

Delimitação de áreas para plantio de eucalipto utilizando regressões logísticas

Delimitation of areas for planting eucalyptus trees using logistic regressions

Rodrigo Teske^I Elvio Giasson^{II}

- NOTA -

RESUMO

A área útil efetiva é um parâmetro importante na aquisição de terras e planejamento do florestamento. A finalidade desta pesquisa foi gerar mapas preditores de áreas aptas ao plantio de eucalipto usando regressões logísticas binárias e variáveis geomorfológicas. As relações entre as variáveis preditoras e as áreas aptas para plantio de eucalipto foram modeladas e a variável que melhor explicou a ocorrência de áreas para plantio foi a distância dos rios. O mapa gerado apresentando as áreas aptas para plantio mostrou alta capacidade de reproduzir o mapa original de plantio de eucalipto. As regressões logísticas demonstraram viabilidade do uso para o mapeamento da aptidão para o plantio de eucalipto.

Palavras-chave: florestamento, SIG, mineração de dados.

ABSTRACT

Effective usable area is a key parameter in land acquisition and afforestation planning. The purpose of this research was to generate predictive maps of areas suitable for planting eucalyptus trees using binary logistic regressions and geomorphometric variables. The relationships between the predicting variables and suitable areas for planting eucalyptus trees were modeled and the variable that best explained occurrence of suitable lands was distance from rivers. The generated map showing areas suitable for planting had a high ability to reproduce the original planting map. Logistic regressions demonstrated the feasibility of use this approach to map suitability for eucalyptus forestation.

Key words: afforestation, GIS, data mining.

Em muitos casos, as áreas adquiridas para o florestamento com eucalipto não são avaliadas de forma adequada no momento da aquisição, com equívocos na avaliação de seu potencial de utilização. Erros nessa avaliação podem acarretar prejuízos econômicos e empecilhos ao bom desenvolvimento das atividades florestadoras.

Na aquisição de terras e planejamento do florestamento, é importante avaliar qual porcentagem da área pode ser plantada, bem como a localização exata das glebas aptas ao plantio, determinando assim a viabilidade de uso para este fim e o valor da terra. Posteriormente, as equipes de campo podem avaliar localmente a real aptidão para o florestamento, decidindo pelo plantio ou não. Muitas vezes, áreas efetivamente plantadas nesta etapa diferem da estimativa do potencial de plantio definido na aquisição das terras.

Como muitas variáveis que determinam a aptidão para plantio de eucalipto podem variar espacialmente, acredita-se que o uso de modelagem e predição de ocorrência de áreas aptas para plantio seja possível utilizando-se variáveis do terreno como variáveis preditoras e mapeando a ocorrência de áreas aptas para plantio utilizando-se regressões logísticas. As regressões logísticas permitem estimar a probabilidade de ocorrência de classes de uma variável dependente (neste caso, ser ou não área apta para o plantio de eucalipto) baseado em variáveis independentes (HOSMER & LEMESHOW, 1989). Tais

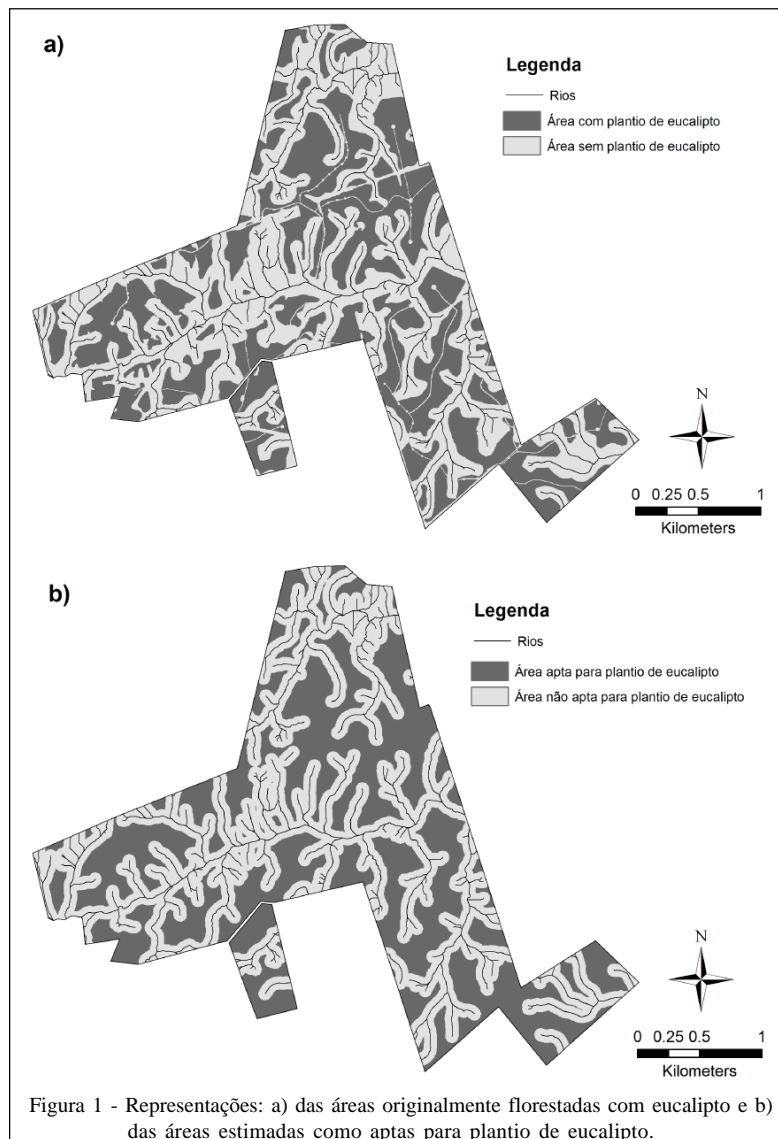
^IPrograma de Pós-graduação em Ciência do Solo, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul (UFRGS). Av. Bento Gonçalves, 7712, 91501-970, Porto Alegre, RS, Brasil. E-mail: rodrigoteske.agr@gmail.com. Autor para correspondência.

^{II}Departamento de Solos, Faculdade de Agronomia, UFRGS, Porto Alegre, RS, Brasil.

regressões ainda têm tido pouco uso nas ciências agrárias e florestais, porém, pode-se citar seu uso em alguns trabalhos como de LONGHI et al. (2009), estudando a ocorrência de *Araucaria angustifolia* e em diversos estudos sobre mapeamento digital de solos (GIASSON et al., 2006; LAGACHERIE & MCBRATNEY, 2007; TEN CATEN et al., 2011). O objetivo deste trabalho foi gerar mapas preditores de áreas aptas ao plantio de eucalipto utilizando-se sistemas de informações geográficas e regressões logísticas binomiais com variáveis geomorfométricas geradas a partir do modelo numérico do terreno (MNT) e do mapa de hidrografia, a fim de servirem como subsídio para a tomada de decisões sobre aquisição de terras para florestamento e para o planejamento do plantio de eucalipto.

A área em estudo está localizada ao sul do Rio Grande do Sul, dentro dos limites municipais de Arroio Grande (UTM 22S Leste de 294.218m a 289.340m, e Norte 6.426.395m a 6.430.466m). O clima regional, segundo Köppen, é subtropical Cfa com média anual de temperatura de 17,5°C e precipitação média anual de 1.379mm. A área estudada constitui-se de propriedade rural com extensão de 754,5 hectares com relevo plano a ondulado. Na ocasião deste estudo, as áreas possíveis de plantio já tinham sido plantadas e representavam a aptidão para o plantio de eucalipto avaliada a campo (Figura 1a).

Utilizando-se o aplicativo computacional ArcGis 9.2 (ESRI, 2006), foi estabelecida base cartográfica constando de hidrografia e curvas de nível



espaçadas verticalmente 3m, obtidas previamente por estação total. O MNT foi gerado a partir da interpolação das curvas de nível e utilizado para geração das variáveis geomorfométricas (SCULL et al., 2003) declividade, curvatura do terreno, direção do fluxo, acúmulo do fluxo, comprimento do fluxo, distância euclidiana dos rios, aspecto (direção angular de plantio) e Índice de Umidade Topográfica (BEVEN & KIRKBY, 1979). Também foi gerado mapa de ocorrência de áreas de proteção ambiental (APP) nas áreas próximas aos cursos d'água, conforme Lei Federal n.4.771/65.

A fim de correlacionar as variáveis preditoras com o plantio de eucalipto, foi feita amostragem aleatória constando de 3.500 pontos. Em cada ponto de amostragem, foram coletadas informações sobre as variáveis preditoras e se era ou não uma área plantada com eucalipto. Esses dados foram exportados em forma de tabela e analisados utilizando o software Weka 3.5.8 (WITTEN & FRANK, 2005), em que a equação obtida na forma de *logit* possibilitou a estimativa da probabilidade de ser uma área apta ao plantio. Dessa forma, foi gerado mapa no qual cada pixel recebeu uma probabilidade de ser área apta ao plantio. A fim de gerar um mapa binário, mostrando as áreas aptas e não aptas ao plantio de eucalipto, houve a necessidade de estabelecer-se a partir de qual probabilidade efetivamente se consideraria uma área como apta ao

plantio. Não existindo referências a esse respeito, explorou-se o comportamento do modelo, testando probabilidade de plantio entre 5 e 100%, gerando-se 20 diferentes mapas de áreas aptas ao plantio, os quais foram comparados com o mapa original de plantio por meio de matrizes de erros (CONGALTON, 1991). A partir dessas matrizes, diversos indicadores de acurácia e de erros foram calculados e avaliados: a) acurácia geral; b) acurácia do mapeador; c) acurácia do usuário; d) erros de inclusão: $EI(\%)=1-AU$; e) erros de omissão: $EO(\%)=1-AM$; e f) índice Kappa de Cohen (K) (COHEN, 1960).

Das variáveis geomorfométricas testadas no algoritmo de regressão logística binomial, a distância dos rios e a ocorrência de APP foram as que explicaram o padrão de plantio de eucalipto, tendo sido selecionadas para a regressão logística resultante:

$$\text{logit} = (-0,84 + APP) \times (-0,8 + 0,02d) \quad (\text{Eq. 1})$$

em que: *logit* = função logística; APP: variável binária com valores de 0 para “não ocorrência de APP” e 1 para “ocorrência de APP”; d = distância dos rios (m).

Conforme a tabela 1, os resultados mostram índices Kappa considerados muito bons, de acordo com a classificação de LANDIS & KOCH (1977) para os valores obtidos ao se utilizar as probabilidades de 55, 60 e 65%, que resultaram em AG de 85,0, 86,3 e 85,9% e índice Kappa de 0,69, 0,72 e 0,72,

Tabela 1 - Valores das acurácias e erros para os mapas de predição de áreas aptas para plantio de eucalipto produzidos usando diversos valores de probabilidade (AG = acurácia geral; AU = acurácia do usuário; AM = acurácia do mapeador; EI = erro de inclusão; EO = erro de omissão; K = índice Kappa de Cohen).

Probabilidade (%)	AU (%)	AM (%)	EI (%)	EO (%)	AG (%)	K
5	70,6	0,6	29,4	99,4	57,4	0,00
10	80,0	4,3	20,0	95,7	58,6	0,04
15	83,5	9,5	16,5	90,5	60,5	0,09
20	85,6	16,4	14,4	83,6	63,1	0,16
25	86,7	24,4	13,3	75,6	66,1	0,24
30	87,2	33,2	12,8	66,8	69,4	0,32
35	87,6	42,5	12,4	57,5	72,8	0,41
40	87,5	51,5	12,5	48,5	76,1	0,49
45	86,9	61,1	13,1	38,9	79,4	0,56
50	85,9	70,3	14,1	29,7	82,4	0,63
55	84,3	79,8	15,7	20,2	85,0	0,69
60	81,8	87,5	18,2	12,5	86,3	0,72
65	77,4	94,6	22,6	5,4	85,9	0,72
70	71,7	98,3	28,3	1,7	82,7	0,66
75	64,8	99,7	35,2	0,3	76,7	0,56
80	64,3	99,7	35,7	0,3	76,2	0,55
85	64,3	99,7	35,7	0,3	76,2	0,55
90	64,3	99,7	35,7	0,3	76,2	0,55
95	51,8	100,0	48,2	0,0	60,3	0,27
100	43,8	100,0	56,2	0,0	45,1	0,04

respectivamente. Comparando-se esses valores com resultados do uso de regressões logísticas no mapeamento digital de solos, percebe-se que os valores de índice Kappa são superiores aos obtidos por GIASSON et al. (2006), que obtiveram índices Kappa de 0,51 a 0,54.

Observa-se que com valores baixos de probabilidade de ser área apta ao plantio de eucalipto, os erros de omissão são grandes, enquanto o aumento das probabilidades aumenta os erros de inclusão. Os erros de omissão são baixos e os erros de inclusão aumentam rapidamente com o aumento da probabilidade, a partir da probabilidade 60%, por essa razão, os melhores resultados gerais foram encontrados para o modelo logístico do mapa gerado a partir do uso de probabilidades de 60% (Figura 1b). Nesse mapa, 86,3% das áreas estão mapeadas corretamente, sendo que as áreas mapeadas como aptas estão 81,8% das vezes mapeadas corretamente e 87,5% das áreas aptas foram mapeadas corretamente.

Eventualmente, outras variáveis podem mostrar-se mais relevantes na predição de áreas de plantio em situações de recursos naturais distintos, como áreas mais declivosas ou mais úmidas. Assim, o método aqui testado mostra-se capaz de prever áreas de plantio com bastante eficiência para qualquer situação de recurso natural, desde que o modelo seja devidamente calibrado para a situação de recurso natural em avaliação, tendo uma área teste como referência. A regressão logística demonstrou viabilidade de seu uso para a estimativa de áreas aptas para o plantio de eucalipto, possibilitando seu uso na avaliação de terras para análises iniciais de aquisição de terras, quantificação e qualificação de áreas para florestamento com eucalipto.

REFERÊNCIAS

- BEVEN, K.; KIRKBY, N. A physically based variable contributing area model of basin hydrology. *Hydrological sciences. Bulletin des Sciences Hydrologiques*, v.24, p.43-69, 1979. Disponível em: <<http://dx.doi.org/10.1080/02626667909491834>>. Acesso em: 03 jun. 2011. doi: 10.1080/02626667909491834.
- COHEN, J. A coefficient of agreement for nominal scales. *Journal of Educational and Measurement*, v.20, n.1, p.37-46, 1960. Disponível em: <<http://dx.doi.org/10.1177/001316446002000104>>. Acesso em: 15 maio, 2011. doi: 10.1177/001316446002000104.
- CONGALTON, R.G. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, v.37, p.35-46, 1991. Disponível em: <[http://dx.doi.org/10.1016/0034-4257\(91\)90048-B](http://dx.doi.org/10.1016/0034-4257(91)90048-B)>. Acesso em: 21 maio, 2011. doi: 10.1016/0034-4257(91)90048-B.
- ESRI. **ArcGis 9.2**. Redland, California, 2006. (Software).
- GIASSON, E. et al. Digital soil mapping using multiple logistic regressions on terrain parameters in Southern Brazil. *Scientia Agricola*, v.63, n.3, p.262-268, 2006. Disponível em: <<http://dx.doi.org/10.1590/S0103-90162006000300008>>. Acesso em: 03 ago. 2011. doi: 10.1590/S0103-90162006000300008.
- HOSMER, D.W.; LEMESHOW, S. **Applied logistic regression**. New York: John Wiley & Sons, 1989. 307p.
- LAGACHERIE, P.; MCBRATNEY, A.B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P. et al. (Ed.). **Digital soil mapping: an introductory perspective**. Amsterdam: Elsevier, 2007. p.3-24.
- LANDIS, J; KOCH, G.G. The measurements of agreement for categorical data. *Biometrics*, v.33, n.3, p.159-179, 1977.
- LONGHI, S.J. et al. Fatores ecológicos determinantes na ocorrência de *Araucaria angustifolia* e *Podocarpus lambertii*, na Floresta Ombrófila Mista da FLONA de São Francisco de Paula, RS, Brasil. *Ciência Rural*, v.40, n.1, p.57-63, 2010. Disponível em: <<http://dx.doi.org/10.1590/S0103-84782009005000220>>. Acesso em: 21 maio, 2011. doi: 10.1590/S0103-84782009005000220.
- SCULL, P. et al. Predictive soil mapping: a review. *Progress in Physical Geography*, v.27, p.171-197, 2003.
- TEN CATEN, A. et al. Regressões logísticas múltiplas: fatores que influenciam sua aplicação na predição de classes de solos. *Revista Brasileira de Ciência do Solo*, v.35, n.1, p.53-62, 2011.
- WITTEN, A.H.; FRANK, E. **Data mining: practical machine learning tools and techniques**. 2.ed. San Francisco: Morgan Kaufmann, 2005. 560p.