Estimation of Shrub Biomass: Development and Evaluation of Allometric Models Leading to Innovative Teaching Methods

Tek Narayan Maraseni[1], Geoff Cockfield[2], Armando Apan[3], Nicole Mathers[4]

[1]*PhD Candidate, University of Southern Queensland (USQ), Toowoomba QLD 4350, Australia, Email: w0007649@mail.connect.usq.edu.au*

[2]*Head of Dept, Economics & Resources Management, USQ*

[3] *Senior Lecturer, Remote Sensing & GIS, USQ*

[4]*Soil Scientist, CRC Greenhouse Accounting, DNRM, Queensland, Australia*

Abstract

Accurate estimation of biomass is becoming vital for selling carbon into national and international markets. Being a dry continent, Australia's natural forest has several shrub species. However, because of limited availability of methodology and difficulty in estimation they are unaccounted for in many cases. This paper has three objectives: (a) to address the major problem in multiple regressions, (b) to develop the best allometric equation for the biomass estimation of a popular shrub species, wild raspberry *(Rubus probus)* and (c) to prepare a teaching tool, by following systematic and logical steps, for biomass estimation using ForecastX™ software. We identified the possible explanatory variables, by discussing with experts and citing literature, for shrub biomass and then measured them by destructive sampling at Taabinga, near Kingaroy, Queensland. Our research suggests that careful analysis of correlation matrices gives very important clues to which variables we should select and which we should not for the models. High multicollinearity among the independent variables is a major problem in multiple regressions. This study shows that this problem could easily be solved by using basic scientific formula and applying a single variable instead of applying many highly correlated variables in the model. Unlike most statistical books, our analysis does not suggest to reject that variable from the model whose coefficient is not statistically significantly different from zero as it could be highly influential in another set of combination. Similarly, we recommend using the 'intercept' even if its value is not significantly different with zero as it does not cost extra money to be included but it does help the predictive power of the model. Although we developed a range of biomass prediction models (for wild raspberry) that can be used in different circumstances, our first recommendation is for the model which is based on girth and crown volume. Where cost is the major issues, we prefer the model which employs girth and crown area, as it gives a good result and needs only three variables to be measured. These findings can be helpful in teaching the practical applications of multiple regression in courses such as Data Analysis and Business Forecasting.

Key words: shrub, wild raspberry, biomass, multiple regression, multicollinearity.

Introduction

Biomass is an important attribute of vegetation for several reasons. Biomass of rangelands and pasturelands is necessary for carrying capacity estimation. Biomass of crops is essential for land productivity prediction. Similarly, biomass of legume is crucial for biological nitrogen fixation and nitrous oxide emission estimations, and biomass of herbs, shrubs and trees is vital for herbivores and therefore for their subsequent predators and ecosystem management. More importantly, after the Kyoto Protocol entered into force, in 2005, the demand for 'carbon credits' has been escalating in the international market. In this circumstance, an accurate estimation of biomass is becoming vital for selling carbon into national and international markets.

There are a range of methods (from aerial photography and imagery to destructive sampling) for biomass estimation. Several general and local biomass tables and species-specific allometric equations are developed for this reason. However, these are limited to tree species. Shrub biomass is an important component of the total biomass, especially in the natural forest. However, because of the lack of methodology and difficulty in calculation, in some cases, they are omitted (Karki, 2002; Khanal, 2001), which results in underestimation of the total biomass. In other cases, the basal area of shrub (π x diameter^2/4) is estimated as cross-sectional area of a tree (by measuring diameter) and then biomass is calculated by applying the same formula of tree biomass estimation (Palm, 2003), which may over or under estimate the biomass depending on species attributes.

In Australia, there are dozens of species-specific allometric equations available. For example, Specht and West (2003) developed six different equations for six popular tree species (*Eucalyptus microcorys, E grandis, E. saligna, E. nitens, Grevillea robusta and Pinus radiata*) of New South Wale and Margules Poyry Pty Ltd (1997) developed 11 equations for 11 morphological groups of tree species found in the South East Queensland Regional Forest Agreement (SEQRFA) region. Similarly, Scanlan (1991) developed an allometric equation for *Acacia harpophylla*, Burrows *et al*. (1999) developed one for *Eucalyptus melanopholoia* and Harrington (1979) developed one for *Eucalyptus populnea.* Some equations are based on diameter at breast height (DBH) as an independent variable whereas others are based on both DBH and height.

There are some limitations to the application of these equations. Firstly, equations developed for the same species in different sites would be different because of different climatic and edaphic conditions (Mohns *et al*., 1988). For instance, the equations developed for the same species, *Eucalyptus resinifera,* at two different sites are different and so is the case of *Acacia aneura* (see Harrington, 1979 and Burrows et al., 1999 for Acacia species and Ward and Pickersgill, 1985 for *Eucalyptus* species). Secondly, in some cases, the range of applicable diameters (and/or heights) is not quantified (e.g., Specht and West, 2003). In others, the range of diameters (and/or heights) is too narrow to apply even for the same species found in the same locality (e.g., Burrows et al., 1999; Margules Poyry Pty Ltd, 1997).

There are also some allometric equations developed for biomass estimation of some shrub species. Since they are based on only one response variable, for example, total height in the case of *Geijera parviflora*, and shoot height in the cases of *Casea nemophila*, *Dodonaea viscose, Eremophila bowmannii* and *Acacia aneura*, their accuracy may be questionable (Harrington, 1979). Developing more accurate allometric equations for shrub biomass estimation in Australia is essential for at least two reasons. Firstly, being a dry shrub species are dominant in the forests. The accelerating ‘woody thickening’, grasslands are becoming woodlands and woodlands are becoming thicker (Mitchell and Skjemstad, 2004; Burrows et al., 2002), has heightened its importance. Secondly, payments to land managers for different local and regional environmental services have also been initiated in Australia (Cacho *et al*., 2003).

This paper presents a method for the development and improvement of allometric equations for the estimation of biomass of a popular shrub species, wild raspberry’s (*Rubus probus*). We have followed all steps of model development and improvement sequentially and logically. Thus, it could be used as a teaching tool for biomass estimation using multiple regression and ForescastX^{TM} software. This software has been adopted in statistics related courses in more then 25 universities around the world (Wilson *et al*., 2001) including the University of Southern Queensland, Australia (Pensiero *et al*., 2005). The lesson learned will help with teaching the practical applications of multiple regression in courses such as Data Analysis and Business Forecasting.

Study area overview

The study area, shrub-land (E 375842 and N 70500067 at 501 m asl) in the Marshall Property at Taabinga, near Kingaroy, Queensland, lies below the Tropic of Capricorn. Therefore, it is classified as subtropical. Surprisingly, more than 90% of the shrub in the study area is of the research species, wild raspberry. The study area has hot and long summers and mild winters. Mean daily maximum temperature ranges from 27^0 to 33^0C in December to 18^0 to 22^0C in July (Smith and Kent, 1993). Rainfall is highly variable but dominant in summers with about 70% falling between October and March (Mills and Schmidt, 2000). Frost is a common phenomenon. June, July and August are the coldest months averaging 24 heavy plus 22 light frosts a year (Mills and Schmidt, 2000). The soils were sampled in April 2005, using the methodology of McKenzie et al (2000). The analysis shows that the soil in the study area is 'Krasnozem' (Red Ferrosol, Isbell, 2002) and is clayey (33-55 percent clay) and acidic (4.5-5.5 pH) in nature.

Model development and evaluation

Under this Section, a series of logical and stepwise procedures followed for the development and evaluation of multiple regression models for shrub (wild raspberry) biomass estimation are presented. In order to give a clear picture to readers, in many instances, a great deal of detailed discussion is necessary.

Identification of independent variables

The biomass of shrub (dependent variable) depends on many independent (explanatory) variables. We cited many publications and discussed with experts to explore the range of explanatory variables. While exploring, we were fully aware that there should be a 'cause and effect relationship' between the dependent and independent variables. The identified independent variables were total height (ht, height from the ground or collar region to the top of the leading shoot), girth at 10cm height (G10), crown height (Cr-ht, height from ground to the height of first crown origination point), crown length (Cr-L, diameter of crown at larger side), crown breadth (Cr-B, diameter of crown at smaller side) and crown depth (Cr-D, total height of tree minus the crown height). See picture of spotted gum tree (Figure 1) for general idea of different attributes.

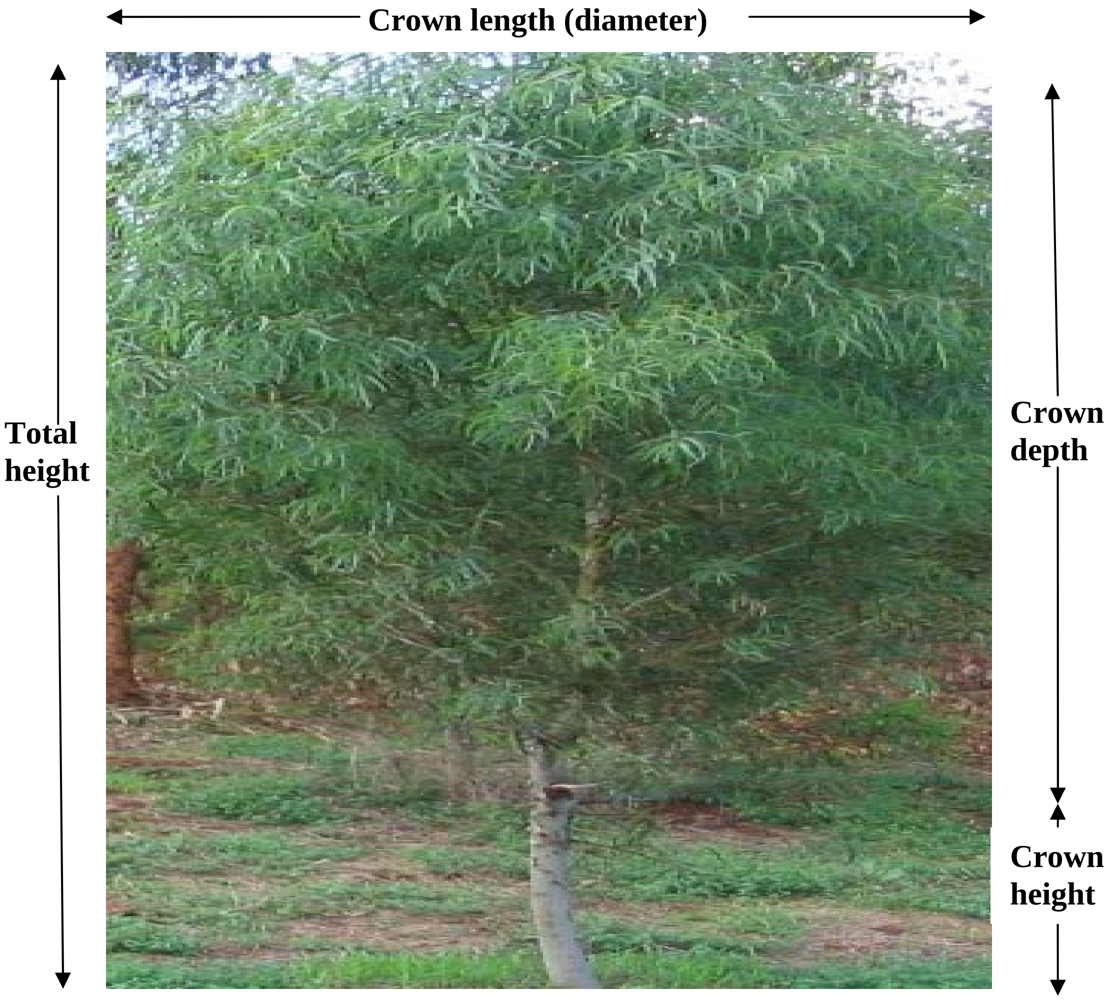

Fig 1: Showing different attributes

Field measurements of dependent and independent variables

Wild raspberry (shrub) is a shade-loving woody perennial plant. Unlike tree species, they have no well-defined stem. A range of methods could be used for the estimation of its biomass, such as line and planner intersect method, calibrated visual estimation method and double sampling using regression estimators method (Catchpole and Wheeler, 1992), but destructive sampling is the most promising method (Snowdon et al., 2002). Since it is the only method by which we can get all the required independent variables, we used this method.

Table 1: Measured and calculated attributes of wild raspberry

ht	G10	Cr-ht	Cr-L	Cr-B	Cr-A	Cr-D	Cr-V	W
270	14.3	46	190	140	26600	224	5958400	3950
205	6.8	20	115	80	9200	185	1702000	1050
140	3.8	35	86	86	7396	105	776580	150
201	6.3	40	101	91	9191	161	1479751	750
131	7.3	40	131	92	12052	91	1096732	1250
155	4.8	12	129	61	7869	143	1125267	350
145	4.3	12	59	49	2891	133	384503	175
103	2.8	15	58	49	2842	88	250096	75
168	4.7	12	79	61	4819	156	751764	201
185	6.3	44	122	73	8906	141	1255746	625
101	2.3	24	40	31	1240	77	95480	100
331	10.3	49	179	140	25060	282	7066920	3690

Note: ‘G10’ is girth (in cm) at 10 cm, ‘W’ fresh weight in gram, ‘ht’ is total height in cm, and ‘Cr-ht’, ‘Cr-L’, ‘Cr-B’, ‘Cr-D’, ‘Cr-A’, ‘Cr-V’ are crown height (cm), crown length (cm), crown breadth (cm), crown depth (cm), crown area (cm^2) and crown volume (cm^3) of wild raspberry respectively

We cut the different sizes of wild raspberry and measured total height (ht), girth at 10cm height (G10), crown height (Cr-ht), crown length (Cr-L), crown breadth (Cr-B) and total weight (W) of each. Crown depth (Cr-D) was calculated later from crown height and total height (Table 1). Similarly, the crown area and crown volume were calculated later by using basic scientific relationship between crown length, crown breadth and crown depth (crown area, Cr-A =crown length X crown breadth, and crown volume, Cr-V= crown length X crown breadth X crown depth).

Software selection

We selected ForecastX^{TM} software (Wilson et al. (2002) for development and evaluation of different regression equations due to two main reasons. Firstly, this software has been used in many universities and one of the objectives of this paper is to teach students how to develop good regression model using this software. Secondly, it gives many accuracy measures (such as Mean Error, Mean Absolute Percentage Error, Mean Percent Error, Sum of Squares Error, Mean Square Error, Root Mean Square Error, Coefficient of Determination, Theils' U-statistics etc) of each regression model, thus, it could be easier to select the best model.

Scatter diagram (using Excel)

After precisely measuring the required data for dependent and independent variables and selection of software, we entered all data in Excel. A scatter diagram was plotted and the trends (relationship) of each independent variable with the dependent variables were found. We found good relationships between each of the measured and calculated variables with the biomass (weight) of wild raspberry. The scatter diagram informed us, in advance, whether the relationship between dependent and independent variable was positive (or negative) and what sign (positive or negative) should be expected for the coefficient of that particular variable in the model.

Correlation matrix between dependent and independent variables

Correlation coefficients show the degree of relationship between two variables. Plotting correlation coefficient values of each independent variable with a dependent variable and between independent variables themselves, gives a matrix, which is called a correlation matrix. This is the most important step that gives us valuable clues for selecting the appropriate explanatory variables and, therefore, to develop the most promising regression model. After entering the data into Excel we ran the software and found the correlation coefficient between each independent variable with dependent variable (Table 2). The correlation coefficient values under the column 'W' show that there is good positive correlation between all independent variables with the dependent variables. However, the explanatory power of G10 (girth at 10 cm height), Cr-A (crown area) and Cr-V (crown volume) to the dependent variable are significantly higher than others. It means, if there is no multicollinearity between these three variables then the combination of these three variables should give the best regression model.

Table 2: Correlation matrix

variables	ht	G10	Cr-ht	Cr-L	Cr-B	Cr-A	Cr-D	Cr-V	W
ht	1.00	0.84	0.60	0.83	0.87	0.88	0.98	0.94	0.88
G10	0.84	1.00	0.67	0.92	0.91	0.95	0.78	0.89	0.95
Cr-ht	0.60	0.67	1.00	0.67	0.78	0.73	0.43	0.65	0.68
Cr-L	0.83	0.92	0.67	1.00	0.91	0.94	0.77	0.86	0.88
Cr-B	0.87	0.91	0.78	0.91	1.00	0.97	0.79	0.91	0.91
Cr-A	0.88	0.95	0.73	0.94	0.97	1.00	0.82	0.96	0.98
Cr-D	0.98	0.78	0.43	0.77	0.79	0.82	1.00	0.90	0.82
Cr-V	0.94	0.89	0.65	0.86	0.91	0.96	0.90	1.00	0.97

Test for individual coefficients (t-test)

Sometimes, especially when the sample size is small, even if we get high correlation between dependent and independent variables the relationships may not be statistically significant. For this particular reason, we performed a t-test (p-value test) which indicated to us whether or not the correlation between a particular independent variable with the dependent variable was statistically significant. In all cases, except in one variable (crown height), we obtained very high calculated t-values and lower p-values (lower than 0.05). Therefore, our alternative hypothesis 'correlation between dependent and independent variable is statistically significantly different with zero' was accepted.

Test for multicollinearity

Multicollinearity is the strong correlation between the independent variables**.** Some correlation between them is highly expected as all variables are 'causing factors' to shrub biomass. However, 'strong correlation' causes 'strong multicollinearity' by which the true effect of estimated regression coefficients would be lost. As a result, we can no longer rely on standard statistical tests. If that is the case, that variable which has less explanatory power with the dependent variable should be removed from the regression model. But while doing so, there should not have strong collinearity between the remaining variable with other potential explanatory variables.

In this case, there was strong collinearity between crown area and crown volume (r=0.96). As a principle, we would have not considered the crown-volume from consideration as its correlation with weight (0.97) is lower than that of the correlation of crown area and weight (0.98). But girth at 10cm height (G10) is another potential explanatory variable (as its correlation with weight is 0.95) and there is better correlation between G10 and area (r=0.95) than G10 and volume (r=0.89). So, by principle, we should have rejected crown-area and not crown-volume. However, since both variables have highly competitive correlation with weight and also have minor differences in correlation with G10 we decided to develop all possible models, selecting the best one on the basis of analysis of all accuracy measures.

Evaluation of the models

We ran the software and developed many possible regression models (Table 3). Then, the following measures were applied to select the best model.

Check the sign of each coefficient

We checked the sign of each of the coefficients of the independent variables to know whether the sign on the regression model was contradicting with our earlier expectation (compared with scatter plot and correlation). In some models, the sign of coefficient of total height, crown height, crown length and crown depth were not matching. On these grounds, we rejected the models from first to sixth (Table 3).

Analysis of the accuracy measures

Accuracy measures play a major role in the selection of the best model as they deal with the error term. In any prediction, the aim is to reduce the error by reducing the gap between the actual and predicted values. In regression, this is dealt with using least square error. That means the regression line should be plotted in such a way so that the sum of the deviation of actual values from the regression line comes to a minimum. ForecastX^{TM} gives many accuracy measures, but in this case, we are using three reliable accuracy measures; root mean square error, Theils' U-statistics and the adjusted coefficient of determination. We choose root mean square error because it is easier to interpret and explain and is also most compatible with statistical concepts of standard deviation. Theils' U-statistic tells us how good our model is in comparison to the naïve approach.

The coefficient of determination (R^2) tells us the percentage of the variation in the dependent variables explained by the independent variables. In fact, every additional explanatory variable increases R^2 (Makridakis *et al*., 1998), as it does not take the degrees of freedom into account. It is notable that every additional variable would cost money to measure/collect, and therefore, all explanatory variables should not be included in the model. Therefore, in the case of multiple regressions, adjusted R^2 (rather than R^2) would be used, as it considers the degrees of freedom (Wilson et al., 2002). The regression model is considered more accurate if it gives a lower value of root mean square error (RMSE) and Theils' U-statistics (close to the zero) accompanied by a higher value of adjusted R^2.

In this case, when we developed the model using all measured independent variables (total height, crown height, girth at 10cm height, crown length, crown breadth and crown depth), we found very poor adjusted R^2 (close to zero), a high value of root mean square error (1,660 gm) and higher Theils' U-statistics (1.03). These figures strongly suggest us not to use this model. Therefore, on these grounds we rejected the first model (Table 3).

Overall F-test

Sometimes, especially when the sample size is too small, even if we get a high adjusted R^2, our model may not be statistically sound. In the context of multiple regressions, we can test this hypothesis by the overall F-value or by the p-value. Our alternative hypothesis is that 'the slope of the regression line is statistically significantly different from zero at the generally accepted confidence level of 95%. If the calculated overall F-value is higher than the tabulated F-value at a given degrees of freedom and confidence level then the alternative hypothesis is accepted. Alternatively, we will come to the same conclusion if the calculated p-value is less than 0.05. In the case of opposite results, the null hypothesis (slope is not significantly different with zero) is accepted and the regression model is rejected. In this case, the calculated F-values of all the models (except the first) were greater than their respective tabulated values of F at the 95% confidence level. Therefore, on these grounds we accept all models, except the first in Table 3.

Table 3: Regression models for the estimation of biomass of wild raspberry (shrub) and their accuracy measures

Regression Model	Accuracy measures			F
	A-R^2	RMSE	U	
W = -1,776.11 - 43.5(ht) + 269.5(G10) - 342(Cr-ht) - 3.08(Cr-L) + 7.21(Cr-B) - 271.5(Cr-D)	0.0	1,660	1.03	-0.8
W = -1,769.41+5.74(ht)+290.17(G10)+6.3(Cr-ht)-1.8(Cr-L)	91.3	307	0.23	20.3
W = -1,776.11+ 4.77(ht) +269.5(G10) +1.9(Cr-ht) -3.08(Cr-L) + 7.2(Cr-B)	90.8	291	0.23	14.4
W = -1,569.12+ 287.52(G10) - 2.27 (Cr-L) +13.48 (Cr-B)	91.4	325	0.26	27.5
W=-1732.8+268.02(G10)-3.17(Cr-L)+10.78(Cr-B)+4(Cr-D)	92.4	287	0.23	20.6
W = -1,799.09 + 11.33(ht) + 272.58(G10) -5.83 (Cr-D)	92.6	303	0.22	30.7
W = -1,327.61 + 382.4(G10)	*90.5*	*382*	*0.33*	*84.9*
W = -1,747.03 + 5.67(ht) + 286.78(G10)	*93.3*	*305*	*0.24*	*50.4*
W = -1,799.09 + 5.5(ht) + 272.58(G10)+ 5.83(Cr-ht)	*92.6*	*303*	*0.22*	*30.7*
W = -43.96 + 0.000588(Cr-V)	*94.3*	*295*	*0.14*	*184*
W = 682.31 -5.74 (ht) + 0.000749(Cr-V)	*94.9*	*265*	*0.12*	*104*
W = -707.68 + 56.12(G10) + 0.141493(Cr-A)	*96.7*	*213*	*0.10*	*102*
W = -625.03 + 155.16(G10) + 0.000382 (Cr-V)	*98.3*	*155*	*0.08*	*218*

Note: 'G10' is girth (in cm) at 10 cm; 'ht' is total height in cm, and 'Cr-ht', 'Cr-L', 'Cr-B', 'Cr-D', 'Cr-A', 'Cr-V' are crown height (cm), crown length (cm), crown breadth (cm), crown depth (cm), crown area (cm^2) and crown volume (cm^3), respectively. Similarly, 'W' is the fresh weight[1] of wild raspberry in gram, 'A-R^2' is adjusted coefficient of determination, 'U' is Theils' U-statistics, 'RMSE' is root mean square error and 'F' is overall F-calculated value.

Test for individual coefficient and intercept (t-test)

Higher 'F' and adjusted R^2 values, and lower root mean square and Theils' U-statistics values sometimes do not guarantee the robustness of the model. There may be some room to improve the model by removing the independent variables that have no statistically significant contribution to the model. T-test (or p-value tests) on the individual coefficient of each independent variable is a powerful test that tells us whether or not that coefficient is statistically significantly different from zero in the

[1] Dry weight is necessary for several reasons, especially for the carbon amount estimation. Fresh weight and dry weight ratio of the wild raspberry biomass is 1.1.

presence of other independent variables. The independent variable whose coefficient value is statistically not different from zero could, by principle, be removed from the model.

In our case, we found very interesting results. Some independent variables had significantly higher influencing power (very high t-values and lower than 0.05 p-values) to the dependent variables in one set of combinations of independent variables, but in another set of combination of independent variables the influencing power of the same variable was very poor (very low t-values and higher than 0.05 p-values). Therefore, we decided not to drop any independent variables. Instead, we tried to develop as many models as we could by using all possible combinations of the major independent variables and then the same test was applied to the best model for further improvement.

We can apply the t-test to know whether the 'intercept' is statistically significantly different from zero or not. If it is not, from a statistical point of view, we can omit the intercept from our model. Technically, we recommend using the 'intercept' even if its value is not significantly different with zero as it does not cost extra money to be included but it does help the prediction power of the model.

Discussion

The accuracy measures show that the weight or biomass of wild raspberry is explained well by one (girth) or two (girth and total height) variables (Table 3). However, there is room to improve the explanatory power of the model by including other explanatory variables. But adding more variables does not always help to improve the model. For example, the model that employed two independent variables (girth and total height) had higher adjusted R^2 than the model that employed three independent variables (G10, total height and crown height). This is because the addition of crown height increased the degrees of freedom which reduces the adjusted R^2 (Adjusted $R^2 = 1 - (1-R^2)$ X total degree of freedom/error degree of freedom), as it does not significantly increase the coefficient of determination (R^2) due to its poor correlation (r=0.68) with shrub weight. This suggests to us that the thorough evaluation of the model is necessary after adding each explanatory variable.

The correlation matrix (Table 2) shows that all our measured independent variables have very good explanatory power to the shrub biomass. Acceptance of alternative hypothesis (correlation is significantly different from zero) by t-test and p-value test, also statistically validates this verdict. In spite of this obvious relationship, we could not develop a good model by employing all measured variables at the same time. This is simply because of multicollinearity among the independent (measured) variables. Because of this problem, some models came with unexpected signs of coefficients. In many others, the influencing power of the individual independent variable on the dependent variable was lost. One independent variable in one set of combinations of independent variables displayed very good influencing power (by t-test and p-value) on biomass, so, we decided to include that variable but with another set of combination that variable proved to be useless in that combination.

Initially, we had only used measured independent variables to develop the regression model, but we could not develop a reliable model. For example, we used girth, crown height and crown breadth as independent variables and developed a model, but that model resulted in an unexpected negative sign on the coefficient of crown length (Table 4). Definitely, we could not use these models as negative signs on crown lengths and crown depth are unusual and can give a wrong estimation. This does not mean that the crown length and crown depth have no explanatory power. The high correlation between crown height and weight (r=0.88) and crown depth and weight (r=0.82) shows that these are potential explanatory variables. These problems are due to strong multicollinearity among the independent variables.

To avoid this problem, we calculated crown area (Cr-A=crown length X crown breadth) and crown volume (Cr-V= crown length crown X breadth X crown depth) from measured independent variables and used these variables instead of using individual variables. As a result, our models improved significantly and have the expected signs in all coefficients with high values of adjusted R^2, and low values of root mean squire error and Theils' U-statistics (Table 4).

Table 4: Regression models for the estimation of biomass of wild raspberry (shrub) and their accuracy measures

Regression Model	Accuracy measures	F

	A-R^2	RMSE	U	
W = -1,569.12+ 287.52(G10) - 2.27 (Cr-L) +13.48 (Cr-B)	91.4	325	0.26	27.5
W=-1732.8+268.02(G10)-3.17(Cr-L)+10.78(Cr-B)+4(Cr-D)	92.4	287	0.23	20.6
W = -707.68 + 56.12(G10) + 0.141493(Cr-A)	96.7	213	0.10	102
W = -625.03 + 155.16(G10) + 0.000382 (Cr-V)	98.3	155	0.08	218

Note: ‘G10’ is girth of shrub (in cm) at 10 cm; ‘ht’ is total height of shrub in cm, and ‘Cr-ht’, ‘Cr-L’, ‘Cr-B’, ‘Cr-D’, ‘Cr-A’, ‘Cr-V’ are crown height (cm), crown length (cm), crown breadth (cm), crown depth (cm), crown area (cm^2) and crown volume (cm^3), respectively. Similarly, ‘W’ is the fresh weight of wild raspberry (shrub) in gram, ‘A-R^2’ is adjusted coefficient of determination, ‘U’ is Theil’s U-statistics, ‘RMSE’ is root mean square error and ‘F’ is overall F-calculated value.

Conclusion and recommendation

Plotting scatter diagrams between each independent variable and dependent variable notifies, in advance, what sign we should expect on the coefficient of that particular variable in the regression model. Careful analysis of correlation coefficient values between each independent variable with dependent variable gives very important clues as to which variables should be selected and which should not for the models. However, while doing so, there should not be multicollinearity between the remaining independent variables. Many statistical textbooks suggest to reject that variable from the model whose coefficient is not statistically significantly different from zero. But our analysis shows that those variables could have a very high influencing power in another set of combination of the independent variables. Therefore, before deciding to reject independent variables, especially those which have good correlation with the dependent variable, thorough examination of several other models by using that particular variable is strongly recommended.

Because of high collinearity between the independent variables we could not develop a reliable model by employing all the measured variables (girth, crown length, crown breadth and crown height) although these are obviously the major explanatory variables (by correlation, t-test, p-value test). Our important conclusion is that this problem could be solved by using scientific formulae and applying a single variable instead of applying highly correlated independent variables in the model. In our case, instead of using crown length and crown breadth as two separate independent variables we use crown area. Similarly, instead of using three independent variables (crown length, crown breadth and crown depth) we use crown volume.

Depending on the availability of data, we can apply any model written in italic letters in Table 3, but our first recommendation is for the last model {W = -625.03 + 155.16(G10) + 0.000382 (Cr-V)}, which is based on girth and crown volume (Table 3 and 4). Where the cost is the major problem, our recommendation is for the second last model {W = -707.68 + 56.12(G10) + 0.141493(Cr-A)}, which employs girth and crown area as they give a good result and it needs only three variables to be measured (Table 3 and 4). These models are only applicable to the wild raspberry found in a similar climatic and edaphic region of the study area. Since we have followed all the steps of model development and improvement sequentially and logically, it could be helpful in teaching the practical applications of Multiple Regression in courses such as Data Analysis and Business Forecasting.

Acknowledgement

This research was funded by the Faculty of Business, University of Southern Queensland. The authors would like to thank two anonymous reviewers and Dr Mehryar Nooriafshar for their valuable comments and suggestions.

References

Burrows, W., B. K. Henry, et al. (2002). "Growth and Carbon Stock Change in Eucalypt Woodlands in Northeast Australia: Ecological and Greenhouse Sink Implication." Global Change Biology **8**: 769-84.

Cacho, O. J., R. L. Hean, et al. (2003). "Carbon Accounting Methods and Reforestation Incentives." The Australian Journal of Agricultural and Resource Economics **47**(2): 153-179

Catchpole, W. R. and C. J. Wheeler (1992). "Estimating Plant Biomass: A Review of Allometric Techniques." Australian Journal of Ecology **17**: 121-131

Harrington, G. (1979). "Estimation of Above-Ground Biomass of Trees and Shrubs in a *Eucalyptus populnea* F. Mull. Woodland by Regression of Mass on Trunk Diameter and Plant Height." Australian Journal of Botany **27**: 135-143

Isbell, R.F. (2002). The Australian Soil Classification. 2nd ed. CSIRO Publishing, Melbourne

Karki, D. (2002). Economic Assessment of Community Forestry in Inner *Terai* of Nepal: A Case Study from Chitwan District, Asian Institute of Technology, Thailand

Khanal, K. P. (2001). Economic Evaluation of Community Forestry in Nepal and its Equity Distribution Effect, The Royal Veterinary and Agricultural University, Denmark

Makridakis, S., S. C. Wheelwright, et al. (1998). Forecasting Method and Application, John Willey and Sons, Inc. Third Edition, America

Margules Poyry Pty Ltd (1998). Wood Fibre Yields For South East Queensland RFA Region (Draft Report), Prepared for CRA Unit, Department of Natural Resources, Brisbane and BRS, DPIE, Canberra, Australia

Mitchell, C. and J. Skjemstad (2004). "Greenhouse Chemistry not only in the Atmosphere" Chemistry in Australia **17**: 12-16

Mills, G. and D. Schmidt (2000). Overview of Cropping in the South Burnett Crops Grown and Climate, Department of Primary Industries and Fisheries, Queensland Government, http://www.dpi.gov.au./fieldcrops, Cited on 11/05/2005

Mohns, B., G. B. Applegate, et al. (1989). "Biomass and Productivity Estimation for Community Forest Management: A Case Study from the Hills of Nepal-II. Dry Matter Production in Mixed Young Stands of Chir Pine (*Pinus roxburghai*) and Broad-leaved Species." Biomass **17**: 165-184

Palm, M. (2003). Clean Development Mechanism and Local Sustainable Development in Karnataka, India, Goteborgs Universitet, Sweden

Pensiero, D., M. Nooriafshar, et al. (2005). MGT 2101 Business Forecasting-Study Book, University of Southern Queensland, Toowoomba, Queensland, Australia

Scalan, J. C. (1991). "Woody Observatory and Herbaceous Understorey Biomass in *Acacia harpophyla* (Briglow) Woodlands." Australian Journal of Ecology **16**: 521-529

Smith, G. and D. Kent (1993). Understanding and Managing Soils in the Inland Burnett District, Department of Primary Industries, Brisbane

Snowdon, P., J. Raison, et al. (2002). Protocol for Sampling Tree and Stand Biomass, Technical Report no 31, National Carbon Accounting System, Australian Greenhouse Office, Canberra, Australia

Specht, A. and P. W. West (2003). "Estimation of Biomass and Sequestered Carbon on Farm Forest Plantation in Northern New South Wales, Australia." Biomass and Bioenergy **25**: 363-379

Wilson, J. H., B. Keating, et al. (2001). Business Forecasting with Accompanying Excel-Based ForecastX™ Software, Fourth Edition, McGraw-Hill Companies, Inc, New York, America