Use of qualitative variables in the choice of volumetric models in the forest inventory.

Vinicius Pizzo Ferreira, João Luís Ferreira Batista
University of São Paulo, Superior School of Agriculture “Luiz de Queiroz” – ESALQ/USP

Introduction

Forest volume prediction is essential for a variety of logistical purposes, such as transportation, harvesting, quantity of timber to the plant and other purposes. Therefore the measurement of such a variable is indispensable for a forest enterprise. For such measurement, mathematical models are used to quantify wood in the field. The aim is to evaluate the impact of incorporating locality information as indicator variables on the predictive accuracy of volumetric equation models, and to verify whether this incorporation influences the choice of the most appropriate model.

Methods

A total of 11 volumetric models were used, divided into double-entry and local models (Table 1). For the qualitative variables we used the information of Region, Farm, Stratum and Stand. It was observed the coefficient of determination, standard error of the estimate and various indexes related to the residuals, such as the mean absolute deviation, interquartile distance, range of variation, among others. In addition to the indices, the scatter plots, quantile-quantiles and boxplot were analyzed. To evaluate the heteroscedasticity of the models, the Spearman correlation coefficient and the adjusted values of the models were calculated. The qualitative variables adjusted as intercept and in a second moment the models were interacted with the qualitative variables.

Table 1 – Double entry and local volumetric models, respectively, used in the estimations

Modelo	Fórmula
Spurr	$v_i=\beta_0+\beta_1(d^2\cdot h)+\varepsilon_i$
Schumacher-Hall	$\ln(v_i)=\beta_0+\beta_1\cdot\ln(d_i)+\beta_2\cdot\ln(h_i)+\varepsilon_i$
Baden	$\ln(v_i)=\beta_0+\beta_1\cdot\ln(d_i)+\beta_2\cdot\ln^2(d_i)+\beta_3\cdot\ln(h_i)+\beta_4\cdot\ln^2(h_i)+\varepsilon_i$
Meyer	$v_i=\beta_0+\beta_1\cdot d_i+\beta_2\cdot d_i^2+\beta_3(d_i\cdot h_i)+\beta_4(d_i^2\cdot h_i)+\beta_5\cdot h_i+\varepsilon_i$
Stoate	$v_i=\beta_0+\beta_1\cdot d_i^2+\beta_2(d_i^2\cdot h_i)+\beta_3\cdot h_i+\varepsilon_i$
Husch	$\ln(v_i)=\beta_0+\beta_1\cdot\ln(d_i)+\varepsilon_i$
Baden Local	$\ln(v_i)=\beta_0+\beta_1\cdot\ln(d_i)+\beta_2\cdot\ln^2(d_i)+\varepsilon_i$
Área Transversal	$v_i=\beta_0+\beta_2\cdot d_i^2+\varepsilon_i$
Parabólico	$v_i=\beta_0+\beta_1\cdot d_i+\beta_2\cdot d_i^2+\varepsilon_i$
Schumacher-Hall Local	$\ln(v_i)=\beta_0+\beta_1(1/d_i)+\varepsilon_i$
Brenac	$\ln(v_i)=\beta_0+\beta_1\cdot\ln(d_i)+\beta_2\cdot(1/d_i)+\varepsilon_i$

Results

The results of the double entry models were better compared to the local models. The improvement of the results with the use of Location information was visible in all models used, such improvement was greater with the greater specification of the variable Region to Stand, thus obtaining the best performance of the models when interacting with Stand.

Conclusion

The use of Location information proved to improve the performance of the mathematical models used in the study, so the conclusion is that the usage of these information were more efficient than to choose the better model, when used the Stand information, because it is the most specific Location information used.

Table 2 - Indices used in the comparison of double-entry equations in their general model and in the presence of qualitative variables, with interaction of these variables.

	N	R2	PSE	MAD	Range	GSkew	PSkew	PCor	
vtot ~ I(dap^2 * ht)	1482	99,4	7,17	3,75	129,88	16	20,3	56,1	Geral
log(vtot) ~ log(dap) + log(ht)	1482	99,3	7,43	3,72	126,63	0,5	22,5	62,3	
log(vtot) ~ log(dap) + I(log(dap)^2) + log(ht) + I(log(ht)^2)	1482	99,4	7,28	3,64	134,07	2,8	1,3	61,4	
vtot ~ dap + I(dap^2) + I(dap * ht) + I(dap^2 * ht) + ht	1482	99,4	6,93	3,72	119,99	4,2	1,1	58,1	
vtot ~ I(dap^2) + I(dap^2 * ht) + ht	1482	99,4	7,01	3,7	123,4	-9,4	-3,7	57	
vtot ~ (I(dap^2 * ht)) * regiao	1482	99,4	7,11	3,81	128,94	9,7	16,9	53,6	Região
log(vtot) ~ (log(dap) + log(ht)) * regiao	1482	99,3	7,4	3,61	119,36	3,3	20,3	62,9	
log(vtot) ~ (log(dap) + I(log(dap)^2) + log(ht) + I(log(ht)^2)) * regiao	1482	99,4	7,09	3,37	131,23	5,3	2,6	62,2	
vtot ~ (dap + I(dap^2) + I(dap * ht) + I(dap^2 * ht) + ht) * regiao	1482	99,5	6,71	3,75	117,49	5,5	-0,5	57,8	
vtot ~ (I(dap^2) + I(dap^2 * ht) + ht) * regiao	1482	99,4	6,85	3,76	117,96	-0,3	-0,8	56,2	
vtot ~ (I(dap^2 * ht)) * fazenda	1482	99,5	6,67	3,44	135,32	6,7	15,5	54,5	Fazenda
log(vtot) ~ (log(dap) + log(ht)) * fazenda	1482	99,5	6,55	3,22	111,88	2,4	11	60,3	
log(vtot) ~ (log(dap) + I(log(dap)^2) + log(ht) + I(log(ht)^2)) * fazenda	1482	99,5	6,87	3,08	152,85	5,9	-2,3	59,5	
vtot ~ (dap + I(dap^2) + I(dap * ht) + I(dap^2 * ht) + ht) * fazenda	1482	99,6	6,23	3,31	104,08	-1,8	-5,3	54	
vtot ~ (I(dap^2) + I(dap^2 * ht) + ht) * fazenda	1482	99,5	6,38	3,42	125,19	-2,6	-5,4	52,5	
vtot ~ (I(dap^2 * ht)) * estrato	1482	99,6	6,16	3,27	123,52	0,3	3,8	48,2	Estrato
log(vtot) ~ (log(dap) + log(ht)) * estrato	1482	99,6	6,27	2,75	128,94	2,9	3,6	58,8	
log(vtot) ~ (log(dap) + I(log(dap)^2) + log(ht) + I(log(ht)^2)) * estrato	1482	99,7	5,68	2,42	80,17	8,8	4,2	53,6	
vtot ~ (dap + I(dap^2) + I(dap * ht) + I(dap^2 * ht) + ht) * estrato	1482	99,7	5,44	2,62	66,16	1,2	-0,5	45,1	
vtot ~ (I(dap^2) + I(dap^2 * ht) + ht) * estrato	1482	99,7	5,83	2,89	103,65	-2,5	-2,3	46,3	
vtot ~ (I(dap^2 * ht)) * talhao	1482	99,6	5,96	2,87	122,58	-5,8	0	43,6	Talhão
log(vtot) ~ (log(dap) + log(ht)) * talhao	1482	99,6	6,67	2,44	155,97	4,6	4	53,3	
log(vtot) ~ (log(dap) + I(log(dap)^2) + log(ht) + I(log(ht)^2)) * talhao	1482	99,8	5,45	1,71	70,71	4,7	4,4	39,5	
vtot ~ (dap + I(dap^2) + I(dap * ht) + I(dap^2 * ht) + ht) * talhao	1482	99,8	5,06	1,79	52,72	3,5	0	25,3	
vtot ~ (I(dap^2) + I(dap^2 * ht) + ht) * talhao	1482	99,8	5,43	2,34	77,9	-0,2	0	34,4	

Table 3 - Indices used in the comparison of double-entry equations in their general model and in the presence of qualitative variables, with interaction of these variables.

	N	R2	PSE	MAD	Range	GSkew	PSkew	PCor	
log(vtot) ~ log(dap)	1482	96,2	17,61	10,31	228,08	0,3	25,7	58,8	Geral
log(vtot) ~ log(dap) + I(log(dap)^2)	1482	96,1	17,95	10,59	273,11	2,9	4,6	59,6	
vtot ~ I(dap^2)	1482	94,8	20,66	17,74	240,96	0,1	1,7	14,6	
vtot ~ dap + I(dap^2)	1482	96,7	16,51	10,56	223,24	3,1	3	50,7	
log(vtot) ~ I(1/dap)	1482	68,7	50,75	13,48	527,12	-7,4	68,9	62,1	
log(vtot) ~ log(dap) + I(1/dap)	1482	96,3	17,38	10,48	260,97	4,7	8,5	58,8	
log(vtot) ~ (log(dap)) * regiao	1482	96,6	16,81	9,99	204,9	5	24,8	60	Região
log(vtot) ~ (log(dap) + I(log(dap)^2)) * regiao	1482	96,2	17,69	10,03	280,03	7	6	61,3	
vtot ~ (I(dap^2)) * regiao	1482	95,5	19,19	15,26	222,53	-0,4	-5,6	21,9	
vtot ~ (dap + I(dap^2)) * regiao	1482	96,9	16,1	10,2	204,02	-1,1	-4,6	53,6	
log(vtot) ~ (I(1/dap)) * regiao	1482	72,4	47,75	13,33	521,65	5,7	66,9	61,5	
log(vtot) ~ (log(dap) + I(1/dap)) * regiao	1482	96,4	17,34	10,05	267,82	3,9	6,1	61,5	
log(vtot) ~ (log(dap)) * fazenda	1482	97,3	14,98	7,41	225,44	7	10,3	61	Fazenda
log(vtot) ~ (log(dap) + I(log(dap)^2)) * fazenda	1482	97,5	14,65	7,27	204,05	8,6	5,8	60,3	
vtot ~ (I(dap^2)) * fazenda	1482	96,8	16,39	10,7	194,11	2,1	-8,7	33,2	
vtot ~ (dap + I(dap^2)) * fazenda	1482	97,7	13,9	7,79	201,83	6,2	-1,3	53,2	
log(vtot) ~ (I(1/dap)) * fazenda	1482	81,5	39,44	10,36	530,68	4,9	59	64	
log(vtot) ~ (log(dap) + I(1/dap)) * fazenda	1482	97,5	14,59	7,33	194,54	8,4	7	60	
log(vtot) ~ (log(dap)) * estrato	1482	98,2	12,62	4,78	201,78	4,4	-6,4	59,4	Estrato
log(vtot) ~ (log(dap) + I(log(dap)^2)) * estrato	1482	98,6	11,66	4,71	188,11	-0,5	3,8	60,2	
vtot ~ (I(dap^2)) * estrato	1482	98,4	11,81	5,67	193,74	-5,4	-2	42,6	
vtot ~ (dap + I(dap^2)) * estrato	1482	98,7	11,09	5,04	189,96	-1	1,7	52,2	
log(vtot) ~ (I(1/dap)) * estrato	1482	89,3	30,98	6,94	427,89	-5	48	60,8	
log(vtot) ~ (log(dap) + I(1/dap)) * estrato	1482	98,5	11,88	4,78	190,86	-0,7	2	60,5	
log(vtot) ~ (log(dap)) * talhao	1482	99	9,86	3,71	167,96	6,5	-5,8	53,5	Talhão
log(vtot) ~ (log(dap) + I(log(dap)^2)) * talhao	1482	99,3	8,91	3,31	125,42	3,7	4,5	50,2	
vtot ~ (I(dap^2)) * talhao	1482	99,1	9,22	4,34	125,39	-4,5	-1,3	33,8	
vtot ~ (dap + I(dap^2)) * talhao	1482	99,4	8,17	3,57	108,03	2,8	0	38,9	
log(vtot) ~ (I(1/dap)) * talhao	1482	91,7	28,61	5,89	367,71	-3	44,6	53,5	
log(vtot) ~ (log(dap) + I(1/dap)) * talhao	1482	99,2	9,2	3,32	137,44	4,4	2,9	50,3	

1- vinicius.pizzo.ferreira@usp.br; 2- batista.jlf@usp.br