

ÁRVORE DE DECISÃO PARA CLASSIFICAÇÃO DE OCORRÊNCIAS DE FERRUGEM ASIÁTICA EM LAVOURAS COMERCIAIS COM BASE EM VARIÁVEIS METEOROLÓGICAS

GUILHERME A. S. MEGETO¹, STANLEY R. DE M. OLIVEIRA²,
EMERSON M. DEL PONTE³, CARLOS A. A. MEIRA⁴

RESUMO: A ferrugem asiática é a mais importante doença da soja no Brasil. Apesar de sua epidemiologia ser conhecida, são escassos os estudos sobre os fatores que desencadeiam a doença com base em dados de campo. Este trabalho objetivou modelar a influência de variáveis meteorológicas a partir de um conjunto extenso de dados de ocorrência da ferrugem, por meio da técnica de indução de árvores de decisão. Os modelos foram desenvolvidos com dados de data de ocorrência da doença em quatro safras (2007/08 a 2010/11) e variáveis de temperatura e chuva em diferentes janelas de tempo prévias à data de detecção. Para cada registro de ocorrência, foi gerado um correspondente de “não ocorrência” como sendo o trigésimo dia anterior ao dia da detecção, assumindo-se a presença de inóculo, mas condições meteorológicas desfavoráveis à doença. O conjunto de treinamento para a modelagem foi composto de 45 variáveis de chuva e temperatura e 12.591 registros. O modelo preditivo escolhido resultou em uma árvore de decisão com, aproximadamente, 78% de taxa de acerto e 108 regras, determinadas por validação cruzada. O modelo interpretado, com 28 regras, considerou variáveis de temperatura como mais importantes, sendo que temperaturas abaixo de 15 °C e acima de 30 °C foram relacionadas com eventos de não ocorrência, enquanto temperaturas dentro da faixa favorável foram associadas com eventos de ocorrência, mostrando coerência com a literatura.

PALAVRAS-CHAVE: *Phakopsora pachyrhizi*, mineração de dados, previsão de doenças de plantas, epidemiologia, sistemas de suporte à decisão.

DECISION TREE FOR CLASSIFICATION OF SOYBEAN RUST OCCURENCE IN COMMERCIAL CROPS BASED ON WEATHER VARIABLES

ABSTRACT: Soybean rust is the most aggressive soybean disease in Brazil. Despite its epidemiology is known, there are few studies about factors that cause it based on field data. This paper aimed to report influence of weather variables on rust occurrence using the decision tree technique. The models were developed based on disease detection dataset during harvests (2007/08 to 2010/11), temperature and rainfall variables at varied time windows prior to disease detection. For each disease “occurrence” record, a corresponding “non-occurrence” was generated based on the assumption that disease was not present at the thirtieth day prior to the report date, due to unfavorable weather conditions. The training set for modeling consisted of 45 rainfall and temperature variables and 12,591 records. The chosen predictive model resulted in a decision tree with approximately 78% of accuracy and 108 rules, determined by cross-validation. The interpreted model, with 28 rules, considered the temperature variables as more important, of which temperatures below 15 °C and above 30 °C were related to events of non-occurrence, while temperatures within the favorable range have been associated with events of occurrence, showing consistency with the literature.

KEYWORDS: *Phakopsora pachyrhizi*, data mining, plant disease forecast, epidemiology, decision support systems.

¹ Mestre em Eng. Agrícola, Faculdade de Eng. Agrícola, UNICAMP/Campinas - SP, guilherme.megeto@colaborador.embrapa.br

² Dr. em Ciência da Computação, Pesquisador da Embrapa Informática Agropecuária, Prof. do Programa de Pós-Graduação da Feagri/UNICAMP, stanley.oliveira@embrapa.br.

³ Eng.º Agrônomo, Prof. Dr., Depto. de Fitossanidade, UFRGS, Porto Alegre - RS, emerson.delponte@ufrgs.br.

⁴ Dr. em Eng. Agrícola, Pesquisador, Embrapa Informática Agropecuária, EMBRAPA, Campinas - SP, carlos.meira@embrapa.br

Recebido pelo Conselho Editorial em: 22-11-2012

Aprovado pelo Conselho Editorial em: 6-1-2014

INTRODUÇÃO

A ferrugem asiática, causada pelo fungo *Phakopsora pachyrhizi* Syd. & P. Syd., atualmente é considerada uma das doenças mais importantes da soja. No Brasil, sua primeira ocorrência foi observada em 2001 (YORINORI, 2005). A partir da safra de 2007/2008, os danos em produtividade ocasionados pela ferrugem têm-se mantido relativamente mais baixos, especialmente devido à aplicação massiva de fungicidas para o controle da doença, o que representa a maior parte do custo de controle, o qual se tem mantido acima de dois bilhões de dólares anuais (CONSÓRCIO ANTIFERRUGEM, 2012).

O controle da doença é realizado com aplicações de fungicidas, em sua maioria, feitas de forma calendarizada, sem levar em conta os fatores de risco (GODOY et al., 2009). Essa prática pode, por um lado, resultar em aplicações desnecessárias e, por outro, iniciar aplicações com atraso, diminuindo a eficiência do controle. Nesse sentido, sistemas de alerta poderiam indicar a necessidade e o momento mais apropriado para as aplicações de fungicidas.

Estudos sobre fases específicas do ciclo da doença proporcionaram um conhecimento básico e importante sobre a epidemiologia da ferrugem asiática da soja (MARCHETTI et al., 1976; KOCHMAN, 1979; ALVES et al., 2006; BONDE et al., 2007). Diversos modelos e sistemas de alerta de ferrugem asiática foram desenvolvidos, em sua maioria baseados em abordagens mecanística ou empírica, ou seja, por meio de ajustes de modelos aos dados obtidos em ambiente controlado ou em condições de campo (ver revisão sobre os modelos em DEL PONTE et al., 2006a). Entretanto, alguns modelos utilizaram informações muito específicas sobre o hospedeiro ou sobre o patógeno, o que pode restringir a interpretação e a aplicação dos modelos.

Além disso, foram utilizadas mais informações provenientes de ambientes controlados do que de situações de campo, porém as interações entre os fatores presentes em ambientes naturais podem ser mais complexas (DEL PONTE & ESKER, 2008). Assim, são necessários mais estudos com dados provenientes de situações de campo e também utilizar técnicas ainda não exploradas.

A indução de árvores de decisão é uma técnica de mineração de dados que fornece uma abordagem de modelagem alternativa à estatística clássica. A representação simbólica e interpretável do modelo obtido pode ser utilizada tanto para predição como conhecer a estrutura preditiva do problema. É um modelo representado graficamente por nós e ramos, parecido com uma árvore, mas no sentido invertido. Os nós de decisão - o nó-raiz, no topo da estrutura, e os nós internos - contêm um teste de valor sobre um dos atributos mais relevantes, e os resultados desses testes formam os ramos. Os nós-folha, nas extremidades, referem-se às classes da variável resposta e representam o resultado da predição obtida pelo modelo (WITTEN et al., 2011).

Modelos em árvore de decisão vêm sendo aplicados em diversas áreas do conhecimento, inclusive na área agrícola, como os modelos de classificação de área plantada de cana-de-açúcar por meio de sensoriamento remoto (DELGADO et al., 2012). Epidemias da ferrugem do cafeeiro foram analisadas com o auxílio de uma árvore de decisão (MEIRA et al., 2008), e modelos em árvore de decisão foram desenvolvidos, com 81% de taxa de acerto, para prever a taxa de infecção da ferrugem igual ou superior a 5 pontos percentuais para o prazo de um mês, utilizando informações meteorológicas e o espaçamento entre plantas (MEIRA et al., 2009).

Considerando a escassez de trabalhos utilizando dados de campo e a existência de uma extensa base de dados com registros de ocorrência da ferrugem asiática da soja, hipotetiza-se ser possível utilizar modelos em árvores de decisão e obter resultados coerentes com conhecimento atual. O objetivo do trabalho foi construir e avaliar modelos em árvore de decisão, com regras baseadas em variáveis meteorológicas, para prever situações de ocorrência ou não ocorrência da ferrugem asiática, a partir de um extenso conjunto de dados de monitoramento da doença no Brasil.

MATERIAL E MÉTODOS

Base de dados da doença

Os dados de ocorrência da ferrugem da soja foram obtidos junto ao Consórcio Antiferrugem disponibilizados em sua página na Internet (CONSÓRCIO ANTIFERRUGEM, 2013). Embora a base de dados contivesse registros de ocorrências desde a safra de 2004/2005, devido à falta de dados meteorológicos para todos os anos, foram utilizadas as ocorrências de ferrugem da soja relatadas em lavouras comerciais entre as safras de 2007/2008 e 2010/2011 (4 safras), perfazendo um total de 7.810 registros, incluindo-se ocorrências múltiplas para um mesmo município, em uma mesma safra. Os atributos disponíveis para cada registro eram: data de ocorrência, safra, mês de plantio, quinzena de plantio, código do município segundo o IBGE (Instituto Brasileiro de Geografia e Estatística), nome do município, unidade da federação, latitude e longitude do município, nome da cultivar, estágio fenológico da cultura no momento da detecção e o tipo de área (comercial, soja voluntária, etc.).

Base de dados meteorológicos

Para cada município com ocorrência, foram obtidos dados estimados de temperatura mínima, temperatura máxima e chuva, na escala diária, junto ao sistema Agritempo (Sistema de Monitoramento Agrometeorológico). Uma vez que, dos 582 municípios com ocorrência da doença apenas 134 continham dados meteorológicos de estações de superfície, foram utilizados dados simulados de temperatura por interpolação dos dados das estações de superfície (estações virtuais, segundo ROMANI et al., 2007) e dados de chuva estimada pelo radar a bordo do satélite TRMM - *Tropical Rainfall Measuring Mission* (PINTO et al., 2009; COLLISCHONN et al., 2007). Esses dados estavam disponíveis apenas a partir da safra de 2007/2008.

Variáveis preditoras e dependente

Cada registro da doença constituiu uma ocorrência verdadeira (confirmada pelo laboratório), sendo então criada uma variável resposta (atributo meta) definida como “Oc” relacionada à data da ocorrência. Para cada uma das ocorrências, foi criado um registro correspondente de não ocorrência “NaoOc”, definido como sendo o trigésimo dia anterior ao dia do relato da ocorrência. Foi assumida a presença do inóculo, mas condições meteorológicas desfavoráveis à doença.

Foram criadas nove variáveis preditoras com base em número de eventos, sumário e limiares mínimos ou máximos de temperatura (T) e precipitação pluvial (P), descritas na Tabela 1, conforme informações sobre o efeito de fatores meteorológicos no ciclo da doença (DEL PONTE & ESKER, 2008). Cada uma das nove variáveis foi sumarizada em cinco janelas de tempo (período latente; 5; 10; 15 e 20 dias) anteriores à data de infecção (Figura 1), simulada pela eq.(1) do período latente de ALVES et al. (2006). Ao todo, foram criadas 45 variáveis preditoras, as quais foram agregadas ao conjunto de ocorrências e não ocorrências, totalizando 12.591 registros.

Indução e validação das árvores de decisão

Os dados foram preparados com rotinas desenvolvidas em linguagem de programação *Python* ou no software R. As árvores de decisão foram induzidas no software Weka 3.6, por meio do classificador J48, que é uma versão adaptada do conceituado algoritmo C4.5 (QUINLAN, 1993). Buscando-se minimizar um possível efeito de *overfitting* (superajuste) do modelo aos dados, foram removidos os valores discrepantes (*outliers*), usando-se o filtro *InterquartileRange*, embutido no Weka, e o procedimento de pré-poda, com a variação do número mínimo de objetos por folha.

TABELA 1. Variáveis sumário de número de eventos, média ou acumulado, criadas com base na temperatura e precipitação pluvial diária para cada município com ocorrência de ferrugem asiática, visando à indução de árvores de decisão. Summary variables of number of events, average or cumulative based on temperature and rainfall daily data for each municipality with soybean rust occurrence in order to induce decision trees.

Atributo	Descrição
TMIN	Média da temperatura mínima
DTMIN<15	Número de dias com temperatura mínima menor que 15 °C.
TMAX	Média da temperatura máxima.
DTMAX>30	Número de dias com temperatura máxima maior que 30 °C.
TMED	Média da temperatura média diária.
PMED	Média da precipitação (em mm).
PTOTAL	Precipitação (em mm) total ou acumulada.
DP>1	Número de dias com precipitação maior do que 1mm.
PMAX	O máximo evento de precipitação pluvial (em mm).

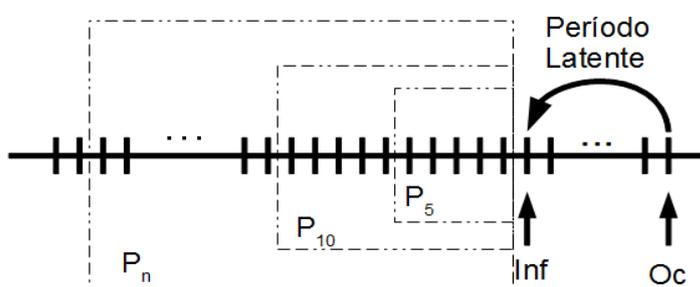


FIGURA 1. Períodos utilizados para compor os atributos preditivos meteorológicos. *Oc* - Dia da ocorrência; *Período Latente* - período latente calculado a partir da eq.(1); *Inf* - data estimada de infecção; *P_n* generaliza o procedimento aplicado para *P₅* e *P₁₀* para os outros períodos de 15 e 20 dias. **Time-frame used to define the meteorological-based variables. *Oc* - Day of occurrence; *Período Latente* (latent period) - obtained from eq.(1) in the text; *Inf* - estimated date of infection; *P_n* procedure applied to *P₅* and *P₁₀* for other periods of 15 and 20 days.**

$$Y = 0,11 T^2 - 5,20 T + 69,53 \quad (1)$$

em que,

Y - número de dias do período latente, e

T - temperatura em °C.

Os modelos induzidos, com variação do número de objetos por folha, foram avaliados utilizando o método de validação cruzada em 10 partes e três métricas: (i) taxa de acerto (acurácia); (ii) o número de folhas (número de regras) geradas, que em geral são associadas à facilidade de interpretação do modelo; (iii) a estatística kappa, que é uma medida de concordância entre as classes preditas e observadas, que deduz o número esperado de acertos (utilizando uma classificação ao acaso) do número real de acertos do classificador.

Após realizar os experimentos para determinar o conjunto de treinamento e os parâmetros iniciais, foi gerado o modelo preditivo. O modelo foi avaliado com base em 10 repetições, sendo que em cada uma foi utilizada a validação cruzada em 10 partes. Com isso, foram obtidas 100 medidas das métricas de avaliação, destacando-se a média e o desvio-padrão.

Ao conjunto de dados que gerou o modelo preditivo, foi aplicado o filtro *RemoveMisclassified*, também embutido no Weka, para remover os registros que foram classificados incorretamente (WITTEN et al., 2011), gerando um conjunto com as mesmas 45

variáveis preditoras do modelo preditivo, mas com 7.129 registros. Para interpretar o modelo, as regras geradas foram analisadas de forma geral, buscando descrever um cenário capaz de explicar a influência de fatores meteorológicos nas ocorrências ou não ocorrências da doença.

RESULTADOS E DISCUSSÃO

Uma primeira avaliação dos modelos gerados envolveu a detecção de valores discrepantes e a pré-poda da árvore, especialmente no que se refere à relação entre o número mínimo de objetos por folhas com o número de folhas e as respectivas taxa de acerto e estatística kappa (Figura 2). A análise dos parâmetros da Figura 2 mostra que a faixa de valores entre 30 e 50 objetos por folha pode ser considerada adequada, uma vez que afetou menos o número de folhas (regras) e apresentou um valor kappa ao redor de 0,6, o que representa um bom índice de concordância e com taxa de acerto (acurácia) ao redor de 80%.

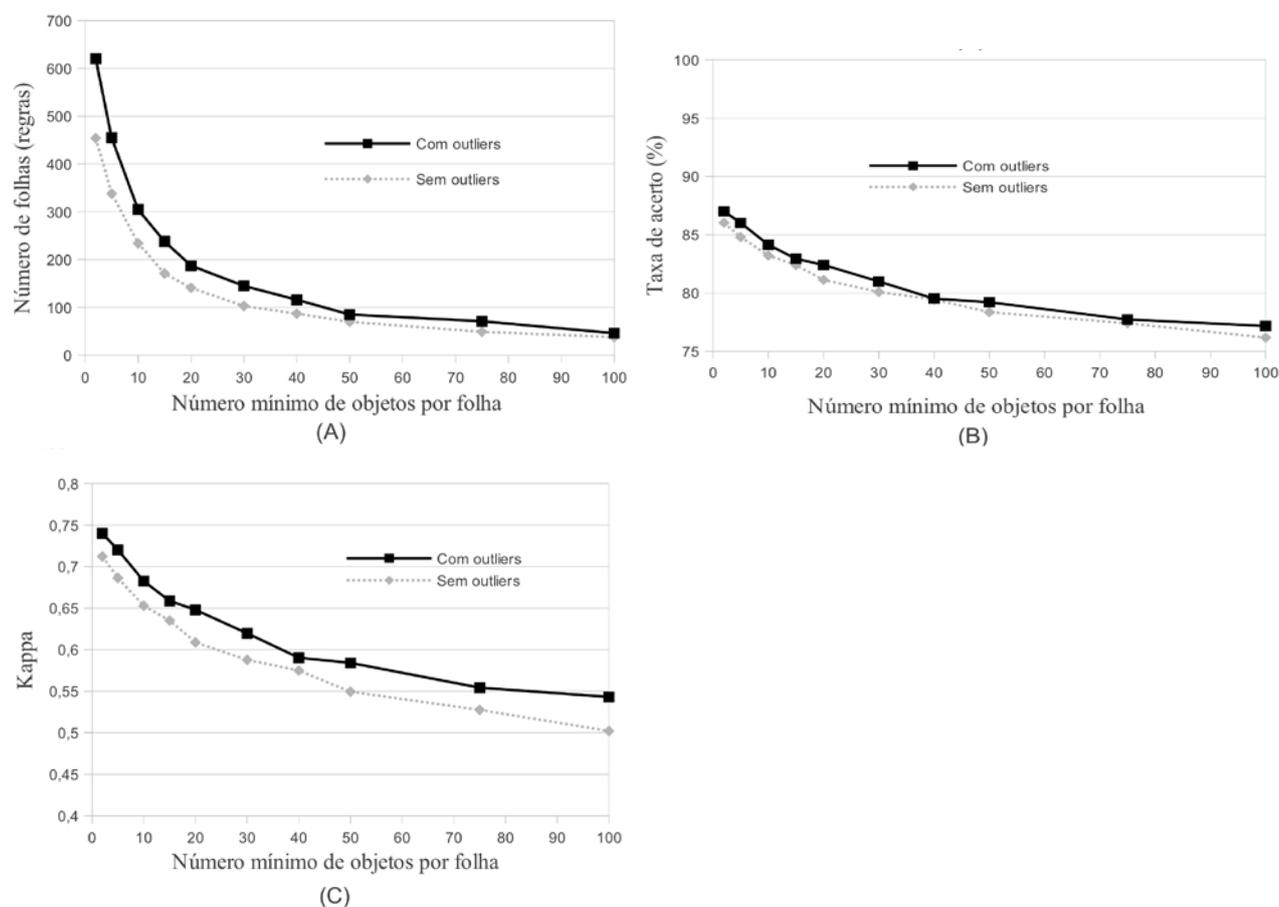


FIGURA 2. Relação entre o número mínimo de objetos por folha e o número de folhas (ou regras de decisão) (A), taxa de acerto (B) e estatística kappa (C). **Relation between the minimum number of objects per leaf and the number of leaves (number of rules)**

O modelo selecionado foi aquele com o número mínimo de 40 objetos por folha. Por não apresentarem diferenças significativas nos resultados em relação aos três aspectos avaliados, foi escolhido o conjunto com o maior número de registros (com *outliers*), corroborando o conceito de que algoritmos de árvores de decisão são robustos o suficiente para minimizar o efeito da presença de ruídos nos dados (TAN et al., 2009).

Desta forma, usando-se o número mínimo de 40 objetos por folha, o modelo preditivo para os casos de ocorrência e não ocorrência da ferrugem asiática da soja apresentou uma taxa de acerto de $78,26 \pm 1,20\%$, estatística kappa de $0,56 \pm 0,02$ e uma média de 108 ± 5 folhas.

Após aplicar o filtro *RemoveMisclassified* (embutido no Weka) para a remoção dos registros classificados incorretamente no modelo preditivo e gerar o modelo interpretativo, houve redução do

número de folhas (ou regras) para 28. O novo modelo (Figura 3) representa o principal padrão encontrado nos registros classificados corretamente.

A variável mais importante (nó 1) foi o número de dias com temperatura mínima abaixo de 15 °C em 20 dias antes da possível infecção. De acordo com QUINLAN (1993), o nó-raiz é o atributo com menor entropia e, como consequência, apresenta o maior ganho de informação, otimizando o processo de geração de árvores de decisão.

No caso de nenhum dia atender ao critério acima, a próxima variável avaliada foi o número de dias com temperatura máxima maior que 30 °C no mesmo período de 20 dias (nó 2). No caso de ter havido apenas um ou dois dias, em 20 dias, com temperatura acima do limiar da regra do segundo nó, 1.370 casos foram classificados como ocorrência.

A partir do quinto nó, para temperaturas mínimas médias abaixo de 18,3 °C, em 15 dias anteriores à possível data de infecção, foram classificados 433 casos de não ocorrência (nó-folha 8). Para temperaturas maiores que esse limiar, e com temperaturas máximas médias acima de 31,9 °C para esse mesmo período (nó 9), foram classificados 170 casos de não ocorrência (nó-folha 13). Segundo experimentos realizados em ambientes controlados, temperaturas abaixo de 15 ou acima de 30 °C foram desfavoráveis à ocorrência da doença (MARCHETTI et al., 1976; KOCHMAN, 1979; ALVES et al., 2006).

No décimo segundo nó, que avaliou o número de dias com temperatura máxima acima de 30 °C no período latente estimado, para nenhum dia, foram classificados 385 casos de ocorrência da doença (nó 16). Seguindo com o nó que avaliou a média da temperatura mínima no período latente (nó 17), foram classificados 180 casos de não ocorrência quando a média de temperatura no período latente foi abaixo de 18,7 °C (nó 22). Segundo dados da literatura, a temperatura tem um efeito significativo na duração do período latente, que é mais curto (6 a 8 dias) na faixa ótima (ao redor de 22 °C) e aumenta consideravelmente quando se distancia do valor ótimo, podendo chegar a 15 ou 20 dias com temperaturas entre 15 e 17 °C (KOCHMAN, 1979; ALVES et al., 2006). Uma vez que os casos correspondentes de não ocorrência precediam os casos de ocorrência, é possível que a temperatura tenha sido o principal limitante na ocorrência da doença, fazendo com que essa fosse detectada somente após os eventos de condições favoráveis, especialmente de temperatura.

Segundo HAN et al. (2011), quanto mais a variável estiver posicionada na porção inferior da árvore (próxima ao nó-folha), menor será sua importância na classificação. Assim, serão destacadas algumas variáveis com base no número de casos classificados ou resultados contraditórios com a literatura.

Por exemplo, o nó 39 avalia o número de dias com precipitação (>1 mm) no período de 5 dias antes da possível infecção, resultando uma classificação de 380 casos de não ocorrência quando houve pelo menos um dia de chuva no período, e 57 casos de ocorrência com nenhum dia de chuva (nós-folha 42 e 43). É possível que tais eventos de precipitação tenham contribuído para a instalação da doença na lavoura, porém 5 dias antes da possível infecção, somado com o período latente, resulta em um período curto para observar os sintomas da doença no campo. Diferentemente de observações realizadas em ambiente controlado, considerando práticas de campo e situações ideais, são necessários, pelo menos, 20 dias após a infecção para que a doença atinja um nível detectável (DEL PONTE & ESKER, 2008). Outros aspectos da influência da precipitação sobre a ferrugem da soja serão vistos mais adiante nesta seção.

Voltando ao primeiro nó e considerando a subárvore à direita, há um significativo maior número de casos de não ocorrência (2001 casos) do que ocorrência (305 casos), evidenciando o efeito prejudicial das baixas temperaturas (<15 °C) na germinação de esporos e eventos de penetração nas plantas (MARCHETTI et al., 1976; KOCHMAN, 1979; ALVES et al., 2006; BONDE et al., 2007).

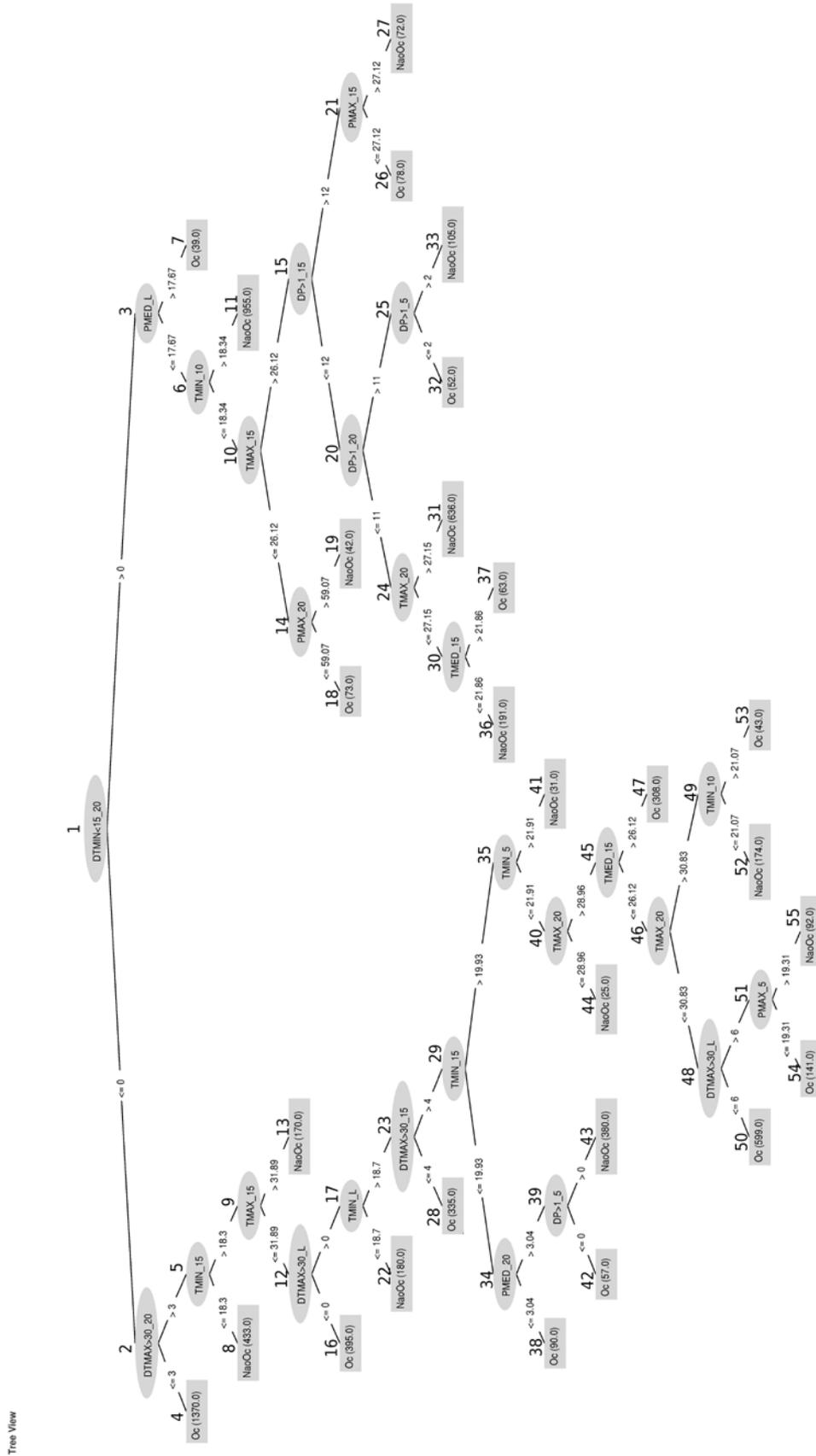


FIGURA 3. Modelo em árvore de decisão para interpretação. **Decision tree model for interpretation.**

O nó-folha 31 foi o que classificou o maior número de casos de não ocorrência (636) e teve sua origem a partir do nó 24, que determinou o limiar de 27,1 °C para a média da temperatura máxima, no período de 20 dias antes do evento. Salienta-se, no entanto, a existência de outros ramos do modelo com casos com temperaturas superiores a esse limiar e ainda assim foram detectados casos de ocorrência, como, por exemplo, a partir do nó 40.

Nesse sentido, considerando estudos de eventos após a ocorrência da doença, DEL PONTE et al. (2006b) não verificaram ganhos de predição nos modelos empíricos (com dados de epidemias no Brasil) com a inclusão de variáveis de temperatura no modelo linear para a predição da severidade máxima da doença. Ou seja, mesmo em locais onde temperaturas acima desses limiares (>30 °C) usualmente ocorrem no verão, as epidemias não foram limitadas pela temperatura, uma vez que ocorressem eventos de precipitação. Os autores discutem que, sob condições de umidade, pode haver diferenças marcantes entre a temperatura do ar e a temperatura no interior do dossel da cultura, que tende a ser mais baixa, especialmente quando a planta está com alta taxa de transpiração. Isso pode explicar alguns casos de ocorrência da doença mesmo sob condições de temperatura do ar consideradas acima dos limiares determinados em condições controladas.

Considerando toda a subárvore a partir do nó 3, observam-se seis variáveis relacionadas com a precipitação e quatro com a temperatura, diferentemente da subárvore a partir do nó 2, onde foram utilizadas 13 variáveis de temperatura e três de precipitação. De forma geral, os testes sobre os atributos de precipitação relacionados na subárvore a partir do nó 3 tendem a estabelecer valores que indicam que, para menores quantidades de chuva, tanto em número de dias com precipitação acima de 1 mm (nós 15; 20 e 25), quanto para o maior volume de chuva em um único dia (nós 14 e 21), há mais casos classificados como de ocorrência.

Um dos possíveis efeitos prejudiciais da precipitação sobre a ocorrência da doença foi evidenciado por DFAULT et al. (2010), em que a precipitação de intensidades de 45 ou 85 mm h⁻¹, em seguida (de 1 a 30 minutos) à deposição dos esporos causada pela chuva, removeu de 38 a 91% dos esporos depositados nas folhas. Porém, como os dados de intensidade e duração da chuva não estavam disponíveis nos dados meteorológicos, não foi possível verificar o efeito negativo da precipitação sobre as ocorrências da ferrugem asiática.

CONCLUSÕES

O presente estudo destaca o potencial da técnica de indução de árvores de decisão para classificar casos de ocorrência e não ocorrência da ferrugem da soja e identificar as variáveis meteorológicas associadas a partir de um extenso banco de dados de monitoramento da ferrugem asiática da soja no Brasil. Destaca-se que os resultados demonstram coerência com dados da literatura para um grande número de casos, especialmente relacionados a temperaturas abaixo ou acima dos limiares favoráveis à infecção.

É importante salientar que o estudo utilizou dados e técnicas ainda pouco explorados e possui limitações. Os registros de não ocorrência foram estimados com base nos registros de ocorrências e não foram confirmados por laboratórios. Além disso, os dados meteorológicos podem não representar as situações microclimáticas das lavouras nos períodos estudados.

Mesmo assim, este trabalho representa uma nova contribuição na utilização de um processo de mineração de dados e da técnica de árvore de decisão para explorar fatores meteorológicos que influenciam as epidemias de doenças de plantas.

AGRADECIMENTOS

À CAPES, pelo financiamento da bolsa de estudos de mestrado ao primeiro autor. Ao CNPq, pelo apoio financeiro (Edital CNPq/MAPA/SDA n° 64/2008).

REFERÊNCIAS

- ALVES, S.A.M.; FURTADO, G.Q.; BERGAMIN FILHO, A. Influência das condições climáticas sobre a ferrugem da soja. In: ZAMBOLIM, L. (ed.). *Ferrugem asiática da soja*. Viçosa: Universidade Federal de Viçosa, Departamento de Fitopatologia, 2006. p.37-59.
- BONDE, M.R.; BERNER, D.K.; NESTER, S.E.; FREDERICK, R.D. Effects of temperature on urediniospore germination, germ tube growth, and initiation of infection in soybean by *Phakopsora* isolates. *Phytopathology*, Saint Paul, v. 97, n. 8, p. 997-1003, ago. 2007.
- COLLISCHONN, B.; ALLASIA, D.; COLLISCHONN, W.; TUCCI, C.E.M. Desempenho do satélite TRMM na estimativa de precipitação sobre a bacia do Paraguai superior. *Revista Brasileira de Cartografia*, Rio de Janeiro, v. 59, n. 1, p. 93-99, 2007.
- CONSÓRCIO ANTIFERRUGEM. *Custo ferrugem asiática da soja*. Disponível em: <http://www.consorcioantiferrugem.net/portal/?page_id=1347>. Acesso em: set. 2012.
- CONSÓRCIO ANTIFERRUGEM. *Ferrugem da soja*. Disponível em: <<http://www.consorcioantiferrugem.net/portal/>>. Acesso em: 30 set. 2013.
- DEL PONTE, E.M.; ESKER, P.D. Meteorological factors and Asian soybean rust epidemics: a systems approach and implications for risk assessment. *Scientia Agricola*, Piracicaba, v. 65, n. spe, p. 88-97, dez. 2008.
- DEL PONTE, E.M.; GODOY, C.V.; CANTERI, M.G.; REIS, E.M.; YANG, X.B. Models and applications for risk assessment and prediction of Asian soybean rust epidemics. *Fitopatologia Brasileira*, Brasília, v. 31, n. 6, p. 533-544, 2006a.
- DEL PONTE, E.M.; GODOY, C.V.; LI, X.; YANG, X.B. Predicting severity of asian soybean rust epidemics with empirical rainfall models. *Phytopathology*, Saint Paul, v. 96, n. 7, p. 797-803, 2006b.
- DELGADO, R.C.; SEDIYAMA, G.C.; COSTA, M.H.; SOARES, V.P.; ANDRADE, R.G. Classificação espectral de área plantada com a cultura da cana-de-açúcar por meio da árvore de decisão. *Engenharia Agrícola*, Jaboticabal, v. 32, n. 2, p. 369-380, 2012.
- DUFAULT, N.S.; ISARD, S.A.; MAROIS, J.J.; WRIGHT, D.L. Removal of wet deposited *Phakopsora pachyrhizi* urediniospores from soybean leaves by subsequent rainfall. *Plant Disease*, St. Paul, v. 94, n. 11, p. 1336-1340, 2010.
- GODOY, C.V.; FLAUSINO, A.M.; SANTOS, L.C.M.; DEL PONTE, E.M. Eficiência do controle da ferrugem asiática da soja em função do momento de aplicação sob condições de epidemia em Londrina - PR. *Tropical Plant Pathology*, Brasília, v. 34, n. 1, p. 56-61, 2009.
- HAN, J.; KAMBER, M.; PEI, J. *Data mining: concepts and techniques*. 3rd ed. San Francisco: Morgan Kaufmann Publishers, 2011.
- KOCHMAN, J.K. The effect of temperature on development of soybean rust (*Phakopsora pachyrhizi*). *Australian Journal of Agricultural Research*, Victoria, n. 30, p. 273-277, 1979.
- MARCHETTI, M.A.; MELCHING, J.S.; BROMFIELD, K.R. The effects of temperature and dew period on germination and infection by uredospores of *Phakopsora pachyrhizi*. *Phytopathology*, Saint Paul, n. 66, p. 461-463, 1976.
- MEIRA, C.A.A.; RODRIGUES, L.H.A.; MORAES, S.A. Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. *Tropical Plant Pathology*, Brasília, v. 33, n. 2, p. 114-124, 2008.
- MEIRA, C.A.A.; RODRIGUES, L.H.A.; MORAES, S.A.D. Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. *Pesquisa Agropecuária Brasileira*, Brasília, v. 44, n. 3, p. 233-242, mar. 2009.

- PINTO, L. I. C.; COSTA, M. H.; DE LIMA, F. Z. Comparação de produtos de precipitação para a América do Sul. *Revista Brasileira de Meteorologia*, São José dos Campos, v. 24, n. 4, p. 461-472, 2009.
- QUINLAN, J.R. *C4.5: Programs for machine learning*. San Francisco: Morgan Kaufmann, 1993.
- ROMANI, L.A.S.; OTAVIAN, A.F.; EVANGELISTA, S.R.M.; ASSAD, E.D. Modelo de estações virtuais com estimativa de precipitação e temperatura para aprimoramento dos mapas no Agritempo. In: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 15., 2007, Aracaju. *Anais...*
- TAN, P.N.; STEINBACH, M.; KUMAR, V. *Introdução ao data mining: Mineração de dados*. Rio de Janeiro: Ciência Moderna, 2009. 932p.
- YORINORI, J. T.; PAIVA, W. M.; FREDERICK, R. D.; COSTAMILAN, L. M. et al. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. *Plant Disease*, Saint Paul, v. 89, n. 6, p. 675-677, jun. 2005.
- WITTEN, I.H.; FRANK, E.; HALL, M.A. *Data mining: practical machine learning tools and techniques*. 3rd ed. San Francisco: Morgan Kaufmann, 2011.