

Modelo Probabilístico Independente da Distância para Implementação de Desbastes em Povoamentos Regulares de Sobreiro

Paulo* J. A. e Tomé*, M.

Instituto Superior de Agronomia, Tapada da Ajuda, 1399 LISBOA Codex.

Resumo. O objectivo deste trabalho foi desenvolver um modelo do tipo logístico e independente da distância, capaz de calcular a probabilidade de uma árvore ser desbastada num povoamento regular de sobreiro. O modelo será dependente de variáveis dendrométricas da árvore, de índices de competição independentes da distância, que indicam a sua posição na estrutura do povoamento, e da intensidade do desbaste definida pelo coeficiente de espaçamento do povoamento.

Introdução

A maioria dos algoritmos desenvolvidos para selecção de árvores para desbaste está dependente do conhecimento das coordenadas de cada árvore (Pukkala 1998). Isto acontece porque a competição é tida como um factor condicionante no crescimento das árvores, pelo que a selecção das árvores a desbastar deve ponderar, não só as características da árvore, como também a dimensão e a posição das árvores envolventes.

No caso do sobreiro, a maioria dos inventários realizados não contempla a medição das coordenadas das árvores, pelo que algoritmos como os acima referidos não são aplicáveis. Neste trabalho pretende-se desenvolver um modelo probabilístico para a selecção de árvores para desbaste, mas, neste caso, independente da distância. No entanto, pela importância conhecida da competição, tanto no crescimento da árvore, como na prática da selecção das árvores para desbaste, foi testada a incorporação no modelo de índices de competição, sendo estes do tipo independente da distância.

Material e métodos

Tomando as medições de parcelas de área considerável existentes na base de dados SUBERDATA2, foram realizadas simulações para largos períodos de tempo com o programa SUBER (Tomé, *et. al* 1990), a partir do qual se definiram condições determinadas para a gestão do povoamento em termos de desbastes.

O período de tempo da simulação dependeu da idade do povoamento na altura da medição, de forma a que, no final da simulação, a idade do povoamento se aproximasse dos 130 anos de idade, no caso desta ser conhecida. Caso não o seja, definiu-se que seria simulado o crescimento ao longo de 45 anos, no caso de povoamentos adultos, e 90 anos, no caso de povoamentos juvenis.

Na Tabela 1 encontra-se informação sobre os povoamentos simulados, assim como das parcelas consideradas em cada povoamento.

O crescimento de todas as 20 parcelas foi simulado 6 vezes, cada uma delas correspondente a uma das alternativas de gestão consideradas. Estas alternativas foram definidas em termos de dois factores: coeficiente de espaçamento e índice de desbaste.

No caso do coeficiente de espaçamento, considerou-se, como uma das alternativas, o valor de 1.2, indicado como “ideal” por Natividade (1950) para o montado de sobreiro. Definiram-se ainda as alternativas de 1.0 e 1.4, por representarem os casos em que o montado está demasiado denso ou demasiado ralo, respectivamente.

Tabela 1. Informação sobre os povoamentos simulados. N - número de árvores por hectare; G – área basal; ddom – diâmetro das árvores dominantes; * idade desconhecida (povoamento adulto); ** idade desconhecida (povoamento jovem)

Povoamento	Idade (anos)	Número de parcelas	Área por parcela (ha)	Período de simulação (anos)	Caracterização das parcelas		
					N	G	ddom
Mora I	*	6	4	45	132	8.7	36.8
					127	8.5	43.3
					121	8.4	46.4
					87	10.9	47.2
					100	11.0	42.6
					90	7.6	44.0
Palmela	70	1	4	60	69	10.7	48.8
Coruche	*	1	4	45	93	9.0	49.8
Mora II	*	2	2	45	73	10.1	53.3
					87	9.9	51.1
Estremoz	45	4	0.2	90	509	3.2	27.8
					386	2.0	24.9
					343	2.2	26.4
					495	2.9	26.6
Chamusca	**	6	0.2	90	871	5.1	18.9
					938	5.2	22.3
					790	4.2	21.7
					819	5.2	22.9
					905	5.5	20.7
					771	5.4	22.7

No que diz respeito ao índice de desbaste, este considera duas características para decidir quais as árvores que saem em desbaste: a competição entre árvores e a qualidade da cortiça. A introdução destes factores no programa faz-se através de uma ponderação para a importância da competição vs. qualidade da cortiça, cujo somatório é igual a 1. Foram consideradas duas situações: o desbaste é feito apenas considerando a competição (1 para o factor da competição e 0 para o da qualidade da cortiça), ou o desbaste é feito tendo também em atenção a qualidade da cortiça (0.75 para o factor da competição e 0.25 para o da qualidade da cortiça).

Multiplicando as alternativas possíveis para cada um dos dois factores, ficamos com 6 (2x3) alternativas de gestão diferentes para a realização dos desbastes ao longo das simulações.

O modelo SUBER utilizado nas simulações apenas realiza desbastes nos anos de descortiçamento do povoamento, e desde que o coeficiente de espaçamento do povoamento

esteja inferior ao coeficiente de espaçamento limite imposto para a simulação. Nos casos em que tal não acontece, o povoamento não necessita de ser desbastado, pelo que o modelo não procede a essa operação. Desta forma, no final das simulações foram retirados todos os dados referentes aos anos de desbaste em que o povoamento não foi sujeito a esta intervenção.

No final, recolheram-se 74620 registos ao nível da árvore, entre as $6 \times 20 = 120$ simulações realizadas, para as quais foram registadas, para além de múltiplas variáveis do povoamento e da árvore, uma variável dicotómica que representa o acontecimento de a árvore ter sido desbastada no ano a que se refere o registo ou não. Essa variável, do tipo binomial, toma o valor 0 no caso da árvore não ser descortiçada, e 1 no caso de o ser.

O modelo aqui desenvolvido, do tipo logístico, pretende modelar a probabilidade de uma árvore, para determinadas condições do povoamento e de gestão do mesmo, não ser desbastada. A variável resposta será designada CodDesb (código de desbaste), cujo valor é 0 ou 1 como anteriormente indicado.

A forma assumida pelo modelo logístico é a seguinte, onde $\pi(x)$ é a probabilidade estimada, e $g(x)$ é um modelo linear (Hosmer 1989):

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} .$$

A transformação aplicada a $\pi(x)$ é designada por transformação logística, e é definida da seguinte forma:

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} .$$

A importância desta transformação é que $g(x)$ adquire muitas das propriedades do modelo de regressão linear. Entre outras, $g(x)$ pode ser contínua e pode variar entre $-\infty$ e $+\infty$.

As variáveis independentes testadas, todas do tipo contínuo, estão indicadas na lista que se segue.

Variáveis da árvore:

- d – diâmetro à altura do peito medido sobre a cortiça
- du – diâmetro à altura do peito medido sob a cortiça
- h – altura total da árvore
- APB – índice de competição independente da distância: $\left(\frac{10000}{N} \right) \frac{du}{dg}$
- Rcopa – razão entre o diâmetro da copa da árvore e da copa média das árvores
- dcopa – diâmetro da copa da árvore
- Rdmax – índice de competição independente da distância: $\frac{du}{d \max}$
- Rdmi – índice de competição independente da distância: $\frac{du}{dg}$

Variáveis do povoamento:

- Nha – número de árvores por hectare
- Gsc – área basal por hectare sob a cortiça
- Gcc – área basal por hectare sobre a cortiça
- Acob – área coberta do da parcela
- Pacob – percentagem de área coberta da parcela
- Acopam – área da copa média das árvores
- dmax – diâmetro máximo da parcela

- dg – diâmetro quadrático médio da parcela
- CoefEsp – coeficiente de espaçamento real da parcela

Variáveis de gestão do povoamento:

- CE – coeficiente de espaçamento limite imposto
- RCE – razão: $\frac{CE - CoefEsp}{CE}$
- Comp – factor de competição (1 ou 0.75)
- Cort – factor de qualidade da cortiça (0 ou 0.25)

Inicialmente, estudou-se o relacionamento de cada uma das variáveis individualmente com a variável resposta. Pretende-se, assim, observar desde logo quais são as variáveis mais significativas para a variável resposta, sem a interacção de diferentes variáveis entre si. Tanto nesta fase, como nas fases posteriores do ajustamento do modelo, a significância das variáveis foi avaliada com base em três testes estatísticos indicados por Hosmer (1989):

- teste da razão de verosimilhanças (*likelihood ratio test*) – informa acerca da significância das variáveis em conjunto no modelo;
- teste Wald – testa a significância individual de cada variável;
- razão de probabilidades (*Odds Ratio*) – fornece, para cada variável qualitativa, a variação na probabilidade de sucesso da variável resposta, para uma unidade de variação da variável independente.

Posteriormente realizou-se o ajustamento do modelo com todas as variáveis que anteriormente se mostraram mais significativas, seleccionando as que ficariam no modelo através do método de selecção *Stepwise*.

Várias alternativas foram investigadas até se encontrar o modelo que fornece maior percentagem de concordância entre os valores observados e preditos, assim como o modelo mais simples e sem colinearidade entre as variáveis independentes.

Resultados e conclusão

Primeiramente optou-se por estudar o relacionamento com a variável resposta de cada uma das variáveis *per se*, tentando encontrar, logo de início, quais destas seriam significativas e quais as mais significativas sem a interacção das outras. Tal como em qualquer outro tipo de regressão, depois de estimar os valores dos coeficientes associados a cada variável, atendeu-se à significância dos mesmos, tendo-se, para isso, recorrido aos testes estatísticos anteriormente referidos.

Na Tabela 2 poder-se-ão encontrar os resultados do ajustamento do modelo logístico com cada uma das variáveis individualmente. Para cada variável podemos ver o valor de *Odds Ratio*, e as percentagens de concordância e discordância encontradas para o conjunto das observações. Foram eliminadas as colunas referentes aos testes da razão de verosimilhanças e do teste de Wald, porque a sua conclusão é, em todos os casos, no sentido da significância da variável em questão.

Pela tabela 2 podemos observar quais as variáveis que apresentaram maiores valores de concordância final, as quais pertencem à categoria de variáveis da árvore. Entre estas evidenciam-se o diâmetro e a altura da árvore, o diâmetro da copa e o índice de competição, designado Rdmi.

Tabela 2. resultado do ajustamento do modelo logístico individualmente para cada uma das variáveis

Tipo	Variável	Odds Ratio	Concordância (%)	Discordância (%)
Variáveis da árvore	du	1.07	72.9	26.4
	d	1.07	72.5	26.8
	h	2.63	75.4	23.7
	APB	1.01	67.3	31.7
	rcopa	3.86	58.9	29.8
	dcopa	1.58	76.5	24.8
	Rdmax	40.05	66.7	32.3
	Rdmi	265.8	77.6	21.9
Variáveis do povoamento	Nha	0.99	60.4	36.1
	Gsc	0.98	54.6	39.9
	Gcc	0.98	57.2	38.6
	Acob	1.00	57.8	38.4
	Pacob	0.99	55	40
	Acopam	1.02	61.3	37
	dmax	1.03	61.7	36.7
	dg	1.04	61.7	36.6
	CoefEsp	5.94	57.4	38.9
Variáveis de gestão	RCE	<0.0001	67.3	29.1
	CE	0.81	32.2	30.8
	Comp	1.08	25.2	24.7
	Cort	0.92	25.2	24.7

Embora a variável altura total apresentasse um bom resultado no ajustamento individual, quando em conjunto com outras variáveis o sinal do seu parâmetro toma um significado que não se consegue explicar em termos biológicos, o mesmo acontecendo a outras variáveis incluídas no modelo.

Por outro lado, as variáveis do povoamento, quando colocadas no modelo em conjunto com variáveis da árvore, não levam a melhorias significativas da qualidade do modelo, e provocam mesmo, em alguns dos casos, a inversão do sinal de alguns parâmetros que anteriormente apresentavam parâmetros com um correcto significado biológico.

No que diz respeito às variáveis designadas como “de gestão”, entre elas apenas a variável RCE demonstrou resultados significativos que levem a suspeitar da sua inclusão no modelo final. A fraca percentagem de concordância das variáveis Comp e Cort indica que na selecção das árvores para desbaste não é significativo o facto de estarmos a considerar apenas a competição, ou, também, a considerar a qualidade da cortiça entre árvores.

O modelo final seleccionado apresenta três variáveis significativas. São elas: o diâmetro da árvore medido sob a cortiça, o índice de competição independente da distância, e a variável RCE, que nos indica uma medida da intensidade do desbaste a que o povoamento vai ser sujeito. O modelo é apresentado na Tabela 3.

Tabela 3- valores dos parâmetros do modelo logístico ajustado

	Estimativa do parâmetro	Desvio padrão	p value	Odds ratio
Intercepção	-3.9800	0.078	<0.0001	
du	0.0078	0.002	<0.0001	1.0
Rdmi	6.8988	0.121	<0.0001	991.1
RCE	-12.6742	0.255	<0.0001	<0.001

Observando o valor ajustado para o parâmetro associado ao diâmetro, podemos concluir que, quanto maior o seu valor, maior será a probabilidade da árvore não ser descortçada.

No que diz respeito ao parâmetro associado a R_{dmi}, pelo facto de este ser a razão entre o diâmetro da árvore e o valor do diâmetro quadrático médio da parcela, podemos concluir que quanto mais a árvore se aproximar da árvore média, maior será a probabilidade de não ser removida no desbaste.

Os dois parâmetros atrás referidos permitem confirmar a sensibilidade do modelo à árvore, uma vez que estima tanto menor probabilidade de desbaste para uma árvore, quanto mais afastada esta é da árvore média do povoamento, o que nos conduzirá, na generalidade, a árvores dominadas e a árvores mais velhas.

Entre o conjunto das observações, este modelo apresenta 82% de percentagem de concordância no que diz respeito ao valor resultante para a variável dicotómica (61263 observações em que o modelo indica correctamente o desbaste, ou não, da árvore), e 18% de discordância (13059 observações em que o modelo indica erradamente o desbaste, ou não, da árvore).

Por fim, foi feita a validação sugerida por Vasconcelos (2001). Para esta validação, cada variável incluída no modelo foi classificada num determinado número de classes que abrangem todas as observações. A variável diâmetro foi dividida em classes de 5 centímetros, e as variáveis R_{dmi} e RCE em 5 classes cada. Cada observação do conjunto de dados foi classificada para estas variáveis, de acordo com o valor da variável que as caracteriza. No conjunto das três variáveis independentes do modelo, foram encontradas 50 combinações entre as classes das variáveis.

Para cada uma das combinações, tomando o valor da classe de cada variável, foi calculada a probabilidade de uma árvore não ser desbastada (probabilidade estimada pelo modelo ajustado). Com base nos valores da variável resposta obtidos nas simulações, para cada combinação de classes foi ainda calculada a probabilidade de uma árvore não ser desbastada. Esta probabilidade será obtida pela razão entre o número de árvores não desbastadas e o número total de árvores classificadas na respectiva combinação de classes.

Foi, posteriormente, observado o gráfico que compara as probabilidades estimadas pelo modelo e as probabilidades observadas, calculadas como foi acima indicado. Este gráfico pode ser observado na figura 1, que se segue.

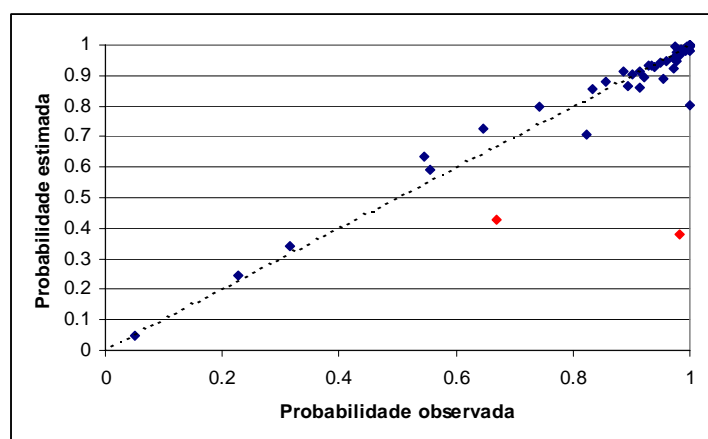


Figura 1. Probabilidade estimada pelo modelo *versus* probabilidade observada. A vermelho, pontos que indicam combinações das variáveis do modelo, em que as estimativas são fracas.

Como se pode observar na figura 1, as estimativas são fracas (diferença entre o valor observado e estimado superior a 0.25) em apenas duas combinações. Estes dois pontos correspondem a: num dos casos, árvores com diâmetros na classe dos 10 centímetros, R_{dmi}

entre 1 e 2, e RCE entre -0.14 e 0.03, e, no segundo, árvores com diâmetros na classe dos 15 centímetros, Rdmi entre 1.0 e 1.5, e RCE superior a 0.2.

Conclusão

O modelo logístico ajustado permite calcular a probabilidade de uma árvore não ser desbastada, dependendo das suas características, da sua posição no povoamento, e da intensidade do desbaste.

As estimativas do modelo são bastante aproximadas dos valores encontrados nas simulações, aumentando tanto mais a probabilidade de uma árvore não ser desbastada quanto mais próxima esta está da árvore média do povoamento. Desta forma o modelo tende, após o desbaste, a conduzir o povoamento para uma situação do tipo regular.

Na expressão seguinte pode-se ver como calcular a probabilidade de uma árvore não ser desbastada.

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}, \text{ onde } g(x) = 3.980 + 0.0078(\text{du}) + 6.899(\text{Rdmi}) - 12.674(\text{RCE})$$

Referências

- Hosmer, D., Lomeshow, S., 1989. *Applied Logistic Regression*. Wiley. NewYork.
- Natividade, J. V., 1950. *Subericultura*. Lisboa.
- Pukkala, T. e Miina, J. 1998. Tree-selection algorithms for optimizing thinning using a distance-dependent growth model. *Can. J. For. Res.* 28: 693-702.
- Tomé, M., Coelho, M. B., Pereira, H. & Lopes, F., 1999. *A management oriented growth and yield model for cork oak and cork oak stands in Portugal*. In: A. Amaro e - - M. Tomé (ed). Empirical and process – based models for forest tree and stand growth simulation. Edições Salamandra. Novas tecnologias. Lisboa. Portugal. pp 271 – 189.
- Vasconcelos, M. J., Silva, S., Tomé, M., Alvim, M. & Pereira, J. M. C., 2001. Spatial prediction of fire ignition probabilities: comparing logistic regression and neural networks. *Photogrametric Engineering and Remote Sensing*, 67(1): 73-81.