

Tạp chí

NÔNG NGHIỆP & PHÁT TRIỂN NÔNG THÔN

*Science and Technology Journal
of Agriculture & Rural Development*

MINISTRY OF AGRICULTURE AND RURAL DEVELOPMENT, VIETNAM

Tạp chí Khoa học và Công nghệ

BỘ NÔNG NGHIỆP VÀ PHÁT TRIỂN NÔNG THÔN

2

2014

TẠP CHÍ

NÔNG NGHIỆP
& PHÁT TRIỂN NÔNG THÔN

ISSN 1859 - 4581

NĂM THỨ MƯỜI BỐN

SỐ 233 NĂM 2014
XUẤT BẢN 1 THÁNG 2 KỲ

TỔNG BIÊN TẬP
TS. BÙI HUY HIỀN
ĐT: 04.38345457

PHÓ TỔNG BIÊN TẬP
PHẠM HÀ THÁI
ĐT: 04.37711070

TOÀ SOẠN - TRỊ SỰ
Số 10 Nguyễn Công Hoan
Quận Ba Đình - Hà Nội
ĐT: 04.37711072
Fax: 04.37711073
E-mail: tapchinongnghiep@vnn.vn

BỘ PHẬN THƯỜNG TRỰC
135 Pasteur
Quận 3 - TP. Hồ Chí Minh
ĐT/Fax: 08.38274089

Giấy phép số:
400/GP - BVHTT
Bộ Văn hoá - Thông tin cấp ngày 28
tháng 12 năm 2000.

In tại Công ty Cổ phần KH&CN Hải Đăng
81/30/1 Lạc Long Quân, Cầu Giấy, Hà Nội

Giá: 20.000đ

MỤC LỤC

- | | |
|--|---------|
| ❑ VÕ THỊ THANH LỘC. Rủi ro và tính tổn thương của người nuôi cá tra vùng đồng bằng sông Cửu Long | 3-9 |
| ❑ MAI DUY MINH, PHÙNG BẢY, NGUYỄN THỊ XUÂN THU. Nghiên cứu tạo hầu đa bội loài <i>Crassostrea angulata</i> (Lamarck, 1819) bằng Cytochanlasin B | 10-14 |
| ❑ TRẦN THỊ THUÝ HÀ, LÊ MINH TOÁN, NGUYỄN THỊ HƯƠNG, NGUYỄN HỮU NINH. Đa hình microsatellite vùng promoter gen prolactin 1 (PRL – 1) với sinh trưởng cá rô phi vằn (<i>Oreochromis niloticus</i>) trong điều kiện nước lợ | 15-20 |
| ❑ ĐÌNH VIỆT HÙNG, WOO JUNG CHOI. Một cách nhìn tổng quan giữa đồng vị phóng xạ trong vòng năm của cây rừng với những biến đổi môi trường | 21-26 |
| ❑ NGUYỄN HOÀNG QUANG, LẠI PHƯƠNG LIÊN, ĐỖ THỊ HẠNH, HÀ VIẾT CƯỜNG, PHẠM XUÂN HỘI. Nghiên cứu đa dạng di truyền phân đoạn S7 của virút lùn sọc đen phương Nam ở Việt Nam | 27-32 |
| ❑ HOÀNG MANH CƯỜNG, LÊ NGỌC BÁU, ĐỖ NĂNG VINH. Đặc tính nông sinh học các giống bơ (<i>Persea americana</i> Mills.) chọn lọc tại Tây Nguyên | 33-40 |
| ❑ NGÔ HỒNG THANH, HOÀNG MINH TÂM, CAO NGỌC ĐIẾP. Xác định đặc tính vi khuẩn nội sinh trong cây mía trồng ở tỉnh Bến Tre và Long An | 41-48 |
| ❑ CHU TRUNG KIẾN, TỬ THỊ MỸ THUẬN. Nghiên cứu phòng trừ bệnh mốc đen lá hại cà chua vụ mùa mưa tại huyện Chợ Gạo, tỉnh Tiền Giang | 49-53 |
| ❑ ĐÀO MINH SÔ, ĐỖ KHẮC THỊNH, VÕ NGỌC VŨ. Hiệu lực phòng trừ rầy nâu hại lúa của chế phẩm sinh học chứa nấm xanh Ometar tại Long An và Tiền Giang | 54-59 |
| ❑ ĐOÀN XUÂN CẢNH. Nghiên cứu đặc điểm nông sinh học và kiểm tra có mặt gen kháng bệnh virút xoắn vàng lá TY ở một số dòng cà chua | 60-65 |
| ❑ NGUYỄN HAY, LÊ ANH ĐỨC, PHẠM CÂN. Nghiên cứu thiết kế chế tạo và thực nghiệm xác định chế độ sấy bông atisô | 66-70 |
| ❑ ĐẶNG XUÂN CƯỜNG, BÙI MINH LÝ, TRẦN THỊ THANH VÂN, VŨ NGỌC BỘI, LÊ NHƯ HẬU, ĐÀO TRỌNG HIẾU. Bước đầu nghiên cứu tinh chế hoạt chất chống ôxi hoá phlorotanin tách chiết từ rong nâu <i>Sargassum serratum</i> của vùng biển Nha Trang | 71-76 |
| ❑ LÊ HỮU HOÀNG, ĐẶNG THUÝ BÌNH, NGUYỄN THỊ ANH THƯ. Nghiên cứu đặc điểm di truyền phân loài chim yến (<i>Aerodramus fuciphagus</i> .) tại Việt Nam | 77-82 |
| ❑ VÕ THÀNH THÌN, ĐÀO DUY HÙNG, PHẠM TRUNG HIẾU, ĐẶNG VĂN TUẤN, NGUYỄN THỊ THẮM, ĐẶNG THỊ SAO MAI, LÊ LẬP. Khảo sát một số đặc điểm dịch tễ bệnh <i>Leptospirosis</i> trên lợn tại khu vực Nam Trung bộ – Tây Nguyên | 83-88 |
| ❑ LÊ THỊ DIỄN, PHAN TRỌNG TRÍ. Đa dạng sinh học và mức độ tương đồng về thành phần loài bướm mắt rằn (<i>Lepidoptera: Satyridae</i>) ở Vườn Quốc gia Bạch Mã | 89-94 |
| ❑ NGUYỄN KHẮC KHÔI, NGUYỄN THỊ YẾN, ĐỖ HỮU THƯ. Tính đa dạng thực vật bậc cao có mạch ở phân ban Khe Rổ, Khu Bảo tồn Thiên nhiên Tây Yên Tử, Bắc Giang | 95-100 |
| ❑ SIMELTON ELISABETH, ĐÀM VIỆT BẮC, ĐỖ TRỌNG HOÀN, TRALDI REBECCA, CATA CUTAN DELIA, ĐÌNH NGỌC LAN. Khuyến nghị chương trình nâng cao năng lực các bên liên quan về chi trả dịch vụ môi trường rừng Việt Nam | 101-109 |
| ❑ HUỖNH NHÂN TRÍ, BẢO HUY. Cơ sở khoa học xây dựng mô hình sinh trắc (<i>Allometric equations</i>) để ước tính sinh khối và ácbon rừng | 110-120 |

CƠ SỞ KHOA HỌC XÂY DỰNG MÔ HÌNH SINH TRẮC (ALLOMETRIC EQUATIONS) ĐỂ ƯỚC TÍNH SINH KHỐI VÀ CÁC BON RỪNG

Huỳnh Nhân Trí¹, Bảo Huy²

TÓM TẮT

Để ước tính sinh khối, các bon rừng thì xây dựng mô hình sinh trắc là nội dung quan trọng. Tuy nhiên, cơ sở khoa học để xây dựng mô hình chưa được hệ thống và đánh giá để đưa ra tiêu chuẩn chung. Nghiên cứu này trên cơ sở thu thập số liệu sinh khối, phân tích các bon từ chặt hạ cây rừng ở rừng lá rộng thường xanh vùng Tây Nguyên, đã phân tích để đưa ra các cơ sở cho việc xây dựng mô hình sinh trắc ước tính sinh khối và các bon rừng. Kết quả cho thấy: i) Cần lựa chọn số biến số tối ưu cho mô hình theo chỉ tiêu C_p của Mallow và để đạt được độ tin cậy thì cần có bốn biến số là đường kính (DBH), chiều cao (H), khối lượng thể tích gỗ (WD) và diện tích tán lá (CA); ii) ước lượng các tham số của mô hình tốt bằng phương pháp bình phương tối thiểu, đồng thời nếu có trọng số sẽ giúp cho việc nắn mô hình sát với số liệu quan sát hơn; iii) lựa chọn hàm số thích hợp dựa vào 7 tiêu chí $R^2(\%)$ cao, tiêu chuẩn T kiểm tra tham số với $P < 0.05$, CF tiến đến 1, AIC bé nhất, S% bé nhất, đồ thị Normal P-P bám sát đường chéo có tọa độ (0,0) và (1,1), đồ thị biến động phần dư (residuals) là hằng số theo giá trị ước lượng qua hàm.

Từ khóa: Các bon rừng, mô hình sinh trắc, rừng lá rộng thường xanh, sinh khối.

1. ĐẶT VẤN ĐỀ

Từ năm 2009 dưới sự hỗ trợ của FAO và các tổ chức quốc tế, Việt Nam đã khởi động chương trình UN-REDD+ quốc gia. Để tham gia chương trình REDD+, cần có những nghiên cứu về phương pháp đo tính, giám sát các bon để cung cấp thông tin, dữ liệu có cơ sở khoa học, đáng tin cậy về sự thay đổi của các bể chứa các bon trong các hệ sinh thái rừng làm cơ sở tính toán chi trả dịch vụ môi trường. Một trong nội dung kỹ thuật cần thiết là lập các mô hình ước tính sinh khối và các bon rừng (allometric equations) cho từng kiểu rừng, vùng sinh thái của Việt Nam.

Với yêu cầu đó chỉ trong vài năm gần đây đã có một số nhà nghiên cứu Việt Nam bắt đầu lập các mô hình ước tính khối và các bon rừng như: Bảo Huy và cộng sự (2008, 2009, 2012, 2013), Võ Đại Hải và cộng sự (2008), Vũ Tấn Phương (2012) và một số trường đại học, viện nghiên cứu..... Đây là khởi đầu cho việc hoàn chỉnh hệ thống mô hình ước tính sinh khối, các bon rừng cho Việt Nam. Tuy nhiên vẫn còn hạn chế như ít mô hình ước tính các bon, phương pháp thiết lập hàm đôi khi chưa đáp ứng được tiêu chuẩn lập mô hình sinh trắc theo các hướng dẫn, tiêu chuẩn quốc tế.

Xuất phát từ những lỗ hổng về mặt khoa học và thực tiễn của các nghiên cứu trong và ngoài nước; cần có nghiên cứu cơ sở khoa học xây dựng các mô hình toán sinh học (Allometric Equations) để ước tính sinh khối và các bon rừng nhằm bổ sung về phương pháp xây dựng mô hình một cách hệ thống và bảo đảm độ tin cậy.

2. PHƯƠNG PHÁP VÀ VẬT LIỆU NGHIÊN CỨU

Căn cứ vào khái niệm mô hình sinh trắc là mô hình biểu thị mối quan hệ giữa sinh khối, các bon trong từng bộ phận cây và toàn bộ cây với các nhân tố điều tra cây rừng; thiết lập các mô hình toán mô phỏng mối quan hệ này để làm cơ sở cho việc ứng dụng điều tra giám sát sinh khối, các bon rừng.

2.1. Đối tượng nghiên cứu, cơ sở dữ liệu sinh khối các bon cây rừng

Đối tượng nghiên cứu là kiểu rừng lá rộng thường xanh ở Tây Nguyên (ở các tỉnh Gia Lai, Đắk Lắk và Đắk Nông). Cơ sở dữ liệu sinh khối và các bon từng bộ phận của cây rừng bao gồm thân, cành, lá, vỏ và rễ được thu thập thông qua phương pháp chặt hạ cây (destructive measurement). Đã bố trí 20 ô mẫu diện tích 2.000 m² ở các trạng thái rừng, trữ lượng rừng khác nhau và chọn cây chặt tỷ lệ theo cấp kính; tổng số cây chặt hạ đo tính sinh khối 4 bộ phận cây trên mặt đất (thân, cành, lá, vỏ) là 224 cây, số cây đào gốc rễ cây để đo tính sinh khối rễ dưới mặt đất là 143 cây. Tổng số mẫu của 5 bộ phận cây chặt hạ là 3.117 mẫu để phân tích sinh khối khô và hàm lượng các bon.

2.2. Mô hình sinh trắc tổng quát

Mô hình sinh trắc có dạng tổng quát là:

$$y_i = f(x_j) \quad (2.1)$$

Trong đó:

y_i (biến số phụ thuộc): Sinh khối, lượng các bon tích lũy trong từng bộ phận thân cây gỗ (thân, cành, lá, vỏ); tổng sinh khối của 4 bộ phận cây trên mặt đất (AGB), dưới mặt đất (BGB); tổng các bon

¹ Trường Trung học Lâm nghiệp Tây Nguyên

² Trường Đại học Tây Nguyên

trong các bộ phận cây trên mặt đất C(AGB) và dưới mặt đất C(BGB).

x_j (các biến số độc lập): Gồm 4 biến số chủ yếu được nghiên cứu là đường kính tại vị trí 1,3 m (DBH), chiều cao cây (H), khối lượng thể tích gỗ (WD) và diện tích tán lá (CA).

2.3. Xác định biến số độc lập x_j ảnh hưởng đến sinh khối, các bon yi

Sử dụng tiêu chuẩn Mallows' C_p (1973). Chỉ số

C_p càng gần với số biến số P thì mô hình càng phù hợp; dựa vào đây để xác định số biến số P tham gia mô hình khi có quá nhiều biến số được giả định là có ảnh hưởng đến yi.

Lúc này các biến số yi và x_j cũng được đổi biến số theo các dạng hàm phi tuyến chuyển sang tuyến tính như $\log(y, x)$, $\sqrt{y, x}$, $1/y, x$, $(y, x)^2$, $\exp(y, x)$; mỗi trường hợp đổi biến số hoặc tổ hợp biến, giá trị C_p được tính toán để xác định p biến số tối ưu cho từng mô hình.

2.4. Các dạng hàm thử nghiệm

Từ mô hình sinh trắc có dạng tổng quát là: $y_i = f(x_j)$, tiến hành đổi biến số yi và x_j theo các hàm $\ln(y_i, x_j)$, $\sqrt{y_i, x_j}$, y_i^2 , x_j^2 , $1/y_i$, $1/x_j$, trong đó biến x_j có thể là biến đơn hay là tổ hợp biến ví dụ DBH^2H , DBH^2WD , DBH^2H*WD , DBH^2CA ...

Tổ hợp tất cả các trường hợp đổi biến số để dò tìm hàm tối ưu. Các dạng hàm được thử nghiệm ở bảng 1 là ví dụ minh họa trong trường hợp một biến số, nếu nhiều biến số thì các biến số mới cũng được đổi biến và mở rộng.

Bảng 1. Các dạng hàm thử nghiệm

TT	Mô hình	TT	Mô hình	TT	Mô hình
1	Multiplicative $\ln(Y) = a + b \cdot \ln(X)$	9	Double reciprocal $1/Y = a + b/X$	17	Square root-Y reciprocal-X $\sqrt{Y} = a + b/X$
2	Square root-Y $\sqrt{Y} = a + b \cdot X$	10	Square root-Y logarithmic-X $\sqrt{Y} = a + b \cdot \ln(X)$	18	Squared-Y $Y^2 = a + b \cdot X$
3	Logarithmic-Y square root-X $\ln(Y) = a + b \cdot \sqrt{X}$	11	Linear $Y = a + b \cdot X$	19	Squared-Y square root-X $Y^2 = a + b \cdot \sqrt{X}$
4	Square root-Y squared-X $\sqrt{Y} = a + b \cdot X^2$	12	Logarithmic-Y squared-X $\ln(Y) = a + b \cdot X^2$	20	Reciprocal-X $Y = a + b/X$
5	Squared-X $Y = a + b \cdot X^2$	13	Square root-X $Y = a + b \cdot \sqrt{X}$	21	Squared-Y logarithmic-X $Y^2 = a + b \cdot \ln(X)$
6	Double square root $\sqrt{Y} = a + b \cdot \sqrt{X}$	14	Double squared $Y^2 = a + b \cdot X^2$	22	Reciprocal-Y squared-X $1/Y = a + b \cdot X^2$
7	Exponential $\ln(Y) = a + b \cdot X$	15	Reciprocal-Y logarithmic-X $1/Y = a + b \cdot \ln(X)$	23	Squared-Y reciprocal-X $Y^2 = a + b/X$
8	S-curve model $\ln(Y) = a + b/X$	16	Logarithmic-X $Y = a + b \cdot \ln(X)$		

2.5. Phương pháp ước lượng các mô hình sinh trắc

Thử nghiệm các phương pháp ước lượng mô hình khác nhau như sau:

Phương pháp bình phương tối thiểu ước lượng hàm phi tuyến được tuyến tính hóa.

Phương pháp bình phương tối thiểu ước lượng hàm phi tuyến được tuyến tính hóa có trọng số (Weight).

Phương pháp Marquardt ước lượng hàm phi tuyến.

Phương pháp Marquardt ước lượng hàm phi tuyến có trọng số (Weight).

2.6. Các tiêu chuẩn lựa chọn mô hình, đánh giá độ tin cậy của hàm

Các chỉ tiêu thống kê, biểu đồ để lựa chọn biến tham gia và hàm tối ưu:

Hệ số quan hệ, xác định R^2 : Về tổng quát thì hàm tốt nhất khi R^2 đạt max và tồn tại ở mức sai $P < 0,05$. Tuy nhiên có trường hợp R^2 đạt max nhưng chưa phải là hàm phù hợp nhất, do vậy cần dựa thêm các chỉ tiêu thống kê khác.

Tiêu chuẩn t kiểm tra sự tồn tại của các tham số của mô hình: Với giả thuyết $H_0: b_i = 0$, giả thuyết bị bác bỏ khi $P < 0,05$; có nghĩa là các tham số tồn tại và khác 0 rõ rệt. Chỉ tiêu này chỉ áp dụng cho hàm đa biến.

Tiêu chuẩn AIC (Akaike Information Criterion): AIC được sử dụng khi cần lựa chọn mô hình tốt nhất với các biến số ảnh hưởng khác nhau (Burnham và Anderson, 2002; Johnson và Omland, 2004 dẫn theo Chave, 2005):

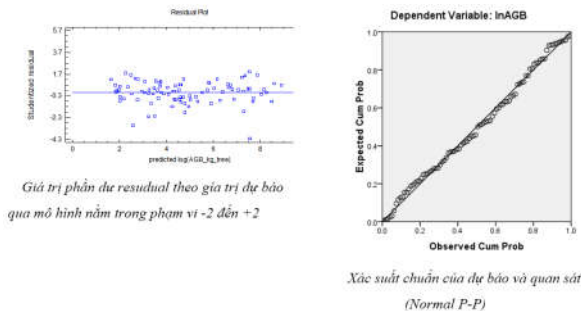
$$AIC = n \cdot \ln(RSS/n) + 2K = -\ln(L) + 2K \quad (2.2)$$

Mô hình tối ưu với các biến số thích hợp khi giá trị đại số của AIC là bé nhất. Trong đó n: số mẫu, RSS (the residual sums of squares) là tổng bình phương phần dư, K: số tham số của mô hình bao gồm tham số sai số ước lượng.

Nhân tố quan hệ (CF - Correction factor) (Baskerville, 1972; Duan, 1983; Parresol, 1999 dẫn theo Chave, 2005):

$$CF = \exp(RSE^2/2) \quad (2.3)$$

CF luôn lớn hơn 1. Trong đó RSE (Residual standard error) là sai tiêu chuẩn của phần dư hay là sai số của mô hình. Khi RSE càng lớn thì CF càng lớn, có nghĩa mô hình càng có độ tin cậy thấp. Mô hình tốt khi CF càng tiến dần đến 1. Tiêu chuẩn này chỉ sử dụng cho mô hình dạng đổi biến số logarit.



Hình 1. Biểu đồ đánh giá sự thích hợp và tin cậy của mô hình lựa chọn theo biến động residual và Normal P-P

Biểu đồ biến động phần dư (residual) ứng với các giá trị dự báo y của mô hình lựa chọn: Mô hình tốt khi biến động residual tập trung trong phạm vi giá trị -2 đến +2 ứng với các giá trị dự báo y. **Biểu đồ xác suất chuẩn Normal P-P:** Mô hình đạt độ tin cậy cao khi xác suất phân bố của giá trị quan sát và lý thuyết nằm trên đường chéo của tọa độ (0, 0) và (1, 1). Biểu diễn ở hình 1.

Biến động trung bình S% để đánh giá mức độ sai lệch, biến động trung bình của giá trị ước lượng qua mô hình với thực tế quan sát: (Brand và Smith, 1985; Cairns et al., 2003; Chave et al., 2005; Nelson et al., 1999 dẫn theo Basuki et al., 2009):

$$S\% = \frac{100}{n} \sum_{i=1}^n \frac{|Y_{ilt} - Y_i|}{Y_i} \quad (2.4)$$

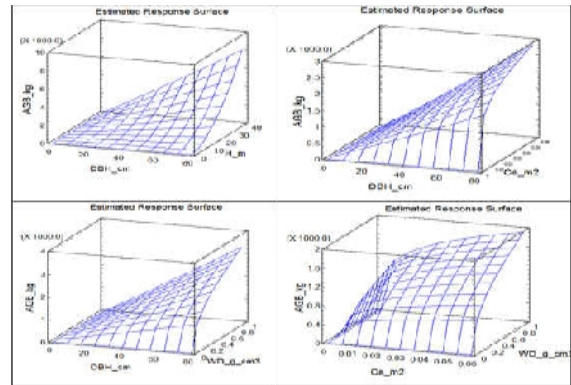
Trong đó: Y_{ilt} : Giá trị dự báo qua mô hình; Y_i : Giá trị thực của sinh khối, các bon, n: Số cây quan sát.

S% được sử dụng như là một tiêu chuẩn để đánh giá độ tin cậy của mô hình; S% càng nhỏ thì biến động giữa mô hình và số liệu thực tế càng nhỏ, hàm có độ tin cậy cao.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

Trong nghiên cứu này, các cơ sở khoa học sau đã được nghiên cứu, đánh giá để làm cơ sở cho việc thiết lập các mô hình ước tính sinh khối, các bon rừng.

3.1. Kết quả chọn biến số cho mô hình sinh trắc



Hình 2. Quan hệ giữa AGB với các biến số DBH, H, WD và CA theo mô hình không gian 3 chiều

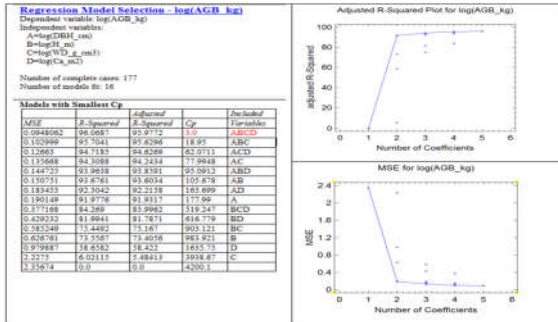
Để chọn biến số và số biến số cho mô hình, tiêu chuẩn Mallows' Cp. Chỉ số Cp thay đổi với số biến số trong mô hình khác nhau và khi Cp bé nhất và gần với số biến số p (bao gồm cả biến số là hằng số của mô hình) nào đó thì mô hình có p biến số đó là tối ưu. Lúc này tương đương với hệ số xác định R^2_{max} cao nhất và trung bình bình phương của sai số ước lượng là bé nhất (MSE_{min}).

Để tính Cp trước hết cần đổi biến số yi và xj theo nhiều trường hợp khác nhau như logarit, sqrt, và từ đó xem xét thay đổi của Cp nhằm xác định được số biến số ảnh hưởng nhiều nhất.

Trong trường hợp này, bốn biến số xi bao gồm DBH, H, WD và CA được đưa vào đánh giá để lựa chọn biến số ảnh hưởng đến sinh khối của cây rừng trên mặt đất AGB. Với việc đổi biến số logarit của tất cả biến thì quy luật thay đổi Cp rõ nhất biểu thị ở hình 2.

Kết quả ở hình 2 cho thấy Cp bé nhất bằng 5.0 và tương ứng với số biến số tối ưu là p = 5 biến số, trong đó có 1 biến số là hằng số của mô hình, do đó thực tế là 4 biến số là thích hợp nhất để đưa vào mô hình ước tính sinh khối AGB. Bốn biến số đó là A = ln(DBH), B = ln(H), C = ln WD) và D = ln(CA). Tương ứng với Cp bé nhất và gần với p = 5 biến số, thì MSE là bé nhất = 0,095 và R^2_{adj} là lớn nhất = 95,977%. Nếu xét giá trị Cp lớn hơn liền kề là 18,95 với ba biến số trong mô hình là ABC, tức

không có biến D của diện tích tán lá ($D = \ln(CA)$), các giá trị MSE tăng và R^2 giảm rõ rệt. Như vậy có nghĩa là biến diện tích tán lá CA ảnh hưởng rõ rệt đến sinh khối của cây AGB, hay nói khác nếu gia tăng thêm biến CA thì mô hình sẽ gần với thực tế hơn, có độ tin cậy cao hơn. Hình 3 biểu diễn đồ thị 3 chiều quan hệ giữa AGB với từng cặp biến số trong 4 biến DBH, H, WD và CA; các đồ thị 3 chiều này cho thấy các biến số này có quan hệ rõ rệt với AGB.



Hình 3. Giá trị Cp của Mallows theo số biến số và đồ thị thay đổi R^2 và MSE theo số biến số tham gia mô hình

Như vậy có thể nói sử dụng tiêu chuẩn Cp của Mallows sẽ giúp cho việc xác định biến số nào ảnh hưởng đến biến phụ thuộc và số biến số tối ưu trong mô hình để cho MSE là bé nhất. Phương pháp này khách quan, tuy vậy hầu như chưa có tác giả nào nghiên cứu, các biến được lựa chọn hầu như dựa vào kinh nghiệm về điều tra thống kê kinh điển. Với việc sử dụng tiêu chuẩn Cp nghiên cứu đã phát hiện thêm và khẳng định biến số CA là quan trọng và ảnh hưởng đến việc nâng cao độ chính xác của ước tính sinh khối, các bon rừng mà nhiều tác giả trong và ngoài nước chưa đề cập đến.

Vì vậy nghiên cứu đã sử dụng chỉ số Cp để xác định các biến số ảnh hưởng trong lập các mô hình sinh khối, các bon cho từng bộ phận, trên dưới mặt đất và cho lâm phần.

Cp chỉ ra có bốn biến số là DBH, H, WD và CA ảnh hưởng đến AGB, về lý thuyết thống kê thì chỉ cần lập một mô hình theo 4 biến số này để cho độ tin cậy cao nhất. Tuy nhiên trong thực tế khi số biến số của mô hình càng nhiều thì sẽ tăng chi phí điều tra rừng để đo tính các biến số. Do đó sử dụng Cp để xác định biến số nào thực sự ảnh hưởng đến sinh khối, các bon cây rừng và bao nhiêu biến số là tối đa. Còn trong xây dựng mô hình sinh trắc thì cần lần lượt lập mô hình ứng với 1, 2, 3 và tối đa là p biến số (với Cp bé nhất) và như vậy tùy theo yêu cầu độ tin cậy, khả năng áp dụng để chọn lựa số biến số thích hợp, nhưng tối ưu vẫn là p biến số.

3.2. Kết quả ước lượng mô hình

Để xác định cơ sở khoa học trong việc chọn lựa phương pháp ước lượng mô hình hồi quy; thử nghiệm 4 phương pháp khác nhau để ước lượng mô hình sinh khối của cây trên mặt đất (AGB) với một đến bốn biến số có ảnh hưởng đến AGB là DBH, H, WD và CA. Bao gồm:

Ước lượng mô hình bằng phương pháp bình phương tối thiểu, trong đó hàm phi tuyến được tuyến tính hóa.

Ước lượng mô hình bằng phương pháp bình phương tối thiểu, trong đó hàm phi tuyến được tuyến tính hóa có trọng số (Weight)

Ước lượng trực tiếp mô hình phi tuyến theo phương pháp Marquardt.

Ước lượng trực tiếp mô hình phi tuyến theo Marquardt có trọng số Weight.

3.2.1. Ước lượng mô hình phi tuyến theo Marquardt:

Đối với phương pháp ước lượng hàm phi tuyến theo Marquardt, vấn đề quan trọng là xác định tham số đầu vào khởi đầu cho mô hình, thông thường việc xác định các tham số đầu vào thường theo kinh nghiệm do đó sẽ cho kết quả ước lượng khác nhau. Nghiên cứu thực hiện chọn tham số đầu vào theo Picard et al. (2012), trong đó tham số đầu vào được thăm dò thông qua ước lượng chính dạng hàm đó bằng phương pháp bình phương tối thiểu, mô hình lúc này được tuyến tính hóa.

3.2.2. Ước lượng mô hình có trọng số Weight theo hai phương pháp bình phương tối thiểu và Marquardt:

Để ước lượng mô hình theo phương pháp có trọng số Weight áp dụng cho cả hai phương pháp bình phương tối thiểu và Marquardt, trong đó mô hình lựa chọn dựa vào thay đổi trọng số weight như sau (Picard et al. (2012):

- Trọng số **Weight** = $\frac{1}{X_i^{2b_i}}$. trong đó X_i là biến

quan trọng ảnh hưởng cao để điều chỉnh tham số mô hình và b_i là tham số gắn biến X_i . Thông thường dò tìm weight tối ưu với $2b_i = -4$ đến $+4$ để có được hàm tối ưu hoặc với các giá trị $-2b_i, 2b_i$ và $4b_i$.

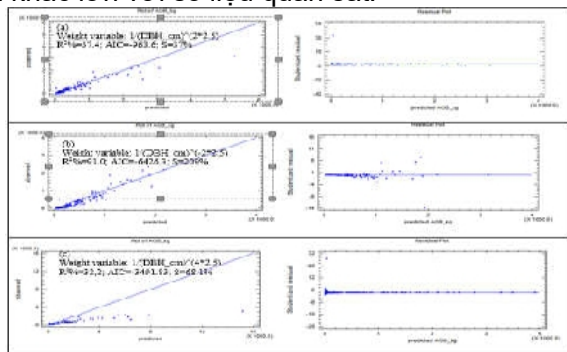
- Với mỗi giá trị trọng số weight, mô hình được đánh giá, lựa chọn thông qua: i) Hệ số quan hệ $R^2(\%)$ cao, ii) đồ thị residuals có biến động là hằng số quanh giá trị ước tính; iii) các tiêu chuẩn AIC_{min} và $S\%_{min}$.

Quá trình dò tìm hàm bằng phương pháp Marquardt với weight khác nhau, cho thấy:

- Với $Weight = 1/X_i^{2b_i}$ thì mô hình có $R^2\%$ thấp, tuy nhiên các giá trị AIC và $S\%$ là nhỏ nhất (tốt nhất) và biến động phần dư residuals quanh giá trị ước lượng là rất thấp và là hằng số. Đây là mô

hình có giá trị ước lượng bám sát số liệu quan sát. Điều này cho thấy để lựa chọn mô hình tối ưu không nên chỉ dựa vào hệ số quan hệ $R^2(\%)$ đặc biệt là đối với ước lượng có trọng số.

- Với Weight = $1/X_i^{2b_i}$ thì mô hình có R% cao nhất, tuy nhiên các giá trị AIC và S% là lớn nhất (kém nhất) và biến động phần dư residuals quanh giá trị ước lượng là cao và mở rộng khi giá trị ước lượng gia tăng. Đây là mô hình có giá trị ước lượng sai khác lớn với số liệu quan sát.



Hình 4. Đồ thị quan hệ giữa giá trị quan sát - ước tính và biến động residuals quanh giá trị ước tính của mô hình $AGB=a*DBH^b$ theo Marquardt với trọng số thay đổi

- Với Weight = $1/X_i^{4b_i}$ thì mô hình có R% thấp nhất, các giá trị AIC và S% ở mức trung bình và biến động phần dư residuals quanh giá trị ước lượng là cao và mở rộng khi giá trị ước lượng nhỏ. Đây là mô hình có giá trị ước lượng sai khác khá lớn với số liệu quan sát.

Nghiên cứu đã thử nghiệm 4 phương pháp ước lượng hàm gồm bình phương tối thiểu, Marquardt và cả hai trường hợp thực hiện có hay không có trọng số. Bảng 2 là minh họa kết quả ước lượng mô hình quan hệ $AGB = f(DBH)$ theo dạng hàm mũ Power theo 4 phương pháp. Mô hình tối ưu được lựa chọn trên cơ sở đánh giá tổng hợp các tiêu chuẩn thống kê R^2 , CF, AIC và S% và đồ thị quan hệ giữa lý thuyết và quan sát và residuals. Trong đó trọng số được lấy tối ưu như phân tích trên là Weight = $1/X_i^{2b_i}$.

Bảng 2. Kết quả dò tìm mô hình ước tính sinh khối cây rừng ($AGB = a*DBH^b$) theo 4 phương pháp ước lượng hàm là bình phương tối thiểu, Marquardt và có hay không có trọng số Weight

Hàm ước lượng theo các phương pháp	R^2 adj (%)	P	n	Pbi	Weight	CF	AIC	S %	Phương pháp ước lượng hàm
$\ln(AGB) = -2,25438 + 2,49193*\ln(DBH)$	93,7	0,0	224	0,0		1,08 9	7266, 3	32 %	Bình phương tối thiểu
$\ln(AGB) = -2,49106 + 2,59253*\ln(DBH)$	87,0	0,0	224	0,0	$1/\ln(DBH)$ ($2^{*}2,5$)	1,00 1	- 1068, 9	34 %	Bình phương tối thiểu có trọng số
$AGB = 0,69229*(DBH)^{1,96595}$	92,8	0,0	224	0,0			217,4	95 %	Marquardt
$AGB = 0,133737*(DBH)^{2,44406}$	37,4	0,0	224	0,0	$1/DBH$ ($2^{*}2,5$)		- 963,6	37 %	Marquardt có trọng số

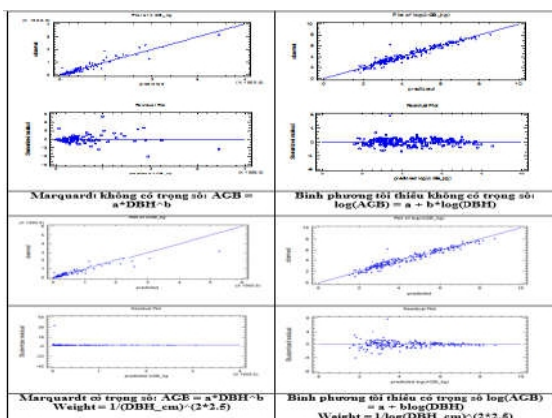
Ghi chú: Pi: Xác suất tồn tại tham bố bi

Bảng 2 cho thấy:

- **Xét hệ số quan hệ có điều chỉnh: R^2_{adj} :** Phương pháp bình phương tối thiểu cao nhất là 93,7%, tiếp đến là phương pháp Marquardt không có trọng số là 92,8%, phương pháp bình phương tối thiểu có trọng số là 87,0% và thấp nhất là Marquardt có trọng số là 37,4%. Kết quả này cho thấy khi có trọng số tham gia vào ước lượng mô hình thì R^2 có xu hướng giảm rõ rệt. Trong đó phương pháp bình phương tối thiểu không có trọng số có R^2 đạt max.

- **Xét nhân tố điều chỉnh CF:** Chỉ xét cho hàm đổi biến số về dạng logarit, trong trường hợp ước lượng bình phương tối thiểu có và không có trọng số. Kết quả cho thấy CF đều gần tiến về 1, có nghĩa hàm có sai số ước lượng nhỏ; trong đó khi có trọng số thì CF thấp hơn = 1,001, tức là tốt hơn không có trọng số.

- **Xét tiêu chuẩn AIC:** Ước lượng hàm có trọng số trong cả hai phương pháp bình phương tối thiểu và Marquardt đều cho AIC tốt hơn, tức là nhỏ hơn rất nhiều. AIC nhỏ nhất ở hàm ước lượng theo phương pháp bình phương tối thiểu có trọng số là -1068; trong khi đó Marquardt có trọng số là -963; trong khi đó với phương pháp không có trọng số thì AIC = 7266 đối với phương pháp bình phương tối thiểu và 2171 đối với Marquardt. Như vậy khi đưa trọng số vào ước lượng mô hình, thì cả hai phương pháp đều cho giá trị AIC tối ưu.



Hình 5. Đồ thị quan hệ giữa giá trị ước lượng qua hàm và quan sát – Đồ thị biến động Residuals cho 4 phương pháp ước lượng hàm

- Xét trên đồ thị quan hệ giữa giá trị lý thuyết với giá trị quan sát cũng như biến động residuals theo giá trị ước lượng: Đồ thị của 4 phương pháp (hình 5) cho thấy trường hợp ước lượng theo phương pháp bình phương tối thiểu có trọng số là tốt nhất, cho đồ thị ước lượng và quan sát bám sát nhau và residuals biến động nhỏ và là hằng số; tiếp đến là phương pháp bình phương tối thiểu không trọng số; phương pháp Marquardt có trọng số cho biến động residuals nhỏ nhưng biến động lớn giữa lý thuyết và thực tế; kém nhất là ước lượng theo Marquardt không có trọng số, biến động residuals mở rộng, phân tán và giá trị lý thuyết với thực tế khá rộng.

- **Xét biến động S%:** Đây là biến động tương đối trung bình của giá trị ước lượng qua hàm với quan sát. Kết quả cho thấy ước lượng theo phương pháp bình phương tối thiểu có S% bé nhất là 32%, S% của ước lượng có trọng số của hai phương pháp có lớn hơn nhưng không đáng kể là 34 – 37%; cuối cùng S% lớn nhất ở phương pháp Marquardt không có trọng số là 95%.

Đánh giá tổng hợp 4 tiêu chuẩn R^2 , CF, AIC và S% và các đồ thị biến động residuals để lựa chọn phương pháp ước lượng hàm, cho thấy:

- Ước lượng hàm có sự tham gia của trọng số $Weight = 1/X_i^{2b_i}$ luôn cho AIC và CF bé nhất; có nghĩa là sai số ước lượng hàm (RSE) hoặc tổng bình phương phần dư (RSS) là nhỏ nhất; cho dù là hệ số quan hệ R^2 có thể là thấp hơn nhiều khi không có trọng số. Bên cạnh đó giá trị biến động giữa lý thuyết và thực tế S% cũng nhỏ và không sai khác nhiều khi không có trọng số. Điều này có kết luận rằng khi lựa chọn hàm và phương pháp ước lượng không nên dừng lại ở việc so sánh hệ số xác định R^2 , vì R^2 max nhưng biến động bình phương

phần dư (RSS) cũng rất lớn, làm cho hàm không phù hợp với thực tế. Do vậy khuyến nghị cần xem xét sử dụng phương pháp bình phương tối thiểu có trọng số trong xây dựng mô hình sinh trắc.

- Ước lượng theo phương pháp bình phương tối thiểu không có trọng số cho $R^2_{max} = 93,7\%$ và S% bé nhất trong tất cả các trường hợp = 32%; tuy nhiên AIC = 7266 là lớn nhất và lớn hơn rất nhiều so với có trọng số là -1068. Điều này cho thấy phương pháp này có hạn chế là làm cho RSS tăng (hay AIC) tăng, tức là tăng tổng biến động bình phương giữa quan sát và thực tế. Tuy vậy đây là phương pháp ước lượng hàm kinh điển, cho R^2_{max} , CF gần bằng 1 và S% là bé nhất. Do đó phương pháp này nên được sử dụng khi ước lượng các mô hình các biến số có quan hệ chặt, đối với mô hình quan hệ phức tạp, nhiều biến số thì nên áp dụng có trọng số để điều chỉnh sai số của mô hình.

- Ước lượng hàm phi tuyến theo phương pháp Marquardt với các tham số đầu vào dựa vào mô hình tuyến tính và không có trọng số cho thấy dù $R^2 = 92,8\%$ là cao, tuy nhiên AIC cũng cao = 2171 và đặc biệt là S% rất lớn = 95%. Do vậy khuyến cáo không nên sử dụng phương pháp này mà không có trọng số. Khi có trọng số cho dù R^2 giảm rất mạnh, chỉ còn 37,4%, nhưng AIC rất thấp = -963 và biến động giữa lý thuyết và thực tế rất tốt là 37%.

3.3. Kết quả lựa chọn hàm sinh trắc

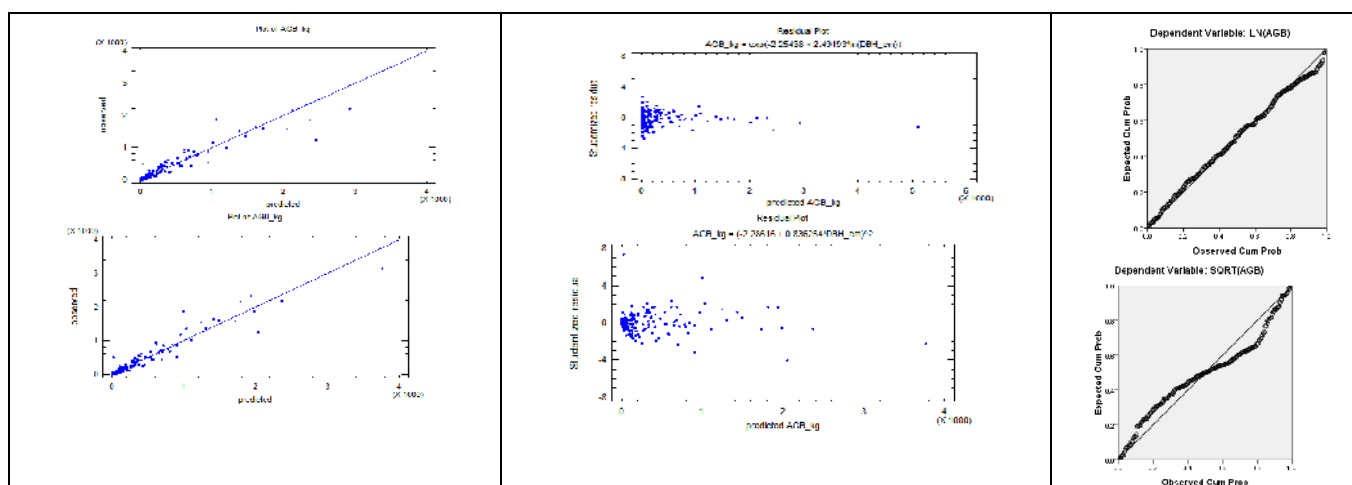
3.3.1. Hàm một biến số hoặc một tổ hợp tạo một biến

Đối với hàm một biến số hoặc một tổ hợp một biến, đầu tiên sử dụng $R^2_{adjusted}$ (điều chỉnh) là cơ sở để chọn các hàm có khả năng mô phỏng quan hệ, kiểm tra sự tồn tại các tham số phương trình với mức Pvalue < 0,05. Mỗi mô hình chọn 2-3 hàm có R^2 cao nhất, từ đó kết hợp các chỉ tiêu CF tiến đến 1, AIC bé nhất về đại số và S% bé nhất để lựa chọn hàm. Từ các chỉ tiêu tổng hợp trên, dựa vào chỉ tiêu S% làm chỉ tiêu chính kết hợp với thứ tự ưu tiên là AIC, CF và cuối cùng là $R^2_{adjusted}$ để chọn ra hàm tối ưu cho mô hình. Ngoài ra 2 đồ thị residuals và Normal P-P cũng được khảo sát để đánh giá sự phù hợp cũng như ít biến động của mô hình lựa chọn.

Kết quả thử nghiệm mô hình một biến $AGB = f(DBH)$ và tổ hợp thành một biến $DBH^2 \cdot H$ của mô hình ước tính các bon trong lá: $Cl = f(DBH^2 \cdot H)$ được trình bày ở bảng 3 và hình 6.

Bảng 3. Kết quả thử nghiệm lựa chọn mô hình ước tính sinh khối AGB dạng một biến (DBH) hoặc các biến trong lá (CI) theo tổ hợp một biến (DBH²*H)

Dạng hàm	Hàm	R ² (%)	R ² adj (%)	P	n	Pbi	CF	AIC	S%	Mã hàm
AGB = f(DBH)	$AGB = (-2,28616 + 0,836254*DBH)^2$	94,4	94,3	0,00	224	0,00		2063,3	34,1%	3.1
	$AGB = \exp(-2,25438 + 2,49193*\ln(DBH))$	93,7	93,7	0,00	224	0,00	1,089	7266,3	32,1%	3.2
CI = f(DBH, H)	$\ln(CI) = -11,0979 + 4,10521*\sqrt{\ln(H*DBH^2)}$	77,6	77,4	0,00	149	0,00	1,266	9593,64	66,0%	3.3
	$\ln(CI) = -5,36351 + 0,725866*\ln(H*DBH^2)$	77,1	77,0	0,00	149	0,00	1,272	9796,14	67,4%	3.4

**Hình 6. Đồ thị quan hệ lý thuyết – quan sát, residuals và Normal P=P của 2 mô hình AGB=f(DBH)**

Kết quả thử nghiệm 2 dạng mô hình AGB = f(DBH) ở bảng 3 như sau:

$$\sqrt{AGB} = a + b \cdot DBH \quad 3.1$$

$$\ln(AGB) = a + b \cdot \ln(DBH) \quad 3.2$$

Cho thấy:

- Hệ số R²_{adj} của mô hình 3.1 cao hơn 3.2 và cả hai đều tồn tại ở mức P < 0,05.

- Tiêu chuẩn CF và AIC không được sử dụng vì biến số y của hai hàm khác nhau: sqrt(AGB) và ln(AGB).

- S% của mô hình 3.2 thấp nhất là 32,1%.

- So sánh đồ thị sai số phần dư của mô hình 3.2 cho thấy đồng đều hơn mô hình 3.1. Đồ thị Normal P-P cũng chỉ ra mô hình 3.2 bám sát đường chéo (0,1) hơn so với mô hình 3.1. Có nghĩa mô hình 3.2 mô phỏng tốt hơn so với mô hình 3.1.

Như vậy mặc dù mô hình 3.1 có hệ số quan hệ R² cao hơn, nhưng qua đánh giá thêm các tiêu chuẩn S%, các đồ thị residuals và Normal P-P cho thấy hàm 3.2 dạng logarit là tốt hơn vì vậy hàm này được lựa chọn để ước tính AGB qua biến số DBH.

Cũng ở bảng 3, trường hợp tổ hợp DBH²*H tạo thành một biến ở dạng hàm CI=f(DBH, H), lựa chọn

được 2 mô hình có R² cao nhất và thứ hai để so sánh. Vì hai hàm đều có biến y là dạng logarit nên CF và AIC được đưa vào so sánh. Kết quả chỉ ra mô hình 3.3 đều có R_{max}, CF_{min}, AIC_{min}, S%_{min} so với mô hình 3.4, vì vậy mô hình lựa chọn tối ưu là mô hình 3.3.

Tóm lại, nguyên tắc cơ bản để lựa chọn mô hình một biến cần phải dựa vào nhiều chỉ tiêu thống kê, trong đó biến động trung bình của mô hình S% là chỉ tiêu quan trọng để kiểm tra mức sai số của mô hình và quyết định lựa chọn mô hình tối ưu, các chỉ tiêu R²_{adj} cao nhất, CF_{min} và AIC_{min} là các chỉ tiêu cần để phối hợp lựa chọn mô hình tối ưu tiệm cận với dữ liệu thực tế. Cùng với nó là 2 đồ thị residuals và Normal P-P để trực quan hóa biến động sai số của mô hình lựa chọn.

3.3.2. Hàm nhiều biến số hoặc nhiều tổ hợp biến

Về nguyên tắc hàm nhiều biến, tổ hợp biến đều được lựa chọn như hàm một biến, chỉ khác là cần kiểm tra sự tồn tại của từng tham số bị gắn từng biến số theo tiêu chuẩn t ở mức P < 0,05; đồng thời các dạng hàm được tạo thành trên cơ sở tổ hợp đa dạng với nhiều dạng đổi biến số của hai

nhóm biến y_i và x_j theo các dạng cơ bản là logarit (y_i, x_j), $\sqrt{y_i, x_j}$, $x_i^2, y_j^2, 1/x_j, 1/y_i, \dots$

Quá trình này được chạy thử nghiệm nhiều hàm với nhiều tổ hợp đổi biến số, sau đó lựa chọn 2-3 mô hình tiềm năng với R^2_{adj} cao nhất, các tham số tồn tại ở mức $P < 0,05$; sau đó sử dụng các tiêu chuẩn AIC, CF, S% và các đồ thị để quyết định lựa chọn mô hình phù hợp nhất.

Kết quả lựa chọn hàm so sánh ở các mô hình tổng hợp trong bảng 4. Quan hệ $AGB = f(DBH, H)$ dò tìm được hai hàm tốt nhất để so sánh và lựa chọn là mô hình 3.5 và 3.6:

$$\ln(AGB) = -3,07831 + 2,01893 \cdot \ln(DBH) + 0,8262 \cdot \ln(H) \quad (3.5)$$

$$\ln(AGB) = -1,83496 + 2,03093 \cdot \ln(DBH) + 0,0593816 \cdot H \quad (3.6)$$

Bảng 4. Kết quả thử nghiệm lựa chọn mô hình ước tính sinh khối AGB dạng nhiều biến đổi biến số

Dạng hàm	Hàm	R^2_{adj} (%)	P	n	Pbi	CF	AIC	S%	Mã hàm
AGB = f(DBH, H)	$\ln(AGB) = -3,07831 + 2,01893 \cdot \ln(DBH) + 0,8262 \cdot \ln(H)$	94,9	0,0	224	0,0	1,071	5592,4	28,2%	3,5
	$\ln(AGB) = -1,83496 + 2,03093 \cdot \ln(DBH) + 0,0593816 \cdot H$	94,7	0,0	224	0,0	1,075	5910,2	29,0%	3,6

Thảo luận về cơ sở khoa học để xây dựng mô hình sinh trắc:

Trên cơ sở nghiên cứu lựa chọn biến số tham gia mô hình, đánh giá các phương pháp ước lượng mô

Mô hình 3.5 có R^2_{adj} (%) lớn hơn mô hình 3.6, hệ số CF của hai mô hình xấp xỉ nhau, qua tính toán tiêu chuẩn AIC cho kết quả AIC của mô hình 3.5 nhỏ hơn mô hình 3.6 về đại số, do đó lựa chọn mô hình tối ưu ở đây sẽ là mô hình 3.5. Kết quả kiểm tra sai số S% cho thấy biến động sai số mô hình 3.5 cũng thấp hơn.

Quan sát Normal P-P của hai mô hình cho thấy, phân bố dữ liệu thực nghiệm và lý thuyết bám đều trên đường chéo (0, 0) và (1, 1) của đồ thị, sai lệch biến động ở hai mô hình thấp và biến động như hằng số; nên hai đồ thị gần giống nhau.

Kết quả cho thấy hai hàm khá tương đồng về các chỉ tiêu thống kê, tuy nhiên mô hình 3.5 được lựa chọn vì toàn bộ các tiêu chuẩn đánh giá đều đạt hơn.

hình và lựa chọn dạng hàm cho quan hệ sinh khối, các bon với các nhân tố điều tra; tổng hợp thành cơ sở để tiếp cận xây dựng mô hình sinh trắc ở hình 7 được tóm tắt như sau:

- **Lựa chọn biến số tham gia mô hình:** Để lựa chọn số biến số tham gia mô hình tốt nhất trong trường hợp có nhiều biến chưa rõ ảnh hưởng đến y thì cần sử dụng tiêu chuẩn Mallows' Cp. Chỉ số Cp càng gần với số biến số p thì mô hình càng phù hợp; dựa vào đây để xác định số biến số p tham gia mô hình khi có quá nhiều biến số được giả định là có ảnh hưởng đến y .

- **Phương pháp ước lượng hàm quan hệ:** Để có được một mô hình với các tham số có độ tin cậy, phù hợp với thực tế thì khi ước lượng hàm cần thử nghiệm cả 4 phương pháp ước lượng là bình phương tối thiểu, bình phương tối thiểu có trọng số, ước lượng hàm phi tuyến tính Marquardt và phương pháp Marquardt có trọng số. Riêng trường hợp nghiên cứu này của nghiên cứu, bốn phương pháp được thử nghiệm cho lập mô hình sinh trắc của đối tượng nghiên cứu cho thấy: i) Phương pháp bình phương tối thiểu không có trọng số có R^2_{adj} cao nhất và S% thấp nhất, tỏ ra phù hợp với quy luật phát triển sinh khối cây rừng; ii) Phương pháp ước lượng hàm có trọng số cả bình phương tối thiểu lẫn Marquardt trong nhiều trường hợp cho AIC bé nhất và biến động phần dư residuals thấp

và là hằng số; do đó ước lượng có trọng số cũng là một phương pháp tốt giúp cho việc điều chỉnh các tham số của mô hình thích hợp với dữ liệu quan sát, giảm biến động.

- *Lựa chọn dạng hàm trên cơ sở các tiêu chuẩn thống kê:*

Dạng hàm cần được xây dựng trên cơ sở tổ hợp các biến số phụ thuộc và các biến độc lập được chuyển đổi thành các dạng hàm logarit, sqrt, mũ, $1/x$, ... sẽ hình thành nên nhiều dạng hàm để thử nghiệm nhằm lựa chọn được mô hình thích hợp, hơn là sử dụng những hàm cố định, có sẵn như power, polynomial, logarit, ...

Việc lựa chọn hàm thích hợp cần dựa vào tổng hợp 7 tiêu chuẩn (Được khái quát ở hình 7) :

- R^2 adjusted (%): Về nguyên tắc R^2 max thì có thể định hướng đây là hàm tốt, tuy nhiên nhiều trường hợp có trọng số thì R^2 max lại cho hàm có biến động lớn hơn. Trong khi đó trong thực tế lập hàm phổ biến, thường chỉ dựa vào R^2 . Do đó cần khảo sát các tiêu chuẩn và đồ thị biến động để cân nhắc chọn mô hình phù hợp là điều cần thiết.

- Tiêu chuẩn t kiểm tra sự tồn tại của các tham số đối với hàm đa biến, với $P < 0,05$.

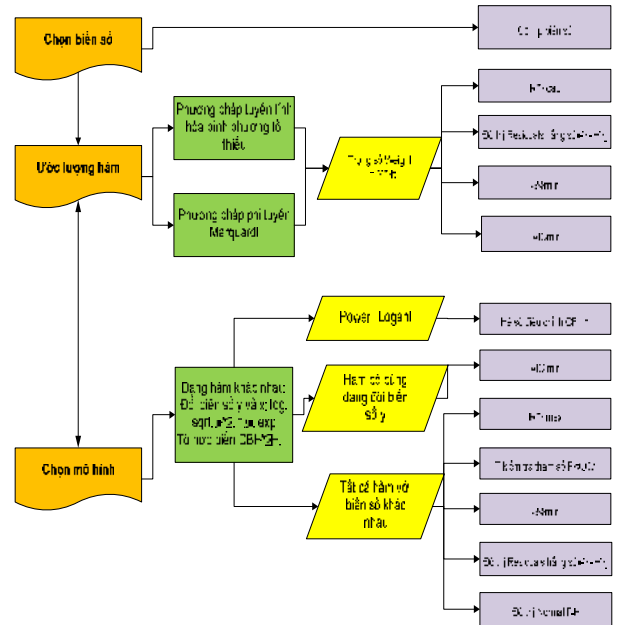
- Nhân tố điều chỉnh (CF- Correction factor): CF luôn lớn hơn 1. Khi sai tiêu chuẩn của phần dư RSE tiến đến 0 thì CF tiến đến 1, mô hình có ước lượng tốt. Mặc khác CF là hàm phân bố chuẩn của hàm logarit neper, do đó chỉ sử dụng tiêu chuẩn này trong trường hợp biến y của mô hình cùng dạng đổi biến số là logarit neper ($\ln(y)$).

- Tiêu chuẩn AIC: Khi cần lựa chọn mô hình tốt nhất với nhiều mô hình khác nhau với các biến số ảnh hưởng khác nhau, mô hình tối ưu với các biến số thích hợp khi giá trị đại số của AIC là bé nhất. Tiêu chuẩn này áp dụng trong trường hợp biến y đồng nhất cách đổi biến.

- Biến động trung bình giữa lý thuyết và quan sát S%: Là chỉ tiêu cho biết mô hình có sai khác với thực tế cao hay thấp và mô hình tối ưu khi sai khác này bé nhất.

- Đồ thị Normal P-P cho biết khả năng mô phỏng của các mô hình với dữ liệu thực tế. Mô hình tối ưu là mô hình có phân bố chuẩn của giá trị lý thuyết và quan sát bám sát nhau và nằm trên đường chéo có tọa độ (0, 0) và (1,1) của đồ thị.

- Đồ thị biến động phần dư residuals theo giá trị lý thuyết: Biến động này tốt khi nó là hằng số nằm trong phạm vi $-2 + 2$ theo dãy giá trị lý thuyết.



Hình 7. Sơ đồ tiếp cận xây dựng mô hình sinh trắc

Việc chọn lựa một hàm cần theo 7 tiêu chí nói trên, trong đó ưu tiên là giá trị S% bé nhất, vì nó phản ánh sai lệch giữa quan sát và ước lượng qua hàm nhỏ nhất; tiếp theo cần ưu tiên tiêu chuẩn AIC_{min} và CF = 1; R^2 adj chỉ là tiêu chí tham khảo đầu tiên để định hướng chọn hàm hơn là chỉ tiêu quyết định; hai đồ thị Normal P-P và residuals là khảo sát trực quan nhất để thấy rõ sự thích hợp của hàm ước lượng.

4. KẾT LUẬN VÀ KIẾN NGHỊ

4.1. Kết luận

Lựa chọn biến số tham gia mô hình sinh trắc theo chỉ số Cp tiến đến p biến số, đó là số biến số tối ưu của mô hình. Đối với ước tính sinh khối và các bon rừng lá rộng thường xanh với sự tham gia của 4 biến số DBH, H, WD và Ca sẽ đạt độ tin cậy cao.

Ước lượng hàm số: Phương pháp bình phương tối thiểu có R^2 adj cao và S% thấp, tỏ ra phù hợp với quy luật phát triển sinh khối cây rừng. Đồng thời ước lượng có trọng số cũng là một phương pháp tốt giúp cho việc điều chỉnh các tham số của mô hình thích hợp với dữ liệu quan sát, giảm biến động.

Lựa chọn dạng hàm thích hợp trên cơ sở 07 tiêu chuẩn thống kê: R^2 adj đạt cao, T kiểm tra tham số với $P < 0,05$, CF tiến đến 1, AIC bé nhất, S% bé nhất, đồ thị Normal P-P bám sát đường chéo có tọa độ (0,0) và (1,1), đồ thị biến động phần dư (residuals) là hằng số theo giá trị ước lượng qua hàm. Trong đó ưu tiên giá trị S% bé nhất, tiếp theo là AIC_{min} và CF = 1; R^2 adj chỉ là tiêu chí tham khảo đầu tiên; hai đồ thị Normal P-P

và residuals là khảo sát trực quan nhất để thấy rõ sự thích hợp của hàm ước lượng.

4.2. Kiến nghị

- Sử dụng cơ sở khoa học của nghiên cứu để thiết lập các mô hình sinh trắc cho các kiểu rừng, vùng sinh thái khác ở Việt Nam.

- Đánh giá độ tin cậy của các mô hình thiết lập ở ngoài vùng thu thập số liệu để đánh giá độ tin cậy khi áp dụng.

TÀI LIỆU THAM KHẢO

1. Võ Đại Hải, 2008. *Nghiên cứu khả năng hấp thụ và giá trị thương mại các bon của một số dạng rừng trồng chủ yếu ở Việt Nam*. Báo cáo tổng kết đề tài Viện Khoa học Lâm nghiệp Việt Nam.
2. Bảo Huy, 2009. *Phương pháp nghiên cứu ước tính trữ lượng các bon của rừng tự nhiên làm cơ sở tính toán lượng CO₂ phát thải từ suy thoái và mất rừng ở Việt Nam*. Tạp chí Nông nghiệp và Phát triển nông thôn. Bộ Nông nghiệp và PTNT số 1/2009: 85-91.
3. Bảo Huy, Nguyễn Thị Thanh Hương, Võ Hùng, Cao Thị Lý, Nguyễn Đức Định, Huỳnh Nhân Trí và cộng tác viên, 2012. *Xác định lượng CO₂ hấp thụ của rừng lá rộng thường xanh làm cơ sở tham gia chương trình giảm thiểu phát thải từ suy thoái và mất rừng*. Báo cáo tổng kết đề tài khoa học công nghệ trọng điểm cấp Bộ. Bộ Giáo dục và Đào tạo.
4. Bảo Huy, Nguyễn Thị Thanh Hương, Võ Hùng, Cao Thị lý, Nguyễn Đức Định, 2012. *Xây dựng mô hình sinh trắc ước tính sinh khối và các bon cho rừng lá rộng thường xanh vùng Tây Nguyên*. Tạp chí Rừng và Môi trường. Số 51 (2012): 21-30. ISSN 1859 – 12483.
5. Bảo Huy, 2013. *Mô hình sinh trắc và viễn thám – GIS để xác định CO₂ hấp thụ của rừng lá rộng thường xanh vùng Tây Nguyên*. Nxb Khoa học và Kỹ thuật.
6. Vũ Tấn Phương, 2012. *Xác định trữ lượng các bon và phân tích hiệu quả kinh tế rừng trồng thông ba lá (Pinus keysia Royle ex Gordon) theo cơ chế phát triển sạch ở Việt Nam*. Luận án Tiến sĩ.
7. Basuki, T. M., P. E. van Laake, Skidmore, A. K., Hussi, Y. A., 2009. *Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests*. Forest Ecology and Management 257 (2009) 1684–1694.
8. Huy, B., Anh, P. T., 2008. *Estimating CO₂ sequestration in natural broad-leaved evergreen forests in Vietnam*. Asia-Pacific Agroforestry Newsletter – APANews, FAO, SEANAFE; No.32, May 2008, ISSN 0859-9742.
9. Huy, B., Hung, V., Huong, N. T. T., Ly, C. T., Dinh, N. D., 2012. *Tree allometric equations in Evergreen Broadleaf Forests in the South Central Coastal region, Viet Nam*. In (Eds) Inoguchi, A., Henry, M. Birigazzi, L. Sola, G., Tree allometric equation development for estimation of forest above-ground biomass in Viet Nam. UN-REDD Programme, Hanoi, Viet Nam.
10. Chave, J., Andalo, C., et al., 2005. *Tree allometry and improved estimation of carbon stocks and balance in tropical forests*. Oecologia (2005) 145: 87-99. DOI 10.1007/s00442-005-0100-x.
11. Mallow, C. L., 1973. *Some Comments on CP. Technometrics* 15 (4): 661–675. doi:10.2307/1267380. JSTOR 1267380.
12. Picard, N., Saint-Andre, L., Henry, M., 2012. *Manual for buiding tree volume and biomass allometric equations*. CIRAD, FAO.
13. Phuong, V. T., Inoguchi, A., Birigazzi, L., Henry, M., Sola, G., *Introduction and Background of the study, Vietnam*. In (Eds) Inoguchi, A., Henry, M. Birigazzi, L. Sola, G. Tree allometric equation development for estimation of forest above-ground biomass in Viet Nam (Part A). UN-REDD Programme, Hanoi, Viet Nam.

METHODOLOGY FOR DEVELOPING FOREST BIOMASS – CARBON ALLOMETRIC EQUATIONS

Huynh Nhan Tri, Bao Huy**Summary**

To estimate biomass, forest carbon, the allometric equations are an important content. However, the scientific basis for developing models has not been evaluated systematically to provide common standards. This study based on the collection of biomass, carbon analysis from felling trees in the evergreen broad-leaved forest in the Central Highlands, was analyzed to provide the basis for the construction of biometric models. The results show that: i) Select the optimal number of variables for the model under criteria of Mallow Cp, indicating to achieve the required reliability has four variables DBH, H, WD (wood density) and CA (crown area); ii) estimate the parameters with least square method, and if the weight used for estimating, models can fit better with observational data; iii) select the appropriate function based on 7 criteria R^2 (%) high, T standard testing parameters with $P < 0.05$, CF approaching 1, the smallest AIC, S% smallest, Normal P-P plot is line with coordinates (0.0) and (1.1), the scatter of residual is constant over the estimated value.

Key word: *Allometric equation, biomass, evergreen broad-leaved forest, forest carbon.*

Người phản biện: PGS.TS. Trần Quang Bảo

Ngày nhận bài: 6/12/2013

Ngày thông qua phản biện: 6/01/2013

Ngày duyệt đăng: 14/01/2013