age at first marriage) were comparatively flat across lags, and $X_3$ fluctuated only slightly, whereas $X_2$ and $X_8$ fluctuated substantially at particular lags, a pattern reflecting unstable estimation in poorly populated lag classes as much as genuine heterogeneity. Overall, the determinants display distinct spatial characteristics, providing evidence of spatially structured variation that supports the use of multivariate geostatistical modeling for stunting prevalence in Indonesia. The estimated cross-semivariograms between stunting prevalence and each determinant are summarized in Table 4.

Based on Table 4, each explanatory variable showed a distinct pattern of cross-spatial dependence with stunting prevalence. A cross-semivariogram measures how two variables jointly vary with separation distance. It describes co-spatial structure between two mapped surfaces and

**Table 4:** Summary of the estimated cross-semivariograms

| Variable pair | Semivariance range (Gamma) | General pattern | Interpretation of cross-spatial dependence |
|---|---|---|---|
| $Y$–$X_1$ | $-0.47$ to 8.45 | Stable positive | Relatively consistent positive |
| $Y$–$X_2$ | $-32.43$ to 59.38 | Highly fluctuating | Strong but unstable |
| $Y$–$X_3$ | $-18.37$ to 23.05 | Positive–negative | Varies with distance |
| $Y$–$X_4$ | $-3.62$ to 7.15 | Close to zero | Relatively weak |
| $Y$–$X_5$ | $-13.39$ to 11.97 | Weak, mixed in sign | Weak, positive after excluding 0% provinces |
| $Y$–$X_6$ | $-44.62$ to 3.15 | Predominantly negative | Inverse |
| $Y$–$X_7$ | $-1.95$ to 1.08 | Very small | Very weak |
| $Y$–$X_8$ | $-62.04$ to 40.93 | Strongly negative | Moderately strong negative |
| $Y$–$X_9$ | $-69.64$ to 34.16 | Highly fluctuating | Heterogeneous |

cannot on its own identify a causal, protective, or risk relationship, so the interpretations below are confined to co-spatial patterns among provincial aggregates. $Y$–$X_1$ showed a predominantly positive cross-semivariance that grew with lag distance, provinces with a higher share of LBW infants tend to lie in the same parts of the country as those with higher stunting prevalence. $Y$–$X_5$ was weaker and mixed in sign in the full sample (negative at short lags, positive at longer ones), and the positive poverty–stunting pattern sharpens once the two 0%-prevalence provinces are set aside (Subsection 3.3). By contrast, $Y$–$X_6$, $Y$–$X_8$, and $Y$–$X_9$ were predominantly negative, indicating that higher coverage of exclusive breastfeeding, antenatal care, and pediatric health services tends to coincide with lower stunting prevalence, though no causal claim is made. $Y$–$X_2$ and $Y$–$X_3$ fluctuated in sign across lags (spatial heterogeneity), while $Y$–$X_4$ and $Y$–$X_7$ stayed near zero (weak association). This heterogeneity is precisely what the joint LMC fit is designed to accommodate. The corresponding cross-semivariogram plots are shown in Fig. 3.

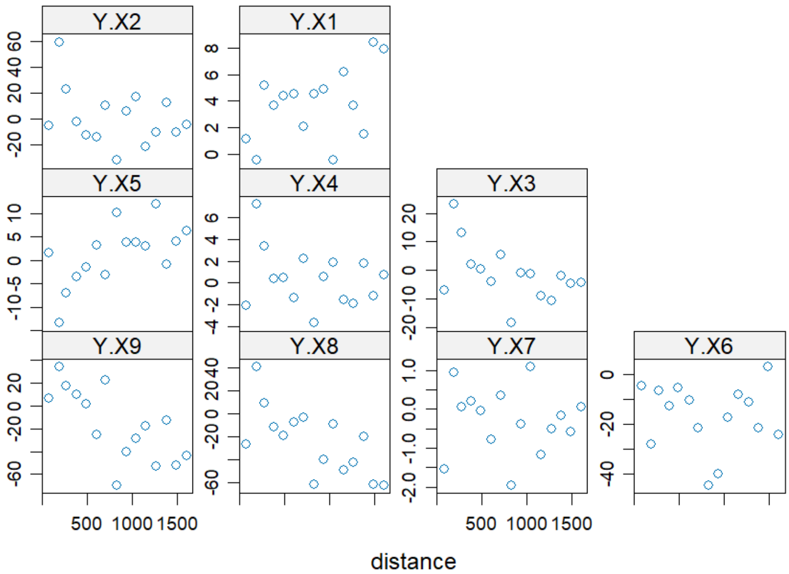
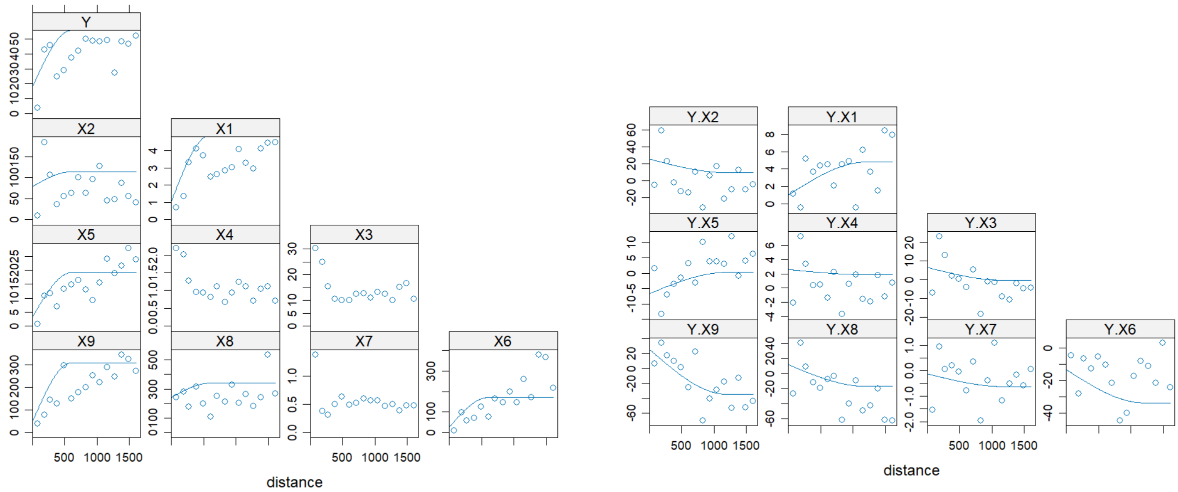



**Fig. 3:** Cross-semivariogram plot between stunting prevalence and each explanatory variable

Fig. 3 shows that most empirical direct and cross-semivariograms increased with lag distance before tending to stabilize, indicating spatial dependence. The selected model was therefore LMC = Nug + Sph(600km), a nugget plus a spherical structure with a 600 km effective range. The spherical structure was preferred over the exponential and Gaussian alternatives because it gave the smallest weighted residual sum of squares across the direct and cross semivariograms and the best leave-one-out cross-validation performance, consistent with the gradual approach to a sill seen in most semivariograms. The range of about 600 km, expressed as great-circle distance between province centroids, was chosen because semivariance stabilized beyond it. Overall, the model adequately represented the multivariate spatial dependence and provided a valid basis for co-kriging interpolation. The fitted direct and cross-semivariograms are presented in Fig. 4.




**(a)** Direct semivariograms **(b)** Cross-semivariograms

**Fig. 4:** Fitted LMC = Nug + Sph(600km)

Fig. 4 presents the fitted direct and cross-semivariograms from the LMC = Nug + Sph(600km) model. The spherical model adequately captured most variables, particularly $Y$, $X_1$, $X_5$, $X_6$, and $X_9$, whose semivariance rose gradually and approached the sill at about 600 km, indicating substantial spatial similarity among provinces up to this range. In contrast, $X_2$ and $X_8$ fluctuated more, so the model captured only their general trend, while $X_3$, $X_4$, and $X_7$ were relatively flat, indicating weaker spatial dependence. Every variable retained a non-zero nugget; for stunting prevalence the nugget accounts for 18.36 of a total sill of 58.81 (about 31% of the modelled variance). Among the cross-semivariograms, $Y$–$X_1$ was positive and $Y$–$X_6$, $Y$–$X_8$, and $Y$–$X_9$ predominantly negative, $Y$–$X_5$ changed sign between structures (negative nugget, positive spherical), therefore its full-sample pattern is mixed rather than clearly positive, $Y$–$X_3$ changed sign across lag distances (scale-dependent), and $Y$–$X_4$ and $Y$–$X_7$ stayed near zero. The coregionalization matrices then partition this structure, with diagonal elements giving each variable's spatial variance and off-diagonal elements the cross-spatial associations (positive or negative), split into nugget ($\mathbf{C}_0$) and spherical ($\mathbf{C}_1$) components.

$$\mathbf{C}_0 = \begin{bmatrix} 18.36 & 0.24 & 27.39 & 3.34 & 1.55 & -5.58 & -18.46 & -1.01 & 6.25 & 18.63 \\ 0.24 & 1.09 & -2.53 & -0.33 & -0.16 & 0.18 & 1.65 & 0.20 & -2.08 & -0.63 \\ 27.39 & -2.53 & 82.37 & 35.49 & 10.53 & -14.48 & -26.38 & 3.87 & 95.83 & 6.18 \\ 3.34 & -0.33 & 35.49 & 33.65 & 9.30 & -6.72 & 4.46 & 6.11 & 87.59 & -23.57 \\ 1.55 & -0.16 & 10.53 & 9.30 & 2.94 & -2.01 & 1.05 & 1.82 & 25.05 & -6.35 \\ -5.58 & 0.18 & -14.48 & -6.72 & -2.01 & 3.21 & 4.50 & -0.64 & -17.71 & -1.04 \\ -18.46 & 1.65 & -26.38 & 4.46 & 1.05 & 4.50 & 27.11 & 3.06 & 13.13 & -29.21 \\ -1.01 & 0.20 & 3.87 & 6.11 & 1.82 & -0.64 & 3.06 & 1.54 & 16.68 & -6.73 \\ 6.25 & -2.08 & 95.83 & 87.59 & 25.05 & -17.71 & 13.13 & 16.68 & 253.19 & -66.61 \\ 18.63 & -0.63 & 6.18 & -23.57 & -6.35 & -1.04 & -29.21 & -6.73 & -66.61 & 47.16 \end{bmatrix}$$

The coregionalization matrix $\mathbf{C}_0$ represents the contribution of the nugget effect to the spatial variance and covariance among variables at very short distances. The diagonal elements are the nugget sills of the individual variables. They are reported in the units of each variable squared and so are not comparable across variables, $X_8$ and $X_2$ carry the largest numbers because they are measured on percentage scales with large sample variances, not because their short-range variation is stronger in any dimensionless sense. The off-diagonal elements represent local spatial covariances between variables. Positive values, such as those between $Y$ and $X_1$ and between $Y$ and $X_3$, indicate positive local spatial associations, whereas negative values, including those between $Y$ and $X_5$, $Y$ and $X_6$, and $Y$ and $X_9$, reflect inverse spatial relationships at the local scale. None of the nugget sills is zero, so short-range variation is present for every variable. For stunting prevalence the nugget sill is 18.36 against a partial sill of 40.45, and the corresponding share of spatially structured variation is quantified by the SDP index in Table 6.

$$\mathbf{C}_1 = \begin{bmatrix} 40.45 & 4.20 & -19.54 & -3.90 & -0.31 & 4.27 & -12.73 & -0.22 & -18.79 & -40.79 \\ 4.20 & 4.06 & -1.49 & 2.00 & 0.31 & 0.57 & -6.51 & 0.76 & -10.14 & -14.41 \\ -19.54 & -1.49 & 37.28 & 2.51 & 0.59 & 5.81 & 45.01 & 2.42 & -2.17 & 21.67 \\ -3.90 & 2.00 & 2.51 & 32.90 & 2.21 & -4.11 & 10.88 & 1.98 & -8.94 & 16.75 \\ -0.31 & 0.31 & 0.59 & 2.21 & 2.82 & -0.91 & 1.81 & 0.79 & -0.90 & 1.89 \\ 4.27 & 0.57 & 5.81 & -4.11 & -0.91 & 16.89 & -6.73 & 0.71 & -16.55 & -26.62 \\ -12.73 & -6.51 & 45.01 & 10.88 & 1.81 & -6.73 & 151.72 & 1.86 & 64.97 & 148.12 \\ -0.22 & 0.76 & 2.42 & 1.98 & 0.79 & 0.71 & 1.86 & 2.25 & -3.42 & -1.39 \\ -18.79 & -10.14 & -2.17 & -8.94 & -0.90 & -16.55 & 64.97 & -3.42 & 104.25 & 147.10 \\ -40.79 & -14.41 & 21.67 & 16.75 & 1.89 & -26.62 & 148.12 & -1.39 & 147.10 & 276.58 \end{bmatrix}$$

The coregionalization matrix $\mathbf{C}_1$ describes the contribution of the spherical structure to the overall spatial variance and covariance. The diagonal elements are the partial sills of the spherical structure, which acts up to the effective range of approximately 600 km. As with $\mathbf{C}_0$, these entries carry the squared units of each variable and are not compared across variables here. The off-diagonal elements further reveal clear cross-spatial relationships among variables. At the spherical (regional) scale, positive cross-covariance entries link stunting prevalence ($Y$) to $X_1$ and $X_5$, whereas negative entries link $Y$ to $X_2$, $X_6$, and $X_9$. For $Y$–$X_5$ the matching nugget covariance is negative, so the full-sample pattern is scale-dependent rather than uniformly positive. These entries describe co-spatial structure and are not effect estimates. Overall, the $\mathbf{C}_1$ matrix indicates that the regional spatial structure contributes more substantially to the total spatial variability than local random effects, implying that broad-scale spatial processes play a dominant role in explaining variations in stunting prevalence and its associated determinants across Indonesia.

Validity of the fitted LMC requires that every coregionalization matrix be positive semidefinite, so that each matrix has only non-negative eigenvalues. This is weaker than positive definiteness, which requires all eigenvalues to be strictly positive, and it is positive semidefiniteness alone that the Goulard and Voltz projection enforces. Table 5 reports the eigenvalues of both fitted matrices, obtained by symmetric eigen-decomposition of $\mathbf{C}_0$ and $\mathbf{C}_1$ as reported above.

Both matrices satisfy the required condition. $\mathbf{C}_1$ has ten strictly positive eigenvalues, the smallest being 1.44, and is therefore positive definite. $\mathbf{C}_0$ has nine clearly positive eigenvalues and a tenth of 0.002, which is zero to the precision at which the matrix entries are reported. $\mathbf{C}_0$ is therefore positive semidefinite but not positive definite: it sits on the boundary of the positive semidefinite cone, which is the expected outcome when the eigenvalue-truncation step of the fitting algorithm sets a negative eigenvalue to zero. The sum $\mathbf{C}_0 + \mathbf{C}_1$, which gives the total sill matrix and governs the co-kriging system at large separations, has a smallest eigenvalue of 1.54 and is positive definite. The model is consequently admissible for co-kriging, and the distinction between the two properties is maintained throughout: positive semidefiniteness is what the LMC requires and what $\mathbf{C}_0$ attains, while positive definiteness is the stronger property attained by $\mathbf{C}_1$ and by $\mathbf{C}_0 + \mathbf{C}_1$.

Subsequently, the proportion of spatial variability among regions was quantified using the Spatially Dependent Proportion (SDP) index, and the resulting measures of spatial variability for each variable are presented in Table 6.

**Table 5:** Eigenvalues of the fitted coregionalization matrices

| Order | Eigenvalues of $\mathbf{C}_0$ (nugget) | Eigenvalues of $\mathbf{C}_1$ (spherical) |
|---|---|---|
| 1 | 345.73 | 460.39 |
| 2 | 110.04 | 92.55 |
| 3 | 5.28 | 47.27 |
| 4 | 4.34 | 39.61 |
| 5 | 1.96 | 12.73 |
| 6 | 1.72 | 6.72 |
| 7 | 0.571 | 3.29 |
| 8 | 0.482 | 2.72 |
| 9 | 0.220 | 2.49 |
| 10 | 0.002 | 1.44 |
| Smallest eigenvalue | 0.002 ($\geq 0$) | 1.44 ($> 0$) |
| Status | Positive semidefinite | Positive definite |

**Table 6:** Proportion of spatial variability for each variable

| Variable | Nugget | Partial sill | Total sill | SDP (%) | Category |
|---|---|---|---|---|---|
| $Y$ | 18.36 | 40.45 | 58.81 | 68.78 | Moderate |
| $X_1$ | 1.09 | 4.06 | 5.15 | 78.83 | Strong |
| $X_2$ | 82.37 | 37.28 | 119.65 | 31.16 | Moderate |
| $X_3$ | 33.65 | 32.90 | 66.55 | 49.44 | Moderate |
| $X_4$ | 2.94 | 2.82 | 5.76 | 48.96 | Moderate |
| $X_5$ | 3.21 | 16.89 | 20.10 | 84.03 | Strong |
| $X_6$ | 27.11 | 151.72 | 178.83 | 84.84 | Strong |
| $X_7$ | 1.54 | 2.25 | 3.79 | 59.37 | Moderate |
| $X_8$ | 253.19 | 104.25 | 357.44 | 29.16 | Moderate |
| $X_9$ | 47.16 | 276.58 | 323.74 | 85.43 | Strong |

Based on Table 6, the SDP analysis indicates that stunting prevalence ($Y$) exhibits a moderate degree of spatial dependence, with an SDP value of 68.78%, suggesting that its distribution is influenced by both spatially structured and local non-spatial factors. Because the SDP is a ratio of sills, it is dimensionless and is the appropriate basis for comparing the strength of spatial dependence across variables measured on different scales. Among the explanatory variables, pediatric health services coverage ($X_9$), exclusive breastfeeding coverage ($X_6$), the percentage of the population living below the poverty line ($X_5$), and the LBW infants ($X_1$) exhibited strong spatial dependence, with SDP values of 85.43%, 84.84%, 84.03%, and 78.83%, respectively. These findings indicate that the spatial distributions of these variables are strongly structured across regions. In contrast, access to safe drinking water ($X_2$), Human Development Index ($X_3$), mean years of schooling ($X_4$), mean maternal age at first marriage ($X_7$), and antenatal care coverage ($X_8$) exhibited moderate spatial dependence, suggesting that their spatial patterns are governed by a combination of regional spatial processes and local random effects.

Thematic maps derived from the LMC visualise co-kriging predictions and thereby indicate where stunting prevalence is most strongly associated with its determinants across Indonesia. Fig. 5 presents the co-kriging interpolation obtained under the fitted LMC = Nug + Sph(600 km), in which prevalence at unsampled locations is predicted by exploiting the spatial dependence between the response and the covariates; darker shades denote higher predicted values. The surface is smooth and continuous, without abrupt transitions between neighbouring regions, indicating that the model adequately captures the underlying spatial structure, while the effective range of approximately 600 km implies that inter-provincial influence remains substantial within this distance before gradually attenuating. One qualification governs interpretation, the inputs are areal, province-level figures represented by 38 centroids, so interior values are interpolations between provincial centroids rather than district-level or local prevalence estimates.

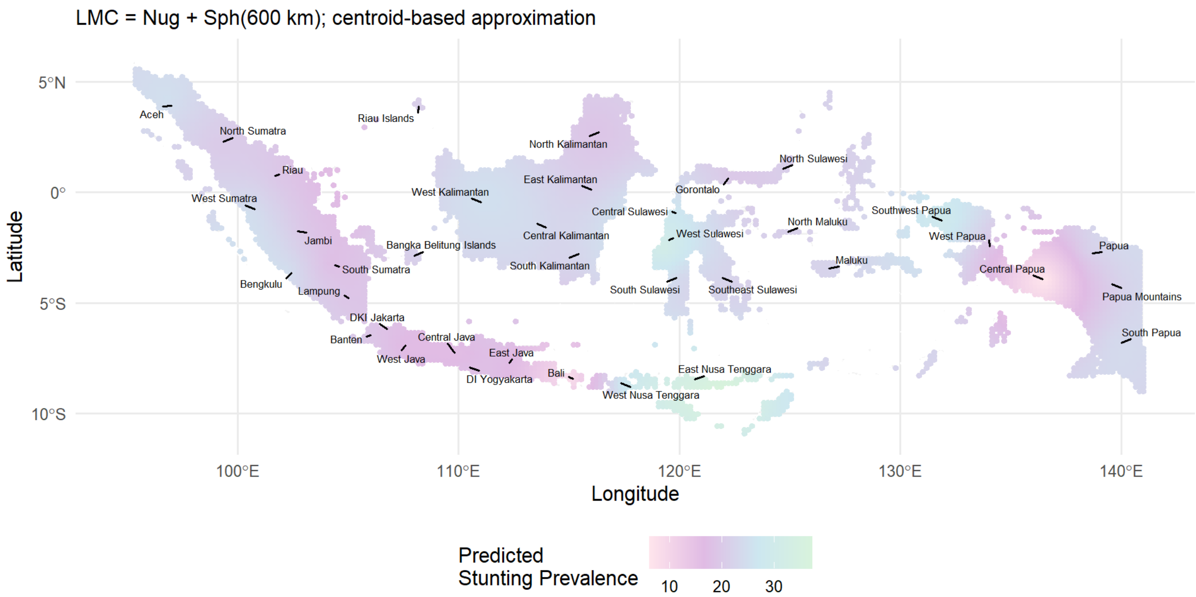



**Fig. 5:** Predicted spatial distribution of stunting prevalence using co-kriging

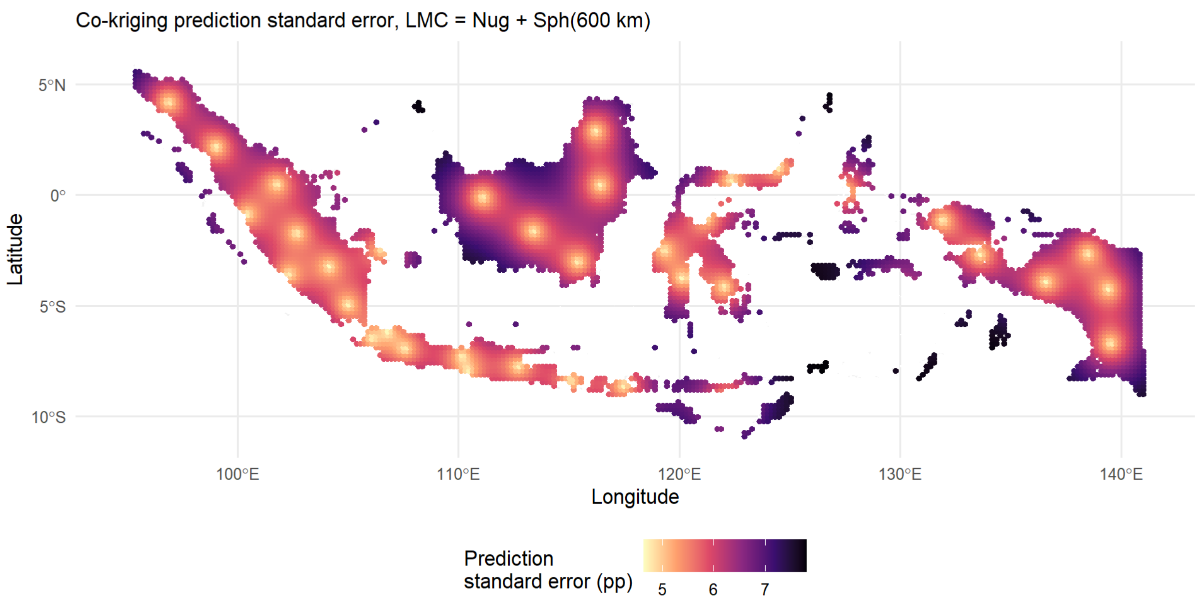



**Fig. 6:** Co-kriging prediction standard error of stunting prevalence, LMC = Nug + Sph(600km)

The prediction map reveals distinct spatial clustering, with eastern Indonesia generally showing higher predicted stunting prevalence than western Indonesia. East Nusa Tenggara, several provinces in Papua, and parts of Sulawesi have relatively high predicted values, whereas northern Sumatra, Kalimantan, and parts of Java are lower to moderate. Stunting prevalence is therefore spatially clustered rather than randomly distributed, and the high-prevalence areas indicate where geographically targeted interventions could be prioritised. The leave-one-out prediction standard error ranged from about 5.6 to 7.7 percentage points (mean 6.8), smallest in Java, where provincial centroids lie close together, and largest in Aceh, East Nusa Tenggara, Maluku, and southern Papua, where the nearest centroid is several hundred kilometres away. Fig. 6 maps this pattern of prediction uncertainty across the country.

In this study, stunting prevalence ($Y$) was treated as the primary variable, whereas variables $X_1$–$X_9$ were considered secondary variables. Accordingly, the co-kriging estimator can be expressed as in Eq. (11). The weights $\lambda$ and $\beta$ in Eq. (11) were obtained by solving the co-kriging system constructed from the semivariogram and cross-semivariogram models defined by the fitted LMC = Nug + Sph(600km). These weights quantify the relative contribution of the primary and secondary variables, as well as the spatial proximity among regions, to the prediction of stunting prevalence. Consequently, neighboring provinces and explanatory variables with stronger spatial correlations exert greater influence on the predicted values. By incorporating both direct and cross-spatial dependence, the co-kriging estimator uses the multivariate spatial structure of the data.

$$\begin{aligned} \widehat{Y}(u) &= \sum_{i=1}^{38} \lambda_i Y(u_i) + \sum_{i=1}^{38}\sum_{j=1}^{9} \beta_{ij} X_j(u_i) \\ &= \sum_{i=1}^{38} \lambda_i Y(u_i) + \sum_{i=1}^{38} \beta_{i1} X_1(u_i) + \sum_{i=1}^{38} \beta_{i2} X_2(u_i) + \sum_{i=1}^{38} \beta_{i3} X_3(u_i) + \sum_{i=1}^{38} \beta_{i4} X_4(u_i) \\ &\quad + \sum_{i=1}^{38} \beta_{i5} X_5(u_i) + \sum_{i=1}^{38} \beta_{i6} X_6(u_i) + \sum_{i=1}^{38} \beta_{i7} X_7(u_i) + \sum_{i=1}^{38} \beta_{i8} X_8(u_i) + \sum_{i=1}^{38} \beta_{i9} X_9(u_i)\,. \end{aligned} \tag{11}$$

Subsequently, the predictive performance of the co-kriging model based on the fitted LMC was assessed using the Mean Error (ME) and Root Mean Square Error (RMSE) metrics, as reported in Table 7.

**Table 7:** Co-kriging prediction performance

| Evaluation metric | Value |
|---|---|
| ME | 0.264 |
| RMSE | 6.423 |

Table 7 reports the leave-one-out cross-validation results. The Mean Error of 0.264 is close to zero, indicating negligible systematic bias, while the RMSE of 6.423 percentage points, about 30% of the national mean (21.09%) and below the between-province standard deviation (7.68), shows that the prediction error is comparable to the natural variation among provinces, therefore the co-kriging predictions are only moderately accurate. This RMSE is also close to the mean co-kriging prediction standard error of 6.8 percentage points, confirming that the model-based uncertainty is consistent with the observed errors. For comparison, univariate ordinary kriging using stunting prevalence alone gave a mean error of 0.074 and an RMSE of 7.699 percentage points. The LMC-based co-kriging thus lowered the RMSE by about 16.6% (6.423 versus 7.699), showing that incorporating the spatially correlated determinants improved point-prediction accuracy. Both models were approximately unbiased, and overall the LMC-based co-kriging yielded nearly unbiased but only moderately accurate predictions.

### 3.3. Sensitivity to the Reported Zero Prevalence in Highland Papua and Central Papua

Highland Papua and Central Papua report a stunting prevalence of 0%. As argued in the exploratory analysis, this almost certainly reflects incomplete reporting in two provinces created by administrative division in 2022 rather than an absence of stunted children. Because these two records are influential, both in the far east of the country and at the extreme of the response distribution, the main results should not be allowed to rest on them without a check. The entire analysis was therefore repeated with both observations treated as unavailable, leaving 36 provinces. The reduced data set was carried through the same steps as the main analysis: empirical direct and cross semivariograms, joint LMC fitting by the Goulard and Voltz algorithm with the same nugget-plus-spherical structure and the same 600 km range, SDP computation, ordinary co-kriging, and leave-one-out cross-validation. Table 8 compares the two fits.

The main conclusions are robust to excluding the two zero-prevalence provinces. The spatial dependence of stunting prevalence stays in the moderate class (SDP falling from 68.78% to

**Table 8:** Sensitivity of the main results to the reported zero prevalence values

| Quantity | Full sample (38 provinces) | Zeros treated as unavailable (36 provinces) |
|---|---|---|
| Basic structures | Nug + Sph(600km) | Nug + Sph(600km) |
| SDP of stunting prevalence | 68.78% (moderate) | 63.80% (moderate) |
| Sign of $Y$–$X_1$ cross-covariance | Positive | Positive |
| Sign of $Y$–$X_5$ cross-covariance | Weak, mixed in sign | Positive |
| Sign of $Y$–$X_6$ cross-covariance | Negative | Negative |
| Sign of $Y$–$X_9$ cross-covariance | Negative | Negative |
| Cross-validation ME (pp) | 0.264 | 0.166 |
| Cross-validation RMSE (pp) | 6.423 | 5.039 |
| Mean prediction standard error (pp) | 6.8 | 4.7 |

63.80%), the effective range remains near 600 km therefore the same spherical structure is retained, and the substantively interpreted cross-covariance signs are unchanged, positive for LBW and negative for exclusive breastfeeding and pediatric health services coverage. Cross-validation RMSE falls from 6.423 to 5.039 percentage points with the mean error still near zero, though this improvement simply reflects dropping two records the model cannot reproduce rather than a better-specified model. Two features do change, the poverty–stunting cross-covariance becomes clearly positive on the reduced sample (weak and mixed in sign when the two high-poverty, zero-prevalence provinces are retained), and the SDP of the variables on which these provinces are extreme, safe drinking water, poverty, and pediatric health services coverage, rises. Overall, the substantive conclusions do not depend on the two zero values, and the positive poverty–stunting pattern is strengthened rather than merely preserved.

### 3.4. Discussion

Stunting prevalence in Indonesia in 2024 showed considerable interprovincial variation (mean 21.09%) and moderate spatial dependence (SDP = 68.78%), with relatively high predicted prevalence in eastern Indonesia, East Nusa Tenggara, several Papua provinces, and parts of Sulawesi. Because the data are provincial aggregates, these results describe how prevalence is arranged across the country, not which children within a province are stunted, and province-level patterns should not be read as individual risk. Neighbouring provinces tend to record similar prevalence, plausibly because they share broadly comparable socioeconomic conditions, healthcare access, sanitation, and food availability, consistent with earlier work [32, 33], emphasizing that geographic disparities sustain stunting in low- and middle-income countries.

The fitted LMC showed positive cross-spatial dependence between stunting and LBW infants ($X_1$) and negative dependence with exclusive breastfeeding coverage ($X_6$) and pediatric health services coverage ($X_9$), the population living below the poverty line ($X_5$) was weak and mixed in the full sample but clearly positive once the two 0%-prevalence provinces were set aside (Subsection 3.3). These four variables also showed strong spatial dependence (SDP = 84.03%, 78.83%, 84.84%, and 85.43%). In other words, provinces with high poverty and LBW rates tend to be those with high stunting prevalence, while high-coverage provinces tend to record low prevalence. These are ecological, province-level co-spatial patterns: consistent with, but not by themselves establishing, the individual-level mechanisms documented in the clinical literature, inferring individual effects would be an ecological fallacy. The findings align with global [34], and Indonesian [35] evidence identifying chronic undernutrition, poor socioeconomic conditions, and inadequate maternal and child healthcare as major determinants, and they support incorporating geographic heterogeneity into stunting-reduction strategies.

Because this evidence is about places rather than people, its natural policy use is geographic targeting, deciding where to concentrate programmes already known to work, not what works. Stunting-reduction efforts should therefore adopt a place-based approach in high-prevalence provinces such as East Nusa Tenggara and those in Papua, including strengthened first-1,000-days nutrition, higher exclusive breastfeeding coverage, LBW prevention through improved maternal nutrition, expanded maternal and child healthcare, and reinforced social protection for vulnerable households. Because these determinants cluster geographically rather than varying independently, multisectoral policies integrating health, education, sanitation, and poverty alleviation are appropriate in these provinces. The specific interventions are drawn from individual-level evidence elsewhere; this study's contribution is to indicate where to prioritise them, not to demonstrate that they work. These spatial patterns are consistent with earlier Indonesian studies using alternative techniques, geographically weighted analyses of child malnutrition [1], geographically weighted regression kriging in Malang [8], and multivariate segmentation of nutritional status in Southeast Sulawesi [18, 19], and extend that work by expressing the interdependence as an explicit multivariate spatial covariance structure and producing continuous, uncertainty-aware provincial predictions rather than local regression coefficients alone.

### 3.5. Limitations

This study has several limitations. First, the analysis rests on aggregate data for only 38 provinces, a small sample for geostatistical estimation that limits the resolution of the fitted semivariograms, so the national-scale results should be regarded as exploratory. Second, because the data are provincial aggregates, every association is ecological, the cross-semivariograms and coregionalization matrices describe co-spatial dependence between mapped aggregates and cannot be read as individual-level relationships, causal effects, or risk and protective factors for any child; relatedly, each province is represented by its centroid, so the surface in Fig. 5 is a centroid-based approximation and not local or district-level prevalence. Third, the 0% values for Highland Papua and Central Papua most likely reflect incomplete reporting rather than genuine zero prevalence and, retained as recorded, may bias predicted prevalence for eastern Indonesia downward. The re-analysis without them in Subsection 3.3 leaves the spatial dependence class, effective range, and cross-covariance signs unchanged while improving cross-validation accuracy and sharpening the poverty–stunting pattern. Fourth, the model was assessed only by internal leave-one-out cross-validation, without external or temporal validation. Fifth, no formal significance-based cluster detection was performed, so the high-prevalence areas are not statistically identified hotspots. Sixth, the moderate spatial dependence (SDP = 68.78%) implies that about 31% of the variation is not captured by the spatial structure, pointing to unmodeled local or non-spatial determinants that future work should incorporate, for example through finer district-level data or hybrid regression-kriging approaches.

## 4. Conclusion

Stunting prevalence in Indonesia in 2024 exhibited moderate spatial dependence (SDP = 68.78%), indicating that adjacent provinces tend to share similar levels. The selected model, LMC = Nug + Sph(600km) with a non-zero nugget for every variable, adequately represented the empirical direct and cross-semivariograms, and both coregionalization matrices were verified positive semi-definite. Cross-spatial dependence with stunting was positive for LBW infants ($X_1$) and negative for exclusive breastfeeding coverage ($X_6$) and pediatric health services coverage ($X_9$). The population living below the poverty line ($X_5$) was weak and mixed in the full sample but clearly positive once the two 0%-prevalence provinces were set aside. These are province-level co-spatial patterns, not causal effects, and $X_1$, $X_5$, $X_6$, and $X_9$ showed the strongest spatial dependence, playing a prominent role in shaping the spatial pattern of stunting.

Co-kriging interpolation revealed distinct clustering, with relatively high predicted prevalence in eastern Indonesia, East Nusa Tenggara, several Papua provinces, and parts of Sulawesi, and lower to moderate levels in parts of Sumatra, Java, and Kalimantan. This surface is a centroid-based approximation, not district-level prevalence. Leave-one-out cross-validation gave a Mean Error of 0.264 and an RMSE of 6.423 percentage points (mean prediction standard error 6.8), and because this error is close to the between-province standard deviation (7.68), the predictions are only moderately accurate and should be read as indicative rather than high-precision estimates. A sensitivity analysis excluding the two 0%-prevalence provinces left these conclusions intact. Overall, LMC-based co-kriging is useful for characterizing spatial dependence and generating information to support geographically targeted stunting-reduction strategies in Indonesia.

## CRediT Authorship Contribution Statement

**Zakiyah Mar'ah:** Conceptualization, Methodology, Formal Analysis, Writing–Original Draft. **Isma Muthahharah:** Supervision, Project Administration, Writing–Review & Editing. **Mahendri Deayu Putri:** Validation, Writing–Review & Editing.

## Declaration of Generative AI and AI-assisted technologies

Generative AI tools (specifically ChatGPT) were utilised exclusively for translation from Bahasa to English during the preparation of this work. The analysis, interpretation, and core content were not generated by AI. All substantive results and discussions were entirely developed by the authors themselves.

## Declaration of Competing Interest

The authors declare no competing interests.

## Ethics Approval and Consent to Participate

Ethics approval was not required for this study, since the analysis used publicly available aggregate data at the provincial level.

## Funding and Acknowledgments

This research was funded by the Ministry of Higher Education, Science, and Technology (Kemdiktisaintek) of the Republic of Indonesia under the Early Career Lecturer Research Grant scheme (Penelitian Dosen Pemula, PDP) for the 2026 fiscal year, under contract 2813/DST/UN36.12/TU/2026.

## Data and Code Availability

The data analysed in this study are publicly available. All variables were obtained from Profil Kesehatan Indonesia 2024, published by the Ministry of Health of the Republic of Indonesia.

## References

[1] T. Susanto, T. Saifudin, and N. Chamidah. "Spatial Modeling of Child Malnutrition in Indonesia Using Geographically Weighted Multivariate Regression". In: *BAREKENG: Jurnal Ilmu Matematika dan Terapan* 20.3 (2026), pp. 1837–1854. DOI: 10.30598/barekengvol20iss3pp1837-1854.

[2] A. Meily and W. Nurasih. "Determinan Stunting pada Anak di Indonesia: Sebuah Tinjauan Literatur Sistematis". In: *MEJORA Medical Journal Awatara* 3.4 (2025), pp. 199–205. DOI: 10.61434/mejora.v3i4.287.

[3] Kementerian Kesehatan Republik Indonesia. *Survei Status Gizi Indonesia (SSGI) 2024*. Tech. rep. Jakarta: Badan Kebijakan Pembangunan Kesehatan, Kementerian Kesehatan Republik Indonesia, 2025.

[4] A. J. Tanesib and A. Kusumayati. "Determinan Kejadian Stunting pada Balita di Provinsi Nusa Tenggara Timur: Analisis Data Survei Kesehatan Indonesia 2023". In: *Jurnal Promotif Preventif* 8.6 (2025), pp. 1542–1554. DOI: 10.47650/jpp.v8i6.2121.

[5] A. Widodo and B. Agus Herlambang. "Sistem Informasi Geografis Sebaran Stunting dan Tingkat Kemiskinan di Provinsi Jawa Tengah Tahun 2023". In: *Jurnal Penelitian Sistem Informasi* 2.1 (2024), pp. 55–63. DOI: 10.54066/jpsi.v2i1.1457.

[6] P. A. Dowd and E. Pardo-Igúzquiza. "The Many Forms of Co-kriging: A Diversity of Multivariate Spatial Estimators". In: *Mathematical Geosciences* 56.2 (2024), pp. 387–413. DOI: 10.1007/s11004-023-10104-7.

[7] S. De Iaco, D. E. Myers, and D. Posa. "The Linear Coregionalization Model and the Product-Sum Space-Time Variogram". In: *Mathematical Geology* 35.1 (2003), pp. 25–38. DOI: 10.1023/A:1022425111459.

[8] H. Pramoedyo, M. Mudjiono, A. A. Fernandes, D. Ardianti, and K. Septiani. "Determination of Stunting Risk Factors Using Spatial Interpolation Geographically Weighted Regression Kriging in Malang". In: *Mutiara Medika: Jurnal Kedokteran dan Kesehatan* 20.2 (2020). DOI: 10.18196/mm.200250.

[9] N. A. C. Cressie. *Statistics for Spatial Data*. Revised. New York: John Wiley & Sons, 1993.

[10] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press, 1997.

[11] T. Hengl, G. B. M. Heuvelink, and D. G. Rossiter. "About Regression-Kriging: From Equations to Case Studies". In: *Computers & Geosciences* 33.10 (2007), pp. 1301–1315. DOI: 10.1016/j.cageo.2007.05.001.

[12] L. Anselin. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers, 1988.

[13] J. Besag, J. York, and A. Mollié. "Bayesian Image Restoration, with Two Applications in Spatial Statistics". In: *Annals of the Institute of Statistical Mathematics* 43.1 (1991), pp. 1–20. DOI: 10.1007/BF00116466.

[14] A. S. Fotheringham, C. Brunsdon, and M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley & Sons, 2002.

[15] A. S. Fotheringham, W. Yang, and W. Kang. "Multiscale Geographically Weighted Regression (MGWR)". In: *Annals of the American Association of Geographers* 107.6 (2017), pp. 1247–1265. DOI: 10.1080/24694452.2017.1352480.

[16] H. Wackernagel. *Multivariate Geostatistics: An Introduction with Applications*. 3rd ed. Springer, 2003. DOI: 10.1007/978-3-662-05294-5.

[17] A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. London: Academic Press, 1978.

[18] B. W. Otok, Purhadi, R. Sriningsih, and D. S. Dila. "Segmentation of Toddler Nutritional Status Using REBUS and FIMIX Partial Least Square in Southeast Sulawesi". In: *MethodsX* 12 (2024), p. 102515. DOI: 10.1016/j.mex.2023.102515.

[19] F. R. Fadmi, B. W. Otok, Kuntoro, S. Melaniani, and R. Sriningsih. "Segmentation of Stunting, Wasting, and Underweight in Southeast Sulawesi Using Geographically Weighted Multivariate Poisson Regression". In: *MethodsX* 12 (2024), p. 102736. DOI: 10.1016/j.mex.2024.102736.

[20] D. Payares-Garcia, F. Osei, J. Mateu, and A. Stein. "Multivariate Poisson Cokriging: A Geostatistical Model for Health Count Data". In: *Statistical Methods in Medical Research* 33.9 (2024), pp. 1637–1659. DOI: 10.1177/09622802241268488.

[21] Kementerian Kesehatan Republik Indonesia. *Profil Kesehatan Indonesia 2024*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2025.

[22] M. C. Ribeiro, A. J. Sousa, and M. J. Pereira. "A Coregionalization Model Can Assist Specification of Geographically Weighted Poisson Regression: Application to an Ecological Study". In: *Spatial and Spatio-temporal Epidemiology* 17 (2016), pp. 1–13. DOI: 10.1016/j.sste.2016.02.001.

[23] F. H. Zhao, A. X. Zhu, and C. Z. Qin. "Spatial Distribution Pattern Analysis Using Variograms over Geographic and Feature Space". In: *Geo-spatial Information Science* 28.4 (2025), pp. 1774–1788. DOI: 10.1080/10095020.2024.2439371.

[24] P. Fabijańczyk, J. Zawadzki, and T. Magiera. "Magnetic Susceptibility as a Proxy for Soil Contamination: Cross-Variogram Analysis Across Contrasting Land Use Areas". In: *Environmental Science and Pollution Research* 32.45 (2025), pp. 25917–25931. DOI: 10.1007/s11356-025-37158-w.

[25] S. Yu, L. Zhao, and S. Li. "Leveraging Deep Learning for Automated Experimental Semivariogram Fitting". In: *Atmosphere* 16.2 (2025), p. 191. DOI: 10.3390/atmos16020191.

[26] K. N. Kyzyurova. "On Linear Model of Coregionalization". In: *arXiv preprint* (2019).

[27] Z. Li, X. Zhang, K. C. Clarke, G. Liu, and R. Zhu. "An Automatic Variogram Modeling Method with High Reliability Fitness and Estimates". In: *Computers & Geosciences* 120 (2018), pp. 48–59. DOI: 10.1016/j.cageo.2018.07.011.

[28] E. J. Seidel and M. S. de Oliveira. "A Classification for a Geostatistical Index of Spatial Dependence". In: *Revista Brasileira de Ciência do Solo* 40 (2016). DOI: 10.1590/18069657rbcs20160007.

[29] G. Caringi and P. Secchi. "Fixed Rank Co-Kriging: A Model for Multivariate Spatial Prediction". In: *arXiv preprint* (2026). http://arxiv.org/abs/2603.21748.

[30] D. Khodoli Zangeneh, H. Amanipoor, and S. Battaleb-Looie. "Evaluation of Heavy Metal Contamination Using Cokriging Geostatistical Method (Case Study of Abteymour Oilfield in Southern Iran)". In: *Applied Water Science* 13.10 (2023). DOI: 10.1007/s13201-023-01980-9.

[31] M. Goulard and M. Voltz. "Linear Coregionalization Model: Tools for Estimation and Choice of Cross-Variogram Matrix". In: *Mathematical Geology* 24.3 (1992), pp. 269–286. DOI: 10.1007/BF00893750.

[32] N. Akseer, T. Vaivada, O. Rothschild, K. Ho, and Z. A. Bhutta. "Understanding Multifactorial Drivers of Child Stunting Reduction in Exemplar Countries: A Mixed-Methods Approach". In: *The American Journal of Clinical Nutrition* 112 (2020), 792S–805S. DOI: 10.1093/ajcn/nqaa152.

[33] UNICEF. *The State of the World's Children 2023*. Tech. rep. United Nations Children's Fund, 2023.

[34] M. de Onis and F. Branca. "Childhood Stunting: A Global Perspective". In: *Maternal & Child Nutrition* 12 (2016), pp. 12–26. DOI: 10.1111/mcn.12231.

[35] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, and L. M. Neufeld. "A Review of Child Stunting Determinants in Indonesia". In: *Maternal & Child Nutrition* 14.4 (2018), e12617. DOI: 10.1111/mcn.12617.