

MODELO LINEAR GENERALIZADO MISTO PARA ESTIMAR INTENSIDADE DE DOENÇA EM TOMATEIRO

Carlos Henrique Osório Silva¹, Fabyano Fonseca e Silva¹, Diógenes da Cuz Batista², Antônio Policarpo Souza Carneiro¹, Eduardo Seiti Gomide Mizubuti³

INTRODUÇÃO

Baseado em levantamentos realizados em campos de produção de tomateiro (*Solanum lycopersicum* L. = *Lycopersicon esculentum* Mill.), constatou-se que a pinta preta, causada pelo fungo *Alternaria solani* Sorauer, é a doença prevalente no Brasil (Couto & Tavares, 2002) e acarreta prejuízos consideráveis na produção de frutos. Embora a pinta preta seja de ocorrência freqüente, a intensidade das epidemias varia em função das condições ambientais (Salustiano, 2000). Dentre as variáveis que determinam o desenvolvimento da doença, a temperatura e a umidade, em suas diferentes formas (precipitação, orvalho e irrigação), são as mais importantes.

Para fins de desenvolvimento de um sistema de previsão para a pinta preta adequado às condições brasileiras, é necessário levar em consideração as peculiaridades do patossistema. Dois aspectos merecem ser destacados: disponibilidade de inóculo e ocorrência de condições ambientais favoráveis à epidemia. Devido ao fato de haver ampla disponibilidade de inóculo para epidemias de pinta preta e de se constatar temperaturas favoráveis ao desenvolvimento da doença em todas as regiões produtoras de tomate, conclui-se que o aumento na intensidade da doença depende principalmente dos eventos que propiciam umidade, como a precipitação.

O desenvolvimento de um modelo de análise de risco que auxilie a tomada de decisão e que seja baseado na precipitação constitui uma importante contribuição para o manejo do tomateiro. A modelagem com regressão logística é uma abordagem interessante para este desenvolvimento, pois esta auxilia na identificação de variáveis importantes que interferem em epidemias de diversos patossistemas (Mila et al., 2004) mediante valores de probabilidade associados à ocorrência da doença.

Objetivou-se desenvolver um modelo baseado em regressão logística que relacione nível e horas de precipitação com o aumento da intensidade de pinta preta em tomateiro.

¹ Professor Dep. Informática – Universidade Federal de Viçosa, e-mail:chos@dpi.ufv.br

² Pesquisador EMBRAPA Semi-Árido, Petrolina-PE

³ Professor Dep. Fitopatologia – Universidade Federal de Viçosa

MATERIAL E MÉTODOS

Dois experimentos foram realizados durante 2005-2006. O primeiro foi realizado entre março e junho de 2005 (outono/inverno) e o segundo, entre setembro de 2005 e janeiro de 2006 (primavera/verão). Mudanças de tomateiro, cultivar Santa Clara, foram transplantadas para o campo em estágio de desenvolvimento de 3 a 4 folhas definitivas. Em cada experimento, foram estabelecidos quatro blocos, cada um com cinco parcelas contendo 60 plantas. A severidade da pinta preta foi quantificada a intervalos de três a quatro dias, em 30 plantas localizadas nas duas fileiras centrais da parcela. Em cada planta, estimou-se a porcentagem de área foliar lesionada por *A. solani* com auxílio de escala diagramática (Boff et al., 1991).

As variáveis climáticas foram medidas por sensores de temperatura e umidade relativa, os quais foram programados para registrar os valores de cada variável a cada dois segundos. Os valores médios para cada período de 30 min foram calculados e armazenados em coletor automático de dados instalado na área experimental.

Calcularam-se os incrementos nos valores de severidade entre duas avaliações (anterior e a consecutiva). A variável resposta foi o incremento percentual (I) em área foliar doente entre avaliações. Uma escala do tipo binária foi adotada e classificada conforme o nível de incremento da severidade: $Z=0$ se $I \leq 2,0\%$ e $Z=1$ se $I > 2,0\%$. A variável Z foi indicadora de aumento severo de doença. A probabilidade $P(Z=1)$, isto é, de haver aumento da severidade da pinta preta em mais de 2 pontos percentuais (2pp), foi modelada como uma função das variáveis: volume total de chuva em mm (X_1) e duração total em horas de chuva (X_2), no período de 4 dias anteriores a cada avaliação. O conjunto de dados foi composto por 16200 dados obtidos a partir de 15 (experimento 1) e 12 (experimento 2) avaliações da severidade da pinta preta em 600 plantas, ao longo dos dois experimentos.

Para a modelagem, empregou-se a regressão logística considerando-se a estrutura de medidas repetidas inerentes aos dados pela incorporação de um efeito aleatório no modelo, isto é, um modelo de efeitos mistos (Wolfinger, 1999). Nesta modelagem considerou-se que a distribuição condicional do número de plantas com aumento maior que 2pp é binomial, $Y_{it} / u_i \sim \text{binomial}(n = 30, p_{it})$, em que Y_{it} é o número de plantas com aumento maior que 2pp na t-ésima avaliação da i-ésima parcela e u_i representa o efeito aleatório da i-ésima parcela de plantas. O valor u_i foi estimado com base no conjunto de dados via estimação Bayesiana

empírica (Wolfinger, 1999). Utilizou-se a distribuição normal com média zero e variância s^2 para a distribuição *priori* – $\pi(u_i)$.

Sendo assim, o modelo misto é dado por: $\ln[p_{it}/(1-p_{it})] = E(Y_{it}) = \beta X_{it} + u_i$, o que resulta em $p_{it} = \frac{e^{\beta X_{it} + u_i}}{1 + e^{\beta X_{it} + u_i}}$, em que p_{it} é a probabilidade de haver aumento maior que 2pp ($Z = 1$) no tempo t (ou avaliação) na i -ésima unidade de observação (parcela com 30 plantas). Seja β um vetor linha de parâmetros (coeficientes de regressão) e X_{it} um vetor coluna de covariáveis ou variáveis explicativas (efeitos fixos), então βX_{it} define como incorporar os efeitos das covariáveis. Assim, $\beta X_{it} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \beta_4 X_{1i}^2 + \beta_5 X_{2i}^2$ foi definido como o modelo completo no presente estudo. Foram testados diversos outros modelos dos quais apenas quatro foram promissores.

O modelo logístico estima a probabilidade \hat{p} de ocorrer incremento da doença igual ou superior a 2% ($Z=1$). A utilização prática do modelo requer o estabelecimento de um limiar de decisão, isto é, um valor P_L tal que se $\hat{p} \geq P_L$ então determinada ação deverá ser implementada (ex. aplicação de fungicida). O valor P_L foi definido como aquele que combinou altas sensibilidade e especificidade, duas importantes propriedades do modelo. Sensibilidade é a probabilidade do modelo estimar a ocorrência do evento quando ocorre o evento e a especificidade é a probabilidade de o modelo estimar a não ocorrência do evento quando este evento realmente não ocorre. Estimativas de sensibilidade e especificidade do modelo foram gerados para valores $P = 0,0; 0,02; 0,04$ até 1,00 considerando-se que o modelo estima o evento corretamente quando $\hat{p} > P$ e $p_{OBS} > P$, em que p_{OBS} é a proporção de plantas com incremento de doença igual ou maior que 2% ($Z=1$) observada por parcela. Portanto, os valores de sensibilidade e especificidade foram estimados para o modelo com base em cálculos iterativos de eventos e não eventos utilizando-se o conjunto de dados.

Adotaram-se como critérios para avaliar a qualidade do ajuste do modelo os valores da correlação de Pearson (r) entre a probabilidade observada e predita pelos modelos e o valor BIC, para ambas as probabilidades.

Todas as análises estatísticas foram realizadas com o programa SAS[®] por meio do PROC NL MIXED, cujo método de estimação adotado foi o da quadratura gaussiana.

RESULTADOS E DISCUSSÃO

Verificou-se que o modelo 4 (Tab.1) foi mais adequado de acordo com os valores de r e BIC, portanto, o logit do modelo para se estimar a probabilidade de aumento maior que 2pp da pinta preta foi: $\ln(p/(1-p)) = -0,6208 - 0,1373X_1 + 0,5767X_2 - 0,0413X_1X_2 + 0,0042X_1^2 + 0,0979X_2^2 + u$, sendo que u representa o efeito aleatório da parcela, cuja variância estimada foi de 8,90. Observando-se o gráfico de sensibilidade e especificidade, definiu-se como melhor limiar de decisão o valor $P_L = 0,3$ (Fig. 1), pois os valores de sensibilidade e especificidade decresceram rapidamente quando P_L alcançou valores superiores a 0,5 e inferiores a 0,2 (Fig. 1).

Tabela 1. Coeficientes de correlação de Pearson (r) entre os valores de probabilidade observados e preditos pelo modelo e valores do BIC¹.

Coef. ²	Estimativas (erro-padrão ³)			
	1	2	3	4
b_0	-0,0588 (0,1741)	1,1755**(0,4320)	-0,8769*(0,4277)	-0,6208 (0,4760)
b_1	-0,0224**(0,0011)	-,1565**(0,0031)	-0,1764**(0,0034)	-0,1373**(0,0038)
b_2	0,0991**(0,0068)	0,3831**(0,0104)	0,9771**(0,0216)	0,5767**(0,0267)
b_3	-----	0,0016**(0,00003)	-0,0137**(0,0004)	-0,0413**(0,0018)
b_4	-----	-----	0,0022**(0,00006)	0,0042**(0,0001)
b_5	-----	-----	-----	0,0979**(0,0056)
r	0,4769**	0,7633**	0,8806**	0,8992**
BIC	-7026	-4585	-3104	-2865

^{1/} Bayesian information criteria (referência), quanto maior, melhor o ajuste; ^{2/} Correspondentes ao logit: $\ln(p/1-p) = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2 + b_4X_1^2 + b_5X_2^2 + u$; ^{3/} Número em parênteses corresponde ao erro padrão das estimativas; **/ e */ Significativo a 1% e 5% de probabilidade, respectivamente.

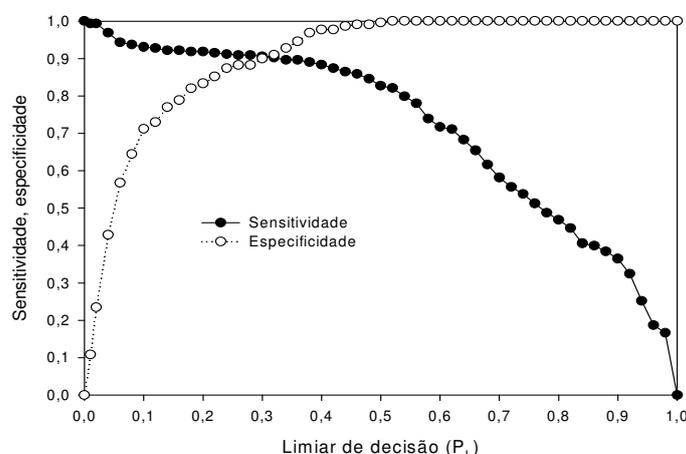


Figura 1. Sensibilidade e especificidade como função do limiar de decisão (se valor P_L é tal que $P \geq P_L$, a medida de controle é recomendada e se $P < P_L$ a medida de controle não é recomendada, onde P é a probabilidade dada pelo modelo 4) apresentado na Tabela 1.

CONCLUSÕES

A utilização de um modelo matemático baseado em probabilidade tem como vantagem a flexibilidade na tomada de decisão. Estimativas pontuais, como os valores numéricos de estimativas de parâmetros, não são facilmente traduzidas em risco de ocorrência de um evento. Melhorias nos sistemas existentes podem ser conseguidas com os modelos fundamentados em regressão logística. A complementação deste estudo consiste em testar o modelo proposto em ensaios de campo e possivelmente incluir novas covariáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

BOFF, P., ZAMBOLIM, L., VALE, F. X. R. Escalas para avaliação de severidade da mancha-de-estenfílio (*Stemphylium solani*) e da pinta preta (*Alternaria solani*) em tomateiro. **Fitopatologia Brasileira** 16: 280-283. 1991.

COUTO, M. E. O.; TAVARES, F. W. Levantamento de problemas fitossanitários no Estado do Rio Grande do Sul realizado pela clínica fitossanitária. **Fitopatologia Brasileira** 27: 181. 2002.

MILA, A. L., CARRIQUIRY, A. L. & YANG, X. B. Logistic regression modeling of prevalence of soybean Sclerotinia stem rot in the north-central region of the United States. *Phytopathology* 94: 102-110. 2004.

SALUSTIANO, M. A. Progresso da pinta preta do tomateiro em cultivares de tomate, em diferentes épocas de plantio. Viçosa, MG. UFV. 2000. 87 p. (Tese de Doutorado) - Universidade Federal de Viçosa, 2000.

Wolfinger, R. D. Fitting Nonlinear Mixed Models with the new NLMIXED Procedure, Proceedings of the 24th annual SAS Users Group International Conference (SUGI 24) Paper 287-24. 1999. Obtido em: <http://ftp.sas.com/techsup/download/v8papers/p287-24.pdf>