

## Modelos multivariados de covariância linear generalizada (McGLM) para a avaliação do efeito de hookpods em pescarias comerciais

Fernando de Pol Mayer<sup>1</sup>, Rodrigo Sant'Ana<sup>2</sup>, Dimas Gianuca<sup>3</sup>, Wagner Hugo Bonat<sup>4</sup>, Paulo Justiniano Ribeiro Jr.<sup>5</sup>

**Resumo:** *A captura de aves em pescarias comerciais gera prejuízos tanto ambientais, como econômicos. Um dispositivo chamado hookpod foi desenvolvido para ser uma alternativa aos chumbos normais, e com a intenção de diminuir esse tipo da captura incidental. A eficácia dos hookpods já é comprovada na diminuição das capturas de aves. No entanto, ainda existiam dúvidas se este dispositivo não poderia diminuir também a captura de espécies de peixes alvo da pescaria. Neste trabalho foram analisados dados de um experimento onde foram utilizados chumbos normais e hookpods, para avaliar se o uso do dispositivo pode interferir nas capturas das espécies alvo. As capturas de duas das principais espécies de atuns foram utilizadas como resposta, e alguns outros fatores, além do tratamento (hookpods), como variáveis explicativas. Para modelar essa resposta bivariada, foram utilizados os Modelos de Covariância Linear Generalizada (McGLM). Essa classe de modelos permite uma série de flexibilizações em relação aos GLMs tradicionais, como lidar com dados não independentes, e a modelagem de múltiplas respostas. A análise dos dados através desta abordagem permitiu verificar que os hookpods não interferem na captura das espécies alvo, tanto conjuntamente, quanto individualmente.*

**Palavras-chave:** Captura incidental; Modelos multivariados; Modelos de Covariância Linear Generalizada.

**Abstract:** *The capture of birds in commercial fisheries generates environmental and economic problems. A device called hookpod was developed with the intention of to reduce this type of bycatch. The effectiveness of hookpods is already proven in the reduction of catches of birds. However, there were still doubts if this device could not reduce also the capture of the main fish species. In this work, we analyzed data from an experiment where normal leads and hookpods were used, in order to evaluate whether the use of this device could affect the catches of target species. Catches of two of the main tuna species were used as response variables, and some other factors, besides treatment (hookpods), were used as explanatory variables. To model this bivariate response, a Multivariate covariance Generalized Linear Model (McGLM) was used. This class of models is more flexible than traditional GLMs. Analyzing the data through this approach allowed to verify that the hookpods do not interfere in the capture of the target species, both when modeled jointly and also individually.*

**Keywords:** Bycatch; Multivariate models; Multivariate covariance Generalized Linear Models.

<sup>1</sup>Laboratório de Estatística e Geoinformação LEG/DEST/UFPR. e-mail: fernando.mayer@ufpr.br

<sup>2</sup>Laboratório de Estudos Marinhos Aplicados LEMA/UNIVALI. e-mail: rsantana@univali.br

<sup>3</sup>Projeto Albatroz. e-mail: dgianuca@projetoalbatroz.org.br

<sup>4</sup>Laboratório de Estatística e Geoinformação LEG/DEST/UFPR. e-mail: wbonat@ufpr.br

<sup>5</sup>Laboratório de Estatística e Geoinformação LEG/DEST/UFPR. e-mail: paulojus@ufpr.br

# 1 Introdução

A coleta de dados experimentais ou observacionais geralmente é composta de uma variável resposta e uma ou mais variáveis explicativas. No entanto, frequentemente são observadas mais de uma variável resposta, que podem ou não ser correlacionadas. No caso univariado, os métodos estatísticos para análise são bem estabelecidos. Já no caso multivariado, pode ocorrer uma das duas opções: (a) análise separada de cada variável resposta, ou (b) análise por métodos estatísticos multivariados. No caso da análise individual de várias respostas, o problema é que nem sempre existem evidências suficientes individualmente, e além disso, as respostas podem ter influências umas às outras, o que não está sendo levado em consideração. Quando se opta por uma análise multivariada, os métodos disponíveis são bem mais restritos, e muitas vezes difíceis de serem determinados para alguns problemas.

Quando diversas variáveis resposta são medidas, a análise multivariada é de fato mais apropriada, pois se existe correlação entre as variáveis, essa informação deve ser explorada. Os modelos multivariados são mais poderosos em detectar efeitos de fatores quando as respostas são correlacionadas. Além disso, pode-se obter uma descrição dos resultados mais compacta e uma conclusão mais global. No entanto, nem sempre é óbvio como fazer uma análise multivariada ou mesmo interpretar seus resultados. Além disso, a maioria dos métodos são baseados na suposição de que as respostas seguem uma distribuição normal multivariada.

A abordagem padrão quando se deseja avaliar a influência de um ou mais fatores em uma resposta é a de modelos lineares. No caso multivariado, um modelo linear multivariado pode ser empregado de maneira similar. Em ambos os casos, quando se deseja fazer inferência sobre os parâmetros estimados, as suposições são as mesmas. Estas suposições são de que as observações são provenientes de uma distribuição normal com variância constante  $\sigma^2$  no caso univariado, e de que as respostas no caso multivariado seguem uma distribuição normal multivariada com matriz de covariância (simétrica e positiva definida)  $\Sigma$ , onde as entradas fora da diagonal principal indicam a covariância (ou correlação) entre as respostas (JOHNSON; WICHERN, 2008).

No caso univariado, os Modelos Lineares Generalizados (GLM) (NELDER; WEDDERBURN, 1972) são uma forma de flexibilizar tanto a suposição de normalidade quanto a de variância constante das observações. Nessa abordagem, dados de contagem e proporções, por exemplo, podem ser modelados com distribuições apropriadas, eliminando a suposição de normalidade. Já no caso multivariado, até então não existia uma abordagem similar nesse sentido.

Atualmente a análise multivariada de dados não normais envolve uma considerável gama de abordagens, indo de Equações de Estimção Generalizadas (GEE) (ZEGER; LIANG; ALBERT, 1988) à modelos lineares generalizados mistos (GLMM). No entanto, cada abordagem permite modelar apenas tipos específicos de dependência ou correlação, sem pertencer à uma abordagem mais geral. Além disso, os *softwares* disponíveis atualmente apresentam uma capacidade limitada em termos de complexidade e dimensão dos dados.

Com a intenção de obter uma abordagem tão flexível quanto os GLMs para o caso univariado, Bonat e Jørgensen (2016) propuseram uma classe de modelos chamada de Modelos Multivariados de Covariância Linear Generalizada (McGLM). Seguindo a ideia de Pourahmadi (1999), estes modelos são especificados por funções de ligação e preditores lineares para um vetor de médias e matriz de covariância respectivamente. Essa me-

todologia permite uma abordagem unificada para dados multivariados correlacionados, levando em consideração respostas com diferentes distribuições e diferentes estruturas de covariância. Os modelos são ajustados por meio de quase-verossimilhança e funções de estimação de Pearson, baseados apenas em suposições de segundo momento.

Neste trabalho, os McGLMs serão utilizados para a análise de dados de um experimento para avaliar se a utilização de um artefato de proteção de captura incidental de aves (*hookpods*) interfere na captura das espécies-alvo de uma pescaria. Em pescarias de espinhel de superfície, a captura incidental de aves causa sérios problemas, tanto ambientais (como a morte dos animais), quanto econômicos (com a inutilização de diversos anzóis). Nesse sentido, os *hookpods* foram desenvolvidos para serem acoplados aos anzóis, evitando a captura das aves, que geralmente ocorre até os primeiros 10 metros de profundidade. Ao passar de 10-15 metros, o dispositivo se solta do anzol e permanece conectado à linha funcionando e substituindo chumbos normais. A eficiência para evitar a captura das aves já é comprovada (SULLIVAN et al., 2018), portanto o objetivo deste trabalho é então avaliar se o uso deste dispositivo pode interferir na captura das espécies de interesse comercial, como os atuns e agulhões.

## 2 Metodologia

### 2.1 Dados

Os dados foram obtidos por meio de observadores de bordo do Projeto Albatroz. Entre os meses de junho e outubro de 2013, os observadores participaram de seis viagens de pesca em barcos espinheleiros da frota nacional. Em cada lançamento do espinhel, uma parte dos anzóis era lançada com o chumbo normal (controle) e outra parte com os *hookpods* (tratamento) (ver Figura 1). O peso do próprio *hookpod* serve como substituto do chumbo normal. Para maiores detalhes sobre o funcionamento deste dispositivo, veja Sullivan et al. (2018). Diversas variáveis foram também coletadas, como profundidade, direção do vento e temperatura da água. As variáveis resposta são então a captura das diversas espécies de peixes, principalmente atuns, agulhões e tubarões. Todas as respostas são em forma de contagens do número de peixes capturados com um determinado número de anzóis.

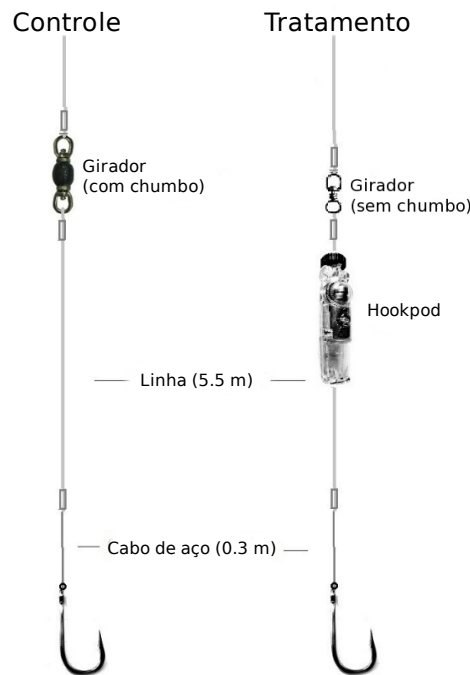


Figura 1: Figura esquemática mostrando as linhas do espinhel utilizadas no experimento. No controle foi utilizado o girador padrão com chumbo, enquanto que nas linhas com o tratamento, foi utilizado o *hookpod* e um girador sem chumbo. (Imagem adaptada de D. Gianuca, Projeto Albatroz).

De particular interesse nesse trabalho são as espécies de atuns, que possuem grande valor comercial e são destinadas quase que em sua totalidade para exportação. Das quatro espécies de atuns capturadas, apenas duas possuíam dados suficientes para análise. Portanto, aqui foram consideradas como respostas a captura da albacora-lage (*Thunnus albacares*) e da albacora-branca (*Thunnus alalunga*). Os diversos fatores coletados foram utilizados como variáveis explicativas, assim como o tratamento (uso de chumbo ou *hookpod*). A descrição de todas as variáveis disponíveis pode ser consultada na Tabela 1.

Tabela 1: Descrição das variáveis utilizadas no estudo.

Tipo	Variável	Descrição
Explicativa	periodo	Período do dia em que foram lançados os anzóis.
	profundidade	Profundidade do local de lançamento (metros).
	beauford	Escala Beauford de vento (varia de 0 a 12, com 0 sendo mar calmo, e 12 sendo furacão).
	dir.vento	Direção do vento.
	temp.agua	Temperatura da água, em graus Celsius.
	tratamento	Se foi utilizado o <i>hookpod</i> (tratamento) ou chumbo normal (controle).
	bloco	Ordem de lançamento do tratamento (ou controle).
Resposta	n.anzois	Número de anzóis lançados ( <i>offset</i> ).
	t.alalunga.cap	Captura (em número) de albacora-branca ( <i>Thunnus alalunga</i> ).
	t.albacares.cap	Captura (em número) de albacora-lage ( <i>Thunnus albacares</i> ).

## 2.2 Modelos

Os Modelos Multivariados de Covariância Linear Generalizada (McGLM) serão apresentados como uma extensão dos GLMs. Seja  $\mathbf{Y}$  um vetor de respostas  $N \times 1$ ,  $\mathbf{X}$  uma matriz de delineamento  $N \times k$ , e  $\boldsymbol{\beta}$  um vetor de parâmetros  $k \times 1$ . Um GLM pode ser escrito da seguinte forma

$$E(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta}) \quad (1a)$$

$$\text{Var}(\mathbf{Y}) = \boldsymbol{\Sigma} = V(\boldsymbol{\mu}; p)^{\frac{1}{2}}(\tau_0 \mathbf{I})V(\boldsymbol{\mu}; p)^{\frac{1}{2}} \quad (1b)$$

onde  $g$  é uma função de ligação monótona e diferenciável,  $V(\boldsymbol{\mu}; p) = \text{diag}(\vartheta(\boldsymbol{\mu}; p))$  é uma matriz diagonal cujas entradas são dadas pela função de variância  $\vartheta(\boldsymbol{\mu}; p)$  aplicada à cada elemento do vetor  $\boldsymbol{\mu}$ . Finalmente,  $p$  e  $\tau_0$  são os parâmetros de potência e dispersão, respectivamente, e  $\mathbf{I}$  é a matriz identidade  $N \times N$ . Note que o modelo linear normal padrão é obtido quando a função de ligação é a identidade e a função de variância é constante.

A grande flexibilidade dos GLMs é devida ao uso de apenas duas funções distintas, a função de ligação e a função de variância. Esta última possui grande importância nos GLMs pois diferentes escolhas implicam em diferentes suposições sobre a distribuição da variável resposta. As funções de potência

$$\vartheta(\boldsymbol{\mu}; p) = \mu^p$$

onde  $p \in (-\infty, 0] \cup [1, \infty)$ , são escolhas frequentes e caracterizam a família de distribuições Tweedie, cujos casos particulares mais importantes são a normal ( $p = 0$ ), Poisson ( $p = 1$ ), gama ( $p = 2$ ), e Gaussiana inversa ( $p = 3$ ) (JØRGENSEN, 1997). Para uma variável aleatória  $Z$  com distribuição Tweedie, escrevemos  $Z \sim \text{Tw}_p(\mu, \phi)$ , onde  $\phi > 0$  é um parâmetro de dispersão.

Apesar de toda a flexibilidade, os GLMs lidam apenas com dados univariados independentes, e a função de variância é assumida como conhecida. Os objetivos do McGLM são então: (a) lidar com dados não independentes, (b) possibilitar a modelagem de múltiplas respostas, e (c) calcular o parâmetro de potência, que funciona como uma seleção de modelos automática.

Para dados de contagem, a família Tweedie pode ser estendida para a família Poisson-Tweedie (EL-SHAARAWI; ZHU; JOE, 2011),  $\text{PTw}_p(\mu, \phi)$ , que surge a partir da seguinte especificação hierárquica (BONAT et al., 2018)

$$\begin{aligned} Y|Z &\sim \text{Poisson}(Z) \\ Z &\sim \text{Tw}_p(\mu, \phi) \end{aligned}$$

assumindo que  $p \geq 1$  para garantir que  $Z$  é não negativa. Em geral, a densidade da distribuição Poisson-Tweedie não possui solução fechada. No entanto, os dois primeiros momentos (média e variância) podem ser facilmente obtidos. Jørgensen e Kokonendji (2016) mostraram que, se  $Y \sim \text{PTw}_p(\mu, \phi)$ , então  $E(Y) = \mu$  e  $\text{Var}(Y) = \mu + \phi\mu^p$ . Esse fato motiva a especificação de um modelo baseado apenas em suposições de momentos de segunda-ordem (BONAT et al., 2018). Sendo assim, a função de variância para esta família de distribuições é

$$\vartheta(\boldsymbol{\mu}; p) = \mu + \phi\mu^p,$$

em que modelos importantes para dados de contagem são casos particulares, como Hermite ( $p = 0$ ), Neyman tipo A ( $p = 1$ ), binomial negativo ( $p = 2$ ), e Poisson-inversa Gaussiana

( $p = 3$ ). Quando utiliza-se a família Poisson-Tweedie, a matriz  $\Sigma$  na equação (1b) adquire a forma especial

$$\Sigma = \text{diag}(\boldsymbol{\mu}) + V(\boldsymbol{\mu}; p)^{\frac{1}{2}}(\tau_0 \mathbf{I})V(\boldsymbol{\mu}; p)^{\frac{1}{2}}$$

que é capaz de acomodar naturalmente dados com sub/sobredispersão e inflacionados de zero.

Ao trocar a matriz identidade na equação (1b), por uma matriz não-diagonal  $\Omega(\boldsymbol{\tau})$ , obtém-se

$$\text{Var}(\mathbf{Y}) = \Sigma = V(\boldsymbol{\mu}; p)^{\frac{1}{2}}(\Omega(\boldsymbol{\tau}))V(\boldsymbol{\mu}; p)^{\frac{1}{2}}$$

que é similar ao conceito de *working correlation matrix* no contexto de GEE (LIANG; ZEGGER, 1986). No entanto, a proposta é modelar  $\Omega(\boldsymbol{\tau})$  como uma combinação linear de matrizes conhecidas

$$h(\Omega(\boldsymbol{\tau})) = \tau_0 Z_0 + \dots + \tau_D Z_D,$$

onde  $h$  é uma função de ligação para as covariâncias,  $Z_d$  com  $d = 0, \dots, D$  são matrizes conhecidas, e  $\boldsymbol{\tau} = (\tau_0, \dots, \tau_D)$  é um vetor de parâmetros de dispersão  $(D + 1) \times 1$ . Essa estrutura é análoga ao preditor linear para a média, e é chamada de preditor linear matricial. Ao incorporar esse preditor linear matricial na equação (1b), obtemos o chamado Modelo de Covariância Linear Generalizado (cGLM). Casos especiais dessa abordagem incluem os modelos mistos generalizados, modelos de médias móveis e modelos condicionais autorregressivos.

A abordagem acima, apesar de ser mais flexível por permitir modelar a covariância entre as observações, é capaz de lidar apenas com respostas univariadas. Agora considere  $\mathbf{Y}_{N \times R} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_R\}$  uma matriz de variáveis respostas, e seja  $\mathbf{M}_{N \times R} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_R\}$  a respectiva matriz de valores esperados. Para indicar que cada resposta  $\mathbf{Y}_R$  possui sua própria matriz de covariância, considere

$$\Sigma_r = V_r(\boldsymbol{\mu}; p)^{\frac{1}{2}}\Omega_r(\boldsymbol{\tau})V_r(\boldsymbol{\mu}; p)^{\frac{1}{2}}$$

como a matriz de covariâncias dentro de cada resposta. Considere também a matriz  $\Sigma_b$  como sendo a matriz de correlações entre as variáveis resposta. Com isso, o Modelo Multivariado de Covariância Linear Generalizada (McGLM) pode ser definido como

$$E(\mathbf{Y}) = \mathbf{M} = \{g_1^{-1}(\mathbf{X}_1\boldsymbol{\beta}_1), \dots, g_R^{-1}(\mathbf{X}_R\boldsymbol{\beta}_R)\}$$

$$\text{Var}(\mathbf{Y}) = \mathbf{C} = \Sigma_R \overset{G}{\otimes} \Sigma_b$$

onde

$$\Sigma_R \overset{G}{\otimes} \Sigma_b = \text{Bdiag}(\tilde{\Sigma}_1, \dots, \tilde{\Sigma}_R)(\Sigma_b \otimes \mathbf{I})\text{Bdiag}(\tilde{\Sigma}_1^\top, \dots, \tilde{\Sigma}_R^\top)$$

é o produto de Kronecker generalizado (MARTINEZ-BENEITO, 2013). A matriz  $\tilde{\Sigma}_r$  é a matriz triangular inferior da decomposição de Cholesky de  $\Sigma_r$ , e o operador  $\text{Bdiag}(\cdot)$  denota uma matriz bloco diagonal.

Os GLMs podem ser vistos como casos particulares dos McGLMs quando a função de ligação é a identidade,  $\mathbf{X}_r = \mathbf{X}$  para todo  $r$ , a função de variância é constante, e  $\Sigma_r = \tau_{0r}\mathbf{I}$  para todo  $r$ .

A especificação do modelo através de momentos de segunda-ordem motiva a dividir os parâmetros em dois subconjuntos  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$ . Nesta notação,  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_R^\top)^\top$  é um vetor  $K \times 1$  de parâmetros de regressão, e  $\boldsymbol{\lambda} = (\rho_1, \dots, \rho_{R(R-1)/2}, p_1, \dots, p_R, \boldsymbol{\tau}_1^\top, \dots, \boldsymbol{\tau}_R^\top)^\top$  um vetor  $Q \times 1$  com os parâmetros de dispersão. As estimativas dos parâmetros de regressão são realizadas pelo método de quase-escore, e os parâmetros de dispersão pelas

funções de estimação de Pearson. A formulação do modelo Poisson-Tweedie é exatamente a mesma daquela para os modelos quase-binomial e quase-Poisson no contexto de GLMs (BONAT et al., 2018). O método de estimação considera então  $\mathcal{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_R^\top)^\top$  o vetor “empilhado”  $NR \times 1$  da matriz de variáveis respostas  $\mathbf{Y}_{N \times R}$ . De maneira similar,  $\mathcal{M} = (\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_R^\top)^\top$  é o vetor empilhado  $NR \times 1$  da matriz de valores esperados  $\mathbf{M}_{N \times R}$ . Os detalhes do método estão bem descritos em Jørgensen e Knudsen (2004), Bonat e Jørgensen (2016) e Bonat et al. (2018).

Testes de hipótese linear multivariada podem ser realizadas da seguinte forma geral

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{c} \quad \text{vs} \quad H_1 : \mathbf{L}\boldsymbol{\beta} \neq \mathbf{c}$$

onde  $\mathbf{L}$  é uma matriz conhecida de contrastes  $JR \times K$ . A estatística de teste de Wald é dada por

$$W_s = (\mathbf{L}\boldsymbol{\beta})^\top (\mathbf{L}\mathbf{J}_\beta^{-1}\mathbf{L})^{-1} (\mathbf{L}\boldsymbol{\beta}),$$

onde  $\mathbf{J}_\beta = \mathbf{D}^\top \mathbf{C}^{-1} \mathbf{D}$ , e  $J$  denota o número de coeficientes de regressão sob teste. Assintoticamente,  $W_s$  possui uma distribuição  $\chi^2$  com  $JR$  graus de liberdade. Nesse caso,  $W_s$  é uma generalização direta da estatística de Hotelling-Lawley para dados multivariados não Gaussianos.

O ajuste de McGLMs podem ser realizados pelo pacote **mcglm** (BONAT, 2018) disponível para o R (R Core Team, 2018). Nesse pacote são implementadas diversas funções auxiliares para a construção das matrizes de covariância, além da função principal, `mcglm()` que possui sintaxe similar às demais funções de funcionalidades similares no R. O pacote estima os parâmetros através de dois algoritmos diferentes, o *modified chaser*, e o *reciprocal likelihood*.

### 3 Resultados e discussão

#### 3.1 Análise exploratória

A distribuição das capturas das duas espécies de interesse pode ser vista na Figura 2 (A e B). Para a albacora-lage, nota-se uma grande dispersão dos valores de captura, sendo aproximadamente 19% de zeros e capturas de até 26 indivíduos. Já para a albacora-branca a quantidade de zeros é de aproximadamente 43%, e ocorreram no máximo a captura de 11 indivíduos por lance de pesca. A Figura 2 (C) mostra que não existe grande associação entre as capturas das duas espécies. O coeficiente de correlação entre as capturas foi de -0.04 ( $p$ -valor = 0.791), indicando que não existem evidências de que as capturas sejam correlacionadas.

Considerando que o objetivo do trabalho é avaliar a eficiência do uso dos *hookpods*, a Figura 3 mostra os boxplots das capturas das duas espécies, separadas por controle (chumbo normal) e tratamento (*hookpods*). No caso da albacora-branca, nota-se que o uso dos *hookpods* aparentemente diminuiu as capturas. No entanto, algumas capturas altas também foram observadas. Já para a albacora-lage, as capturas foram relativamente maiores com os *hookpods*, mas também observam-se valores altos no controle.

#### 3.2 Modelo

Para o ajuste do modelo, inicialmente foram consideradas todas as variáveis explicativas descritas na Tabela 1. No entanto, após uma análise preliminar, verificou-se que

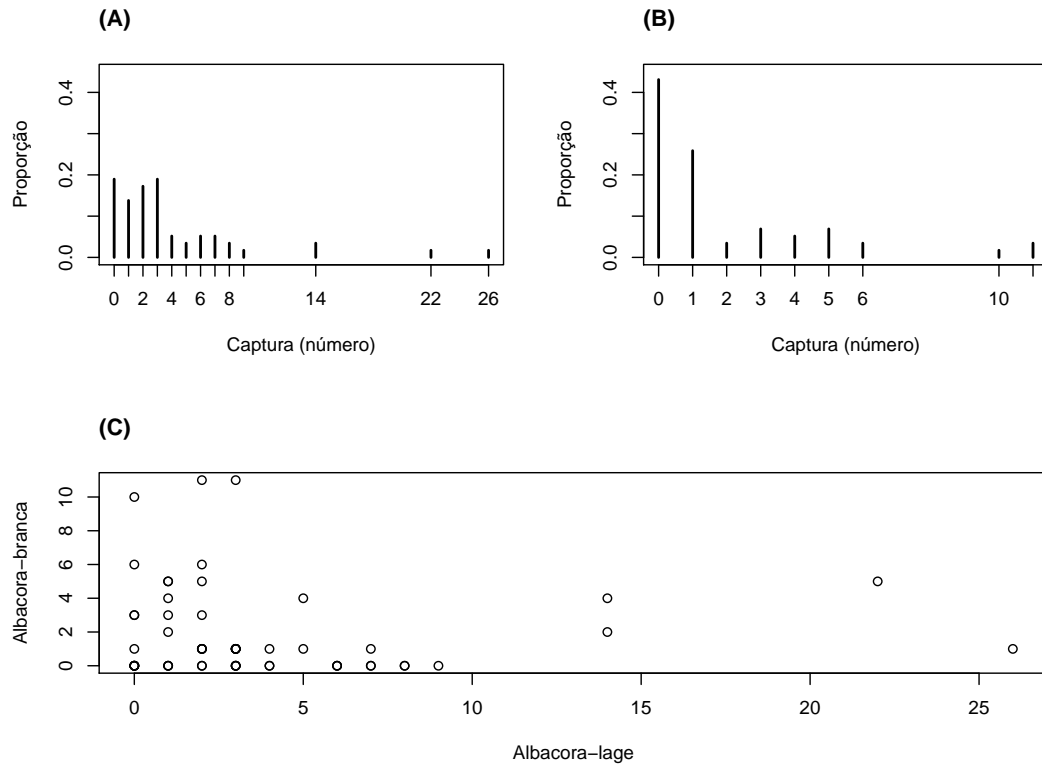


Figura 2: Distribuição das capturas de (A) albacora-lage, e (B) albacora-branca. (C) Associação entre as capturas de albacora-lage e albacora-branca (número de indivíduos de cada espécie por lance de pesca).

a variável **beauford** (escala Beauford de vento) era não significativa, por isso não foi considerada.

Como assumiu-se independência entre as observações de cada variável resposta, as matrizes de covariância para cada uma delas foram especificadas como identidades de ordem  $n$ . Para a relação funcional entre a média e a variância de cada resposta, assumiu-se a distribuição Poisson-Tweedie, ou seja,  $E(Y_i) = \mu_i$ , e  $\text{Var}(Y_i) = \mu_i + \phi\mu_i^p$ , o que determina o primeiro e segundo momento respectivamente. O parâmetro de potência  $p$  da função de variância também foi estimado. A função de ligação entre cada resposta e os preditores lineares utilizada foi a logarítmica. O número de anzóis foi utilizado como um *offset* para especificar o intervalo (variável) onde as contagens foram realizadas. O modelo de covariância multivariado foi especificado através da função **mcglm()** do pacote **mcglm**.

Uma forma de conferir a convergência do modelo é através dos gráficos da Figura 4. Percebe-se que em todos os casos o algoritmo convergiu com 53 iterações.

As estimativas do parâmetro de dispersão  $\phi$  são relevantes pois indicam se existe sub/equi/super-dispersão em cada variável. Por exemplo, se o parâmetro de dispersão for zero, então a variância se iguala à esperança e os dados podem ser descritos por uma distribuição Poisson (equidisperso). Ao ajustar os modelos com os parâmetros de potência à serem estimados, os erros padrão associados às estimativas dos parâmetros de dispersão podem ser inflacionados. Por isso, nestes casos, é aconselhável realizar um teste condicional dos parâmetros de dispersão, dado os valores das estimativas do parâmetro de potência. O resultado deste teste condicional está na Tabela 2. Ambas as estimativas são



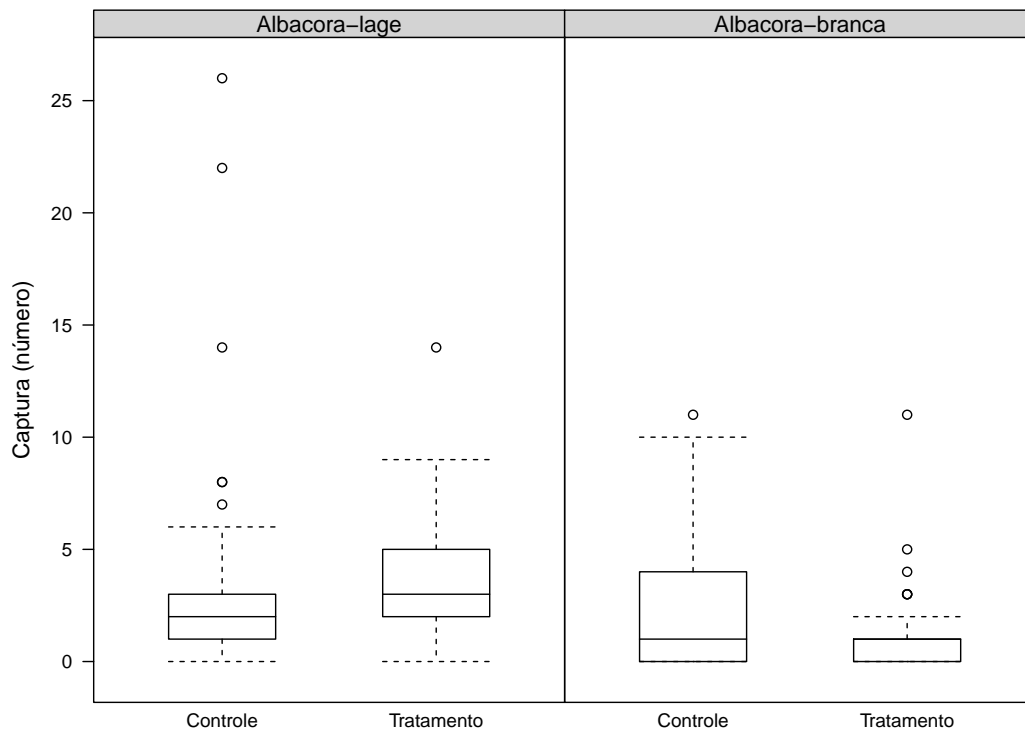


Figura 3: Boxplots das capturas de albacora-branca e albacora-lage, de acordo com uso do chumbo normal (controle), e dos hookpods (tratamento).

positivas, o que indica que existe uma aparente sobredispersão nos dados. No primeiro caso, para a espécie albacora-lage, o erro padrão é relativamente alto, e a estimativa é marginalmente significativa. Para a espécie albacora-branca, o erro padrão condicional é menor, e a estimativa do parâmetro de dispersão se mostrou significativa. Dessa forma, pode-se concluir que a captura de ambas as espécies possuem sobredispersão.

Tabela 2: Estimativas dos parâmetros de dispersão para albacora-lage ( $\phi_1$ ) e albacora-branca ( $\phi_2$ ).

	Estimativa	Erro padrão	Z	$\Pr(>  z )$
$\phi_1$	0,22	0,19	1,17	0,24
$\phi_2$	0,25	0,07	3,65	0,00

Pelo fato de se saber que existe sobredispersão, a estimativa do parâmetro de potência  $p$  é de extrema importância pois pode indicar a distribuição aproximada das observações. As estimativas desse parâmetro são mostradas na Tabela 3. Como pode ser observado, a estimativa deste parâmetro para a espécie albacora-lage foi de 2,67, embora não tenha sido possível estimar seu erro padrão, e conseqüentemente seu nível de significância. De qualquer maneira, pode-se dizer que a distribuição dos dados se aproxima de uma Poisson inversa Gaussiana ( $p = 3$ ). Já a estimativa do parâmetro de potência para a albacora-branca foi de 2,13, significativamente diferente de zero, o que indica que a distribuição dos dados para esta espécie é próxima de uma binomial negativa ( $p = 2$ ).

A correlação entre a captura das duas espécies pode agora ser estimada considerando-se todos os fatores incluídos no modelo. Essa correlação “ajustada”, ou corrigida pelos

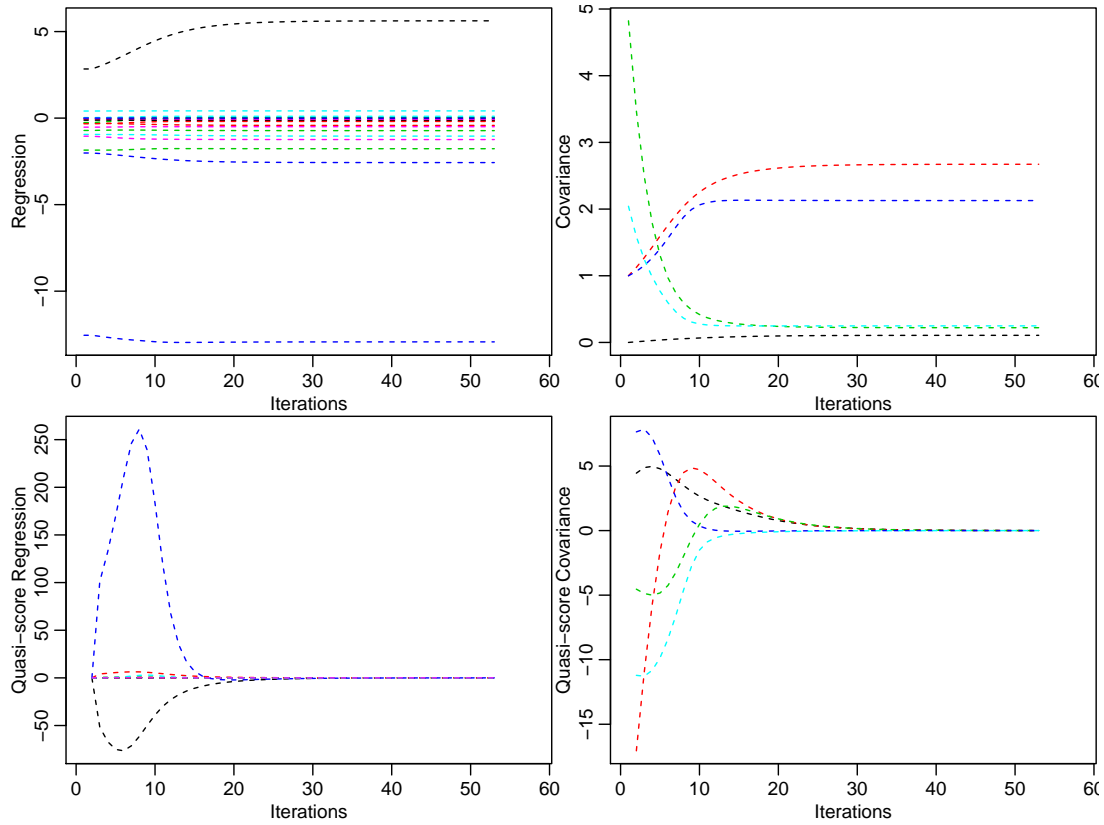


Figura 4: Diagnósticos de convergência do algoritmo.

Tabela 3: Estimativas dos parâmetros de potência para albacora-lage ( $p_1$ ) e albacora-branca ( $p_2$ ).

	Estimativa	Erro padrão	Z	$\Pr(>  z )$
$p_1$	2,67			
$p_2$	2,13	0,84	2,55	0,01

fatores está na Tabela 4. Nota-se que a correlação é positiva e não significativa. Por isso não existem evidências suficientes para supor que existe correlação entre a captura das duas espécies.

Tabela 4: Correlação ajustada entre as capturas das espécies albacora-lage e albacora-branca.

	Estimativa	Erro padrão	Z	$\Pr(>  z )$
$\rho_{12}$	0,11	0,13	0,82	0,41

Um teste de significância conjunto pode ser obtido através de uma tabela de análise de variância multivariada (MANOVA), com os testes de Hotelling-Lawley que segue uma distribuição qui-quadrado (Tabela 5). Os resultados mostram que todas as variáveis são conjuntamente significativas, com exceção de bloco e tratamento. Por ser uma variável de controle apenas, bloco foi mantido no modelo. Como o objetivo era avaliar a eficiência entre controle e tratamento, pode-se concluir que não existem diferenças nas capturas de ambas espécies quando se utiliza o chumbo normal ou o *hookpod*, pelo menos conjuntamente. Esse era o resultado esperado, pois o objetivo do *hookpod* é diminuir a captura

incidental de aves e não afetar a captura de espécies alvo.

Tabela 5: Análise de Variância Multivariada (MANOVA) conjunta para os termos considerados no modelo.

Efeitos	GL	Hotelling-Lawley	$\chi^2$	$\Pr(> \chi^2)$
Intercept	2	0,819	47,508	0,000
bloco	2	0,004	0,255	0,880
periodo	2	0,095	5,509	0,064
dir.vento	6	0,582	33,745	0,000
profundidade	2	0,136	7,880	0,019
temp.agua	2	0,643	37,279	0,000
tratamento	2	0,061	3,562	0,169

A análise individual dos efeitos estimados de cada fator em relação à cada uma das variáveis resposta pode ser vista na Tabela 6. No caso da albacora-lage, nota-se que não existe efeito na captura pelo uso do *hookpod*. Em contraste, as capturas de albacora-branca parecem ter sido influenciadas negativamente pela presença dos *hookpods*, embora a significância seja baixa.

Tabela 6: Efeitos individuais de cada fator considerado no modelo, para as espécies albacora-lage e albacora-branca.

Efeitos	Albacora-lage				Albacora-branca			
	Estimativa	EP	Z	$\Pr(>  z )$	Estimativa	EP	Z	$\Pr(>  z )$
(Intercept)	5,621	1,927	2,917	0,004	-12,931	2,169	-5,961	0,000
bloco2	-0,072	0,255	-0,282	0,778	0,104	0,268	0,389	0,697
periodonoite	-0,725	0,309	-2,347	0,019	-0,079	0,360	-0,221	0,825
dir.ventone	-2,571	0,583	-4,411	0,000	-0,147	0,461	-0,318	0,750
dir.ventose	-1,041	0,534	-1,950	0,051	-0,194	0,437	-0,443	0,658
dir.ventosw	-1,240	0,590	-2,102	0,036	-1,769	0,705	-2,511	0,012
profundidade	-0,000	0,001	-0,283	0,777	0,002	0,001	2,751	0,006
temp.agua	-0,424	0,095	-4,461	0,000	0,417	0,111	3,753	0,000
tratamentopods	0,050	0,255	0,196	0,845	-0,497	0,269	-1,849	0,064

Um modelo alternativo foi ajustado, com a variável **tratamento** excluída. A pseudo-verossimilhança do modelo completo e deste modelo reduzido foram calculadas, e a diferença entre elas utilizada para a realização de um teste aproximado de razão de verossimilhança. O teste resultou em um  $p$ -valor de 0.869, indicando que conjuntamente o tratamento não é significativo, ou seja, o uso do *hookpod* não interfere nas capturas das espécies consideradas.

## 4 Conclusões

Os Modelos Multivariados de Covariância Linear Generalizada (McGLM) demonstraram ser uma alternativa adequada para a modelagem de dados de captura. Normalmente dados de captura de diferentes espécies são altamente correlacionados, devido ao fato de que muitas espécies possuem hábitos de vida similares, e são capturadas pelo mesmo artefato. O McGLM é capaz de medir a correlação entre as espécies, considerando todos os outros fatores que podem influenciar nessa associação.

Dados de captura frequentemente possuem sobredispersão e uma grande proporção de zeros. Esta abordagem é flexível o suficiente para acomodar ambas estas situações. Além disso, a estimativa do parâmetro de potência  $p$ , que define a distribuição mais adequada para cada variável resposta, torna essa classe de modelos ainda mais flexível, pois não fica-se restrito às tradicionais distribuições para dados de contagem, como a Poisson e a binomial negativa.

Através das estimativas dos parâmetros de dispersão, verificou-se que os dados de ambas espécies possuem sobredispersão. Já as distribuições mais adequadas, de acordo com as estimativas do parâmetro de potência  $p$ , foram a Poisson inversa Gaussiana para a albacora-lage, e a binomial negativa para a albacora-branca.

A possibilidade de realização de uma Análise de Variância Multivariada (MANOVA) conjunta faz com que os resultados sejam particularmente úteis na prática. Nesse trabalho, esse teste permitiu verificar que os *hookpods* (tratamento) não interferem na captura conjunta das duas espécies analisadas. Esse resultado foi corroborado quando foram analisados os efeitos individuais dos *hookpods* sobre a captura de cada uma das espécies. Em ambos os casos, o uso dos *hookpods* não demonstrou nenhum efeito significativo na captura das duas espécies.

O uso dos *hookpods* já é comprovadamente eficaz como artefato para evitar a captura incidental de aves em pescarias com espinhel de superfície (SULLIVAN et al., 2018). No entanto, havia uma preocupação por parte dos pescadores de que este artefato também poderia diminuir a captura das espécies alvo da pescaria, diminuindo assim sua rentabilidade. Os resultados obtidos neste estudo, através da análise dos dados de um experimento especificamente planejado para verificar esse questionamento, mostraram que o uso dos *hookpods* não interferem na captura das espécies alvo.

## Referências

- BONAT, W. H. Multiple response variables regression models in R: the mcglm package. *Journal of Statistical Software*, v. 84, n. 4, 2018.
- BONAT, W. H.; JØRGENSEN, B. Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v. 65, n. 5, p. 649–675, nov 2016. ISSN 00359254.
- BONAT, W. H. et al. Extended Poisson–Tweedie: properties and regression models for count data. *Statistical Modelling*, v. 18, n. 1, p. 24–49, 2018.
- EL-SHAARAWI, A. H.; ZHU, R.; JOE, H. Modelling species abundance using the Poisson-Tweedie family. *Environmetrics*, v. 22, n. 2, p. 152–164, 2011.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. London: Pearson, 2008. 800 p. ISBN 9780131877153.
- JØRGENSEN, B. *The theory of dispersion models*. [S.l.]: Chapman & Hall, 1997. 237 p. ISBN 9780412997112.
- JØRGENSEN, B.; KNUDSEN, S. J. Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, v. 31, n. 1, p. 93–114, mar 2004. ISSN 0303-6898.

- JØRGENSEN, B.; KOKONENDJI, C. C. Discrete dispersion models and their Tweedie asymptotics. *AStA Advances in Statistical Analysis*, Springer Berlin Heidelberg, v. 100, n. 1, p. 43–78, 2016. Disponível em: <<http://dx.doi.org/10.1007/s10182-015-0250-z>>.
- LIANG, L.; ZEGER, S. Longitudinal data analysis using generalized linear models. *Biometrika*, v. 73, n. 1, p. 13–22, 1986.
- MARTINEZ-BENEITO, M. A. A general modelling framework for multivariate disease mapping. *Biometrika*, v. 100, n. 3, p. 539–553, 2013.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, v. 135, n. 3, 1972. ISSN 00359238.
- POURAHMADI, M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, v. 86, n. 3, p. 677–690, sep 1999. ISSN 0006-3444.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>.
- SULLIVAN, B. J. et al. At-sea trialling of the Hookpod: a “one-stop” mitigation solution for seabird bycatch in pelagic longline fisheries. *Animal Conservation*, v. 21, n. 2, p. 159–167, 2018.
- ZEGER, S. L.; LIANG, K.-Y.; ALBERT, P. S. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, v. 44, n. 4, p. 1049, dec 1988. ISSN 0006341X.