

# ESTIMAÇÃO DE UM MODELO DE ESPÉCIES DE MACROINVERTEBRADOS BENTÔNICOS VIA ANÁLISE BAYESIANA DO MODELO DE MICHAELIS-MENTEN

Eustáquio José MACHADO<sup>1</sup>  
Joel Augusto MUNIZ<sup>2</sup>  
Thelma SÁFADI<sup>2</sup>  
Taciana Villela SAVIAN<sup>3</sup>

- RESUMO: Neste estudo, estimou-se o número de espécies biológicas denominadas macroinvertebrados bentônicos, por meio do modelo de regressão não-linear conhecido no contexto ecológico como equação de Michaelis-Menten, a partir de amostras coletadas em igarapés (riachos) de águas claras em estudo de Lima (2008). A metodologia utilizada fez uso da inferência bayesiana para a estimação dos parâmetros do modelo. Os resultados encontrados indicaram o aumento considerável da precisão das estimativas dos parâmetros fazendo uso da abordagem Bayesiana com a utilização de *prioris* informativas.
- PALAVRAS-CHAVE: Modelo de regressão não-linear; estimação em espécies; macroinvertebrados bentônicos.

## 1 Introdução

Da Amazônia ao cerrado, da mata atlântica à caatinga, do pantanal aos campos sulinos, o Brasil exibe uma das mais ricas diversidades do mundo em ambientes naturais e espécies de plantas e animais do planeta. No entanto, a informação de quantos (riqueza) e quais organismos (composição) pode ser encontrada em determinados ambientes. Essa informação gera desdobramento que podem fornecer subsídios para a escolha de áreas prioritárias para a conservação e permitem comparações entre comunidades de áreas distintas com vistas a vislumbrar sobre eventos anteriores de incidência sobre as espécies bem como a adaptações evolutivas.

Inevitavelmente, ocorrem na natureza processos de extinção em longo prazo resultante de interações intra e interespecíficas aliadas a eventos naturais com implicações diretas e indiretas em comunidades de espécies dos diferentes ambientes existentes na terra. No entanto, este processo de extinção natural é por demais alterados em virtude da interferência humana a qual promove a quebra deste ciclo evolutivo antes do período

---

<sup>1</sup> Universidade Federal do Acre – UFAC, Rio Branco, AC, Brasil. E-mail: [chefinhoo@globocom](mailto:chefinhoo@globocom)

<sup>2</sup> Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas – DEX, Lavras, MG, Brasil. E-mail: [joamuniz@dex.ufla.br](mailto:joamuniz@dex.ufla.br) / [safadi@dex.ufla.br](mailto:safadi@dex.ufla.br)

<sup>3</sup> Universidade de São Paulo – USP, Escola Superior de Agricultura – ESALQ, Departamento de Ciências Exatas – DCE, Piracicaba, SP, Brasil. E-mail: [tsavian@esalq.usp.br](mailto:tsavian@esalq.usp.br)

proposto pelos eventos naturais, de maneira que muitas espécies desaparecem dos ecossistemas sem sequer serem descritas pela ciência.

A falta de informação sobre o levantamento do número de espécies e conseqüentemente a diversidade biológica implica em estratégias equivocadas, e com isso há a perda de espécies nas condições já abordadas anteriormente. Não é difícil aceitarmos que o aumento de registros de espécies novas em qualquer campanha amostral é resultante da ampliação do esforço de campo. Não obstante, ecólogos com atividades voltadas a pesquisas de campo têm reconhecido as limitações em detectar as espécies existentes em um ambiente utilizando um número limitado de amostras.

A biota de determinado ecossistema é constituída de grupos que apresentam comportamentos diferentes dentro de suas populações. Estas características intensificam as dificuldades na retenção de medidas apropriadas para registros de espécies. Como exemplo, citamos a mobilidade de alguns vertebrados. Enquanto as capivaras (*Hydrochaeris hydrochaeris*) apresentam-se em grupos tanto isolados quanto em pequenos grupos, outros animais como os falcões (família Falconidae) apresentam vôos elevados e técnicas de contagem diferentes. Estas peculiaridades oneram as atividades de inventário faunístico e muitas vezes resultam em trabalhos poucos expressivos.

Considerando essa problemática para invertebrados, as dificuldades de estimação ganham proporções maiores, pois cerca de 90% da fauna conhecida em todo mundo é classificada dentro deste grupo. Neste grupo inserem-se os macroinvertebrados aquáticos (animais de água doce passível de percepção a olho nu), em especial a comunidade bentônica (MIB), caracterizada pelo seu importante papel nas relações ecológicas em fluxo de energia e comprovada utilidade para monitoramento dos diferentes fontes de impactos em sistemas aquáticos.

O presente estudo teve como objetivo estimar o número de espécies em uma população de macroinvertebrados bentônicos a partir de amostras coletadas em igarapés (riachos) de águas claras no Parque Nacional da Serra do Divisor (PNSD), uma Unidade de Conservação (UC), localizada no estado do Acre. Para estimar o número de espécies utilizou-se a curva de acumulação em espécies, ou curva do coletor, a qual foi descrita por meio do ajuste do modelo de regressão não-linear de Michaelis-Menten. Tal ajuste foi realizado via inferência Bayesiana utilizando distribuições a priori informativas e não-informativas.

A curva utilizada nesse estudo foi a equação de Michaelis-Menten um modelo ajustado por técnica de regressão não-linear sob um enfoque Bayesiano.

## **2 Material e Método**

### **2.1 Material**

Os dados utilizados neste estudo foram obtidos no Parque Nacional da Serra do Divisor (PNSD), no município de Mâncio Lima – Acre – Brasil. As coletas foram realizadas em dez igarapés (IG) de primeira e segunda ordem em águas claras da bacia do rio Moa e azul, Lima (2008). Da amostra obtida foram detectados 111 táxons distintos independentes de sua abundância, e baseando-se em algumas premissas com fundamentação relacionada à ecologia de riachos, tornou-se razoável admitir que a

amostra obtida fosse assumida como uma população teórica de referência bastante próxima da população real (Tabela 1).

Tabela 1 - Representação da estrutura dos dados coletados em dez igarapés (IG) de águas claras

Táxons	IG1	IG2	IG3	IG4	IG5	IG6	IG7	IG8	IG9	IG10
<i>Curculionidae gênero ind.</i>	0	0	0	0	0	1	0	0	0	0
<i>Dytiscidae gênero ind.</i>	0	0	0	0	0	0	1	0	0	0
<i>Neollmis</i>	0	2	6	0	1	11	31	48	33	13
...	.	.	.	.	.	.	.	.	.	.
...	.	.	.	.	.	.	.	.	.	.
...	.	.	.	.	.	.	.	.	.	.
<i>Polypsectropus</i>	0	0	0	0	0	0	0	3	0	0
<i>Cyrnellus</i>	0	0	0	1	2	0	0	0	0	0
<i>Seracostomatidae Gênero 1</i>	1	0	0	0	0	0	0	0	0	0
<i>Xiphocentron</i>	6	0	0	0	0	3	0	2	3	0

Fonte: Lima (2008).

Foram realizadas três amostragens a partir da população teórica de referência e a contagem utilizada foi o número de táxons distintos, ou seja, presença ou ausência deles no igarapé, implicando em três amostras; ilustrada para amostra 1 conforme Tabela 2.

Tabela 2 - Representação da estrutura da amostra 1 coletada em seis igarapés (IG) de águas claras

Táxons	IG10	IG8	IG1	IG3	IG5	IG2
<i>Neollmis</i>	13	48	0	6	1	2
<i>Heterelmis</i>	0	0	0	0	0	2
<i>Prionocyphon</i>	0	0	0	0	0	1
<i>Haliplidae gênero ind.</i>	0	11	0	0	0	0
...	.	.	.	.	.	.
...	.	.	.	.	.	.
...	.	.	.	.	.	.
<i>Polypsectropus</i>	0	3	0	0	0	0
<i>Cyrnellus</i>	0	0	0	0	2	0
<i>Seracostomatidae Gênero 1</i>	0	0	1	0	0	0
<i>Xiphocentron</i>	0	2	6	0	0	0

Fonte: Lima (2008).

A partir das amostras 1, 2 e 3 foram geradas três massas de dados denominadas *dados1*, *dados2* e *dados3* visualizadas na Tabela 3, 4 e 5.

Tabela 3 - Táxons acumulados por unidades amostrais (IG) relacionados aos *dados1*, *2* e *3*

Unidades Amostrais	Táxons acumulados por unidades amostrais		
	<i>dados 1</i>	<i>dados 2</i>	<i>dados 3</i>
1	38	37	43
2	52	59	61
3	64	68	70
4	77	79	80
5	85	90	85
6	93	96	94

### 2.2.1 Modelo não-linear de Michaelis-Menten

O modelo de regressão não-linear de Michaelis-Menten (1913), o qual foi inicialmente proposto para descrever o comportamento cinético de reações químicas (Machado, 2006), foi utilizado para descrever a relação entre a quantidade acumulada de espécies ( $y_i$ ) e o número de amostras ( $x_i$ ) em cada um dos três conjuntos de dados (*dados1*, *dados2* e *dados3*) considerado. O modelo em questão é dado por:

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \varepsilon_i, \quad (1)$$

em que  $y_i$  é a quantidade acumulada de espécies;  $x_i$  é o número de amostras;  $\theta_1$  é a quantidade máxima de espécies;  $\theta_2$  é o número de amostras na qual a quantidade acumulada de espécies é metade da quantidade máxima de espécies  $i = 1, \dots, n$ ;  $\mathbf{y} = \{y_1, \dots, y_n\}$ ;  $\varepsilon_i$  é o resíduo na  $i$ -ésima amostra considerado independente e igualmente distribuído, com distribuição  $N(0, \sigma^2)$ .

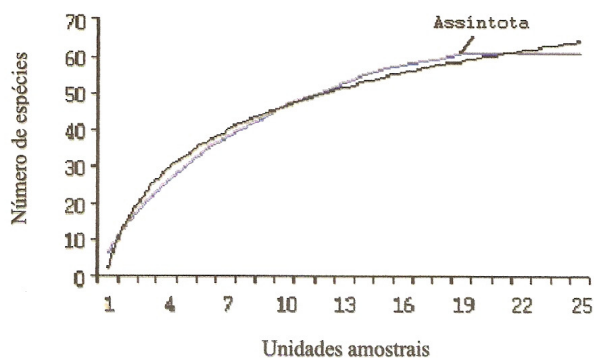


Figura 1 - Curva do coletor.

De acordo com Ehlers (2007), considere uma quantidade de interesse desconhecida  $teta$  (tipicamente não observável). A informação de que dispomos sobre  $teta$ , resumida probabilisticamente através de  $p(teta)$ , pode ser aumentada observando-se uma quantidade aleatória  $X$  relacionada com  $teta$ . A distribuição amostral  $p(x|teta)$  define esta relação. A ideia de que após observar  $X = x$  a quantidade de informação sobre  $teta$  aumenta bastante é intuitiva e o teorema de Bayes é a regra de atualização utilizada para quantificar este aumento de informação, como mostrado a seguir:

$$p(teta | y) = \frac{p(y, teta)}{p(y)} = \frac{p(y | teta)p(teta)}{p(y)}.$$

Note que  $1/p(y)$  que não depende de  $teta$ , funciona como uma constante normalizadora de  $p(teta | x)$ .

Para um valor fixo de  $x$ , a função  $l(teta; y) = p(x|teta)$ , fornece a *plausibilidade* ou *verossimilhança* de cada um dos possíveis valores de  $teta$ , enquanto que  $p(teta)$  é chamada distribuição *a priori* de  $teta$ . Estas duas fontes de informação, priori e verossimilhança, são combinadas levando a distribuição *a posteriori* de  $teta$ ,  $p(teta | x)$ . Assim, a forma usual do teorema de Bayes é:

$$p(teta | y) \propto l(teta, y)p(teta),$$

sendo que  $P(teta|y)$  é *proporcional a*  $l(teta,y)p(teta)$ . Dessa forma no presente estudo será estimado o vetor  $teta=(teta1, teta2$  e  $var\_erro)$ , ou seja, o vetor de parâmetros desconhecidos a serem estimados.

Segundo Bernardo (1989), o método de Jeffreys é uma alternativa à não invariância da priori de Bayes-Laplace. Baseia-se na informação de Fisher sobre  $\theta \in \mathfrak{R}$ . A priori de Jeffreys para o caso uniparamétrico é dada por:

$$\pi(\theta) \propto I(\theta)^{1/2}.$$

No caso multiparamétrico a priori é dada por:

$$\pi(\theta) \propto [\det I(\theta)]^{1/2}.$$

Apesar de utilizar a informação esperada de Fisher, não exige que  $\theta$  pertença a um estado limitado, de acordo com Bernardo (1989) a priori de Jeffreys, possui as seguintes características:

- é invariante, o que é uma condição necessária, mas não suficiente para determinar uma referência;
- a existência da distribuição a priori de Jeffreys requer condições de regularidade fortes. Dessa forma essa priori foi utilizada para calcular a variância do erro.

Segundo Machado (2006), a obtenção das distribuições condicionais completas são importantes para a utilização dos algoritmos MCMC, como será visto na próxima seção percebe-se que para sigma2 o algoritmo utilizado foi o gibbs sampler já para teta1 e teta2, usou-se Metropolis Hastings, as distribuições condicionais completas serão vistas na a partir da próxima seção.

**2.2.2. Obtenção da distribuição *a posteriori* conjunta para os parâmetros do modelo relacionado a *dados1***

Seja a função densidade de probabilidades para o *i*-ésimo dado observacional dada por:

$$f(y_i | \theta_1, \theta_2, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right], \quad (2)$$

a função de verossimilhança para os *dados1*, considerando-se uma amostra aleatória de tamanho *n* é:

$$L_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right] \right\}$$

$$L_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]. \quad (3)$$

Uma densidade a priori para  $\theta = (\theta_1, \theta_2)$  e  $\sigma^2$ , assumindo  $\theta$  e  $\sigma^2$  independentes é dada por:

$$p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) = p_{\text{dados1}}(\theta_1, \theta_2) p_{\text{dados1}}(\sigma^2).$$

Assumiu-se que o pesquisador não tinha informação prévia a respeito dos parâmetros, e para a obtenção de uma priori não informativa para  $\theta_1$  e  $\theta_2$  utilizou a regra multiparamétrica de Jeffreys, dada por:

$$p_{\text{dados1}}(\theta_1, \theta_2) \propto \left| \theta_1^2 q(\theta_2) \right|^{\frac{1}{2}} \Rightarrow$$

$$p_{\text{dados1}}(\theta_1, \theta_2) \propto \theta_1^2 \frac{1}{(\sigma^2)^2} \left\{ \underbrace{\left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2}_{q(\theta_2)} \right\}^{\frac{1}{2}},$$

em que:

$$q(\theta_2) = \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}.$$

Para a variância residual considerou-se a distribuição *a priori* não informativa para  $\sigma^2$  dada por  $p(\sigma^2) \propto \sigma^{-1}$ , obtendo-se então a distribuição *a priori* conjunta para os parâmetros  $0 < \theta_1, \theta_2 < \infty$  e  $\sigma^2 > 0$ , dada por:

$$p_{\text{dados1}}(\theta_1, \theta_2) \propto |\theta_1^2 q(\theta_2)|^{\frac{1}{2}} \Rightarrow$$

$$p_{\text{dados1}}(\theta_1, \theta_2) \propto \theta_1^2 \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{\frac{1}{2}}$$

$q(\theta_2)$

$$p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) \propto \sigma^{-1} |\theta_1^2 q(\theta_2)|^{\frac{1}{2}} \Rightarrow p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) \propto \sigma^{-1} \theta_1 |q(\theta_2)|^{\frac{1}{2}}. \quad (4)$$

A distribuição *a posteriori* conjunta para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo, considerando a distribuição *a priori* (4) e a função de verossimilhança (3) é dada por:

$$p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2) L_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y})$$

$$\propto \left\{ \sigma^{-1} \theta_1 |q(\theta_2)|^{1/2} \right\} \left\{ (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right] \right\}$$

$$\propto \sigma^{-1} \theta_1 |q(\theta_2)|^{1/2} (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]$$

$$\propto (\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-\frac{n}{2}} \theta_1 \left| \frac{1}{(\sigma^2)^2} \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\} \right|^{1/2} \times$$

$$\exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]$$

$$\propto (\sigma^2)^{-\frac{1}{2}} (\sigma^2)^{-\frac{n}{2}} (\sigma^2)^{-1} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times$$

$$\exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right].$$

Efetuada o produto e substituindo

$$p_{\text{dados1}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times$$

$$\exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]. \quad (5)$$

A partir de (5), ou seja, da distribuição *a posteriori* conjunta dos parâmetros obter-se-á as distribuições condicionais completas para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo.

Para  $\theta_1$  tem-se:

$$p_{\text{dados1}}(\theta_1 | \mathbf{y}, \theta_2, \sigma^2) \propto \theta_1 \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right], \quad (6)$$

para  $\theta_2$  tem-se:



$$p_{\text{dados1}}(\theta_2 | y, \sigma^2, \theta_1) \propto \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right], \quad (7)$$

para  $\sigma^2$  tem-se:

$$p_{\text{dados1}}(\sigma^2 | y, \theta_1, \theta_2) \propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right], \quad (8)$$

que é o núcleo de uma gama inversa cuja a função de densidade é da forma:

$$p(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}}, \quad x > 0 \quad X \sim GI(\alpha, \beta),$$

de parâmetros dados por:

$$\alpha = \frac{(n+1)}{2} \quad e \quad \beta = \frac{1}{2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2.$$

### 2.2.3 Obtenção da distribuição *a posteriori* conjunta para os parâmetros do modelo relacionado a *dados2*

A distribuição *a priori* conjunta a ser utilizada será a distribuição *a posteriori* conjunta (5) que foi obtida em *dados1* dada por:

$$p_{\text{dados2}}(\theta_1, \theta_2, \sigma^2) \propto (\sigma^2)^{-\left[\frac{(n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right], \quad (9)$$

e a função de verossimilhança utilizada eq. (3), considerando-se que o modelo e o número de unidades amostrais são os mesmos é:

$$L_{\text{dados2}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right], \quad (10)$$

obtendo-se a distribuição *a posteriori* por meio da combinação de (9) e (10) dada por:

$$p_{\text{dados2}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left[\frac{(2n+1)}{2}+1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left(\frac{x_i}{\theta_2 + x_i}\right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \times \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]. \quad (11)$$

A partir de (11), ou seja, da distribuição *a posteriori* conjunta dos parâmetros obter-se-á as distribuições condicionais completas, para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo ajustado a *dados2*.

Para  $\theta_1$  tem-se:

$$p_{\text{dados2}}(\theta_1 | \mathbf{y}, \theta_2, \sigma^2) \propto \theta_1 \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right], \quad (12)$$

para  $\theta_2$  tem-se:

$$p_{\text{dados2}}(\theta_2 | \mathbf{y}, \theta_1, \sigma^2 | \mathbf{y}) \propto \left\{ \left[ \sum_{i=1}^n \left(\frac{x_i}{\theta_2 + x_i}\right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \\ \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right]. \quad (13)$$

Para  $\sigma^2$  tem-se:

$$p_{\text{dados2}}(\sigma^2 | \mathbf{y}, \theta_1, \theta_2) \propto (\sigma^2)^{-\left[\frac{(2n+1)}{2}+1\right]} \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^n \left(y_i - \frac{\theta_1 x_i}{\theta_2 + x_i}\right)^2\right], \quad (14)$$

que é o núcleo de uma gama inversa:

$$p(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}}, \quad x > 0 \quad X \sim GI(\alpha, \beta),$$

de parâmetros dados por:

$$\alpha = \frac{(2n+1)}{2} \quad \text{e} \quad \beta = \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2.$$

### 2.2.4 Obtenção da distribuição *a posteriori* conjunta para os parâmetros do modelo relacionado a *dados3*

A distribuição *a priori* conjunta a ser utilizada será a distribuição *a posteriori* conjunta (11) que foi obtida em *dados2* dada por:

$$p_{\text{dados3}}(\theta_1, \theta_2, \sigma^2) \propto (\sigma^2)^{-\left[\frac{(2n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right], \quad (15)$$

e a função de verossimilhança que será utilizada é idêntica a utilizada em *dados2* (10) considerando-se o mesmo modelo, tamanho amostral e a conjunto de *dados3*,

$$L_{\text{dados3}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]. \quad (16)$$

Obtendo-se a distribuição *a posteriori* por meio da combinação de (15) e (16) dada por:

$$p_{\text{dados3}}(\theta_1, \theta_2, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-\left[\frac{(3n+1)}{2} + 1\right]} \theta_1 \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]. \quad (17)$$

A partir de (17), ou seja, da distribuição *a posteriori* conjunta dos parâmetros obter-se-á as distribuições condicionais completas para os parâmetros  $\theta_1$ ,  $\theta_2$  e  $\sigma^2$  do modelo.

Para  $\theta_1$  tem-se:

$$p_{\text{dados3}}(\theta_1 | y, \theta_2, \sigma^2) \propto \theta_1 \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]. \quad (18)$$

Para  $\theta_2$  tem-se:

$$p_{\text{dados3}}(\theta_2 | y, \theta_1, \sigma^2 | y) \propto \left\{ \left[ \sum_{i=1}^n \left( \frac{x_i}{\theta_2 + x_i} \right)^2 \right] \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^4} \right] - \left[ \sum_{i=1}^n \frac{x_i^2}{(\theta_2 + x_i)^3} \right]^2 \right\}^{1/2} \times \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right]. \quad (19)$$

Para  $\sigma^2$  tem-se:

$$p_{\text{dados3}}(\sigma^2 | y, \theta_1, \theta_2) \propto (\sigma^2)^{-\left[\frac{(3n+1)}{2} + 1\right]} \exp \left[ -\frac{3}{2\sigma^2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2 \right], \quad (20)$$

que é o núcleo de uma gama inversa:

$$p(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}}, \quad x > 0 \quad X \sim GI(\alpha, \beta),$$

de parâmetros dados por:

$$\alpha = \frac{(3n+1)}{2} \quad \text{e} \quad \beta = \frac{3}{2} \sum_{i=1}^n \left( y_i - \frac{\theta_1 x_i}{\theta_2 + x_i} \right)^2.$$

### 2.2.5. Implementação das análises

Utilizar-se-á o software R (R Development Core Team, 2010) versão 2.10.0 construindo-se rotinas (algoritmos), ou seja, o Amostrador de Gibbs e o Algoritmo de Metropolis Hastings, para estimar os parâmetros do modelo via MCMC do modelo de Michaelis e Menten, a convergência dessas cadeias foi verificada de acordo com os seguintes critérios (Raftery e Lewis (1992), Geweke (1992), Gelman & Rubin (1992) utilizando o pacote do R denominado BOA (Bayesian Output Analysis) e para a análise dos resultados foram implementadas rotinas específicas.

### 3 Resultados e Discussão

#### 3.1 Ajuste do modelo aos *dados1*, *dados2* e *dados3*

Para  $\sigma^2$  por meio dos núcleos das densidades *a posteriori* condicionais completas foram identificadas distribuições gamas inversas respectivamente relacionadas à *dados1*, *dados2* e *dados3*. No entanto, devido à dificuldade em reconhecer o núcleo das *posteriors* condicionais completas para  $\theta_1$  e  $\theta_2$  buscou-se distribuições candidatas normais que pudessem ser utilizadas na geração de cadeias de valores por meio do MCMC, obtendo-se amostras dos parâmetros.

A convergência das cadeias de todos os parâmetros do modelo foi monitorada por meio da visualização gráfica do traço e dos critérios disponíveis no pacote BOA do software R (R Development Core Team, 2010), de acordo com Figuras 2 a 4 e Tabela 4.

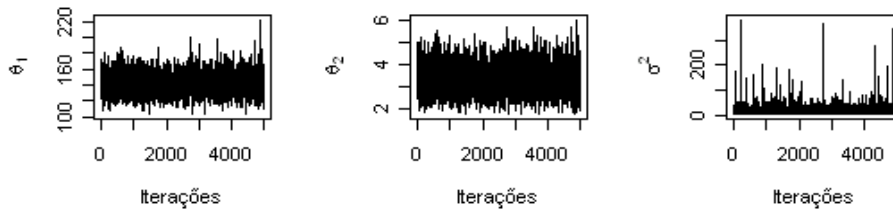


Figura 2 - Gráfico de traços dos parâmetros da cadeia *dados1*.

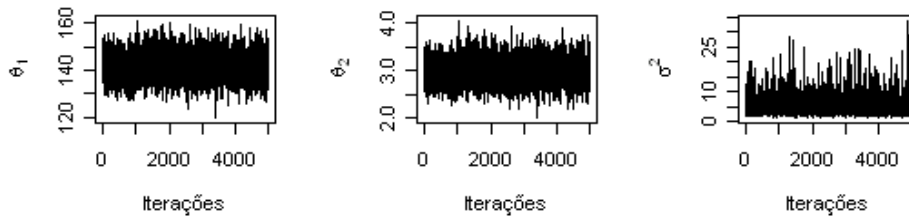


Figura 3 - Gráfico de traços dos parâmetros da cadeia *dados2*.

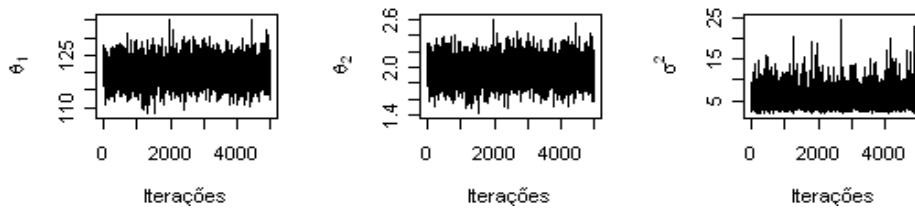


Figura 4 - Gráfico de traços dos parâmetros da cadeia *dados3*.

Tabela 4 - Critérios de convergência relacionados a *dados1*, *dados2*, *dados3*

Critérios de convergência	Parâmetros								
	<i>dados1</i>			<i>dados2</i>			<i>dados3</i>		
	$\theta_1$	$\theta_2$	$\sigma^2$	$\theta_1$	$\theta_2$	$\sigma^2$	$\theta_1$	$\theta_2$	$\sigma^2$
Raftery e Lewis FD	2,2216	2,1949	1,0320	2,1554	2,2103	0,9824	1,0320	1,0320	1,0075
Geweke p-valor	0,5618	0,4866	0,6761	0,9763	0,8934	0,8355	0,0720	0,0853	0,5985
Gelman e Rubin $\hat{R}$	1,0330	1,0339	1,0596	1,0003	1,0003	1,0005	1,0001	1,0002	0,9999

Por meio da visualização gráfica percebeu-se um comportamento que denotou a estabilização da cadeia evidenciando a convergência da mesma para os parâmetros do modelo. Observando-se a tabela, nota-se que o fator de dependência de Raftery e Lewis (1992), não forneceu evidências contrárias a convergência de todos os parâmetros para os dados, pois seu valor foi menor que 5.

O critério de Geweke (1992) possibilitou postular a não existência de evidências contra a convergência porque o *p*-valor foi menor do que o nível de significância determinado pelo pesquisador que nesse estudo foi considerado igual a 5%. Todos os parâmetros analisados nas três massas de dados mostraram não haver nenhuma evidência contra a convergência dos mesmos. Os valores de *R* estão muito próximos ao valor unitário, situação que indica a convergência segundo Gelman e Rubin (1992).

As estimativas dos parâmetros, ou seja, moda, mediana e média acompanhadas dos seus respectivos desvios padrões, erros de Monte Carlo e o HPD encontram-se na Tabela 5.

Tabela 5 - Estimativa do parâmetro que pode ser dada via média, moda e mediana da distribuição marginal *a posteriori* de cada parâmetro (média, moda ou mediana do conjunto de valores gerados das condicionais completas *a posteriori* de cada parâmetro) relacionados aos *dados1*, *dados2*, *dados3*

Medidas	Parâmetros								
	<i>dados1</i>			<i>dados2</i>			<i>dados3</i>		
	$\theta_1$	$\theta_2$	$\sigma^2$	$\theta_1$	$\theta_2$	$\sigma^2$	$\theta_1$	$\theta_2$	$\sigma^2$
Moda	137,92	3,16	6,79	140,78	2,93	3,95	120,08	1,95	4,42
Mediana	139,26	3,19	9,57	141,16	2,95	3,60	120,17	1,95	4,93
Média	140,96	3,28	13,76	141,41	2,97	5,29	120,26	1,95	5,36
DP	13,80	0,70	16,86	5,47	0,26	2,80	3,12	0,14	2,09
EMC	0,579	0,029	0,3334	0,0948	0,0046	0,0416	0,0505	0,0025	0,028
HPD	[113,90; 168,3]	[1,96; 4,66]	[1,99; 35,92]	[131,00; 152,5]	[2,47; 3,49]	[1,67; 10,3]	[114,04; 126,38]	[1,67; 2,22]	[2,22; 9,28]

De acordo com Barbosa (2004), os erros de Monte Carlo são considerados baixo, se forem menores que 0,4987, dessa forma observa-se que os erros de Monte Carlo apresentaram valores baixos, confirmando a boa qualidade das estimativas dos parâmetros.

Os HPD para todos os parâmetros foram consideravelmente reduzidos em seus tamanhos proporcionando alta precisão, em especial o parâmetro  $\theta_1$  que tem uma interpretação ecológica das mais importantes por significar o número de espécies que pode ser encontrada no ambiente onde o estudo foi realizado.

A estimativa pontual dos parâmetros acompanhado dos seus HPD com seus desvios padrões faz com que a estimativas se aproxime da quantidade de espécies contidas na população teórica considerada na massa de *dados3*.

As Figuras 5, 6 e 7 representam por meio de gráficos as densidades das distribuições *a posteriori* de *dados1*, *dados2* e *dados3*.

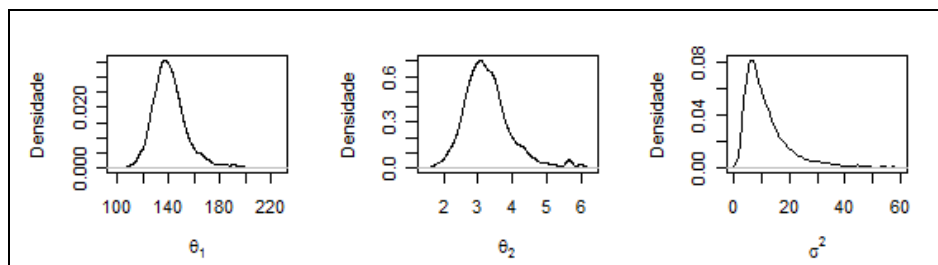


Figura 5 - Gráfico de densidade da distribuição *a posteriori* dos parâmetros *dados1*.

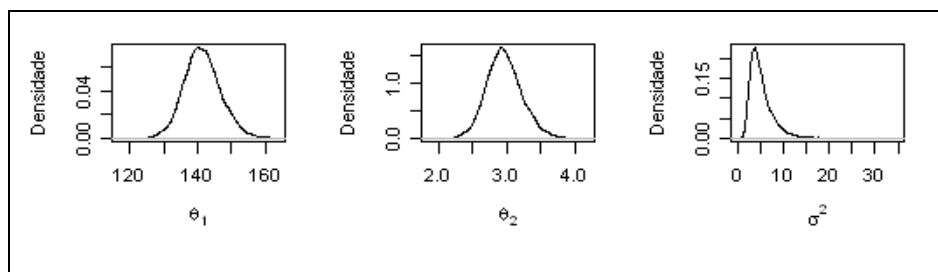


Figura 6 - Gráfico de densidade da distribuição *a posteriori* dos parâmetros *dados2*.

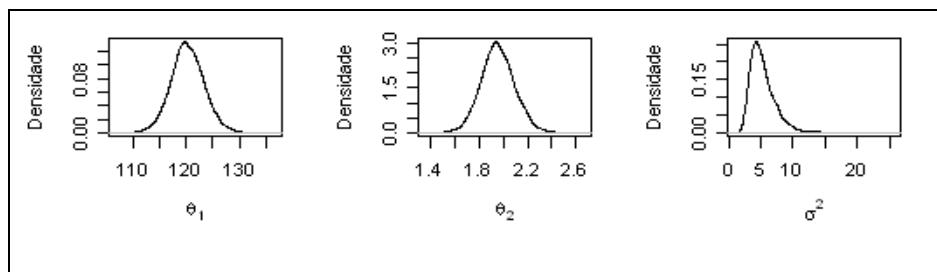


Figura 7 - Gráfico de densidade da distribuição *a posteriori* dos parâmetros *dados3*.

A Figura 8 representa as distribuições de densidades *a posteriori* dos parâmetros do modelo agrupadas em *dados1*, *dados2* e *dados3*.

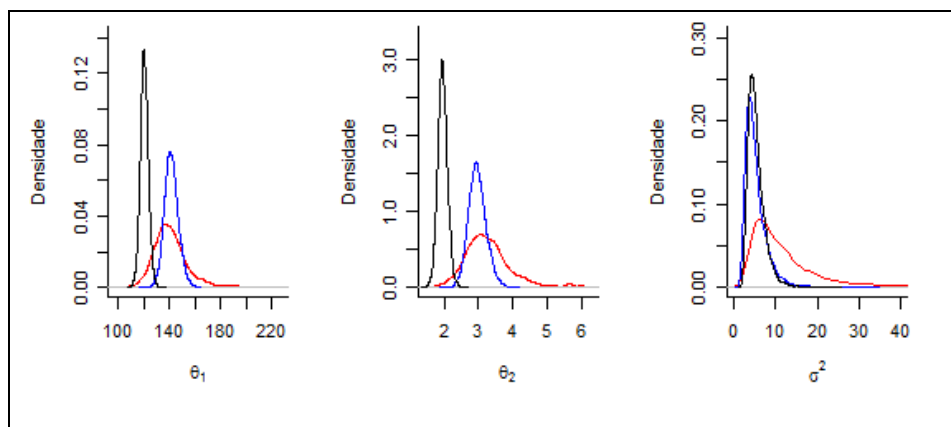


Figura 8 - Gráfico das distribuições marginais *a posteriori* dos parâmetros. As curvas em vermelho, azul e preto são referentes aos dados1, dados2, dados3.

O significado dos resultados dessa sobreposição de *posteriors* é um indicador que a atualização da informação aplicada sucessivamente, evidencia que todos os parâmetros do modelo aumentam a precisão das estimativas. Isso pode ser notado, pelos desvios, uma vez que, são sempre maiores em *dados1*, seguido de *dados2* e *dados3* em especial  $\theta_1$  que fica próximo do número de espécies da população teórica.

## Conclusões

Diante dos resultados obtidos, pode-se concluir que: a metodologia Bayesiana mostrou-se eficiente para estudar a estimação em espécies por meio do ajuste do modelo de Michaelis-Menten, pois produziu estimativas condizentes com a realidade ecológica. A essência deste artigo foi à utilização de um conjunto de distribuições de densidades *a*



*posteriori* para as massas de dados observadas em que se percebeu que as estimativas dos parâmetros de interesse relacionadas ao modelo não-linear apresentaram nesta sequência um melhoramento em relação à precisão e variabilidade. linear utilizado e outros modelos disponíveis na literatura, aprofundando a idéia em *prioris* mais informativas, com base no atual estudo.

MACHADO, E. J.; MUNIZ, J. A.; SÁFADI, T.; SAVIAN, T. V. Estimation of a model of species benthic macroinvertebrate analysis via Bayesian model Michaelis-Menten. *Rev. Bras. Biom.*, São Paulo, v.30, n.1, p.106-123, 2012.

- **ABSTRACT:** *In this study, it was estimated that the number of macroinvertebrate called biological species, using the model of non-linear regression in the context ecological known as Michaelis-Menten equation, from samples collected in streams (streams) clear water study Lima (2008). The methodology made use of Bayesian inference to estimate the model parameters. The results showed a considerable increase in the precision of estimates of parameters using the Bayesian approach with the use of informative priors.*
- **KEYWORDS:** *Regression model non-linear estimation species, benthic macroinvertebrates.*

## Referências

BARBOSA, A. H. *Análise de confiabilidade estrutural utilizando o método de Monte Carlo e redes neurais*. 2004. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal de Ouro Preto, Ouro Preto, 2004.

BERNARDO, J. M. *Bayesian reference analysis*. A postgraduate tutorial course. Valencia, Spain, 1998.

EHLERS, R. S. *Introdução a Inferência Bayesiana*. Departamento de Estatística, UFPR. 2007. Disponível em <<http://leg.est.ufpr.br/ehlers/notas>>.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, Hayward, v.7, n.4, p.457-511, 1992.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: BERNARDO, J. M.; BERGER, J. O.; DAVID A. P.; SMITH, A. F. M. (Ed.). *Bayesian Statistics*, New York: Oxford University, 1992. 526p.

LIMA, D. V. M. L. *Macroinvertebrados em igarapés de águas claras e brancas no oeste Amazônico (Acre, Brasil): implicações para elaboração de estratégias de biomonitoramento*. 2008. Dissertação (Mestrado em Ecologia e Manejo de Recursos Naturais) – Universidade Federal do Acre, Rio Branco, 2008.

MACHADO, E. J. *Uso do 'Bootstrap' na estimação de parâmetros em modelos não lineares: uma aplicação em mecanismos cinéticos de Michaelis-Menten*. 2006. 131f. Dissertação (Mestrado em Agronomia/Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2006.

MICHAELIS, L.; MENTEN, M. L. Die Kinetik der invertinwirkung. *Brochen*, Stuttgart, v.49, p.335-369, 1913.

R DEVELOPMENT CORE TEAM. The R manuals: current version: 2.9 2010. Disponível em: <<http://www.r-project.org>>. Acesso em: 15 jan. 2010.

RAFTERY, A. L.; LEWIS, S. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Stat. Sci.*, Hayward, v.7, n.4, p.493-497, 1992.

Recebido em 19.12.2011

Aprovado após revisão 21.05.2012