

MÉTODOS PROBABILÍSTICOS PARA QUANTIFICAR A INFLUÊNCIA DO EL NINO/OSCILAÇÃO SUL SOBRE O RISCO DE EPIDEMIAS*

ALINE DE HOLANDA N. MAIA¹, EMERSON M. DEL PONTE², THIAGO V. DOS SANTOS³

¹ Pesquisador, Embrapa Meio Ambiente, Jaguariúna-SP, ahmaia@cnpma.embrapa.br

² Professor, Departamento de Fitossanidade, UFRGS, Porto Alegre-RS, emerson.delponte@ufrgs.br

³ Bolsista, Departamento de Fitossanidade, UFRGS, Porto Alegre-RS

RESUMO: Neste trabalho, apresentamos e discutimos as vantagens e limitações de métodos estatísticos paramétricos e semi-paramétricos para a caracterização da influência do fenômeno El Niño/Oscilação Sul (ENOS) sobre o risco de epidemias de doenças de plantas. Foram utilizados dados simulados de um índice de severidade da ferrugem asiática em soja para Rosário do Sul, RS, local previamente identificado como passível da influência de ENOS na precipitação pluvial no período da safra. A influência ENOS sobre a produtividade agrícola e outras variáveis dependentes da precipitação pluviométrica, como a severidade de doenças, é bem conhecida e documentada em todo o mundo. No entanto, o uso rigoroso de abordagens probabilísticas inferenciais para quantificação da influência de indicadores climáticos sobre o de risco de epidemias ainda é raro.

PALAVRAS-CHAVE: ferrugem asiática, risco climático, epidemiologia, modelo de Cox.

INTRODUÇÃO

A influência do fenômeno El Niño/Oscilação Sul (ENOS) sobre a produtividade agrícola e outras variáveis dependentes da precipitação pluviométrica, como a severidade de doenças ou a ocorrência de pragas, é bem conhecida e documentada em todo o mundo (ZHAO and YAO, 1989; WANG et al, 2003; DEL PONTE et al, 2009; DEL PONTE et al, 2010). No entanto, o uso rigoroso de métodos de estatística descritiva e abordagens probabilísticas inferenciais para quantificação de risco e teste de hipóteses sobre a influência de preditores nas variáveis de interesse, ainda é raro (DEL PONTE et al, 2010).

Neste trabalho, apresentamos e discutimos métodos estatísticos paramétricos e semi-paramétricos para caracterização de influência do ENOS sobre o risco de epidemias de doenças de plantas, utilizando como estudo de caso a ferrugem asiática em soja no Rio Grande do Sul. O risco é expresso como a probabilidade do índice de severidade médio (ISM) exceder um determinado valor crítico $\text{Prob}(\text{IS} > y)$, em função das informações sobre o estado do ENOS num período anterior ao período crítico de ocorrência da doença na região.

Três abordagens são apresentadas: comparação de distribuições empíricas, métodos não paramétricos para estimação de densidades (kernel) e modelo de Cox. Essas abordagens utilizam séries históricas de índices de severidade simulados a partir de variáveis meteorológicas e séries de indicadores do ENOS (X) que podem ser contínuos (anomalias de temperaturas da superfície do mar, índice de oscilação sul) ou categorizados (classes derivadas dessas anomalias, como fases 'quente', 'neutra' ou 'fria' do ENOS). As informações são sumarizadas em estimativas da $\text{Prob}(\text{IS} > y)$ em função de valores de X, para qualquer valor de y e suas respectivas incertezas representadas por intervalos de confiança.

O uso de abordagens probabilísticas permite considerar incertezas inerentes às inter-relações ENOS/chuva/epidemias. O conhecimento antecipado do risco de epidemias em função do comportamento do ENOS auxilia a tomadas de decisão sobre o manejo preventivo da doença.

* Trabalho expandido a partir de resumo apresentado como pôster no Workshop de Epidemiologia, Bento Gonçalves, RS.

MATERIAL E MÉTODOS

Dados

Foi utilizada uma série de chuva diária de uma estação localizada em Rosário do Sul, RS, do período de 1979 a 2008. A partir desses dados, foi simulada uma série de índices de severidade (IS) da ferrugem asiática da soja (*Phakopsora pachyrhizi* Syd) para 28 datas hipotéticas de detecção (DHD) a partir de 31/01 (1979-2008). Uma série de anomalias médias da temperatura da superfície do mar no trimestre OND (TSM-OND), na região Niño 3.4 do Pacífico Equatorial, no período de 1978-2007, foi utilizada como indicador do fenômeno de larga escala El Niño/Oscilação Sul (ENOS), relacionado ao aquecimento ou resfriamento do Pacífico Equatorial.

Métodos

Os IS para cada DHD foram simulados utilizando um modelo empírico que estima a severidade máxima da doença, em função do número de dias chuvosos e do total de chuva (DEL PONTE et al, 2006), a partir da detecção hipotética da doença (28 datas hipotéticas a partir de 31/01). Foi calculado um índice de severidade médio (ISM) para cada safra, considerando todas as DHD de fevereiro.

A série dos ISM para o mês de fevereiro foi particionada em três sub-séries de acordo com as anomalias da TSM-OND (período anterior à janela epidêmica), correspondentes às fases 'fria', 'neutra' e 'quente' do ENOS. Os dados relativos à ISM são representados por distribuições de probabilidade acumulada (Prob [ISM ≤ y]); uma distribuição não condicional, que utiliza todos os dados da série, independente da fase do ENOS e uma distribuição condicional para cada fase (Prob [ISM ≤ y] | X=x), onde x é a fase do ENOS. Dependendo do problema de interesse, as distribuições acumuladas são representadas por suas funções complementares, ditas funções probabilidade de excedência (POE; Prob [ISM > y]). Neste estudo, onde o interesse é quantificar risco de epidemias severas, a representação via POE (probabilidade de o ISM exceder limites críticos) facilita a interpretação dos resultados. A influência do ENOS sobre o risco de epidemias de ferrugem foi avaliada usando os seguintes métodos:

a) Comparação de distribuições empíricas (M1)

As POE (probability of exceedance, Prob [ISM>y]) para cada fase foram estimadas pelo método de Kaplan-Meier e comparadas pelo teste não paramétrico Log-Rank. Este teste considera as distâncias entre cada POE condicional e a POE não condicional para todos os valores de ISM no intervalo de dados considerado, ao invés da distância máxima, métrica utilizada para o teste de Kolmogorov-Smirnoff, de uso comum na comparação de distribuições. A comparação de POE empíricas é útil para análise exploratória de um conjunto de preditores candidatos (MAIA et al, 2007). Um exemplo detalhado do uso dessa abordagem para avaliação de risco de epidemias de ferrugem da soja no Rio Grande do Sul é discutido em Del Ponte et al, 2010.

b) Ajuste de distribuições utilizando métodos não paramétricos de suavização (M2)

POE foram ajustadas, para cada fase do ENOS, utilizando métodos kernel para suavização de dados (SILVERMAN, 1986). Dado um conjunto de observações y_1, y_2, \dots, y_n , a expressão matemática do estimador kernel da função densidade é dado pela expressão:

$$\hat{f}_Y(y) = \frac{1}{n \cdot h} \sum_{i=1}^n K\left(\frac{y - y_i}{h}\right)$$

na qual K é a função kernel e h a janela (*bandwidth*) que controla a variância da função kernel. Após a estimação da densidade do ISM para cada fase, as correspondentes distribuições acumuladas (ou POE) são obtidas via métodos de integração. Métodos kernel são de uso comum em análise de risco na área de hidrologia (LALL, 1995, KIM et al, 2003).

c) Modelo de Cox de riscos proporcionais (M3)

Segundo Allison (1995), o modelo [de regressão] de Cox (COX, 1972), refere-se, de modo conjunto, ao próprio modelo, que descreve a relação entre a $\text{Prob}[Y>y]$ e preditores candidatos (covariáveis), e ao método utilizado para estimação dos parâmetros do modelo, da máxima verossimilhança restrita. O modelo de Cox, tradicionalmente utilizado para a modelagem do tempo até a ocorrência de eventos em medicina, sociologia ou engenharia (ex. morte, cura, divórcio, reincidência, falha de componentes), é representado pela seguinte função de sobrevivência:

$$S(t; \mathbf{z}_i) = [S_0(t)] \exp(\mathbf{z}'_i \cdot \boldsymbol{\beta})$$

Onde t é o tempo até a ocorrência do evento de interesse, $S_0(t)$ a função de sobrevivência obtida quando todos os preditores têm valor nulo; $\mathbf{z}'_i = (z_{i1}, z_{i2}, \dots, z_{ik})$, o vetor de preditores do modelo; $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_i, \dots, \beta_k)$, o vetor de parâmetros do modelo correspondentes a cada preditor.

Neste estudo, conforme proposto por Maia e Meinke (2010), utilizamos o modelo de Cox para uma variável influenciada por fatores meteorológicos que não é mensurada na escala de tempo (severidade da ferrugem da soja). Dessa forma, substituímos t pelo ISM, obtendo-se a equação:

$$P(ISM>y|X=x) = [POE_0(y)] \exp(\beta x)$$

na qual $P(ISM>y|X=x)$ é a probabilidade de a ISM exceder um valor crítico y , condicionada à anomalia média do índice Niño 3.4 (X , indicador do ENOS) no trimestre OND do ano imediatamente anterior à janela considerada para as datas hipotéticas de detecção (28 dias após 31/01); POE_0 é a função probabilidade de excedência para um valor nulo de anomalia do ENOS, que corresponde à POE não condicional; e β é o parâmetro a ser estimado no modelo, que quantifica a magnitude da influência do ENOS sobre o índice de severidade médio (ISM). A pressuposição de riscos proporcionais é investigada acrescentando-se ao modelo um termo referente à interação entre o preditor e a variável resposta.

RESULTADOS E DISCUSSÃO

A POE empírica não condicional (Figura 1) corresponde à informação sobre risco de epidemias de ferrugem que não incorpora o conhecimento sobre a fase do ENOS. Quanto maior a divergência entre as POE de cada fase (condicionais), maior a diferença entre os riscos estimados considerando ou não esse indicador de larga escala.

As POE condicionais estimadas pelos três métodos (Figura 2) apresentam padrões similares. Observa-se, no entanto, um afastamento mais pronunciado da POE da fase quente para os métodos M2 e M3, nos quais as POE são estimadas utilizando apenas os dados da respectiva fase do ENOS. Nas estimativas obtidas via modelo de Cox, onde todo o conjunto de dados é utilizado para estimar as POE condicionais, há uma maior homogeneidade entre as divergências em decorrência da pressuposição de riscos proporcionais. Nesse modelo, é possível obter estimativas de risco relativo e razões de risco derivadas do parâmetro β , que são úteis para quantificação e interpretação da magnitude de influência do preditor.

Os testes de hipótese sobre a influência do ENSO, investigados utilizando os métodos M1 e M3, indicaram forte evidência de efeito, com níveis de significância nominal (valor p) de 0,0008 e 0,0017, respectivamente, para o teste Log-rank e o teste da razão de verossimilhanças para β . No método M2, para o qual não há um teste de significância associado, são obtidas apenas as estimativas das distribuições utilizando métodos de suavização de dados.

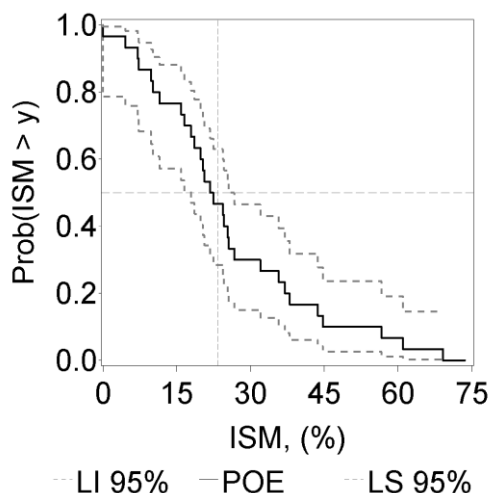


FIGURA 1. POE (Prob [ISM>y]) para o índice de severidade médio (ISM) correspondente às datas hipotéticas de detecção no mês de fevereiro, utilizada para quantificar o risco de epidemias sem considerar a informação do ENOS. As linhas pontilhadas correspondem aos limites inferior (LI) e superior (LS) do intervalo de confiança de 95% para a POE.

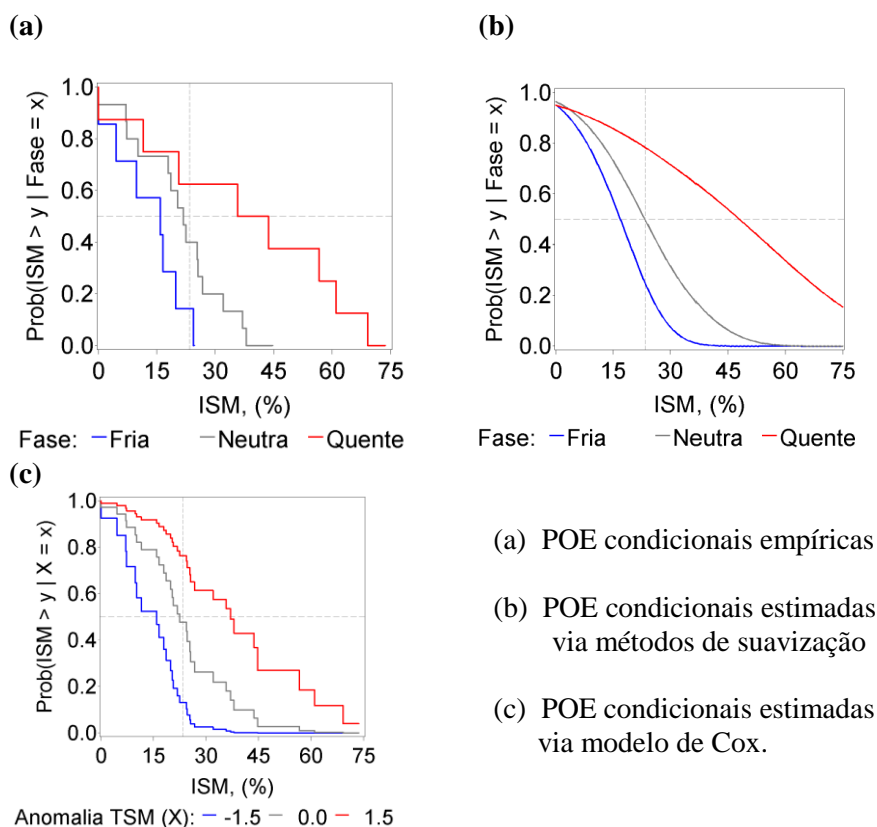


FIGURA 2. POE para o ISM da ferrugem da soja em Rosário do Sul, RS, estimadas utilizando informações sobre fases do ENOS: (a) M1 - distribuições empíricas para cada fase do ENOS (preditor categorizado); (b) M2 - POE condicionais para cada fase do ENOS (preditor categorizado), estimadas via métodos 'kernel'; e (c) M3 - POE condicionais estimadas via modelo de Cox para valores pontuais de anomalias da TSM (preditor contínuo).

A estimativa do parâmetro β do modelo de Cox foi $-0,6730 (\pm 0,2145)$, o que corresponde a uma estimativa da razão de riscos de 0,51. Assim, a cada aumento unitário na TSM-OND, a $P(\text{ISM} > y)$ é multiplicada por 1,66 ($=e^{0,51}$), para qualquer valor de y . Essa tendência é consistente com o favorecimento de chuvas mais intensas durante a janela epidêmica (fevereiro a março) associadas ao

maior aquecimento do pacífico equatorial no trimestre OND (anomalias positivas da TSM_OND) o que implica em maior risco de epidemias da ferrugem da soja no RS. Utilizando as POE condicionais estimadas pelo modelo (Figura 2), obtivemos as seguintes estimativas da $P[ISM > 45\%]$: 0,05% (TSM-OND=-1,5 - fase fria); 3% (TSM-OND=0 - fase neutra); e 37% (TSM-OND=1,5 - fase quente). Este exemplo ilustra o intenso efeito do ENOS sobre o risco de epidemias de ferrugem.

Em M1 e M2 as POE condicionais são estimadas utilizando apenas os dados da sub-série de casa fase do ENOS, uma limitação para séries históricas curtas. Em M3, toda a série é utilizada na estimação, mas existe a pressuposição de riscos proporcionais, que não é válida em algumas situações. Nesses casos há a opção de uso de extensões do modelo de Cox que acomodam riscos não proporcionais. Em M2, as POE são suaves, o que é uma vantagem em relação aos demais métodos cujas estimativas são expressas na forma de funções-escada, mas, por outro lado, testes de hipótese sobre as POE não estão prontamente disponíveis, o que exigiria o uso de métodos computacionalmente intensivos.

Uma comparação criteriosa dos métodos estatísticos disponíveis deve ser feita utilizando dados gerados por simulação, com características conhecidas e que representem diversos cenários para as distribuições do ISM, assim como diferentes magnitudes de influência do ENOS, compatíveis com a realidade da região geográfica objeto de estudo.

CONCLUSÕES

Os métodos de comparação de distribuições empíricas (M1) e ajuste de funções suavizadas (M2) são úteis como ferramentas exploratórias para visualização de divergências entre as POE correspondentes a cada categoria dos preditores candidatos. O modelo M1 ainda é útil na quantificação preliminar da influência desses preditores sobre as variáveis de interesse na análise de risco.

O modelo de Cox, em função da possibilidade de uso de preditores contínuos, é mais adequado quando as séries são relativamente curtas, o que resulta em sub-séries com número limitado de dados para aplicação dos métodos M1 e M2. No entanto, quando houver evidências de riscos não proporcionais, devem ser utilizadas extensões do modelo de Cox que acomodem essa característica do fenômeno a ser modelado.

REFERÊNCIAS

ALLISON, P. D. **Survival analysis using the SAS System: a practical guide**. Cary, NC, SAS Institute Inc., 1995. 292 p.

COX, D.R. Regression Models and Life-Tables (with Discussion), **Journal of the Royal Statistical Society, Series B**, v. 34, p. 187–220, 1972.

DEL PONTE, E. M., GODOY, C. V., LI, X., YANG, X. B. Predicting severity of Asian soybean rust with empirical rainfall models. **Phytopathology**, v. 96, p. 797–803, 2006.

DEL PONTE, E. M.; FERNANDES, J. M. C; PAVAN, W; BAETHGEN, W.E. A model-based assessment of the impacts of climate variability on fusarium head blight seasonal risk in Southern Brazil. **Journal of Phytopathology**, v. 157, p.675–681, 2009.

DEL PONTE, E. M.; MAIA, ALINE de H. N.; SANTOS, T. V.; MARTINS, E. J. ; BAETHGEN, W. E. Early-season warning of soybean rust regional epidemics using El Niño Southern/Oscillation information. **International Journal of Biometeorology**, v. 55, p. 575–583.

KIM, T-W., YOO, C.; VALDÉS, J.B. Nonparametric approach for estimating effects of ENSO on Return periods of droughts. **KSCE Journal of Civil Engineering**, v. 7, n. 5, p. 629-636.

LALL, U. Nonparametric function estimation: Recent hydrologic contributions. **Reviews of Geophysics**, p. 1093-1102, 1995.

MAIA, A. H. N.; H. MEINKE, S. LENNOX; R. C. STONE. Inferential, non-parametric statistics to assess quality of probabilistic forecast systems. **Monthly Weather Review**, v. 135, n. 2, p. 351-362, 2007.

MAIA, A. H. N.; H. MEINKE. Probabilistic methods for seasonal forecasting in a changing climate: Cox-type regression models. **International Journal of Climatology**, v. 30, n. 15, p. 2277-2288.

SILVERMAN, B.W. **Density Estimation for Statistics and Data Analysis**, Chapman and Hall, New York, 1986. 175p.

WANG, E; RYLEY, M; MEINKE, H. Effect of climate variability on event frequency of sorghum ergot in Australia. **Australian Journal of Agricultural Research**, v. 54, n. 5, p. 599–611, 2003.

ZHAO, S.; YAO, C. On the sea temperature prediction models of the prevailing level of wheat scab. **Acta Phytopathologica Sinica** v. 19, p. 229–234. 1989.