Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

Aplicação dos modelos lineares generalizados na análise do número de estômatos em coentro (*Coriandrum sativum L.*): estimação bayesiana utilizando INLA

Everton Batista da Rocha [1,2]
Roseli Aparecida Leandro [1]
Clarice Garcia Borges Demétrio [1]
Simone Silmara Werner Gurgel do Amaral[1]
Paulo Justiniano Ribeiro Jr [3]

1 Introdução

O coentro é uma planta herbácea anual, que cresce de 40 a 50 cm, originária da região do Mediterrâneo, amplamente utilizada na culinária brasileira, e industrialmente, no preparo de perfumes, licores, gim, pães, cervejas, achocolatados e ainda na preparação de fármacos (LORENZI e MATOS, 2002). A produção de coentro está associada com o número de estômatos nas faces de suas folhas, a saber, face inferior (abaxial) e face superior (adaxial).

Os estômatos são estruturas celulares responsáveis pela realização de trocas gasosas na planta, de maneira que a contagem de estômatos é de grande importância, pois tem relação direta com a fotossíntese na planta, assim como na fixação de carbono e, consequentemente, na produtividade. Além disso, sua presença na folha pode ser um indicativo da tolerância às intempéries climáticas (TAIZ e ZEIGER, 2004).

Para a análise do número de estômatos em coentro considerou-se o ajuste dos **M**odelos **G**aussianos **L**atentes (**MGL**), que formam uma rica classe de modelos de regressão de estrutura aditiva. Um caso particular dos MGL, são os **M**odelos **L**ineares **G**eneralizados (**MLG**). O modelo de regressão de Poisson é um caso de MLG largamente empregado para analisar dados de ocorrências de um evento de interesse, por unidade de tempo, comprimento, área ou volume. Dado que uma variável aleatória segue uma distribuição de Poisson de parâmetro λ_i, ou seja $Y_i \sim Poisson(\lambda_i)$, tem-se que $Var(Y_i) = E(Y_i) = \mu_i = \lambda_i$. O MLG associado é dado por, $Y_i \sim Poisson(\lambda_i)$, em que $\log(\lambda_i) = X_i^T \beta$.

Para modelar a incerteza relativa as parâmetros seguindo a abordagem bayesiana, será considerada a aproximação INLA (do inglês, *Integrated Nested Laplace Approximations*), proposta por Rue *et. al* (2009), a qual é uma abordagem computacional completa para a Inferência Bayesiana, baseada basicamente em aproximação de Laplace, matrizes esparsas e campos aleatórios Gaussianos Markovianos.

[1]LCE - ESALQ/USP. e-mail: *e.batista.rocha@usp.br*
[2]Agradecimento a CAPES pelo apoio financeiro.
[3]LEG - UFPR.

Neste trabalho, será exemplificado o ganho computacional utilizando o INLA sobre o MCMC, no ajuste de modelos lineares generalizados, para o ajuste bayesiano. Comentários em relação à abordagem clássica para a mesma modelagem, de estudos anteriores, serão considerados.

2 Material e métodos

Um experimento com plantas de coentro *(Coriandrum sativum L.)* foi conduzido em estufa no Departamento de Produção Vegetal da ESALQ/USP, em um delineamento inteiramente casualizado com esquema de tratamentos fatorial 3×2, sendo três combinações de cálcio e potásio ($K : Ca$): $0,75 : 1,0; 1,25 : 1,0$ e $1,5 : 1,0$ e duas doses de silício (Si): 0 e 56 mg/L, aplicados utilizando solução nutritiva, com quatro repetições em cada nível, em cada face da folha.

Duas variáveis respostas foram medidas, a saber: o número de estômatos, por mm^2, na face abaxial e o número de estômatos na face adaxial da folha. Assume-se para esses dados a distribuição de Poisson, com função de ligação logarítmica $g(\mu) = log(\mu)$ e com preditor linear

$$\eta = log(\mu) = \alpha + \gamma_i + \beta_j + (\gamma\beta)_{ij} \tag{1}$$

em que, α é a média geral, γ_i o efeito associado à i-ésima relação $K : Ca$, $i = 1, ..., 3$, β_j o efeito associado à j-ésima dose de Si, $j = 1, 2$, e $(\gamma\beta)_{ij}$ o efeito associado à interação da i-ésima combinação $K : Ca$ com a j-ésima dose de Si.

Para o ajuste dos dados serão considerados os MGL que formam uma rica classe de modelos constituídos por modelos de regressão de estrutura aditiva em que dada uma variável resposta y_i, a média μ_i está ligada a uma estrutura linear aditiva no preditor linear η_i por meio de uma função de ligação $g(\cdot)$, tal que $g(\mu_i) = \eta_i$. Rue *et al.* (2009) mostram que a estrutura aditiva no preditor linear η_i leva em conta efeito de várias covariáveis na forma

$$\eta_i = \alpha + \sum_{j=1}^{n_h} h^{(j)}(u_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \varepsilon_i \tag{2}$$

em que $\{h^{(j)}(\cdot)\}$ são funções não conhecidas de covariáveis u, n_h é o número de funções $h^{(j)}$ utilizadas na modelagem, $\{\beta_k\}$ representam o efeito linear de covariáveis z, n_β a dimensão do vetor $\{\beta_k\}$ e os ε_i são termos não estruturados. Esta classe de modelos é muito flexível devido as diferentes formas que as funções desconhecidas $\{h^{(j)}(\cdot)\}$ podem assumir. Esta classe de modelos tem ampla aplicação, todavia, neste trabalho será considerada apenas aplicação em MLG.

Segundo Dey *et al.* (2000), o preditor linear dos MLG pode ser escrito por,

$$\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{ki}$$

que é um caso particular da expressão (2).

De acordo com Rue *et. al* (2009), Rasmussen e Williams (2006), modelos que assumem a forma aditiva em (2), considerando distribuições a priori gaussianas para α, $\{h^{(j)}(\cdot)\}$, $\{\beta_k\}$ e $\{\varepsilon_i\}$, são denominados MGL. Neste trabalho, no caso do MLG, que é um caso particular do MGL, serão atribuídas prioris gaussianas para α e $\{\beta_k\}$. Note que sob o preditor linear em (1), $\beta_k = (\gamma_i, \beta_j, (\gamma\beta)_{ij})$.

Utilizando a proposta de Rue et. al (2009) e denotando $\pi(\cdot|\cdot)$, como sendo uma distribuição condicional, e considerando-se $x = \{\alpha, \beta_k\}$ como sendo o vetor de parâmetros de efeitos fixos a serem estimados. A distribuição $x|\theta$ é assumida ser normal com vetor de médias zero e matriz de precisão $Q(\theta)$, com hiperparâmetros θ conhecidos, uma vez que o modelo é apenas de efeitos fixos.

Seja y o vetor de observações, $y_i = \{y_i : i \in I\}$, sua função de probabilidade é denotada por $\pi(y|x,\theta)$ e assumindo-se que $\{y_i : i \in I\}$ são condicionalmente independentes dados x e θ. A distribuição conjunta a posteriori para os parâmetros do MLG, pode ser expressa por,

$$\begin{aligned} \pi(x|y,\theta) &\propto \pi(x|\theta)\prod_{i\in I}\pi(y_i|x_i,\theta) \\ &\propto \exp\left[-\frac{1}{2}x^T Q(\theta)x + \sum_{i\in I}\ln\{\pi(y_i|x_i,\theta)\}\right]. \end{aligned} \tag{3}$$

O ajuste será baseado no modelo bayesiano hierárquico de dois estágios, como descrito por Martins et al. (2013),

$$\begin{aligned} y \mid x,\theta &\sim \pi(y|x,\theta) = \prod_{i\in I}\pi(y_i|x_i,\theta) \qquad (4) \\ x \mid \theta, y &\sim \pi(x|\theta,y) \end{aligned}$$

em que supõe-se a propriedade de independência condicional, dado que x é considerado como um campo aleatório gaussiano markoviano. A aproximação de $\pi(x_i|y,\theta)$, é feita por integração numérica aninhada, como proposto por Rue et al. (2009).

Para medir a divergência entre a aproximação de Laplace (ou simplificada) e Gaussiana, entre as densidades marginais a posteriori, foi consdierado o Critério da Divergência Simétrica de Kullback-Leibler (KLD), como sugerido por Rue *et al.* (2009) e para a comparação de modelos foi ralizada considerando-se o Critério de Informação da Deviance Bayesiana (DIC).

Os resultados obtidos neste trabalho serão comparados com o trabalho de Rocha *et al.* (2010), no qual foi utilizada a abordagem clássica, e por Rocha *et al.* (2011), em que foi considerada a abordagem bayesiana, utilizando-se métodos de Monte Carlo com Cadeias de Markov. O tempo computacional do presente estudo será comparado com o estudo de Rocha et al. (2011).

As análises deste trabalho, considerando a abordagem INLA, foram realizadas no programa livre R.

3 Resultados e discussões

Considerando-se, primeiramente, os dados referentes ao número de estômatos na face abaxial das folhas de coentro, foram ajustados cinco modelos, assumindo-se o modelo hierárquico bayesiano expresso em (4), com diferentes preditores lineares (Tabela 1). Para cada modelo observou-se $KLD = 0$, o que indica aproximação exata. Em seguida, foram observados os valores de DIC segundo os dois métodos de ajuste: INLA (DIC_1) e MCMC (DIC_2) os quais estão apresentados na Tabela 1.

Tabela 1: DIC para a análise dos números de estômatos na face abaxial da folha de coentro.

Modelo	Preditor Linear	DIC_1	DIC_2
1	$\eta = \alpha$	**191,19**	**192,22**
2	$\eta = \alpha + \gamma_i$	**191,67**	**191,66**
3	$\eta = \alpha + \beta_j$	193,74	193,75
4	$\eta = \alpha + \gamma_i + \beta_j$	193,22	193,22
5	$\eta = \alpha + \gamma_i + \beta_j + (\gamma \times \beta)_{ij}$	193,39	193,53

Observa-se que o menor valor de DIC foi obtido considerando o modelo com apenas o efeito individual da relação de $K : Ca$ (Modelo 2). Ajustando-se o modelo 2 aos dados, verificou-se que o valor zero pertence ao intervalo de credibilidade 95% do parâmetro associado ao efeito da relação K:Ca; sendo assim, conclui-se que o efeito da relação de cálcio e potássio não é significativo no modelo. Assim, com base no valor de DIC, o modelo selecionado é o modelo minimal ($\eta = \alpha$), Modelo 1, em que, o respectivo intervalo de credibilidade 95% para o parâmetro γ é dado por $IC(\alpha, 95\%) = (4,69; 4,77)$, considerando a abordagem INLA. Os mesmos resultados foram obtidos por Rocha *et al.* (2011), assim como o mesmo modelo foi adotado na abordagem clássica por Rocha *et al.* (2010).

Para os resultados referentes à face adaxial da folha de coentro foram ajustados cinco modelos, e após a qualidade da aproximação ser verificada ($KLD = 0$ para todos os modelos), observou-se que o menor valor de DIC ocorre para o modelo com interação (Modelo 10), para o qual todos os parâmetros são significativos e portanto este será o modelo adotado (Tabela 2). O mesmo modelo foi selecionado por Rocha *et al.* (2010), Rocha *et al.* (2011).

Tabela 2: DIC para a análise dos números de estômatos na face adaxial da folha de coentro.

Modelo	Preditor Linear	DIC_1	DIC_2
6	$\eta = \alpha$	229,79	229,99
7	$\eta = \alpha + \gamma_i$	215,97	216,06
8	$\eta = \alpha + \beta_j$	227,24	227,38
9	$\eta = \alpha + \gamma_i + \beta_j$	213,42	213,52
10	$\eta = \alpha + \gamma_i + \beta_j + (\gamma \times \beta)_{ij}$	**194,80**	**194,96**

Considerando as médias dos tratamentos, assim como apontado por Rocha *et al.*(2010), observa-se que a aplicação da dose de $56mg/L$ de Silício, combinada com a relação $0,75 : 1$ de $K : Ca$, apresentou o maior número médio de estômatos na face adaxial da folha (172), o que indica que as plantas que receberam tal tratamento, provavelmente, terão uma taxa fotossintética maior.

4 Conclusões

Os resultados encontrados neste trabalho levam às mesmas conclusões obtidas por Rocha *et al.* (2010) e Rocha *et al.* (2011). A aplicação de Silício e combinações $K : Ca$ não teve efeito sobre o número de estômatos na face abaxial das folhas de coentro, entretanto para o número de estômatos na face adaxial, o modelo com a interação destes fatores deve ser considerado.

Em termos de tempo computacional, para o ajuste dos dados da face adaxial, utilizando a função *system.time()* do R, verifica-se que o uso do INLA reduz o tempo computacional em 94,71%, em comparação com o tempo de processamento do método MCMC. Para os dados da face abaxial, essa redução é de 82,22%.

Referências

[1] DEY, D. K.; GHOSH, S. K.; MALLICK, B. K. **Generalized linear models - a bayesian perpective**. New York: Chapman & Hall/CRC. 2000. 423 p.

[2] LORENZI, H. & MATOS, F.J.A. **Plantas medicinais do Brasil: nativas e exóticas**. Instituto Plantarum, Nova Odessa, 2002. 512p.

[3] MARTINS, T.G.; SIMPSON, D.; LINDGREN, F.; RUE, H. Bayesian computing with INLA: new features.**Computational Statistics & Data Analysis**. Elsevier, v.67, p.68-93, 2013.

[4] RASMUSSEN, C. E.; WILLIAMS, C. K. I. **Gaussian Processes for Machine Learning**. Cambridge: MIT Press, 2006, 248p.

[5] ROCHA, E. B.; AMARAL, S. S. W. G.; DEMETRIO, C. G. B.; DONEGA, M. A.; MELLO, S. C. **Aplicação dos modelos lineares generalizados na análise do número de estômatos em coentro (Coriandrum sativum L.)**. In: Anais do 19º SINAPE, 2010, São Pedro-SP.

[6] ROCHA, E. B.; AMARAL, S. S. W. G.; LEANDRO, R. A.; DEMETRIO, C. G. B.; DONEGA, M. A.; MELLO, S. C. **Aplicação dos modelos lineares generalizados na análise do número de estômatos em coentro (Coriandrum sativum L.): uma abordagem bayesiana**. In: Anais da 56ª RBras e 14º SEAGRO, 2011, Maringá-PR.

[7] TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. Porto Alegre: Artmed, 2004. 719p.