

ỨNG DỤNG PHƯƠNG PHÁP THẨM ĐỊNH CHÉO ĐỂ XÂY DỰNG MÔ HÌNH THỂ TÍCH CÂY ĐÚNG MỘT SỐ LOÀI CÂY ƯU THẾ RỪNG KHỘP TẠI TỈNH ĐẮK LẮK

Nguyễn Thanh Tân¹

Ngày nhận bài: 29/9/2025; Ngày phản biện thông qua: 23/10/2025; Ngày duyệt đăng: 25/10/2025

TÓM TẮT

Mục tiêu của nghiên cứu này là xây dựng phương trình thể tích cây đứng tối ưu cho ba loài cây ưu thế rừng khộp tại tỉnh Đắk Lắk gồm Chiêu liêu (*Terminalia corticosa* Pierre), Dầu đồng (*Dipterocarpus tuberculatus* Roxb) và Dầu trà beng (*Dipterocarpus obtusifolius* Teijsm. e. Miq). Số liệu được thu thập từ 156 cây giải tích thuộc rừng khộp tại tỉnh Đắk Lắk. Phương pháp thẩm định chéo Monte Carlo với 200 lần lặp được sử dụng, cho phép đánh giá và lựa chọn mô hình tối ưu một cách khách quan và ổn định. 05 dạng phương trình thể tích thường dùng được kiểm định, so sánh dựa trên các chỉ tiêu R^2 hiệu chỉnh (R^2_{adj}), Akaike Information Criterion (AIC), Bias%, RMSE% và MAPE%. Kết quả cho thấy phương trình dạng $V = a \times (D_{1,3}^{2,2} \times H)^b$ đạt độ chính xác cao nhất đối với cả ba loài nghiên cứu. Tham số của phương trình tối ưu được ước lượng từ toàn bộ số liệu, với $R^2_{adj} > 0,96$, phần dư phân bố đều phân bố quanh giá trị dự báo và có sự phù hợp chặt chẽ giữa giá trị quan sát và giá trị dự đoán. Kết quả của nghiên cứu khẳng định tính hiệu quả của phương pháp thẩm định chéo Monte Carlo trong điều kiện dữ liệu hạn chế, đồng thời cung cấp công cụ đáng tin cậy phục vụ điều tra và quản lý rừng khộp tại tỉnh Đắk Lắk.

Từ khóa: Phương trình thể tích, rừng khộp, thẩm định chéo Monte Carlo, mô hình hóa.

1. MỞ ĐẦU

Trong sản xuất lâm nghiệp, việc xây dựng phương trình thể tích cây đứng với độ chính xác mong muốn có ý nghĩa đặc biệt quan trọng. Ở Việt Nam, nhiều công trình nghiên cứu đã xây dựng phương trình thể tích và biểu sản lượng cho các kiểu rừng khác nhau (Đồng Sĩ Hiền, 1974; Đào Công Khanh, 2001; Trần Hữu Viên, 2002). Tuy nhiên, các nghiên cứu này chủ yếu sử dụng phương pháp truyền thống, đánh giá mô hình thông qua hệ số xác định (R^2) và một số chỉ tiêu cơ bản, mà chưa chú trọng đến việc phân tích sai số mô hình một cách khách quan. Do vậy, mức độ tin cậy và tính khái quát của các phương trình thể tích còn hạn chế, nhất là khi áp dụng cho các loài cây cụ thể ở những vùng sinh thái khác nhau.

Rừng khộp là một trong những hệ sinh thái đặc trưng và có giá trị cao ở Tây Nguyên, trong đó Đắk Lắk là tỉnh có diện tích rừng khộp lớn. Các loài cây như Chiêu liêu (*Shorea obtusa*), Dầu đồng (*Dipterocarpus tuberculatus*) và Dầu trà beng (*Dipterocarpus obtusifolius*) là những loài cây ưu thế trong rừng khộp, đóng vai trò quan trọng trong cơ cấu sản xuất lâm nghiệp. Thực tế việc thu thập số liệu giải tích thân cây được thực hiện thông qua chặt hạ cây, đo đạc toàn bộ kích thước và thể tích là một công việc tốn kém, khó khăn và chi phí lớn. Chính vì thế số liệu giải tích thân cây thường rất hạn chế, vì vậy mà cho đến nay, các phương trình thể tích đáng tin cậy dành riêng cho các loài này

còn chưa được thiết lập đầy đủ, gây khó khăn cho công tác điều tra và quản lý rừng.

Trong bối cảnh nguồn dữ liệu ít ỏi như vậy, việc áp dụng các phương pháp phân tích hiện đại để khai thác tối đa giá trị của số liệu là điều đặc biệt cần thiết. Phương pháp thẩm định chéo Monte Carlo cho phép chia tách dữ liệu để vừa xây dựng, vừa đánh giá mô hình lặp lại nhiều lần. Nhờ đó, các mô hình thể tích được kiểm chứng một cách khách quan, giảm thiểu sai số ngẫu nhiên và tận dụng tối đa số liệu hiện có. Đây chính là ưu điểm nổi bật giúp khắc phục hạn chế về dung lượng dữ liệu trong nghiên cứu lâm nghiệp, đặc biệt là đối với các nghiên cứu dựa trên số liệu giải tích thân cây.

Xuất phát từ thực tiễn trên, nghiên cứu tiến hành phân tích số liệu giải tích thân cây để thiết lập phương trình thể tích cây đứng cho ba loài cây gỗ ưu thế rừng khộp tại tỉnh Đắk Lắk. Kết quả nghiên cứu kỳ vọng sẽ cung cấp cơ sở khoa học và công cụ hữu ích cho công tác điều tra, quản lý và sử dụng bền vững tài nguyên rừng tại địa phương.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu

Số liệu dùng cho nghiên cứu được thu thập từ năm 2009 - 2014 trong thời gian tác giả làm Nghiên cứu sinh và thực hiện đề tài khoa học cấp Bộ về rừng khộp. 156 cây giải tích thuộc ba loài

¹Khoa Nông nghiệp, Trường Đại học Tây Nguyên;

Tác giả liên hệ: Nguyễn Thanh Tân; Email: nttan@ttn.edu.vn.

gỗ ưu thế của rừng khộp tại tỉnh Đắk Lắk, gồm: Chiêu liêu (*Shorea obtuse*), 48 cây, Dầu đồng (*Dipterocarpus tuberculatus*), 56 cây và Dầu trà beng (*Dipterocarpus obtusifolius*), 52 cây. Khu vực thu thập số liệu thuộc rừng khộp trên địa bàn huyện Buôn Đôn, Ea Soup và Ea H'leo (theo địa danh cũ), tỉnh Đắk Lắk. Các cây được chọn giải tích có đường kính ngang ngực ($D_{1.3}$) từ 10 cm đến 77 cm. Cây giải tích có thân thẳng, tròn đều, sinh trưởng bình thường, không cụt ngọn hay phân cành bất thường.

Phương pháp thu thập số liệu cây giải tích như sau: Trước hết, đo chiều cao vút ngọn, đường kính tại vị trí 1,3m, sau đó chặt ngã cây tiêu chuẩn và tiến hành đo chiều dài men thân (L , m), chiều dài thân cây tại vị trí dưới cành lớn nhất còn sống (H_{ac} , m) và chiều cao gốc chặt (H_{gc} , m) bằng thước dây với độ chính xác 0,01m. Tiếp đến, phân chia thân cây ngã thành 10 phân đoạn bằng nhau theo các vị trí 00, 01, 02, 03, 04..., 09L, đo đường kính thân cây có vỏ (D_{cv} , cm) tại các vị trí chiều cao tương đối đã phân chia. Đồng thời, xác định chiều dài thân cây đến vị trí đó.

Tính thể tích gỗ thân cây ($V_{cây}$): là thể tích gỗ toàn bộ thân cây từ mặt đất (cổ rễ) đến đỉnh ngọn cây. Khi phân chia thân cây thành 10 phân đoạn có độ dài tuyệt đối bằng nhau, thể tích toàn bộ thân cây được xác định theo công thức kép sau:

$$V_{cây} = \frac{\pi}{4} \left(\frac{d_{00}^2}{2} + d_{01}^2 + \dots + d_{09}^2 \right) \frac{L}{10}$$

Trong đó: V là thể tích thân cây; d_{00} , d_{01} , d_{02} ... d_{09} là đường kính tại các vị trí 1/10 thân cây; L là chiều dài toàn thân cây.

2.2. Phương pháp nghiên cứu

2.2.1. Lựa chọn các dạng phương trình thể tích để thử nghiệm

Để xây dựng phương trình thể tích, nghiên cứu lựa chọn và kiểm định một số dạng phương trình thường được sử dụng trong lâm nghiệp (Đồng Sĩ Hiền, 1974; Đào Công Khanh, 2001), bao gồm:

$$V = a \times (D_{13}^2 \times H)^b \quad (2.1)$$

$$V = a + b \times (D_{13}^2 \times H) \quad (2.2)$$

$$V = a + b \times D_{13} + c \times H \quad (2.3)$$

$$V = a \times D_{13}^b \quad (2.4)$$

$$V = a \times D_{13}^b \times H^c \quad (2.5)$$

Trong đó: - V là thể tích gỗ thân cây (m^3);

- $D_{1.3}$ là đường kính tại vị trí 1,3m (cm);

- H là chiều cao vút ngọn (m);

- a , b , c là các tham số của phương trình cần ước lượng.

2.2.2. Phương pháp thẩm định chéo

Để đánh giá và lựa chọn mô hình tối ưu, nghiên cứu áp dụng phương pháp thẩm định chéo Monte Carlo (Bảo Huy, 2016; Zhang, 1997; Nguyễn Thanh Tân, 2019; Moore, 2017; Nguyễn Thị Tinh, 2021). Theo phương pháp này, dữ liệu được chia ngẫu nhiên thành hai phần, trong đó 80% số liệu được dùng để xây dựng mô hình và 20% số liệu còn lại dùng để kiểm định sai số. Phần mềm mã nguồn mở R được sử dụng với các câu lệnh (codes) được lập trình để tính toán các chỉ tiêu thống kê đánh giá, so sánh các mô hình (Bảo Huy, 2016). Thực hiện việc phân chia dữ liệu ngẫu nhiên đến 200 lần, qua mỗi lần chạy mô hình sẽ cho kết quả giá trị các chỉ tiêu và sai số và sau đó sẽ xuất kết quả giá trị trung bình của 200 lần chạy mô hình cho 200 bộ dữ liệu được phân chia một cách ngẫu nhiên.

Các chỉ tiêu dùng để đánh giá nhằm lựa chọn các mô hình bao gồm AIC, R^2_{adj} và các sai số thường được áp dụng khi thẩm định chéo là Bias%, RMSE%, MAPE% (Mayer et al, 1993; Zhang, 1997; Temesgen et al, 2014; Huy et al, 2016).

- Chỉ tiêu AIC (Akaike Information Criterion) được tính theo công thức:

$$AIC = n \times \ln(RSS/n) + 2p$$

Trong đó n là số mẫu, RSS tổng bình phương phần dư của mô hình (Residual Sum of Squares), p là số tham số của mô hình.

Các sai số tương đối áp dụng theo phương pháp thẩm định chéo Monte Carlo với R lần lặp lại ngẫu nhiên được tính như sau:

- Sai lệch % giữa quan sát và dự báo qua mô hình (Bias %):

$$Bias (\%) = \frac{1}{R} \sum_{r=1}^R \times \frac{100}{n} \sum_{i=1}^n \frac{(y_i - \hat{y}_i)}{y_i}$$

- Sai số trung phương trung bình tương đối RMSE% (Root Mean Square Error):

$$RMSE (\%) = \frac{1}{R} \sum_{r=1}^R \times 100 \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{y_i}}$$

- Sai số tuyệt đối trung bình MAPE% (Mean Absolute Percent Error):

$$MAPE (\%) = \frac{1}{R} \sum_{r=1}^R \times \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$$

Trong đó, R là số lần phân chia dữ liệu ngẫu nhiên thành hai phần; n là số dữ liệu đánh giá của mỗi lần rút mẫu (20% mẫu được rút ngẫu nhiên); y_i và \hat{y}_i là giá trị quan sát và dự đoán qua mô hình.

Các chỉ tiêu và sai số trên càng bé thì mô hình có độ chính xác càng cao. Phương trình tối ưu được chọn dựa trên tiêu chí là R^2_{adj} cao nhất, AIC và các sai số nhỏ nhất.

2.2.3. Ước lượng tham số của mô hình thể tích tối ưu

Sau khi dùng phương pháp thẩm định chéo sai số mô hình Cross Validation Monte Carlo để lựa chọn được dạng hàm tốt nhất, sẽ sử dụng toàn bộ dữ liệu để xây dựng mô hình thể tích. Việc ước lượng các tham số của mô hình được thực hiện bằng phương pháp tuyến tính bình phương tối thiểu (Linear Least Squares, chương trình “lm” trong phần mềm R) và phi tuyến tính bình phương tối thiểu (Nonlinear Least Squares, chương trình “nls”

trong phần mềm R).

Toàn bộ các phân tích thống kê nói trên được thực hiện bằng phần mềm R (R-Core- Team).

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Kết quả lựa chọn dạng phương trình thể tích tối ưu

Nghiên cứu đã sử dụng phần mềm R để lập mô hình và tính toán các chỉ tiêu thống kê theo phương pháp thẩm định chéo Monte Carlo, nhằm so sánh và đánh giá sai số các mô hình. Trong đó dữ liệu được phân chia ngẫu nhiên với 80% cho thiết lập mô hình và 20% để thẩm định, với số lần lặp lại $R = 200$ lần. Kết quả thẩm định chéo cho thấy cả năm dạng phương trình thử nghiệm đều có hệ số xác định hiệu chỉnh (R^2_{adj}) cao, dao động từ 0,90 đến 0,96. Tuy nhiên, giữa các dạng phương trình có sự khác biệt tương đối rõ rệt về AIC và các chỉ tiêu sai số. Kết quả được thể hiện trong bảng 1 dưới đây.

Bảng 1. Kết quả thẩm định chéo các phương trình thể tích

Loài cây	Dạng phương trình	R^2_{adj}	AIC	Bias%	RMSE%	MAPE%
Chiêu liệu	2.1	0,9722	5,6051	-5,2407	28,8236	22,2343
	2.2	0,9547	24,9326	-33,0086	67,8309	45,6984
	2.3	0,9086	54,1715	38,1522	147,3292	83,6862
	2.4	0,9395	35,5297	-4,2521	32,7419	25,5873
	2.5	0,9721	7,0273	-6,8698	30,5895	23,4829
Dầu đồng	2.1	0,9670	-2,8627	-6,1916	28,1146	18,217
	2.2	0,9358	27,5152	-32,8679	72,9850	46,2028
	2.3	0,9298	33,6222	20,4607	88,8630	52,3122
	2.4	0,9462	18,8193	-8,5857	28,6749	19,4796
	2.5	0,9680	-2,8715	-6,2679	27,5024	18,8490
Dầu trà beng	2.1	0,9588	17,6856	-5,5033	36,0834	25,1536
	2.2	0,9022	54,3522	-88,3924	192,7824	103,1767
	2.3	0,9205	47,6006	52,4282	172,2614	94,3219
	2.4	0,9507	25,0579	-8,8685	37,7619	25,3120
	2.5	0,9605	18,4725	-8,0528	38,7781	26,6759

Trong bảng 1, các chỉ tiêu R^2_{adj} và AIC được tính từ 80% dữ liệu dùng để lập mô hình, còn các sai số Bias%, RMSE% và MAPE% được tính từ 20% dữ liệu đánh giá được rút ngẫu nhiên, độc lập với dữ liệu lập mô hình. Kết quả trong bảng là số liệu được tính trung bình từ kết quả mô phỏng của 200 lần lặp lại tương ứng với 200 bộ dữ liệu khác nhau được rút một cách ngẫu nhiên.

Đối với cả ba loài nghiên cứu (Chiêu liệu, Dầu đồng và Dầu trà beng), phương trình dạng 2.2 và 2.3 có R^2_{adj} thấp nhất đồng thời các chỉ tiêu AIC, Bias%, RMSE%, MAPE% cao hơn

các dạng phương trình còn lại. Trong khi đó dạng phương trình 2.1 và 2.5 có R^2_{adj} cao hơn đồng thời các chỉ tiêu AIC, Bias%, RMSE%, MAPE% thấp. Phương trình chỉ sử dụng một biến độc lập (dạng phương trình 2.4 chỉ có biến đường kính) cho sai số cao hơn và mức độ phù hợp thấp. Kết quả này cho thấy rằng việc kết hợp đồng thời hai biến độc lập ($D_{1,3}$ và Hvn) giúp tăng đáng kể độ chính xác của mô hình, phù hợp với nhận định trước đây của Bảo Huy (2016) và Zhang (1997) về vai trò quan trọng của chiều cao trong mô hình thể tích. Theo Bảo Huy (2016), trong so sánh các mô hình cùng

biến y, chỉ tiêu AIC và các sai số có tầm quan trọng hơn khi đánh giá so với hệ số xác định R^2 . Có trường hợp hàm được lựa chọn dựa vào AIC bé hơn cho dù R^2 của nó có thể bé hơn hàm so sánh; bởi vì AIC phản ánh toàn diện độ tin cậy của giá trị ước lượng so với quan sát. Do vậy khi đánh giá và lựa chọn mô hình, cần thiết phải xem xét đến các chỉ tiêu khác như AIC và các sai số mô hình.

So sánh dạng hàm 2.1 và 2.5 cho thấy chỉ số R^2_{adj} của hàm 2.5 hơi cao hơn một chút so với dạng hàm 2.1, tuy nhiên các chỉ tiêu như AIC, Bias%, RMSE%, MAPE% đều thấp hơn. Căn cứ theo tiêu chuẩn lựa chọn hàm tối ưu, chúng tôi chọn dạng hàm 2.1 để thiết lập phương trình thể tích thân cây cho cả ba loài cây nghiên cứu.

Phương pháp thẩm định chéo Monte Carlo cho phép sử dụng dữ liệu ngẫu nhiên để lập và tính sai số của mô hình rất khách quan. Các phần mềm thống kê chuyên nghiệp như SPSS, Statgraphics, Sata... chỉ cung cấp các sai số của mô hình so với dữ liệu lập mô hình mà lại không cung cấp công cụ để tính toán sai số theo phương pháp thẩm định chéo, trong khi đó, phần mềm R là cơ hội tốt cho việc áp dụng thẩm định chéo các mô hình một cách linh hoạt. Với số lần lặp R đủ lớn thì hầu như tất cả các dữ liệu đều có thể tham gia lập và đánh giá mô hình. Đồng thời, phân bố các chỉ tiêu thống kê và các sai số sẽ tiệm cận chuẩn với sai số trung bình và ổn định. Zhang (1997) đã đề xuất số lần lặp là $R = 500$ lần. Bảo Huy (2016) khi xây dựng mô hình sinh khối cây rừng đã thử nghiệm phương pháp thẩm định chéo Monte Carlo với số lần lặp thay đổi từ 50, 100, 150, 200 và 500 lần. Kết quả cho thấy, với $R = 50$ lần trở lên thì các chỉ tiêu thống kê của mô hình

(AIC, R^2_{adj}) và các sai số Bias, RMSE và MAPE đã ổn định, không có sự khác biệt khi R tăng đến 500 lần. Tuy nhiên, xét thêm phân bố của Bias thì, với $R = 50$ và 100 phân bố có nhiều đỉnh, khi $R \geq 200$ lần, dạng phân bố của Bias đã tiệm cận chuẩn. Vì vậy tác giả đã kết luận sử dụng thẩm định chéo Monte Carlo với $R = 200$ lần là hợp lý.

Phương pháp thẩm định sai số truyền thống có hạn chế lớn là sai số được xác định một lần cho bộ dữ liệu độc lập nhất định, vì vậy sai số có thể khác đi nếu áp dụng theo một bộ dữ liệu độc lập khác, do đó nó thường không cung cấp chính xác sai số trong mọi trường hợp ứng dụng. Nếu áp dụng phương pháp truyền thống để xác định các loại sai số, các giá trị sai số có thể bé hơn hoặc lớn hơn giá trị trung bình rất nhiều và do vậy sẽ ảnh hưởng đến việc lựa chọn phương trình tối ưu. Phương pháp thẩm định chéo Monte Carlo với số lần lặp đủ lớn trên đây sẽ khắc phục được hạn chế này.

3.2. Kết quả xây dựng phương trình thể tích tối ưu

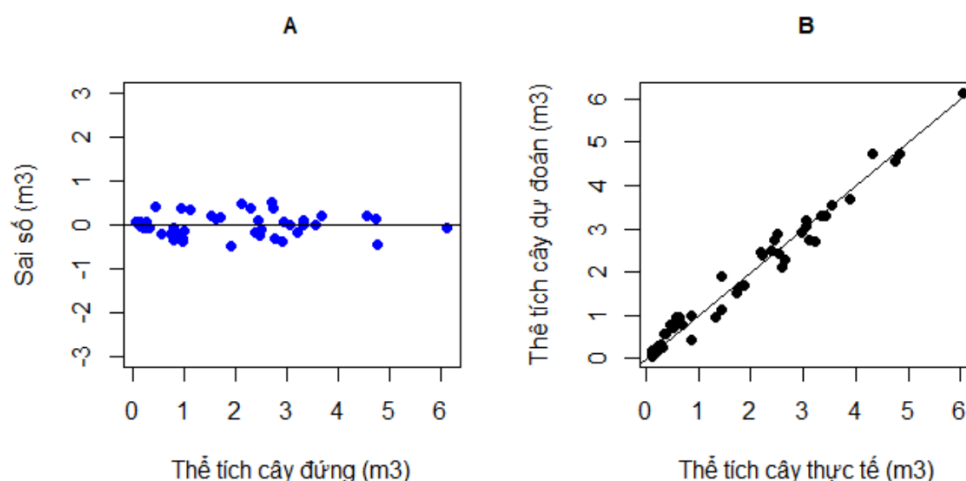
Sau khi xác định dạng phương trình tối ưu, để tối ưu hóa việc sử dụng dữ liệu, toàn bộ bộ số liệu được sử dụng để ước lượng tham số. Đây là ưu điểm của phương pháp thẩm định chéo, bởi nguồn dữ liệu dùng trong nghiên cứu này hạn chế vì việc thu thập rất khó khăn và giá thành cao. Sử dụng phương pháp phi tuyến bình phương tối thiểu để ước lượng các tham số và các chỉ tiêu thống kê, sai số của mô hình thể tích đối với các loài cây nghiên cứu được trình bày trong bảng 3. Tất cả các tham số đều có ý nghĩa thống kê cao ($p < 0,001$). Việc sử dụng toàn bộ dữ liệu giúp cải thiện hệ số R^2_{adj} và giảm các sai số so với khi chỉ sử dụng một phần dữ liệu để lập mô hình cho thấy độ chính xác của mô hình tăng lên.

Bảng 2. Kết quả xây dựng phương trình thể tích tối ưu

Loài cây	Tham số		R^2_{adj}	AIC	Bias%	RMSE%	MAPE%
	a	b					
Chiêu liêu	0,6282	0,7671	0,974	6,326	-5,459	29,504	21,893
Dầu đồng	0,6301	0,6769	0,969	-4,347	-7,287	30,037	18,391
Dầu trà beng	0,7507	0,6318	0,951	15,776	-4,677	34,167	24,319

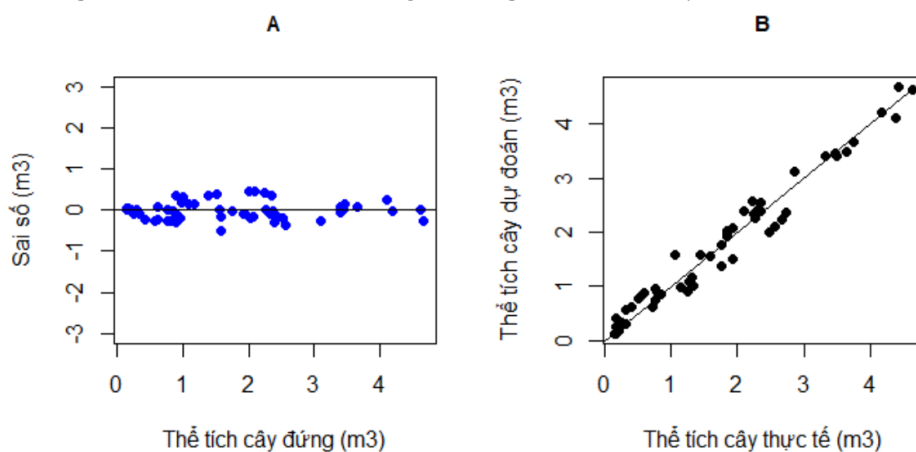
Phương trình thể tích tối ưu của ba loài cây nghiên cứu cụ thể:

- Loài Chiêu liêu: $V_{cây} = 0,6282 \times (D_{13}^2 \times H)^{0,7671}$
- Loài Dầu đồng: $V_{cây} = 0,6301 \times (D_{13}^2 \times H)^{0,6769}$
- Loài Dầu trà beng: $V_{cây} = 0,7507 \times (D_{13}^2 \times H)^{0,6318}$



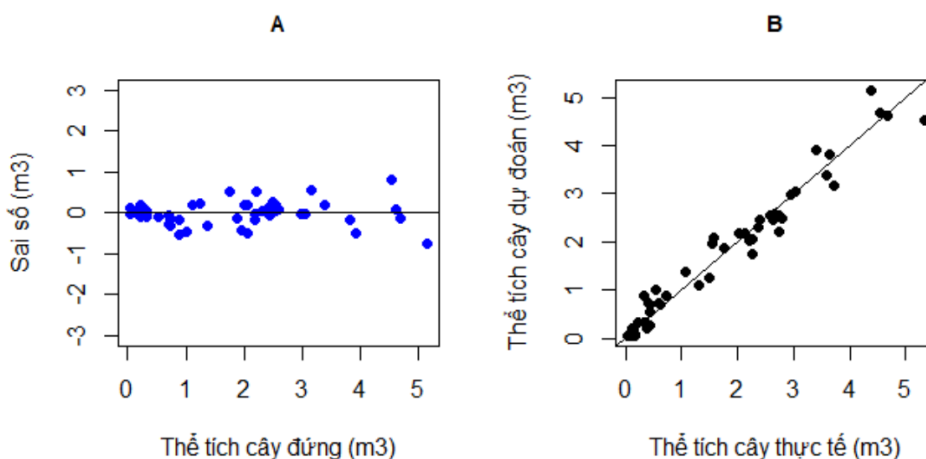
Hình 1. Đồ thị thể hiện sai số loài Chiếu liêu

A: Sai số theo giá trị thể tích dự đoán; B: Quan hệ giữa thể tích cây dự đoán và thực tế



Hình 2. Đồ thị thể hiện sai số loài Dầu đồng

A: Sai số theo giá trị thể tích dự đoán; B: Quan hệ giữa thể tích cây dự đoán và thực tế



Hình 3. Đồ thị thể hiện sai số loài Dầu trà beng

A: Sai số theo giá trị thể tích dự đoán; B: Quan hệ giữa thể tích cây dự đoán và thực tế

Để đánh giá độ chính xác của các mô hình nói trên, ngoài việc đánh giá các chỉ tiêu R^2 , AIC và các sai số Bias%, RMSE% và MAPE% như đã trình bày ở trên, cần xem xét độ tin cậy, sai số, độ bám sát của giá trị quan sát so với giá trị dự đoán

của mô hình.

Các hình 1A, 2A và 3A thể hiện biến động sai số (Residuals) theo các giá trị dự đoán thể tích qua mô hình của loài cây theo thứ tự là loài Chiếu liêu, Dầu đồng và Dầu trà beng.

Kết quả trên đồ thị cho thấy, Biểu đồ phân bố sai số cho thấy phần dư của mô hình phân bố đều quanh trục 0, không xuất hiện xu hướng lệch, đồng thời dao động chủ yếu trong khoảng từ -0,5 đến +0,5. Điều này chứng tỏ mô hình không bị hiện tượng chệch hệ thống và có tính ổn định cao.

Quan hệ giữa giá trị thể tích dự đoán và thực tế được thể hiện ở các hình 1B, 2B và 3B cũng cho thấy các điểm dữ liệu phân bố sát đường 1:1, phản ánh độ khớp rất tốt giữa mô hình và thực tế. Đây là cơ sở quan trọng khẳng định tính chính xác và khả năng ứng dụng của các phương trình thể tích đã xây dựng.

Kết quả nghiên cứu phù hợp với nhiều công trình trong và ngoài nước về mô hình thể tích cây đứng, khi chiều cao được bổ sung như biến giải thích đã làm tăng đáng kể khả năng dự báo (Zhang, 1997; Temesgen et al., 2014). Phương pháp thẩm định chéo Monte Carlo cho phép khai thác hiệu quả bộ dữ liệu hạn chế, đồng thời cung cấp đánh giá khách quan và ổn định về sai số mô hình. Đây là một lợi thế rõ rệt so với các phương pháp kiểm định truyền thống vốn chỉ dựa trên một tập dữ liệu độc lập duy nhất.

4. KẾT LUẬN

Nghiên cứu đã thiết lập thành công phương trình thể tích cây đứng tối ưu cho ba loài cây gỗ ưu thế trong rừng khộp ở tỉnh Đắk Lắk, gồm Chiêu liêu (*Terminalia corticosa* Pierre), Dầu đồng (*Dipterocarpus tuberculatus* Roxb) và Dầu trà beng (*Dipterocarpus obtusifolius* Teijsm. e. Miq). Thông qua việc áp dụng phương pháp thẩm định chéo Monte Carlo với 200 lần lặp, quá trình lựa chọn mô hình được tiến hành một cách khách

quan, đảm bảo giảm thiểu sai số và tăng cường độ tin cậy so với các phương pháp đánh giá truyền thống.

Kết quả cho thấy dạng phương trình $V = a \times (D_{1,3}^{2,2} \times H)^b$ đạt độ chính xác cao nhất đối với cả ba loài nghiên cứu. Các chỉ tiêu thống kê quan trọng như hệ số xác định hiệu chỉnh (R^2_{adj}) đều trên 0,96, giá trị AIC, cùng với các chỉ tiêu sai số (Bias%, RMSE%, MAPE%) ở mức thấp. Đồng thời, phân bố sai số ổn định quanh giá trị trung bình và quan hệ chặt chẽ giữa giá trị dự đoán và quan sát khẳng định sự phù hợp của mô hình. Việc kết hợp đồng thời đường kính ngang ngực ($D_{1,3}$) và chiều cao vút ngọn (H_{vn}) đã nâng cao đáng kể khả năng dự báo, đồng thời củng cố nhận định rằng cần xem xét toàn diện các chỉ tiêu thống kê thay vì chỉ dựa vào hệ số R^2 .

Các phương trình thể tích được xây dựng từ nghiên cứu này là công cụ khoa học đáng tin cậy, có thể ứng dụng hiệu quả trong công tác điều tra, quy hoạch và quản lý rừng khộp tại Đắk Lắk. Việc xác định chính xác thể tích cây đứng sẽ góp phần nâng cao chất lượng điều tra trữ lượng gỗ, phục vụ quản lý rừng bền vững ở địa phương.

Kết quả nghiên cứu khẳng định ưu thế của phương pháp thẩm định chéo Monte Carlo trong xây dựng mô hình thể tích cây đứng so với cách tiếp cận truyền thống, khi vừa tận dụng tốt nguồn số liệu hiện có vừa cung cấp đánh giá khách quan, ổn định về sai số. Cần có các nghiên cứu tiếp theo ứng dụng phương pháp thẩm định chéo để xây dựng hệ thống mô hình thể tích cho các loài cây ưu thế khác của rừng khộp, phục vụ quản lý rừng.

APPLYING METHOD OF CROSS-VALIDATION TO DEVELOP STANDING VOLUME MODELS FOR SOME DOMINANT TREE SPECIES IN DIPTEROCARP FORESTS IN DAK LAK PROVINCE

Nguyen Thanh Tan¹

Received Date: 29/9/2025; Revised Date: 23/10/2025; Accepted for Publication: 25/10/2025

ABSTRACT

The main objective of this study was to develop optimal standing volume equations for three dominant tree species in dry deciduous forests of Dak Lak province, including *Terminalia corticosa* Pierre, *Dipterocarpus tuberculatus* Roxb, and *Dipterocarpus obtusifolius* Teijsm. e. Miq. Data were collected from 156 stem analysis trees belonging to dipterocarp forests in Dak Lak province. The Monte Carlo cross-validation method with 200 iterations was applied, allowing objective and stable evaluation and selection of the optimal model. Five commonly used volume equation forms were tested and compared based on adjusted coefficient of determination (R^2_{adj}), Akaike Information Criterion (AIC), Bias%, RMSE%, and MAPE%. The results showed that the equation form $V = a \times (D1.3^2 \times H)^b$ provided the highest accuracy for all three studied species. Parameters of the optimal equations were estimated from the entire dataset, with $R^2_{adj} > 0.96$, residuals evenly distributed around the predicted values, and a close agreement between observed and predicted tree volumes. The findings of this study confirm the effectiveness of the Monte Carlo cross-validation method under conditions of limited data, while simultaneously providing reliable tools for forest inventory and management of dipterocarp forests in Dak Lak province.

Keywords: volume equation, dipterocarp forest, Monte Carlo cross-validation, modeling.

TÀI LIỆU THAM KHẢO

- Đồng Sĩ Hiền (1974). *Lập biểu thể tích và biểu độ thon cây đứng cho rừng Việt Nam*. Nhà xuất bản Khoa học Kỹ thuật.
- Bảo Huy (2016). *Tin học thống kê trong lâm nghiệp*. Nhà xuất bản Khoa học Kỹ thuật.
- Huy, B. Poudel K.P., and Temesgen, H. (2016). *Aboveground biomass equations for evergreen broadleaf forest in South Central Coastal ecoregion of Viet Nam: Selection of eco-regional or pantropical models*. For. Ecol. And Mgmt. 376: 276-282.
- Đào Công Khanh (2001). *Lập biểu quá trình sinh trưởng và sản lượng cho rừng trồng Keo tai tượng*. Viện Khoa học Lâm nghiệp Việt Nam.
- Mayer DG., and Butler DG. (1993). *Statistic validation*. Ecological modeling. 68(1993): 21-32.
- Moore, A.W. (2017). *Cross validation for detecting and preventing overfitting*. School of Computer Science. Carnegie Mellon University.
- R-Core- Team (2015). *R: A language and environment for statistic computing*. R Foundation for statistic computing. Vienna, Austria. Retrieval from <https://www.R-project.org/>.
- Nguyễn Thanh Tân (2019). *Xây dựng phương trình thể tích cây đứng cho một số loài cây khai thác rừng tự nhiên tại tỉnh Đắk Lắk*. Tạp chí Nông nghiệp & Phát triển nông thôn, số 18/2019: 80 – 87.
- Nguyễn Thanh Tân (2014). *Nghiên cứu xây dựng mô hình dự đoán tăng trưởng và sản lượng rừng Khộp phục vụ kinh doanh rừng ở Tây Nguyên*. Đề tài khoa học cấp Bộ.
- Temesgen, H., Zhang, C.H., and Zhao, X.H. (2014). *Modelling tree height-diameter relationships in multi-species and multi-layered forests: A large observational study from Northeast China*. Journal of Forest Ecology and Management, 316(2014): 78-89.
- Nguyễn Thị Tình (2021). *Thiết lập và thẩm định chéo hệ thống mô hình ước tính sinh khối trên mặt đất cây rừng Khộp ở Việt Nam*. Luận án Tiến sĩ Lâm nghiệp. Trường Đại học Tây Nguyên.
- Zhang, L. (1997). *Cross-validation of Non-linear Growth Functions for modelling Tree Height-Diameter Relationships*. Annals of Botany 79 (1997): 251-257.

¹Faculty of Agriculture, Tay Nguyen University;

Corresponding author: Nguyen Thanh Tan; Email: nttan@ttn.edu.vn.