



Modelo logístico e *Wavelets* para o diagnóstico de doença em clones híbridos de *E. grandis* x *E. urophylla*

Márcia Lorena Alves dos Santos¹, Eniuce Menezes de Souza², José Raimundo de Souza Passos³, João Ricardo Favan⁴, André S. Jim⁵ e Edson Luiz Furtado⁶

¹ Mestre em Bioestatística - Universidade Estadual de Maringá - UEM

² Departamento de Estatística - Universidade Estadual de Maringá - UEM

³ Instituto de Biociências - Departamento de Bioestatística - Universidade Estadual Paulista - UNESP/ Botucatu

⁴ Doutorando no Programa de Pós - Graduação em Ciência Florestal FCA/UNESP/Botucatu

⁵ Doutorando no Programa de Pós - Graduação em Ciência Florestal FCA/UNESP/Botucatu

⁶ Docente do Depto de Proteção Vegetal FCA/UNESP/Botucatu

RESUMO

Os clones híbridos de *E. grandis* x *E. urophylla* predominam as áreas de plantio florestal no Brasil, atividade que gera milhões de empregos nos diversos segmentos da base florestal plantada. A redução da produtividade está associada a bacteriose foliar, doença caracterizada por lesões sobre as folhas. A reflectância foliar é utilizada para detectar a doença, a qual é medida ao longo de comprimentos de onda. Este estudo objetiva investigar as medições das reflectâncias com o propósito de definir quais comprimentos de onda são mais sensíveis à infecção provocada pela bacteriose em mudas de clones híbridos de *E. grandis* x *E. urophylla*, propondo o diagnóstico precoce. Mudas infectadas e sadias foram sorteadas diariamente para a tomada das medidas da reflectância ao longo de 128 comprimentos de onda. Considerando a variável resposta binária, foi investigada a possibilidade de previsão da infecção a partir do modelo logístico. Devido a baixa qualidade preditiva deste modelo, foi construído um modelo logístico híbrido, em que as covariáveis são os coeficientes *wavelets* de detalhes e coeficientes escala suaves das covariáveis de reflectância. A partir da transformada *wavelet* discreta não decimada foi possível extrair informações sobre a doença em diferentes níveis de resolução e prever o diagnóstico. Os modelos foram avaliados neste primeiro momento de acordo com o Pseudo R^2 de Mc Fadden. Os resultados mostraram que apenas os coeficientes de escala do último nível suave possuem informações suficientes para explicar o diagnóstico. Verificou-se que medições da reflectância para comprimentos de onda superiores a 1126 nm são suficientes para a detecção da doença.

Palavras chave: *Wavelets*, Regressão Logística, Coeficiente de determinação, Eucalyptus, Bacteriose Foliar.

1. INTRODUÇÃO

O setor de base florestal brasileiro tem participação significativa na economia nacional com números em torno dos 3,5% do PIB nacional[1]. No ano de 2011 o valor bruto da produção florestal atingiu 53,91 bilhões de reais, gerando 4,73 milhões de empregos diretos nos diversos segmentos da cadeia produtiva de base florestal plantada, cuja área atingiu 6,5 milhões de hectares[2]. Desse total, a área ocupada por plantios florestais de clones híbridos de *E. grandis* x *E. urophylla* e Pinus no Brasil totalizou 6.515.844 ha, sendo 74,8% correspondente à área de plantios de clones híbridos e 25,2% aos plantios de Pinus. Mediante tais considerações, fica evidente a importância de investigar os fatores que contribuem para a redução da produtividade florestal.

Predominante nas áreas de plantios florestais, a redução da produtividade de clones híbridos de *E. grandis* x *E. urophylla* está vinculada a ocorrência da bacteriose foliar, uma doença caracterizada por lesões concentradas nas margens ou distribuídas aleatoriamente sobre o limbo das folhas[3]. Os agentes causais da bacteriose foram identificados como *Xanthomonas axonopodis* e *Pseudomonas cichori*[4]. Os danos causados pela doença são de 3 milhões de mudas por ano, o que acarreta em perdas estimadas de 2,4 milhões de reais por ano[5].

O diagnóstico da bacteriose foliar é realizado pela severidade da mancha foliar bacteriana da planta, a qual resulta na modificação das propriedades óticas dos tecidos das folhas. A reflectância foliar é uma alternativa a detecção da bacteriose foliar, em virtude da sua sensibilidade ao estresse das plantas à mudança na pigmentação, reação de hiperesensibilidade e degradação celular[6]. A resposta fornecida pela reflectância foliar pode se diversificar em dois segmentos: temporal e comprimentos de onda, sendo a primeira resposta associada a evolução da doença e a segunda ao tempo do processo infeccioso. Os comprimentos de onda são resultados da interação do espectro eletromagnético com a estrutura foliar.

Diante disso, o presente trabalho irá investigar as medições das reflectâncias foliar com o propósito de definir quais comprimentos de onda são mais sensíveis à severidade da infecção provocada pela bacteriose foliar em mudas de clones híbridos de *E. grandis* x *E. urophylla*, visando o diagnóstico precoce da doença. Em outras palavras, quer-se identificar a partir de qual comprimento de onda é possível diagnosticar a infecção. A modelagem da reflectância deu-se pelo modelo de regressão logístico e pela combinação deste com a transformada *wavelet* discreta não decimada, com a presença de uma única covariável.

2. METODOLOGIA

Na totalidade, foram utilizadas 700 mudas do clone híbrido de *E. grandis* x *E. urophylla* sem patogenicidades, cedidas pelo viveiro Pirafloira situado em Paranapanema, São Paulo. Dentre elas, 288 mudas foram separadas em dois tratamentos: inoculado e controle. O tratamento inoculado refere-se as mudas com inoculação da estirpe de *Xantomonas* spp 3177, responsável pelo diagnóstico de bacteriose foliar e o tratamento controle diz respeito as mudas sem inoculação da bactéria. Para cada tratamento foram destinadas 6 bandejas com 24 mudas, ou seja, 144 mudas inoculadas e 144 mudas sadias.

Diariamente, 6 mudas de cada tratamento foram sorteadas para serem retiradas 4 folhas, que foram mantidas em um recipiente térmico até a avaliação. A preparação para a tomada de medidas da reflectância iniciou-se com secagem de cada folha seguida pela fixação das mesmas em uma folha de papel com espaçamento pré definido. A reflectância foi tomada a partir de espectroradiômetro da Ocean Optics, modelo Flame - NIR, 950 a 1650 nm, processo este que teve a duração de 8 dias entre o período de 01 a 08 de setembro de 2017. A tomada das medidas foi interrompida devido ao enfraquecimento das folhas decorrente dos fungos.

Portanto, 24 folhas inoculadas e 24 folhas sadias foram coletadas diariamente e a partir destas realizou-se as medições da reflectância foliar perante 128 diferentes valores de comprimentos de onda, variando entre 966 nm até 1685 nm. Neste trabalho, a análise considerou apenas as medições da reflectância tomadas no terceiro dia. Cada faixa de comprimento de onda considerada na medição da reflectância representa uma covariável, de maneira que o conjunto de dados das medidas tomadas no terceiro dia é composto por 128 covariáveis. Por sua vez, cada covariável apresenta 24 observações pertinentes ao tratamento inoculado e 24 observações concernentes ao tratamento controle.

Seguidamente, a transformada *wavelet* discreta não decimada (TWDND) foi aplicada em cada covariável, decompondo-as em 4 níveis de resolução segundo a *wavelet* de Haar. As *wavelets* são funções matemáticas que satisfazem determinados requisitos, necessários para garantir cálculo rápido e fácil da decomposição dos dados nas suas partes constituintes[7]. Além disso, a TWDND é uma derivação da transformada *wavelet* decimada e destaca-se por garantir a invariância por translação do cálculo dos coeficientes *wavelets*, característica essencial para a análise de séries temporais[8].

Por consequência, oito conjuntos de dados semelhantes ao conjunto inicial foram obtidos, respectivos aos níveis da transformada com coeficientes *wavelets* de detalhes d1, d2, d3, d4 e coeficientes escala suaves s1, s2, s3 e s4. Nesse momento, as 24 observações dos tratamentos inoculado e controle foram substituídas por 24 coeficientes *wavelets* ou escala em cada nível. Assim, ao invés de modelar o conjunto de dados inicial composto pelas medidas da reflectância no terceiro dia, o intuito foi "fragmentar" esse conjunto em níveis de resolução por meio da TWDND e modelar os coeficientes *wavelets*, visando reconhecer o nível que possibilita a identificação da bacteriose foliar ao longo das faixas de comprimentos de onda. Em resumo, são cinco conjuntos de dados, um para cada nível, compostos por 128 covariáveis que possuem 24 coeficientes *wavelets* para o tratamento inoculado e 24 coeficientes *wavelets* para o controle.

A partir de agora, resta modelar os coeficientes *wavelets* dos comprimentos de onda pertinentes a cada nível. Lembre que, a finalidade da análise consiste em reconhecer a partir de qual comprimento de onda é possível prever o tratamento que a planta pertence. Nessa perspectiva, a variável de interesse é dicotômica, planta inoculada ou não, constituindo uma variável categórica de natureza nominal. A relação entre uma variável aleatória dicotômica e um conjunto de covariáveis quantitativas é comumente descrita pela regressão logística. Os modelos foram construídos considerando uma covariável de cada vez, isto é, para cada conjunto de dados com 48 observações totais (24 inoculado e 24 controle) foram construídos 128 modelos de regressão logística considerando cada faixa de comprimento de onda individualmente.

Para comparar a capacidade preditiva dos modelos de regressão logística em cada nível, utilizou-se em um primeiro momento o pseudo R^2 de Mc Fadden. Dessa forma, uma sequência de 128 valores de R^2 foram calculados em cada conjunto de dados e por meio destes foi possível identificar qual nível de resolução apresentou modelos que melhor descreveram a variável resposta e, também, verificar a partir de qual faixa de comprimento de onda os modelos logísticos apresentaram coeficientes de determinação desejáveis, indicando o comprimento de onda que melhor viabiliza a predição do diagnóstico de bacteriose foliar em determinada planta.

3. RESULTADOS E DISCUSSÕES

A Figura 1 exibe os valores dos coeficientes de determinação R^2 para os modelos de regressão logística relativos aos níveis de detalhes e suaves, respectivamente, decorrentes da TWDND mediante a *wavelet* de Haar. Para os níveis de detalhes, os 128 modelos de regressão apresentaram valores de R^2 nulos, indicando a ausência de informação nos coeficientes *wavelets* de detalhes para explicar ou prever o diagnóstico da bacteriose foliar. Deste modo, as faixas de comprimentos de onda não influenciam no desempenho preditivo desses modelos.

Por outro lado, a Figura 1 evidencia a presença de informação nos coeficientes *wavelets* suaves em todos os níveis (s1, s2, s3, e s4), sugerindo que os mesmos capturam mais efetivamente as características da bacteriose foliar nas folhas examinadas. Além disso, é possível visualizar a tendência crescente de R^2 à medida que os comprimentos de onda são amplificados. No entanto, os níveis s1 e s2 exibem valores baixos de coeficientes de determinação, tendo valores mínimos iguais a 0.07 e 0.14 e valores máximos iguais a 0.52 e 0.68, respectivamente. Para o nível s3, os coeficientes R^2 atingem valores consideráveis apenas nas duas últimas faixas de comprimento de onda, sendo $R^2 = 0.817$ para a covariável R1679 e $R^2 = 1$ para a covariável R1685. É válido enfatizar que o nível s4 apresentou os melhores valores de coeficientes de determinação para os modelos de regressão logística entre todos os níveis de decomposição, variando de $R^2 = 0.891$ até $R^2 = 1$ e manteve-se constantemente alto a partir da covariável R1126, a qual representa o comprimento de onda igual a 1126 nm. Isso significa que, os coeficientes *wavelets* pertinentes ao último nível suave contém considerável informação para explicar ou prever o diagnóstico da bacteriose foliar, principalmente para comprimentos de onda superiores a 1126 nm.

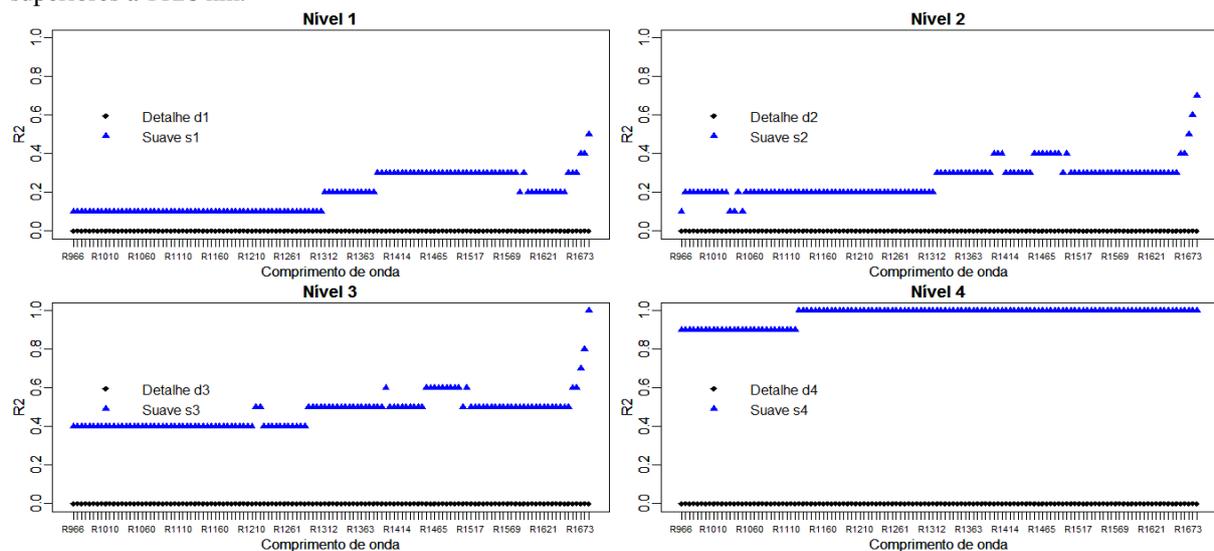


Figura 1: Pseudos R^2 de Mc Fadden para 128 modelos de regressão logística do diagnóstico de bacteriose usando coeficientes *wavelets* de detalhes e suaves ao longo dos comprimentos de onda.

Ainda, a Figura 2 apresenta os valores dos coeficientes de determinação R^2 para os modelos de regressão usuais, quer dizer, sem considerar a decomposição *wavelet* nas covariáveis. De forma análoga aos modelo híbridos, foram construídos 128 modelos de regressão logística usuais considerando cada faixa de comprimento de onda individualmente. A capacidade preditiva de tais modelos manteve-se baixa para todos os comprimentos de onda, atingindo o maior valor $R^2 = 0.4$ para a covariável R1685. Em vista disso, fica notório a importância da decomposição *wavelet* na modelagem híbrida do diagnóstico da bacteriose foliar. Com respeito a análise de resíduos dos modelos de regressão, em todos os casos, os resíduos mostraram-se condizentes com os valores obtidos dos coeficientes de determinação, confirmando a interpretação dos mesmos.

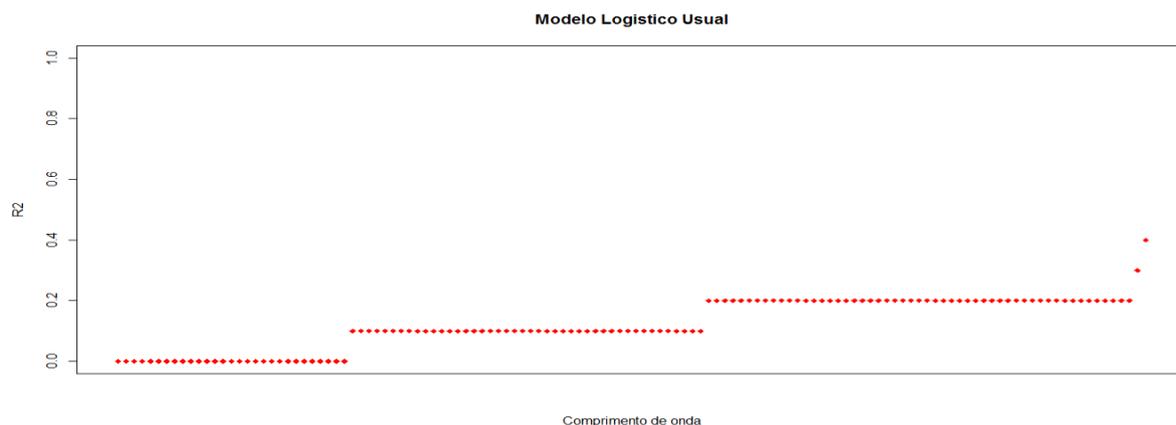


Figura 2: Pseudo R^2 de Mc Fadden para o modelo de regressão logística usual (sem *wavelets*) do diagnóstico de bacteriose foliar.

4. CONCLUSÃO

Conforme os resultados obtidos para os dados em questão, concluímos que os modelos híbridos de regressão logística aplicados aos coeficientes escalas do último nível suave (*s4*) da decomposição *wavelet* são capazes de prever o diagnóstico da bacteriose foliar das mudas de clones híbridos de *E. grandis* x *E. urophylla* ao longo dos comprimentos de ondas considerados neste estudo para a tomada das medidas de reflectância. Enfatizando que, a partir da faixa de comprimento de onda de 1126 nm, os modelos atingem $R^2=1$, apresentando eficiência máxima em explicar a variabilidade da variável dicotômica definida como: muda inoculada ou não (controle). Nessa perspectiva, sugere-se uma restrição das faixas de comprimentos de ondas para a medição da reflectância, sendo suficiente medir para comprimentos superiores a 1126 nm para o diagnóstico da planta. Vale ressaltar que se o modelo logístico é aplicado aos dados de reflectância sem a decomposição *wavelet*, o desempenho preditivo é muito fraco, atingindo no máximo 36 %, aproximadamente. Outras medidas de diagnóstico, embora não apresentadas, também foram verificadas. Ainda é importante destacar a possibilidade de investigar tais resultados considerando *wavelets* de diferentes famílias, além da *wavelet* de Haar aqui abordada. A mesma investigação estende-se ao número de níveis de decomposição da transformada *wavelet*. Os resultados apresentados referem-se a análise no terceiro dia após a infecção. Em trabalhos futuros, será investigado o dia em que a capacidade preditiva do modelo é maximizada.

5. AGRADECIMENTOS

Este trabalho é resultado do Projeto Universal financiado pelo CNPq. Processo: 475968/2013-1.

Referências

- [1] SOCIEDADE BRASILEIRA DE SILVICULTURA - SBS. Fatos e números do Brasil florestal, 2012.
- [2] ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS - ABRAF. Anuário Estatístico ABRAF, 2012.
- [3] Gonçalves, R. C.; Lau D.; Oliveira J.R.; Maffia L. A.; Cascardo J.C.M.; Alfnas A.C. Tropical Etiology of bacterial leaf blight of eucalyptus in Brazil. *Plant Pathology*, vol. 33, 3, 180-188 (2008).
- [4] Furtado E.L.; Dias D. C.D.; Ohto C.T.; Rosa D.D. Doenças do Eucalipto no Brasil. FEPAF/FCA/UNESP, 2009.
- [5] Furtado E.L.; Wilken C.F. Principais doenças e pragas em plantios de eucalipto no Brasil. IPEF/PROTEF Câmara Setorial de Florestas Plantadas no Brasil, maio/2011.
- [6] Mahlein A.K.; Oerke E.C.; Steiner U.; Dehne H.W. Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology*, 133:197–209, 2012.
- [7] VIDAKOVIC, B.; MUELLER, P. *Wavelets for kids*. Instituto de Estadística, Universidad de Duke, 1994.
- [8] NASON, G.; SAPATINAS, T.; SAWCZENKO, A. Statistical modelling of time series using non-decimated *wavelet* representations. submitted for publication, 1997