KEMENTERIAN EKONOMI
JABATAN PERANGKAAN MALAYSIA

Comparison analysis for variogram models of Teak stands volume specific to Solomon-clone in Tawau, Sabah Malaysia; Effect of bin width.

Johannah Jamalul Kiram[1,2], Rossita Mohd Yunus[2], Yani Japarudin[3], Mahadir Lapammu[3], Oliver Monteuuis[4], and Doreen Kim Soh 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

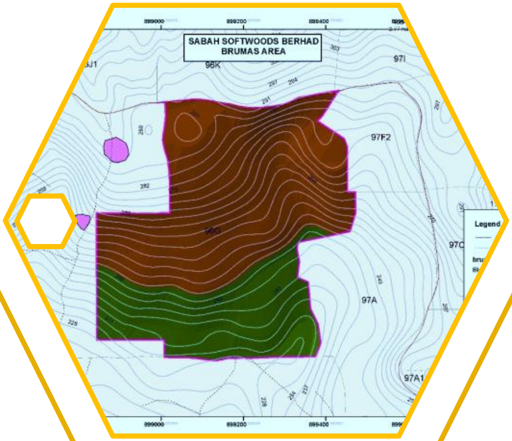
PERSIDANGAN STATISTIK KALI KE-11
"Data dan Kecerdasan Buatan: Memperkasa Masa Depan"
19 September 2024
Dianjurkan oleh:
KEMENTERIAN EKONOMI JABATAN PERANGKAAN MALAYSIA
BANK NEGARA MALAYSIA
MALAYSIA INSTITUTE OF STATISTICS
MyStats 2024

CONTENTS

INTRODUCTION

Scientifically known as *tectona grandis*, planted in Brumas Camp, Tawau, Sabah Malaysia, particularly Solomon Island-derived clones are proven to thrive best in this research of teak trees.

This study focuses on determining what is the best bin width to obtain the best variogram model of the volume of the teak trees.



A variogram's accuracy depends on an adequate amount of data that suits its density, not how large the sample size is, especially if the sample collected does not vary in lag interval.

based on research in 1994, instigated by Innoprise Corporation Sdn Bhd (ICSB) in an investment for mass cloning of teak trees, partnering with the CIRAD Forestry Department (D. K. Goh & Galiana, 2000).

INTRODUCTION

Objective: to determine the correct bin width to obtain the best variogram model for the volume of tectona grandis, specific to the Solomon Island-derived clone.

1. Identifying relationship between physical parameters and spatial information (previously been executed; Kiram et al., 2022, 2023)
2. Examine the effect of bin widths using existing experimental variogram models; Exponential model.
3. Models are then graphed for comparison, calculate root mean square error, cross validation.

METHODOLOGY:

DATA:

one block dedicated to *tectona grandis* that is managed by the research and development team of Sabah Softwood Berhad, at Brumas camp, Tawau, Sabah, Malaysia

observed throughout the span of 12 years, located on the coordinates 4°37’23.85”N and 117°47’05.12” E

randomized complete block with four contiguous replications, comprised two rows each of 30 plants of the 15 different genotypes

spaced 4 x 4 m with 625 stems per hectare, resulting in over 4000 trees

Assessed rows were only the 11th to 20th plants of each row, corresponding to 80 plants per clone in all, 1200 trees were collected,

but considering only Solomon island-derived clone, samples collected are 451 trees → 432 georeferenced individual tree points obtained for the 6th year plot

METHODOLOGY:

EXPERIMENTAL VARIOGRAM:

Table 1. Empirical variogram and theoretical variogram.

Model	VARIOGRAM	
Empirical	$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$	
Theoretical-Exponential	$\gamma(h) = C_0 + C\left[1 - e^{\left(\frac{-h}{a}\right)}\right]$	if, $0 \leq h \leq a$. Otherwise, $C_0 + C$.

- The bin-width that are being compared are 555 metres apart, 333 metres apart, 277.5 metres apart and 222 metres apart. This is using latitude conversion where 1 degree of latitude is approximately 111 kilometers, thus using bin-widths of 0.005, 0.003, 0.0025 and 0.002 respectively on the Rstudio command for fit.variogram under the package ‘gstat’.
- Cross validation is then carried out to ensure reliability and the variogram represents the spatial structure better, graphical observations are done, and the root mean square error(RMSE) of each model is calculated.

RESULTS:

- Figures shows the fitted Exponential variogram to the empirical variogram with different bin widths.
- Graph with the bin width of 0.002 in Figure 2d shows the smoothest fit.
- The bin width of 0.005 in figure 2a shows clear and concise fittings.

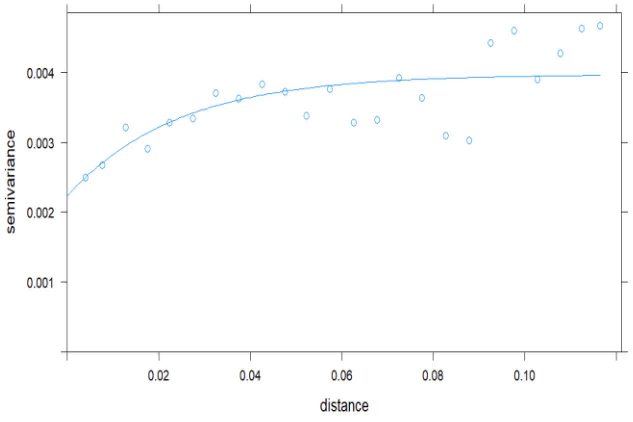


Fig. 2a. Bin-width 0.005

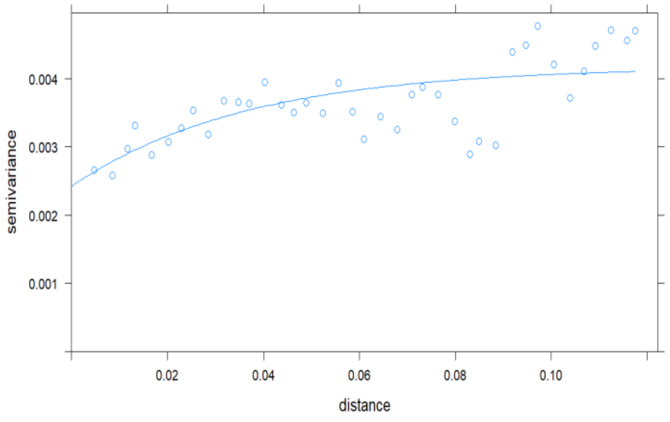


Fig. 2b. Bin-width 0.003

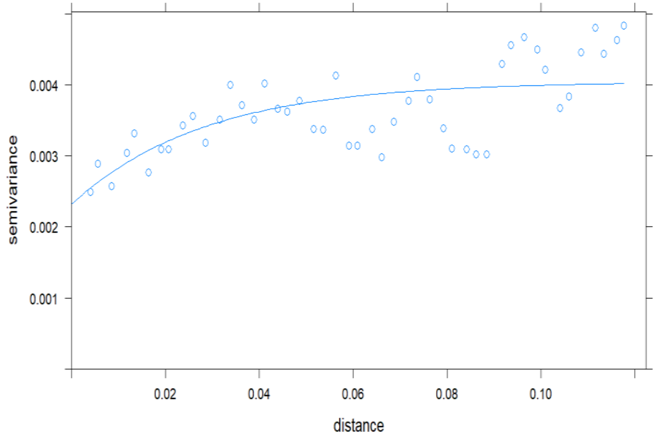


Fig. 2c. Bin-width 0.0025

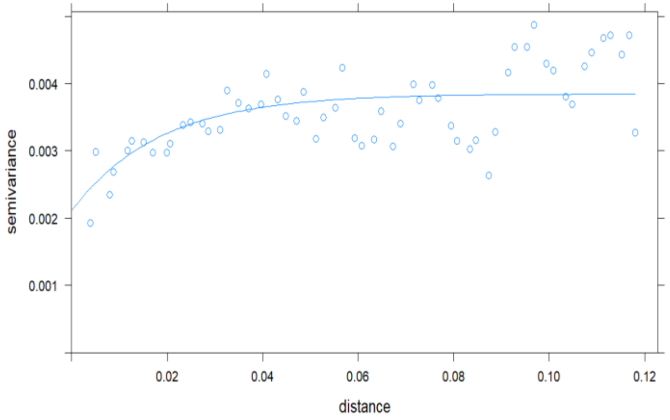


Fig. 2d. Bin-width 0.002

RESULTS:

- Cross-validation : The prediction errors for all four different bin widths are relatively symmetric around zero, which indicates balanced errors without systematic bias.
- All four graphs suggest small errors overall, due to the narrow histograms.
- Close observations do show that the histogram for bin-width 0.003 in figure 3b has a slightly higher concentration of errors around zero, which implies slightly better overall prediction accuracy, indicating that it is the more accurate model.

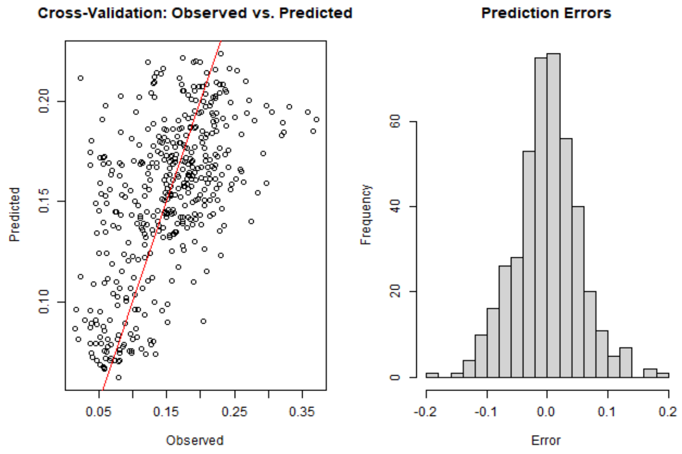


Fig. 3a. Cross-validation of model with bin-width 0.005

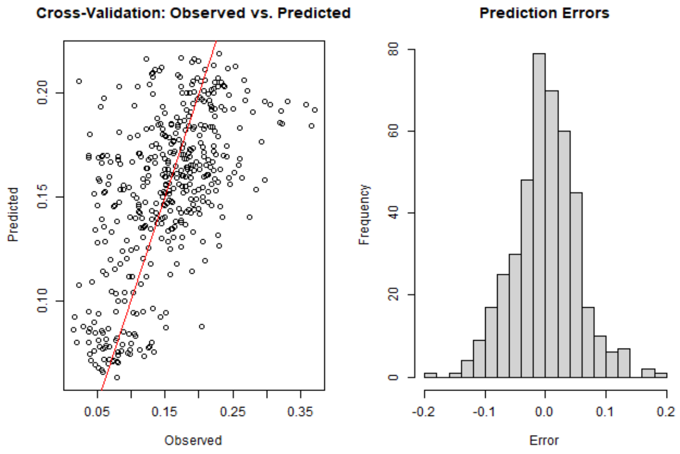


Fig. 3b. Cross-validation of model with bin-width 0.003

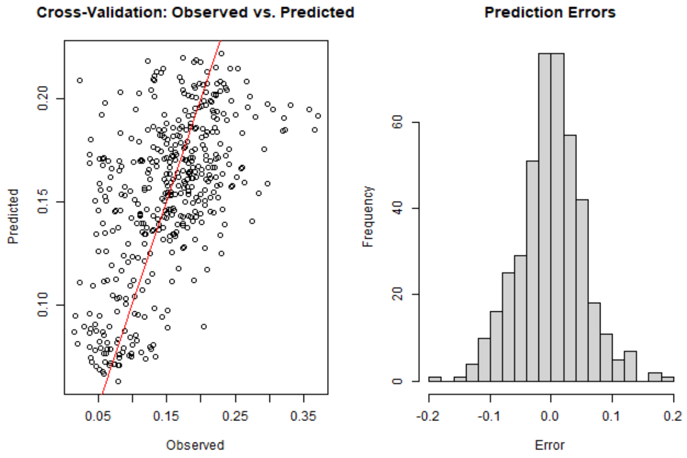


Fig. 3c. Cross-validation of model with bin-width 0.0025

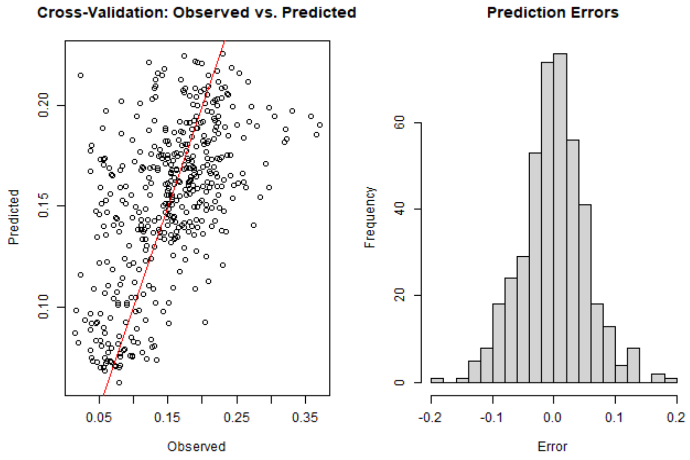


Fig. 3d. Cross-validation of model with bin-width 0.002

RESULTS:

Table 2. Calculated RMSE, Nugget, psill and Range for all models

Bin Width	pairs	RMSE	Nugget	Psill	Range
0.005	24	0.05383854	0.002230198	0.001740518	0.02396327
0.003	39	0.05361541	0.002424124	0.001753697	0.03648523
0.0025	47	0.05372381	0.002320105	0.001727215	0.02840659
0.002	59	0.05404312	0.002111044	0.001736450	0.01845881

- The final calculations of RMSE, nugget, psill and range is shown in Table 2, suggest that the most accurate model is the model with bin width 0.003 with the smallest RMSE calculated of 0.05361541. Thus, making the exponential model of bin width 0.003 (333metres apart) to be the best model compared to the others.

CONCLUSIONS:

- proves that the decision to choose the correct bin width to predict a model highly affects its accuracy.
- Comparison with different bin width must be made first before proceeding with selecting a model. In this study, the best bin width is 0.003, approximately 333 meters apart in lag distance between the tree pairs, which yielded 39 pairs of trees; least amount of error and its prediction errors graphed suggested higher concentration around zero, which indicates better prediction accuracy.
- While most statistical models suggest that the bigger the sample size, the better the model will be, it appears different when it comes to spatial modelling.
- Bin width and lag distance plays a huge role in determining the reliability and accuracy of the spatial model. If the data is sampled in every direction possible, with a variety of lag distances enough to cover the whole research site, then this may be the case.
- However, when it comes to spatial data, it could be costly, difficult and may even be dangerous to ensure that the sample collected is wide enough to cover what is needed.
- Thus, amongst reason why spatial modelling is also crucial is to be able to predict without having the need to go through difficult and dangerous sampling tasks especially when it involves data that are within a forest area where it is the habitat for so much wildlife.

Thank you