AIMS Mathematics, 10(12): 29664−29685.
DOI: 10.3934/math.20251304
Received: 25 September 2025
Revised: 08 December 2025
Accepted: 12 December 2025
Published: 16 December 2025

https://www.aimspress.com/journal/Math

Research article

A modified spherical variogram model with constrained optimization for spatial volume estimation

Johannah Jamalul Kiram[1,2], Rossita Mohamad Yunus[2,*], Yani Japarudin[3], Mahadir Lapammu[3], Olivier Monteuuis[4] and Doreen K. S. Goh[5]

[1] Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, Jalan UMS, Kota Kinabalu, Sabah, Malaysia
[2] Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, Kuala Lumpur, Malaysia
[3] Sabah Softwood Berhad, Tawau, Sabah, Malaysia
[4] CIRAD-BIOS Department – UMR AGAP, TA A-108/03, Avenue Agropolis, F-34398 Montpellier, Cedex 5, France
[5] YSG Biotech Sdn Bhd, Yayasan Sabah Group, Voluntary Association Complex, Mile 2½, off Tuaran Road, Kota Kinabalu, Sabah, Malaysia

* **Correspondence:** Email: rossita@um.edu.my; Tel: +60176316329.

Abstract: In this study, we proposed a modified spherical variogram model aimed at improving the accuracy of spatial modeling in volume estimation. The model enhances the flexibility of the traditional spherical variogram structure by incorporating additional polynomial terms to better capture spatial variability in structured plantation datasets. Parameters such as nugget, sill, range, and the coefficients of the polynomial terms were estimated using the L-BFGS-B optimization algorithm under box constraints, ensuring numerical stability and physically meaningful values. The performance of the modified model was evaluated using real-world volume data from *Tectona grandis Linn. f.* (teak) trees planted in a multiclonal block in Brumas Camp, Tawau, Sabah, Malaysia. To assess model accuracy and generalizability, predicted volumes derived from the fitted variogram model were compared to measured values using three validation strategies: Full dataset fitting, Leave-One-Out Cross-Validation (LOOCV), and K-Fold Cross-Validation. The modified spherical variogram model demonstrated superior performance over the classical version in terms of weighted root mean squared error (RMSE) and coefficient of determination (R^2). These findings highlighted the value of refining variogram structures to improve estimation precision in geostatistical applications, particularly when modeling spatially complex forest data.

Keywords: modified spherical variogram; geostatistics; spatial modelling; constrained optimization; L-BFGS-B; Tectona grandis Linn. F
Mathematics Subject Classification: 62M30, 62H11, 65K10, 62P12

1. Introduction

The accuracy of modeling spatial dependence is a major concern in geostatistics, with broad applications across environmental sciences, engineering, and applied mathematics. The variogram, a key tool for quantifying spatial continuity, plays a central role in geostatistical modeling and spatial prediction [1]. Among the widely used models, the spherical variogram is particularly popular for its simplicity and bounded support, which makes it appealing in applications requiring spatial interpolation with limited range influence [2,3]. However, traditional spherical models often struggle to capture complex spatial structures, particularly when the underlying data exhibit anisotropy, non-stationarity, or scale-dependent variability [4].

While the spherical model is widely used in geostatistical modeling due to its simplicity and ease of interpretation, other models such as exponential, Gaussian, and Matérn variograms have also been widely studied [5]. These models vary in their smoothness and support properties, which influence their suitability in different spatial data. However, even these standard models may fall short when faced with spatially complex, non-stationary, or anisotropic datasets, prompting a need for more adaptable structures [4,6]. Prior research has explored these challenges by proposing new techniques such as nested models, spatial deformation, and spline-based variograms [7–9], but these often increase computational complexity or require additional assumptions. Research has demonstrated the potential of advanced optimization and adaptive modeling techniques in improving spatial prediction accuracy, such as Kriging-assisted reliability analysis [10] and smooth surface fitting via finite-element thin plate splines with adaptive refinement [11], reinforcing the need for more flexible variogram formulations. In this study, we focus on enhancing the flexibility of the spherical model itself through a polynomial-based modification, providing more flexibility and tractability.

We address such limitations by proposing a modified spherical variogram model designed to enhance spatial representation and prediction accuracy. The modified model incorporates additional polynomial terms into the standard spherical structure to enhance flexibility of the model, thus capturing complex spatial patterns. This results in improved fitting capability over heterogeneous spatial domains, which is especially crucial for real-world data such as forestry.

The modified model is constructed to retain essential theoretical properties, including continuity, boundedness, and valid sill behavior. To ensure these properties are preserved, constraints are imposed on the polynomial coefficients so that the variogram transitions smoothly to the sill at $h = a$. To estimate the model parameters, which are nugget, partial sill, range, and polynomial coefficients, we employ the L-BFGS-B optimization algorithm. This quasi-Newton method enables efficient parameter estimation under box constraints, such as non-negativity of the nugget (C_0), partial sill (C), and range ($a > 0$). This constrained optimization framework enables the model to achieve a better empirical fit while maintaining physical and statistical plausibility.

Although the proposed model is demonstrated using spatial volume data from *Tectona grandis Linn. f.* (teak) plantations in Brumas Camp, Tawau, Malaysia, its formulation is grounded in general geostatistical principles and does not rely on domain-specific assumptions. The methodology is therefore applicable to a broad range of spatial datasets across environmental sciences, engineering, and other fields where spatial heterogeneity is present. The main contribution of this work is not only

the improved structure of the variogram model, but also the use of constrained optimization to estimate its parameters. Together, these elements provide a practical alternative to traditional variogram fitting methods, which may give unstable or less accurate results in datasets with complex spatial patterns.

The remainder of this paper is organized as follows: In Section 2, we introduce the mathematical formulation of the modified spherical variogram model, including the continuity constraint and optimization approach. In Section 3, we describe the empirical analysis conducted on the real teak dataset, along with details of the model fitting procedure. In Section 4, we present the results from various cross-validation techniques and compares the performance of the proposed model against the traditional spherical variogram. Finally, in Section 5, we conclude with a summary of findings, discuss the implications for geostatistical modeling, and suggests directions for future research.

2. Materials and methods

2.1. Overview of methodological framework

We present a structured framework for fitting the proposed modified spherical variogram model, as shown in Figure 2.1. The process begins with the mathematical formulation of the model and proceeds to parameter estimation through an optimization approach. To ensure the model remains theoretically valid, a mathematical validity check is conducted, verifying essential conditions such as continuity and boundedness. Finally, the model's predictive performance is evaluated using several cross-validation techniques.

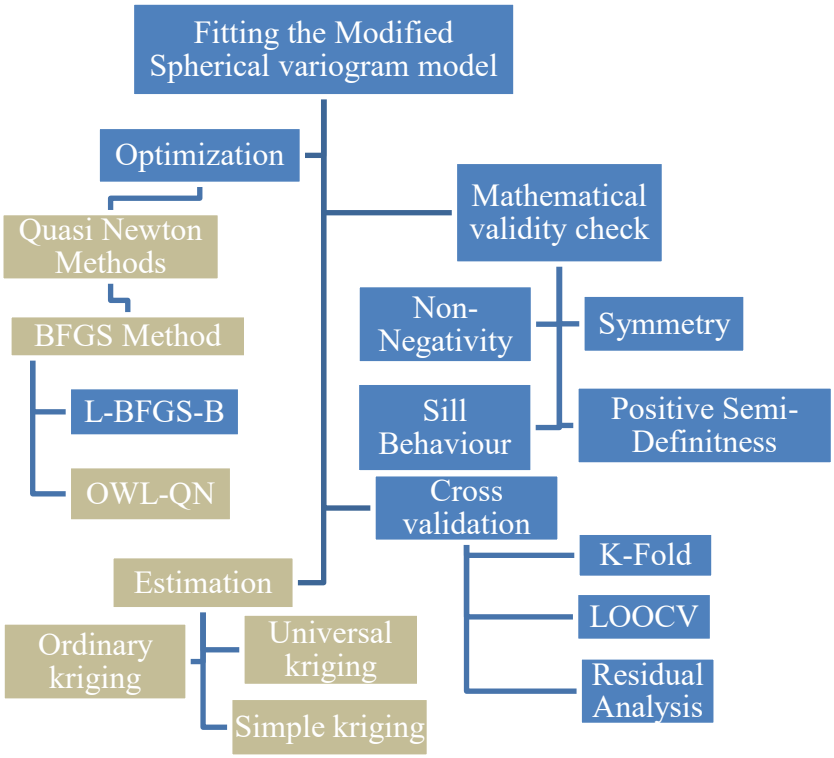


Figure 2.1. Framework for fitting the modified spherical variogram model.

Although Figure 2.1 includes "Estimation" (referring to kriging methods such as ordinary, universal, and simple kriging) as a potential application of the fitted model, this step is acknowledged but not carried out in this paper. The focus remains on variogram fitting and validation. Nonetheless, the model is built to be compatible with these kriging techniques for future use.

2.2. Mathematical formulation of the modified spherical variogram

The modified spherical variogram is defined as the piecewise function:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ C_0 + C\left[x_1\left(\frac{h}{a}\right) + x_2\left(\frac{h}{a}\right)^2 + x_3\left(\frac{h}{a}\right)^3\right], & 0 < h \le a \\ C_0 + C, & h > a \end{cases} \tag{2.1}$$

where:

- C_0 is the nugget effect,
- C is the partial sill variance,
- a is the range parameter,
- h is the lag distance,
- $x_1, x_2,$ and x_3 are coefficients that define the shape of the variogram curve for $0 \le h \le a$.

This definition explicitly satisfies the fundamental condition of a variogram, $\gamma(0) = 0$. The nugget C_0 represents the discontinuity at the origin, i.e.,

$$\lim_{h \downarrow 0} \gamma(h) = C_0, \tag{2.2}$$

consistent with the variogram theory [12]. To guarantee continuity at the range $h = a$ and smooth transition to the sill, the coefficients must satisfy the constraints

$$x_1 + x_2 + x_3 = 1, \text{ and} \tag{2.3}$$

$$x_1 + 2x_2 + 3x_3 = 0. \tag{2.4}$$

Equation (2.3) enforces

$$\gamma(a^-) = C_0 + C,$$

ensuring continuity, while Eq (2.4) ensures

$$\gamma'(a^-) = 0,$$

preventing discontinuity where the variogram reaches its sill.

Rather than estimating (x_1, x_2, x_3) directly under equality constraints, the polynomial is re-parameterised using a single free parameter u, such that

$$x_1 = u, \quad x_2 = 3 - 2u, \quad x_3 = u - 2. \tag{2.5}$$

This analytical substitution ensures that (2.3) and (2.4) are satisfied for all values of u, making the parameterisation both parsimonious and unambiguous.

With this re-parameterisation, the variogram becomes

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ C_0 + C\left[u\left(\frac{h}{a}\right) + (3-2u)\left(\frac{h}{a}\right)^2 + (u-2)\left(\frac{h}{a}\right)^3\right], & 0 < h \le a \\ C_0 + C, & h > a \end{cases} \tag{2.6}$$

which depends only on the four estimable parameters (C_0, C, a, u). Presenting the model in this form eliminates ambiguity regarding non-estimable coefficients and directly aligns with the optimization procedure.

2.2.1. Mathematical validity conditions

Mathematical validity ensures that the variogram model adheres to theoretical conditions essential for reliable spatial prediction, specifically [13]:

1) **Symmetry:** A valid variogram must be symmetric, meaning $\gamma(h) = \gamma(-h)$, which ensures consistent spatial dependence measurement in all directions.
2) **Non-negativity:** Variogram value must never be negative, $\gamma(h) \geq 0$, as negative values are not physically meaningful. This is ensured by; (a) non-negativity constraints on $C_0$and C, and (b) the monotonicity properties of the cubic polynomial under the parameter space used.
3) **Sill Attainment:** A valid variogram should reach a defined sill $(C_0 + C)$ at large distances, indicating the range beyond which spatial autocorrelation becomes negligible.
4) **Conditional Positive Definiteness:** Variograms must be conditionally positive semi-definite to ensure the validity of the kriging interpolation. This condition guarantees the covariance matrix derived from the variogram is invertible and produces non-negative variances.

For kriging to be valid, the covariance matrix

$$K_{ij} = (C_0 + C) - \gamma(|s_i - s_j|)$$

where $|s_i - s_j|$ is the distance between two locations, must be conditionally positive semi-definite. For each fitted model, K was assembled using empirical coordinates, and its eigenvalues were computed. In all cases (6th, 7th, and 10th years), the minimum eigenvalues were non-negative within numerical tolerance, confirming admissibility.

This verification, with the structural form of (2.6), ensures that the model satisfies all necessary validity conditions. Detailed mathematical derivations for the proposed modified spherical variogram model and eigenvalues calculated to prove positive-definiteness are available upon request, as they have not been published.

In terms of monotonicity, a well-behaved variogram should be non-decreasing on the interval $0 \leq h \leq a$. Differentiating (2.6) yields:

$$\gamma'(h) = \frac{C}{a}\left[u + 2(3-2u)\left(\frac{h}{a}\right) + 3(u-2)\left(\frac{h}{a}\right)^2\right]. \tag{2.7}$$

For the estimated values of u obtained in all years, the derivative remains non-negative on $[0, a]$, confirming empirical monotonicity. This monotonic behavior is also visually supported by the fitted curves.

2.2.2. Goodness-of-fit metrics and cross-validation

Goodness-of-fit metrics and cross-validation methods were applied to assess how well the theoretical and modified variogram models describe the empirical spatial structure and predict values at unsampled locations. We employ the following metrics:

1) Weighted Root Mean Squared Error (RMSE): RMSE measures the prediction error magnitude and incorporates bin-level weights based on the number of point pairs:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \omega_i (z_i - \widehat{z_i})^2}{\sum_{i=1}^{n} \omega_i}} \tag{2.7}$$

where ω_i denotes the weight for lag h_i, typically the number of point pairs (np) used in empirical variogram calculation, z_i is the observed value and $\widehat{z_i}$ is the predicted value at location i.

2) Coefficient of Determination (R-squared, R^2): This metric assesses the proportion of variability in observed data explained by predictions:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (z_i - \widehat{z_i})^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2}. \tag{2.8}$$

The closer the value of R^2 to 1, indicates better explanatory ability of the model.

This weighted RMSE and standard R^2 formulation ensures comparability across all models, including the modified spherical variogram, as required for reviewer-consistent evaluation.

To further strengthen the analysis, two cross-validation methods are applied:

1) Leave-One-Out Cross-Validation (LOOCV): Sequentially removes each data point, predicts its value from the remaining data, and evaluates predictive accuracy.
2) K-Fold Cross-Validation: The dataset is partitioned into *k* subsets, training on $k - 1$ folds and testing on the remaining fold, cycling through all folds to robustly assess model performance. We assume k=5.

These validation procedures provide a comprehensive evaluation of each variogram model's suitability for reliable spatial interpolation of teak tree volumes.

2.2.3. Rationale for modifications

The motivation for modifying the spherical variogram model arises from the need to better capture the spatial patterns observed in teak volume data over multiple years and configurations. While the classical spherical model yielded consistently low RMSE and high R^2 among standard models, it still showed residual trends near the range limit, suggesting limited flexibility in fitting real-world variability.

Two major factors guided the modification:

1) **Improved empirical fit**
   Although the spherical model performed well statistically, visual inspection of the empirical variograms revealed that it did not fully capture the smooth transitions in some years, especially near the cutoff distance.
2) **Greater flexibility in curve shape**
   The cubic extension adds shape-controlling coefficients (x_1, x_2, x_3), enabling the model to

better match diverse empirical trends. This is useful when spatial correlation decays non-uniformly or the effective range is unclear.

The modified model maintains key theoretical properties mentioned in Section 2.2.1, making it valid for geostatistical applications. It also includes the classical spherical model as a special case, ensuring compatibility with traditional approaches.

To estimate parameters, we use L-BFGS-B, a constrained optimization method that handles parameter bounds like non-negative nugget and range values. This makes the model more robust and adaptable for spatial modeling in complex environments such as forest plantations commonly studied in tropical regions.

2.3. Parameter estimation via constrained optimization

To estimate the parameters of the modified spherical variogram model, we use the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with box constraints (L-BFGS-B), a quasi-Newton method suited for nonlinear problems with bounded parameters [14,15]. This approach is implemented in R using the optim() function with method "L-BFGS-B".

The cubic component of the model is subject to the continuity at the range and smoothness constraints previously defined in Eqs (2.3) and (2.4). To ensure these conditions are always satisfied during optimization, the polynomial was reparameterised using a single free coefficient u, as defined earlier in Eq (2.5).

This reduces the parameter vector to

$$\theta = \{C_0, C, a, u\},$$

greatly simplifying estimation while guaranteeing mathematical validity of the variogram. Parameters are obtained by minimizing a weighted least-squares objective function:

$$\min_{\theta} Q(\theta) = \sum_{i=1}^{n} w(h_i)\,[\gamma_{emp}(h_i) - \gamma_{model}(h_i;\theta)]^2 \quad \text{subject to: } \theta_{min} \le \theta \le \theta_{max} \tag{2.9}$$

where $\theta = \{C_0, C, a, u\}$ is the parameter vector, h_i are lag distances, $\gamma_{emp}(h_i)$ is the empirical variogram value at lag h_i, $\gamma_{model}(h_i;\theta)$ is the theoretical value from the modified variogram, and $w(h_i)$ is the number of point pairs at lag h_i.

Box constraints ensure non-negativity for the nugget ($C_0 \ge 0$), partial sill ($C \ge 0$), and positivity for the range ($a > 0$). Since all equality constraints are satisfied analytically through the reparameterisation in Eq (2.5), no penalty terms were required. Initial parameter values and bounds were selected based on exploratory variogram inspection and physically interpretable limits.

This estimation framework guarantees that the resulting fitted model is smooth, continuous, and mathematically valid, while retaining the flexibility needed to capture empirical spatial structures.

2.4. Study area and data description

The study is conducted in a clonal teak (*Tectona grandis linn F.*) plantation managed by the Research and Development Division of Sabah Softwood Berhad at Brumas Camp, Tawau, Sabah, Malaysia. Established in 2002, the selected block (Block 96G) spans 5.67 hectares and is at 4°37′23.85″

N, 117°47′05.12″ E. The site is moderately sloped (elevation: 180–370 m), situated in a tropical rainforest climate (Köppen classification), and underlain by Tanjung Lipat soil, which is typically low in nitrogen, potassium, and magnesium.

Among the fifteen teak genotypes propagated via micropropagation, seven were Solomon Island-derived clones, which serve as the primary focus of this study. A randomized complete block design (RCBD) with four replications is implemented, each comprising two rows of 30 trees per genotype, spaced 4×4 meters apart resulting in a planting density of approximately 625 stems per hectare. Only the 11th to 20th trees in each row are sampled, giving 80 trees per genotype.

Although 1,200 trees are initially assessed for height and diameter at breast height (DBH), only the Solomon clones are included in the analysis due to its lucrative outcome. Volume is calculated using the following equation [16]:

$$V = \frac{1}{10}\left[\left[1.3\pi\left(\frac{D}{2}\right)^2\right] + \left[\pi\left(\frac{D}{2}\right)^2\left(\frac{(H-1.3)}{3}\right)\right]\right] \tag{2.10}$$

where V is the tree volume in cubic centimeters (cm^3), D is the diameter at breast height in centimeters, and H is the total tree height in meters. Sampling was conducted at years 6, 7, and 10. Table 2.1 summarizes the total and georeferenced samples. After excluding trees lost to undergrowth and mortality, 801 Solomon-clone trees are retained for analysis in years 6 and 7. Sample sizes decline by year 10 due to natural tree mortality. Tree locations are georeferenced using GPS coordinates recorded in the field and later converted into spatial points using the WGS84 coordinate system (EPSG:4326).

The focus on Solomon Island-derived clones is motivated by their consistently superior growth performance in earlier studies [17]. Topographic details of the site are illustrated in Figure 2.2, provided by the Sabah Softwood Berhad research team.

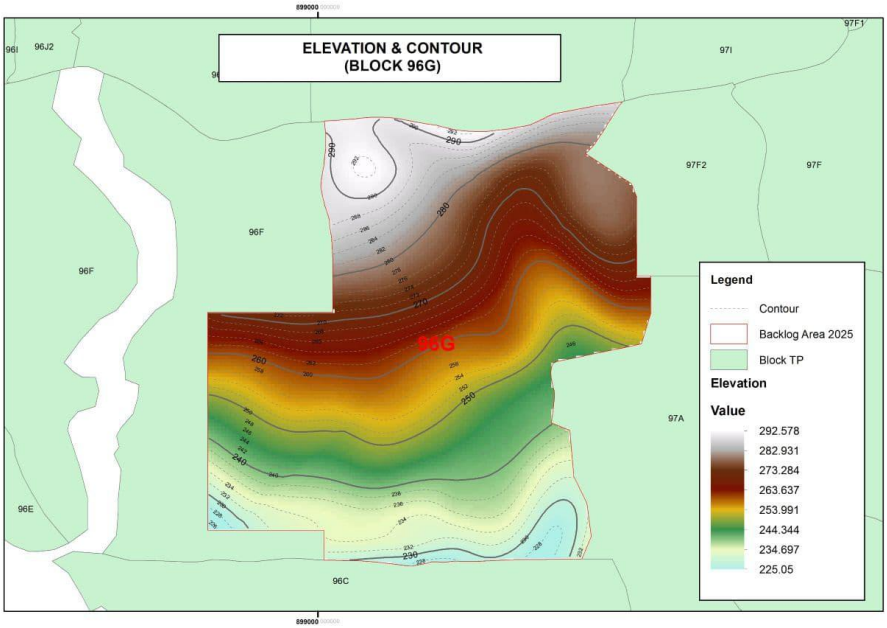


Figure 2.2. The topologic map of block 96G at Brumas Camp, Tawau, Sabah.

Table 2.1. Samples collected according to teak tree's age.

Teak Tree Age (years)	Sample size (n)	Georeferenced samples
6	451	432
7	451	445
10	354	354

2.5. Empirical variogram construction and model fitting

Variograms are essential tools in spatial statistics used to quantify how similarity between observations changes with distance. In forestry applications, they are especially useful for exploring spatial continuity in variables such as tree volume. The empirical variogram, computed from observed data, provides a non-parametric approximation of spatial autocorrelation. It is constructed using a fixed bin width of 0.0035 units, selected to balance resolution and pair count across distance classes. In this study, volume measurements of teak trees serve as the variable of interest, denoted as $Z(x_i)$ and the variogram value is estimated using:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{2.11}$$

where h is the lag distance and $N(h)$ is the number of point pairs at that lag. This bin-based approach offers exploratory insight into spatial patterns within the plantation block.

Once the empirical variogram is computed, valid theoretical models are fitted to ensure smoothness, continuity, and positive semi-definiteness, which are essential for kriging. Several classical models are evaluated, including spherical, exponential, Gaussian, wave, circular, and penta-spherical models. Each model is defined by three key parameters:

- **Nugget,** C_0: Variogram at zero distance, indicating measurement error or micro-scale variation.
- **Sill,** C: Plateau representing the limit of spatial dependence.
- **Range** a: The distance at which the variogram reaches the sill, beyond which spatial correlation diminishes.

These models are fitted to the empirical variogram using least squares, and their performance is evaluated based on visual fit and predictive accuracy. Table 2.2 summarizes the models evaluated in this study. This variogram modeling process lays the foundation for kriging-based spatial prediction, although kriging is not performed in this study. Instead, model validation is conducted via cross-validation techniques to assess the generalizability and robustness of each fitted model.

Table 2.2. Theoretical variogram models evaluated in this study.

Model	VARIOGRAM	
Exponential	$\gamma(h) = C_0 + C\left[1 - e^{\left(\frac{-h}{a}\right)}\right]$	if, $0 \le h \le a$. Otherwise, $C_0 + C$.
Spherical	$\gamma(h) = C_0 + C\left[\left(\frac{3}{2}\right)\left(\frac{h}{a}\right) - \left(\frac{1}{2}\right)\left(\frac{h}{a}\right)^3\right]$	if, $0 \le h \le a$. Otherwise, $C_0 + C$.
Gaussian	$\gamma(h) = C_0 + C\left[1 - e^{\left(-\left(\frac{h}{a}\right)^2\right)}\right]$	for all *h*.
Wave	$\gamma(h) = C_0 + C\left[1 - \frac{\sin\left(\pi\frac{h}{a}\right)}{\pi\frac{h}{a}}\right]$	for all *h*.
Circular	$\gamma(h) = C_0 + C\left[1 + \frac{2}{\pi}\left(\frac{h}{a}\sqrt{1 - \left(\frac{h}{a}\right)^2} - cos^{-1}\frac{h}{a}\right)\right]$	if, $0 \le h \le a$. Otherwise, $C_0 + C$.
Penta spherical	$\gamma(h) = C_0 + C\left[\left(\frac{15}{8}\right)\left(\frac{h}{a}\right) - \left(\frac{5}{4}\right)\left(\frac{h}{a}\right)^3 + \frac{3}{8}\left(\frac{h}{a}\right)^5\right]$	if, $0 \le h \le a$. Otherwise, $C_0 + C$.

3. Results

In this section, we present the outcomes of fitting classical and modified theoretical variogram models to the empirical variogram data derived from teak volume measurements. The results are organized by tree age, in years, starting from 6, 7, and 10, highlighting how well each model captures spatial dependence. Visual comparisons, parameter estimates, and goodness-of-fit metrics are provided to evaluate model performance. Special attention is given to the proposed modified spherical variogram, whose enhanced flexibility is assessed against standard models using cross-validation methods such as weighted RMSE and R^2.

3.1. Empirical variogram modeling and visual fit

Figures 3.1−3.3 present the empirical variograms overlaid with fitted curves for each of the seven theoretical models, including the proposed Modified Spherical Variogram, across the 6th, 7th, and 10th plantation years. Each model is fitted using least squares method based on binned variogram values. These visual comparisons highlight how well each model captures the spatial dependence structure across age stages, particularly around the range threshold.

As shown in the figures, none of the classical models captured the empirical variogram trends perfectly across all years. In the 6th year, the spherical and circular models provided a relatively good visual fit, while the wave model slightly underestimated variogram values at mid-lag distances. The modified spherical variogram model, included for comparison, closely followed the upward curvature across lags, showing improved alignment with empirical points. For the 7th year, the spatial structure appeared more linear, and most classical models struggled to reflect the steady rise in variogram value, particularly across transitional lag zones. Here too, the modified spherical variogram model demonstrated a better visual fit, especially in the mid-to-late lag range. In the 10th year, where a more pronounced increase in variogram was observed, the modified spherical variogram model again offered a smoother transition across lags, capturing the empirical trend more effectively in this later growth stage.

These inconsistencies highlight the limitations of classical variogram forms in capturing the spatial complexity of clonal teak stands at different growth stages. This amplifies the need to introduce

a modified spherical variogram model that is formulated to provide improved flexibility in gradient behavior and curve fitting.

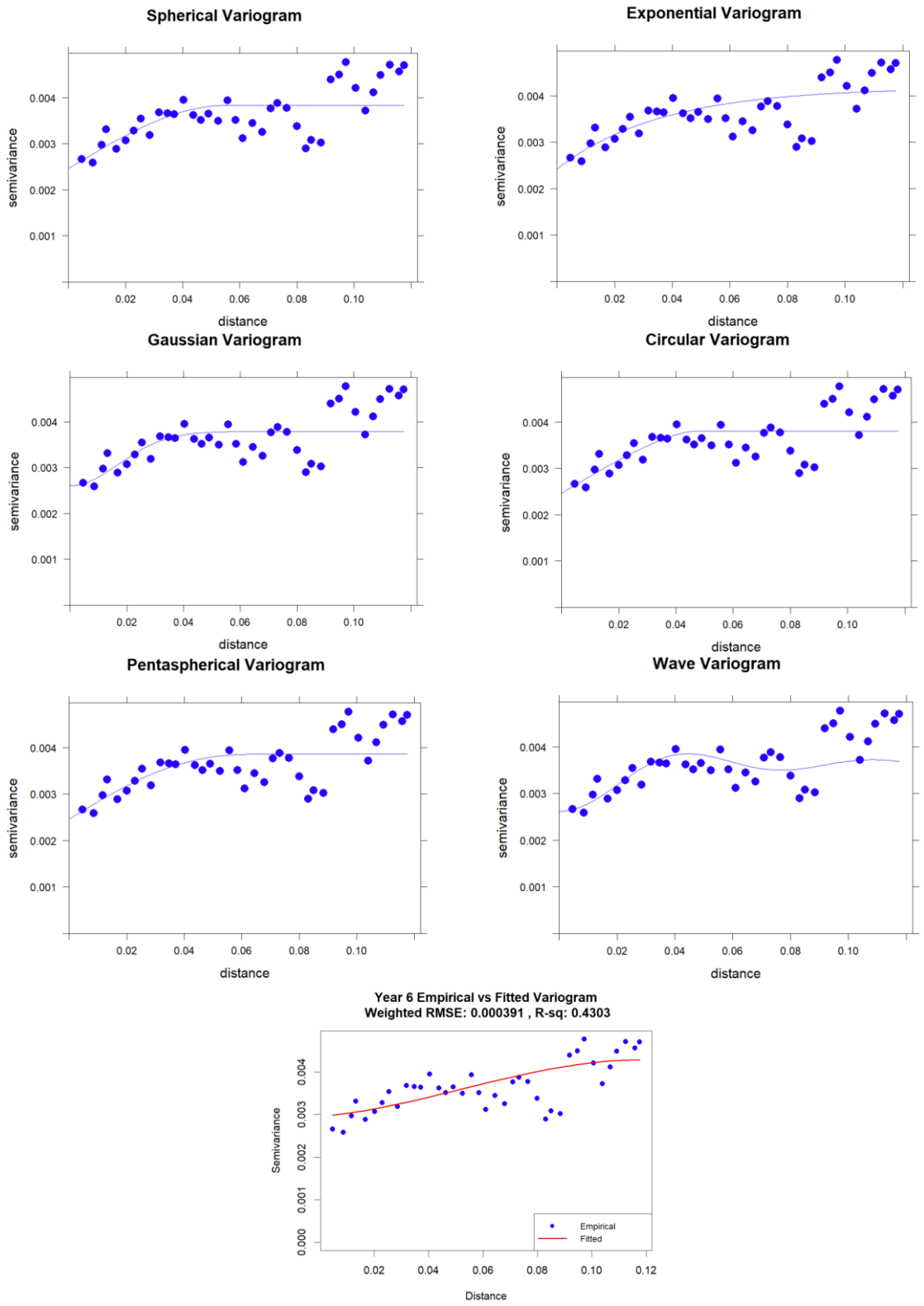


Figure 3.1. Fitted Variogram Models for Year 6.

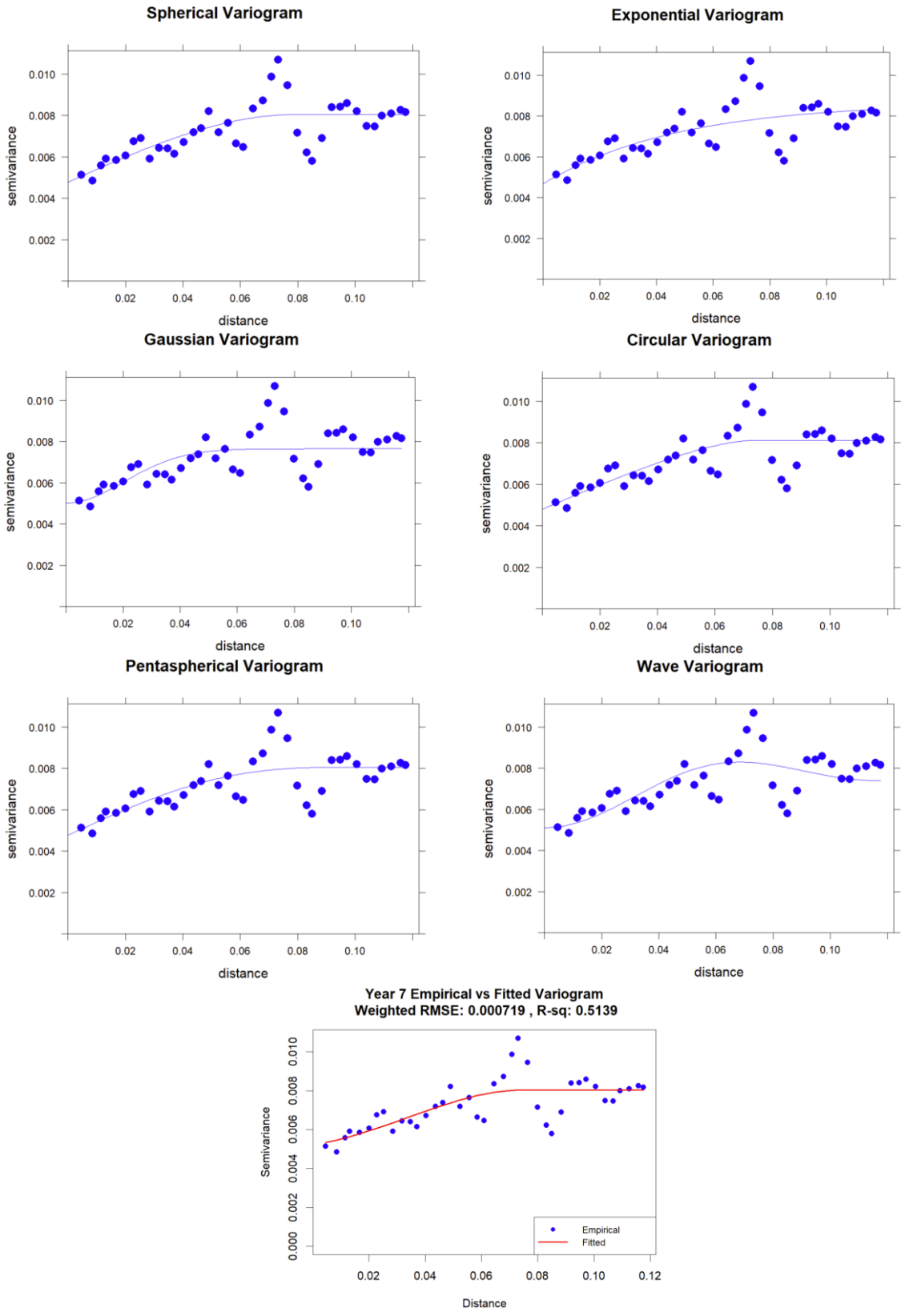


Figure 3.2. Fitted variogram models for Year 7.

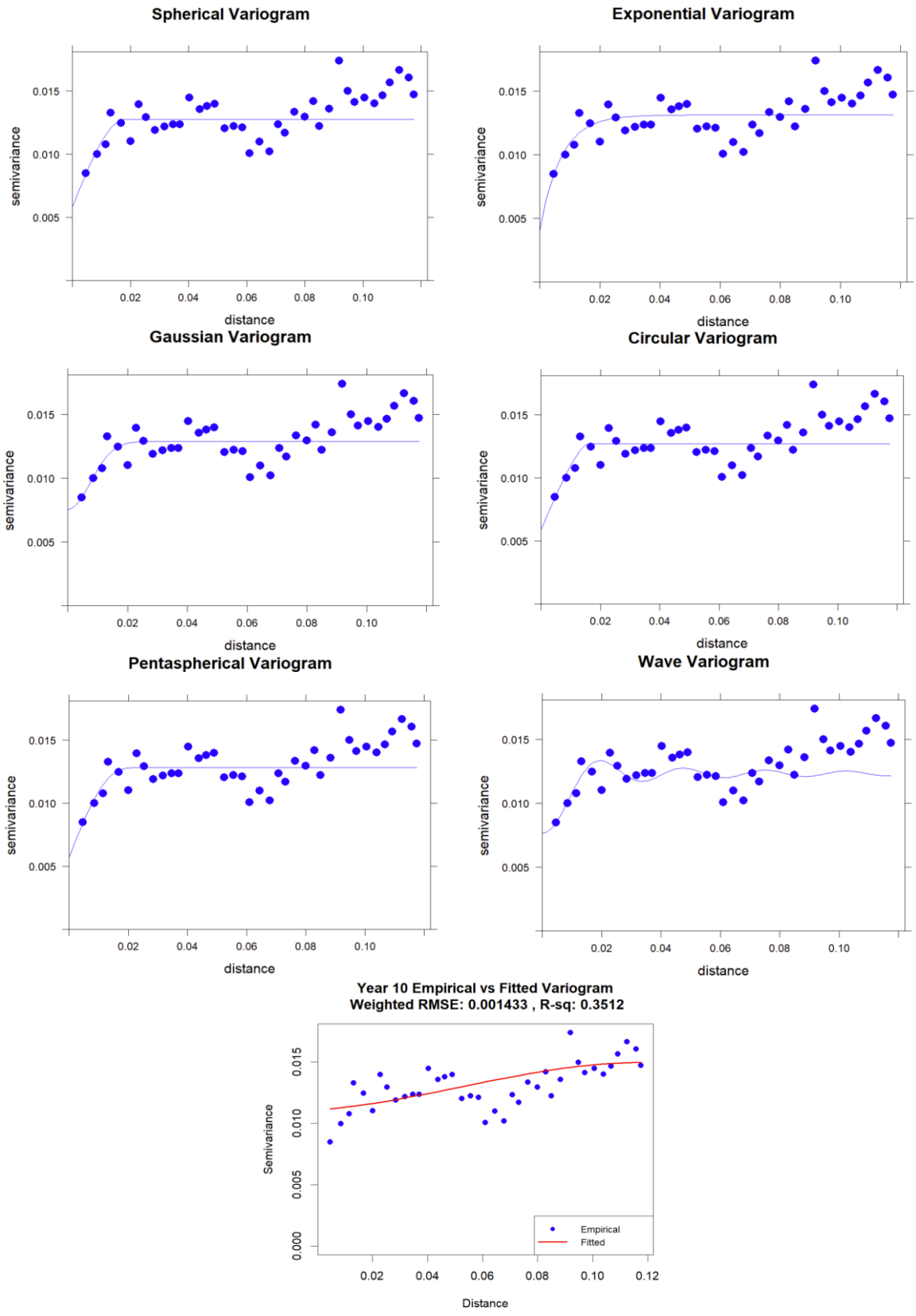


Figure 3.3. Fitted variogram models for Year 10.

3.2. Variogram model fitting and parameter estimates

All theoretical variogram models were fitted using the fit.variogram() function in gstat, and their predictive performances were objectively compared using K-fold and leave-one-out cross-validation (LOOCV). The resulting RMSE and R^2 values serve as numerical indicators of model discrepancy, providing an objective numerical evaluation comparable to automated routines such as autofitVariogram [18]. The modified spherical variogram was estimated using constrained weighted least squares (Section 2.3), ensuring exact satisfaction of all continuity and smoothness constraints. This approach ensures objective model comparison beyond visual assessment, particularly for evaluating the modified spherical variogram alongside the six classical models.

3.2.1. Parameter estimates for classical models

The first six models, Exponential, Spherical, Gaussian, Wave, Circular, and Pentaspherical, were fitted independently for each year. The estimated parameters (nugget, sill, and range) are summarized in Table 3.1, providing insight into the spatial structure captured by each model.

The estimated parameters reveal several interesting trends. Across all years, the Gaussian and wave models tend to produce shorter range estimates, while penta-spherical and spherical models show longer effective ranges. Nugget values were generally higher in the 10th year, indicating increased short-scale variability or measurement error as the trees matured. The exponential model showed relatively high partial sills in the 7th and 10th years, which may be a sign of overestimation of spatial structure near the origin. These differences will be further evaluated in terms of predictive accuracy and visual fit.

Table 3.1. Estimated parameters of theoretical variogram models (excluding Modified Spherical).

		Spherical	Exponential	Gaussian	Circular	Penta spherical	Wave
6^{th}	C_0	0.002462	0.002424	0.002605	0.002464	0.002460	0.002619
	C	0.001370	0.001754	0.001182	0.001337	0.001403	0.001012
	a	0.05506	0.03649	0.02411	0.04618	0.06966	0.03121
7^{th}	C_0	0.004775	0.004668	0.005016	0.004800	0.004762	0.005110
	C	0.003271	0.003988	0.002633	0.003306	0.003294	0.002622
	a	0.07905	0.04719	0.02879	0.07154	0.09657	0.04781
10^{th}	C_0	0.005790	0.004019	0.007559	0.005888	0.005715	0.007644
	C	0.006938	0.009097	0.005327	0.006798	0.007078	0.004669
	a	0.01811	0.007049	0.01069	0.01580	0.02225	0.01369

3.2.2. Parameter estimation for the modified spherical variogram model

To evaluate the performance of the proposed modified spherical variogram, its parameters are estimated for each year using the L-BFGS-B optimization algorithm, as introduced in Section 2.3. The model includes the nugget (C_0), partial sill (C), and range (a), together with a single shape parameter u. The remaining polynomial coefficients (x_1, x_2, x_3) are derived from u using the reparameterisation described in Section 2.3, which guarantees continuity and smoothness at the range while reducing the dimensionality of the optimization.

Table 3.2 presents the estimated parameter values for the modified spherical variogram model

across the 6th, 7th, and 10th plantation years. These results reflect the modified spherical variogram model's ability to adapt its shape to fit various empirical trends, particularly in lag intervals where classical models show limitations.

Table 3.2. Optimized parameters for the modified spherical variogram model.

Year	C_0	C	a	x_1	x_2	x_3
6th	0.00295499	0.00132806	0 .11749	0.499568	2.000864	-1.50043
7th	0.0052383	0.002804838	0.074983	0.500521	1.998959	-1.49948
10th	0.01108929	0.00386093	0.117476	0.499614	2.000771	-1.50039

The modified spherical variogram estimates demonstrate several key advantages:

- **Range flexibility**: Parameter a adapts to the extent of spatial influence, with values consistent yet often smoother than those from classical models.
- **Curvature control**: The cubic terms (x_1, x_2, x_3) enable nuanced adjustments to model the variogram's rate of increase, particularly beneficial in transition zones.
- **Backward compatibility**: In cases where $u = 0$, the modified spherical variogram simplifies to the classical spherical form, reinforcing its generalization capability.

These parameter results further support the visual evidence discussed in Section 3.1, where the modified spherical variogram model consistently produced a better visual fit to the empirical data. The model's adaptability is especially relevant for teak clones where spatial variability evolves with plantation age.

3.3. Cross-validation and predictive accuracy

Cross-validation is performed to evaluate the predictive performance of all seven variogram models across the 6th, 7th, and 10th plantation years. Both 5-fold and LOOCV are conducted to assess the stability and generalizability of each model's predictions. These metrics provide an objective complement to the visual fits described in Section 3.1.

3.3.1. Comparative performance of all models

Table 3.3 summarizes the weighted RMSE and R^2 for each of the seven theoretical models across the three plantation ages under full-data fitting, K-fold CV, and LOOCV. The results highlight several consistent patterns:

6th year: The modified spherical variogram achieved the lowest weighted RMSE and highest R^2 across all validation. This indicates strong predictive accuracy and robustness, and consistent superiority in full-data, K-fold CV, and LOOCV.

7th year: Performance in Year 7 showed a more nuanced pattern where the circular model had the lowest RMSE and highest R^2 under full-dataset fitting, slightly outperforming the modified spherical variogram. However, in K-fold CV and LOOCV, the modified spherical variogram achieved the lowest RMSE and the highest R^2, indicating better generalizability and predictive robustness. Several classical models produced tightly clustered RMSE values, reflecting the relatively linear variogram shape in Year 7.

10th year: The modified spherical variogram again achieved the lowest RMSE and the highest R^2 in full-dataset fitting, K-fold CV, and LOOCV. However, the differences among models were small,

and all R^2 values were lower in general. This reflects the greater biological variability and stand heterogeneity typical of mature teak plantations. In such conditions, predicting volume became inherently more difficult, and even well-behaved models yielded modest R^2 values. Despite this, the modified spherical variogram provided competitive or superior error metrics and maintained a smooth, stable variogram form, supporting its robustness in more complex spatial structures.

Overall, the modified spherical variogram model showed consistent and reliable performance across all validation frameworks, outperforming classical models in younger stands and remaining competitive in later growth stages.

3.3.2. Visual assessment of prediction residuals

To further validate the numerical metrics, visual diagnostics provide critical insight into the quality and bias of spatial predictions. Two primary forms of graphical residual analyses are performed for the modified spherical variogram model across the three plantation years:

a) Histograms of prediction errors

Figures 3.4 displays the distribution of prediction errors (observed minus predicted values) for the 6th, 7th, and 10th years. These histograms help assess the bias and spread of residuals:

- **Year 6**: The histogram was tightly centered around zero with a symmetrical shape, indicating accurate and unbiased predictions.
- **Year 7**: The residuals exhibited a more peaked and concentrated distribution, indicating improved prediction stability and reduced variance.
- **Year 10**: The distribution showed a slight right skew, implying minor overestimation for certain observations, particularly at higher lags. However, the predictions remained acceptably centered.

These plots collectively affirmed the modified spherical variogram model's low-bias prediction behavior and validated its reliability across growth stages.

b) Residuals vs. predicted variogram

Figure 3.5 illustrates residuals plotted against predicted variogram values for each year:

- **Year 6**: Residuals were narrowly distributed around zero with no visible pattern, indicating a well-fitted model and homoscedastic behavior.
- **Year 7**: A more subtle spread was observed, with mild asymmetry across the prediction range, suggesting slight heteroscedasticity likely due to structural complexities in spatial patterning. **Year 10**: Residuals exhibited a broader vertical spread, especially at a lower predicted variogram, reflecting increased biological variability and possible underfitting in mature stand conditions.

Together, these visual assessments reinforce the modified spherical variogram model's predictive strength while highlighting areas where predictive variance grows with plantation age.

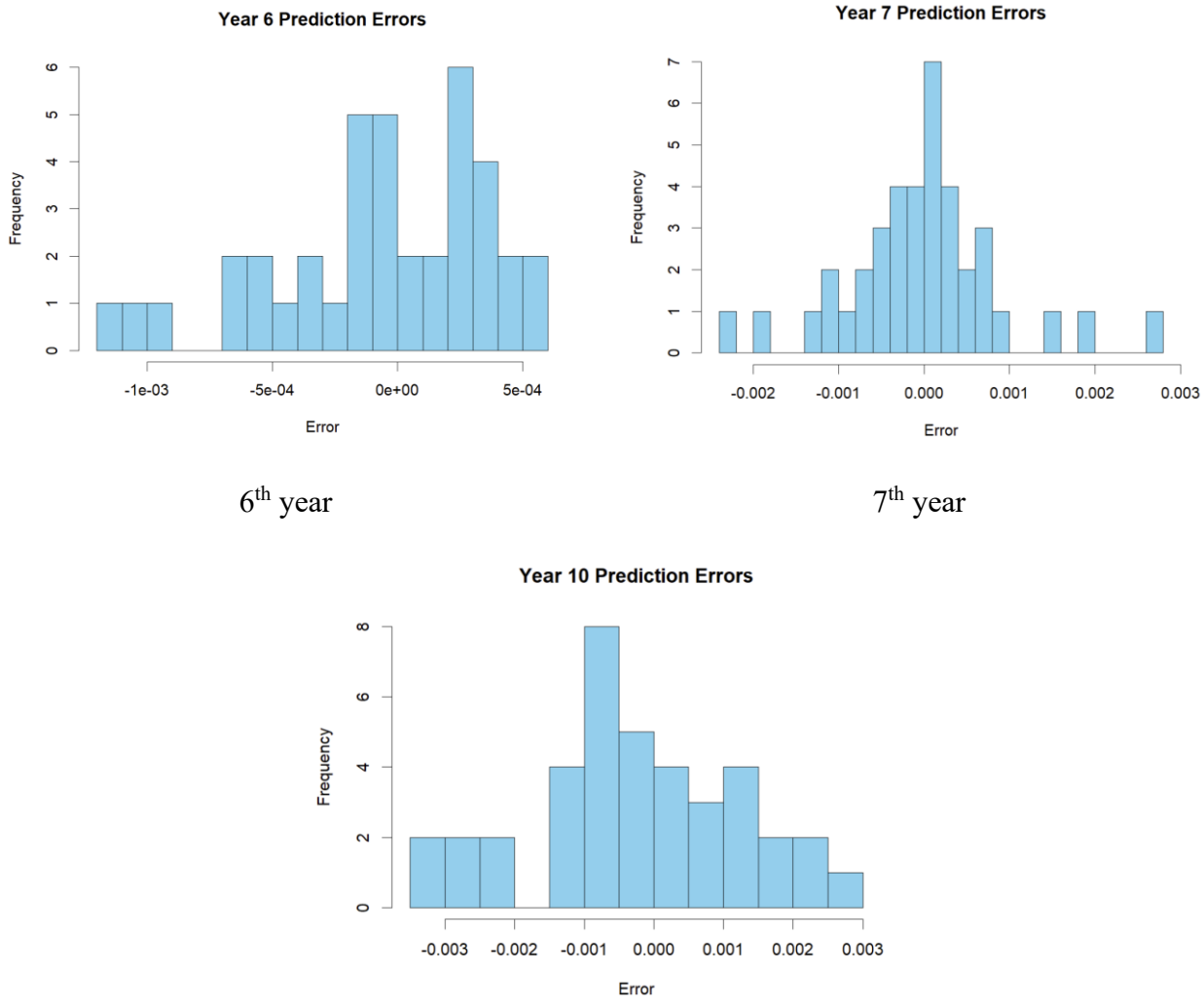
3.4. Summary of model comparison and selection

The results presented across Sections 3.1−3.3 provide a robust comparative evaluation between six classical variogram models and the newly proposed modified spherical variogram model. Visual assessments (Figures 3.1−3.3) showed that the modified spherical variogram model consistently offered better alignment with empirical variogram trends, especially across transition lags where classical models tended to underfit or misrepresent curvature.

Quantitative diagnostics reinforced these visual findings. As seen in Table 3.3, the modified

spherical variogram model produced lower weighted RMSE values and competitive R^2 across all three validation methods, in full dataset, k-fold, and leave-one-out cross-validation, demonstrating improved spatial prediction capability in most years. While the model did not yield the lowest RMSE or highest R^2 for the full-dataset fit in Year 7, it outperformed all classical models under K-fold and LOOCV, indicating stronger generalization. Its performance remained stable and close to the best-performing classical models. Residual histograms in Figure 3.4 further confirmed the model's low bias and stability, while scatterplots of residuals vs. predicted variogram in Figure 3.5 showed no discernible pattern, affirming homoscedastic behavior in most cases.

These collective results affirm the modified spherical variogram model's superior adaptability and predictive reliability for spatial modeling of clonal teak tree volume across growth stages. Therefore, the Modified Spherical Variogram is recommended as the preferred theoretical model in this context.

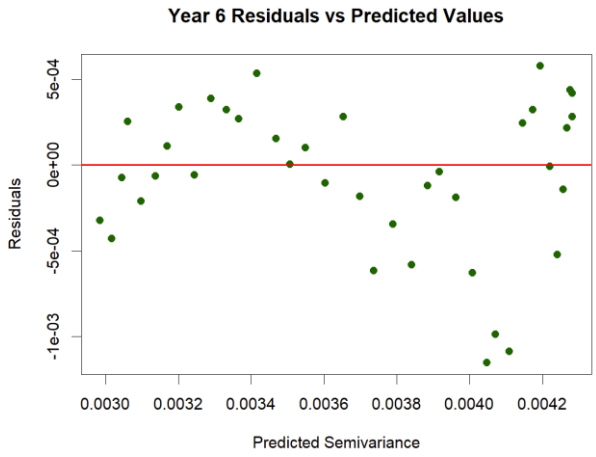


6^{th} year

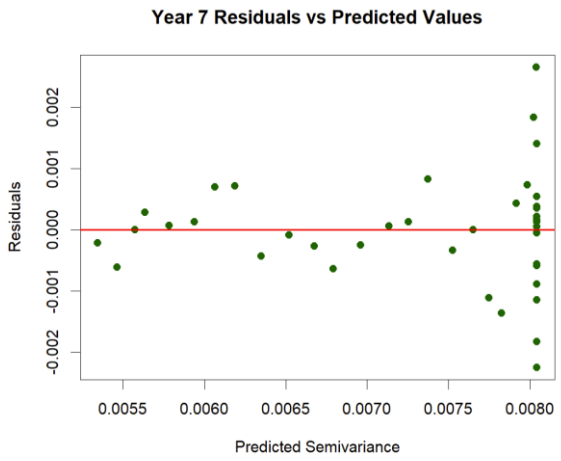
7^{th} year

10^{th} year

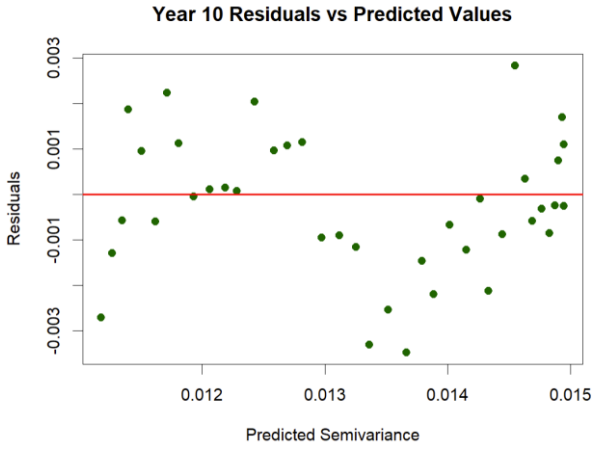
Figure 3.4. Prediction errors for the modified spherical variogram model for the 6^{th}, 7^{th}, and 10^{th} years.



6^{th} year



7^{th} year



10^{th} year

Figure 3.5. Residuals versus predicted variogram values for the modified spherical variogram model for the 6^{th}, 7^{th}, and 10^{th} years.

Table 3.3. Comparative cross-validation results across all models.

			Spherical	Exponential	Gaussian	Circular	Penta spherical	Wave	Modified Spherical
6th Year	Full Dataset	RMSE	0.00047522	0.00040587	0.0004875	0.00048356	0.00046701	0.00051333	0.000391031
		R^2	0. 3490647	0. 4226118	0.342744	0.3442972	0.3539273	0.3263043	0.4302945
	K-Fold	RMSE	0.00049196	0.00049214	0.0005056	0.0004880	0.0004958	0.0004850869	0.0004579814
		R^2	0.2709241	0.2703942	0.2300092	0.2825573	0.2594158	0.2913492	0.3681702
	LOOCV	RMSE	0.00048572	0.00045872	0.0004821	0.0004852	0.0004865	0.0004845	0.0004518318
		R^2	0.2893174	0.3661062	0.2998801	0.2908894	0.2871337	0.2928784	0.385024
7th Year	Full Dataset	RMSE	0.0007201	0.0007517	0.0007898	0.0007176	0.00072627	0.0007955	0.0007189893
		R^2	0.5110018	0.4814087	0.4532553	0.5142804	0.5029523	0.4616261	0.5139201
	K-Fold	RMSE	0.00093977	0.0009666	0.0010964	0.0009475	0.00094476	0.0010192	0.0009216762
		R^2	0.4624117	0.4312217	0.2682273	0.4535296	0.4566888	0.3676417	0.4829209
	LOOCV	RMSE	0.00092905	0.00095478	0.0009882	0.00092988	0.00093616	0.00101014	0.0009268946
		R^2	0.4746044	0.4451006	0.4055578	0.4736724	0.466535	0.3788961	0.4770491
10th Year	Full Dataset	RMSE	0.001922	0.001772	0.001857	0.001939	0.001894	0.002150	0.001433352
		R^2	0.177413	0.262586	0.218637	0.1660478	0.1952701	0.0050117	0.3512021
	K-Fold	RMSE	0.00187148	0.001751824	0.0018992	0.00184962	0.00186196	0.002012513	0.001657315
		R^2	-0.0001627	0.123647732	-0.029983	0.02306437	0.0099966	-0.156578022	0.2156538
	LOOCV	RMSE	0.00177955	0.001667226	0.0017407	0.00176812	0.00176801	0.001924710	0.001577751
		R^2	0.09568683	0.20624462	0.134726	0.10727241	0.10737581	-0.05785958	0.2891558

4. Discussion

In this section, we analyze the results from variogram fitting, parameter estimation, cross-validation, and residual analysis, interpreting the performance of the modified spherical variogram model against classical models. While several classical models showed reasonable fits in some years, none provided consistent accuracy across all plantation stages. The modified spherical variogram model demonstrated competitive full-dataset fits and superior cross-validated performance in most cases, highlighting its flexibility and robustness.

4.1. Model performance across plantation years

The spatial structure of teak clones evolved with plantation age, presenting different modeling challenges. In the 6th year**,** classical spherical and circular models gave visually acceptable fits, but the modified spherical variogram model captured mid-lag curvature better, as evidenced by lower weighted RMSE and tighter residual histograms. By the 7th year, variogram value trends became more linear, and although the modified model did not produce the best full-dataset RMSE or R^2, it achieved the strongest predictive accuracy under K-fold and LOOCV, outperforming all classical models. In the 10th year, some classical models obtained slightly higher R^2 in cross-validation, yet showed increased spread in errors and less stable variogram prediction. The modified spherical variogram model retained smooth curve fitting and balanced residual distribution, offering the most stable performance across all years.

4.2. Strength of the modified spherical variogram

The modified spherical variogram model's strength lies in its flexible cubic structure, which accommodates empirical variogram shapes more precisely than rigid classical forms. The coefficients x_1, x_2, x_3 enable a more detailed control of the variogram curvature, especially at the sill boundary and near the origin. Its performance improvement is not only numerical but also visual, clearly shown in residual plots and error histograms.

Moreover, the bounded optimization using L-BFGS-B ensured parameter constraints were respected, maintaining theoretical validity (e.g., symmetry, non-negativity, sill attainment, and positive definiteness), as mentioned in Section 2.2.1.

4.3. Implications for spatial prediction

Given that kriging predictions depend critically on the quality of the variogram model, the improved accuracy and low bias of modified spherical variogram model have direct implications for practical forest management. Its consistently strong cross-validated performance indicates reliable generalization, which is particularly valuable for operational decision-making. The ability to capture spatial variability more faithfully enables better volume estimation, planning, and clone performance monitoring in plantations. The model's generalizability also suggests applicability in other geospatial domains beyond forestry.

5. Conclusions

We introduced a modified spherical variogram model designed to improve the representation of

spatial dependence structures in clonal teak plantations. By integrating a flexible cubic formulation with constrained optimization, the modified spherical variogram model achieved optimal visual alignment with empirical variograms and competitive or superior predictive performance in classical models in weighted RMSE and R^2 metrics across three plantation years.

These results demonstrate that the modified spherical variogram model effectively adapts to varying spatial patterns associated with tree growth stages. Even in cases where classical models briefly outperform it in isolated validation metrics, the modified spherical variogram model provides the most stable, consistent, and theoretically valid performance across all years studied. This highlights the model's superior generalizability rather than overfitting to any single dataset.

Its mathematical validity and predictive strength make it a robust alternative for spatial interpolation tasks, particularly in complex or evolving environments. In future work, researchers may explore integration with kriging for spatial mapping and assess generalizability across species or geographic settings.

Author contributions

Johannah Jamalul Kiram: Conceptualization, methodology, data analysis, software, investigation, writing original draft; Rossita Mohamad Yunus: Conceptualization, methodology, writing review and editing, supervision; Yani Japarudin and Mahadir Lapammu: Plantation data, resources and data management; Olivier Montuuis and Doreen K. S. Goh: Plantation data and logical reasoning from data finding. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

During the preparation of this work, the author used OpenAI's ChatGPT (GPT-4) to enhance the quality and clarity of the English language. All technical content, mathematical formulations, and interpretations were developed solely by the authors. The final manuscript was reviewed and edited to ensure accuracy, and the authors accept full responsibility for the content and conclusions presented.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. S. D. Iaco, M. Palma, D. Posa, *Geostatistics and the role of variogram in time series analysis: A critical review*, Statistical Methods for Spatial Planning and Monitoring, Milano, Springer, 2013, 47–75. https://doi.org/10.1007/978-88-470-2751-0_3
2. M. A. Oliver, R. Webster, *The variogram and modelling*, Basic Steps in Geostatistics: The Variogram and Kriging, Cham, Springer International Publishing, 2015, 15–42. https://doi.org/10.1007/978-3-319-15865-5_3
3. Z. Arétouyap, P. N. Nouck, R. Nouayou, F. E. G. Kemgang, A. D. P. Toko, J. Asfahani, Lessening the adverse effect of the semivariogram model selection on an interpolative survey using kriging technique, *SpringerPlus*, **5** (2016), 549. https://doi.org/10.1186/s40064-016-2142-4
4. M. Tomasetto, E. Arnone, L. M. Sangalli, Modeling anisotropy and non-stationarity through physics-informed spatial regression, *Environmetrics*, **35** (2024), e2889.

5. M. J. Pyrcz, C. V. Deutsch, *Improved variogram models for more realistic estimation and simulation*, Centre for Computational Geostatistics, University of Alberta, 2003.
6. A. O. Finley, Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence, *Methods Ecol. Evol.*, **2** (2011), 143–154. https://doi.org/10.1111/j.2041-210X.2010.00060.x
7. S. Reda, S. R. Nassif, Accurate spatial estimation and decomposition techniques for variability characterization, *IEEE T. Semiconduct. M.*, **23** (2010), 345–357. https://doi.org/10.1109/TSM.2010.2051752
8. G. A. Qadir, Y. Sun, S. Kurtek, Estimation of spatial deformation for nonstationary processes via variogram alignment, *Technometrics*, **63** (2021), 548–561. https://doi.org/10.1080/00401706.2021.1883481
9. O. Perrin, P. Monestiez, *Modelling of non-stationary spatial structure using parametric radial basis deformations*, Geostatistics for Environmental Applications, Dordrecht, Netherlands, Springer, 1999, 175–186. https://doi.org/10.1007/978-94-015-9297-0_15
10. Y. Ling, Y. Shi, H. Hou, L. Pan, H. Chen, P. Liang, et al., Enhanced dung beetle optimizer for Kriging-assisted time-varying reliability analysis, *AIMS Math.*, **9** (2024), 29296–29332. https://doi.org/10.3934/math.20241420
11. L. Fang, Smooth digital terrain modeling in irregular domains using finite element thin plate splines and adaptive refinement, *AIMS Math.*, **9** (2024), 30015–30042. https://doi.org/10.3934/math.20241450
12. I. Clark, Statistics or geostatistics? Sampling error or nugget effect, *J. S. Afr. I. Min. Metall.*, **110** (2010), 307–312.
13. N. Cressie, *Geostatistical analysis of spatial data*, Spatial statistics and digital image analysis, National Academy Press Washington, DC, USA, 1991, 87–108.
14. R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.*, **16** (1995), 1190–1208. https://doi.org/10.1137/0916069
15. J. Nocedal, S. J. Wright, *Numerical optimization*, Springer, 1999.
16. O. Monteuuis, D. K. S. Goh, C. Garcia, D. Alloysius, J. Gidiman, R. Bacilieri, et al., Genetic variation of growth and tree quality traits among 42 diverse genetic origins of Tectona grandis planted under humid tropical conditions in Sabah, East Malaysia, *Tree Genet. Genomes*, **7** (2011), 1263–1275. https://doi.org/10.1007/s11295-011-0411-5
17. D. K. S. Goh, Y. Japarudin, A. Alwi, M. Lapammu, A. Flori, O. Monteuuis, Growth differences and genetic parameter estimate of 15 teak (Tectona grandis Lf) genotypes of various ages clonally propagated by microcuttings and planted under humid tropical conditions, *Silvae Genet.*, **62** (2013), 196–206.
18. P. H. Hiemstra, E. J. Pebesma, C. J. W. Twenhöfel, G. B. M. Heuvelink, Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network, *Comput. Geosci.*, **35** (2009), 1711–1721. https://doi.org/10.1016/j.cageo.2008.10.011