

UNIVERSIDADE FEDERAL DO RIO GRANDE
PÓS-GRADUAÇÃO EM OCEANOGRAFIA BIOLÓGICA

ESTIMATIVA DA ABUNDÂNCIA DE MAMÍFEROS
MARINHOS ATRAVÉS DA AMOSTRAGEM DE
DISTÂNCIA COM DETECÇÃO IMPERFEITA

FERNANDO ROBERTO GUILHERME SILVEIRA

Tese apresentada ao Pro-
grama de Pós-graduação em
Oceanografia Biológica da
Universidade Federal do Rio
Grande, como requisito par-
cial à obtenção do título de
DOUTOR.

Orientador: Paul Gerhard Kinas

RIO GRANDE

A Deus que me emprestou a
vida e a Dona Maria, Rafaela,
Talita Ana e Rosane, minha
Família.

Agradecimentos

A vida é uma coisa engraçada. Nos é emprestada e temos a ingenuidade de acharmos que é nossa. Até que o credor signatário nos chame a prestar conta dela. Para a grande maioria, o resultado desta prestação de contas termina no vermelho. E assim sendo, precisamos lançar mão de novo empréstimo. Tantas vezes quantas forem necessárias para que percebamos o seu significado, o significado desse empréstimo, o significado da vida, que não nos pertence, mas que porém, nos é emprestada!

Além de compreender o seu significado, este empréstimo, nos serve para que façamos o nosso próprio tesouro, o tesouro das amizades. Ao longo dos últimos cinco anos e meio em que vivi nesta Universidade, tive a oportunidade de aumentar meu tesouro. Conheci a muitos, alguns já partiram desta Instituição, outros ainda não e alguns permanecerão aqui por muito, mas muito tempo. Foram feitas amizades muito boas e algumas nem tanto assim, entretanto, hoje, bem ou mau, todas fazem parte do que sou como pessoa e como profissional. Seria ingenuidade de minha parte querer enumerar a todas que conheci. Certamente cometeria grande equívoco, esquecendo alguém. As vezes não nos damos conta de que além de nossos Orientadores e Professores que contribuíram imensamente para nossa formação, há anônimos, que por vezes convivem diariamente conosco fazendo de nossa existência algo um pouco mais agradável. Refiro-me a infinidade de servidores que compõem esta Instituição e que por vezes nem os cumprimentamos - Bom dia ou até amanhã! Ou daqueles que nem conheci, mas que cruzaram diariamente o meu caminho, seja indo à Biblioteca, ao Centro de Convivência ou ao "Le RU", nome carinhoso que passei a chamar o Restaurante Universitário.

Enriquecemos convivendo com pessoas que não sabemos o nome, mas que, porém sabíamos que estariam lá seguindo para suas rotinas acadêmicas. E quando "faltavam" nos perguntávamos: que será que aconteceu? Será que tais pessoas não fazem mesmo parte do que sou hoje? do nosso tesouro? Mais próximas então, estão as secretárias que limpam nossas salas todos os dias. Por vezes, nem as víamos, pois já haviam passado por lá. As secretárias que cuidam da portaria, que por um certo tempo, estavam a disponibilizar as chaves do laboratório. Fizesse chuva ou sol. As "Tiazinhas" do RU, que trabalharam duro para que todos os dias o café, almoço e jantar fosse servido. E sorridentes, mesmo que fosse uma obrigação profissional. Faltasse ônibus ou não.

Domingo ou feriado. Estavam lá. O "micro", famoso micro e seu Paulo, um dos motoristas da Instituição, que nos acordava toda manhã com sua seleção musical. Coisa que você só vai ver na FURG. E os colegas de aula, os camaradas, que viveram as mesmas coisas que vivi. Não seriam irmão de luta? de jornada? de trabalho? de academia? do jeito que você queira denominar, não seria um irmão? Uma amizade que vai ao ponto de você dizer: um irmão, acho que marca muito a existência, não? Quantos foram esses. Nem todos chamei assim, mas lutaram e sofreram tanto ou mais que eu para finalizar essa etapa acadêmica ou mais uma etapa de vida. Alguns desses camaradas são até Professores! Não que o "até" seja pejorativo, entretanto Professores tem outras atribuições e responsabilidades. Estão em outro momento que não o nosso. Todos eles, tiveram grande peso nas nossas formações. Até os que não foram meus Professores. Pois, muitos dos que não foram, me ajudaram nesse processo de formação. Primeiro, profissional e não menos importante, pessoal. E o que falar do Orientador? Foi quem me pegou pela mão, como fiz com minha filha quando pequena, e me conduziu até onde hoje me encontro profissionalmente falando. Acredito, sem sombra de dúvidas, que deve ter perdido algumas noites de sono pensando em "meus" problemas...

Não sei ao certo, se em relação a este meu último empréstimo de vida, terminarei no vermelho quando chegar a hora de prestar contas. Contudo, no que se refere ao tesouro das amizades, fique mais rico nesta, pelo que vivi nos últimos cinco anos e meio desta Universidade.....

PREFÁCIO

A estrutura desta tese segue a formatação alternativa “Modelo com trabalhos em apêndice”, segundo o Guia de Teses da Coordenação de Curso da Pós-Graduação em Oceanografia Biológica, FURG.

A tese está dividida em duas partes. A primeira constitui-se de uma descrição geral do trabalho, que inclui as seções Introdução, Material e Métodos, Resultados e Conclusões. A segunda, de uma seção Apêndice, onde os resultados apresentados na primeira parte e suas discussões estão descritos detalhadamente em dois artigos científicos formatados de acordo com as revistas onde serão/foram publicados.

O Apêndice I, intitulado **Bayesian Inference in Line Transects with Double Count Sampling and Imperfect on-Line Detection**, trata do desenvolvimento e da avaliação da qualidade do ajuste por meio de simulação, de modelos hierárquicos com uma abordagem bayesiana, que contemplam em uma mesma estrutura o método de transecção linear (*Distance Sampling Method*) com dois observadores independentes e simultâneos no momento da transeção (*Mark and Recapture Method*). Estes modelos estimam a abundância sem considerar a prerrogativa do método "*Distance Sampling*" de que todos os animais sobre a linha de transeção serão observados, ou seja, detecção imperfeita dos animais.

O Appêndice II, intitulado **Viés de Percepção em Estimação da Abundância de Toninhas na Baía da Babitonga, Brasil**, trata da aplicação dos modelos desenvolvidos no primeiro artigo. Estimou-se a abundância, a densidade e conseqüentemente a probabilidade de detecção, bem como o viés de percepção de uma subpopulação de Toninhas, *Pontoporia blainvillei*, espécie ameaçada, num ambiente em pleno desenvolvimento urbano e portuário, a Baía da Babitonga.

Contents

1	RESUMO	8
2	ABSTRACT	9
3	INTRODUÇÃO	10
3.1	Contextualização Metodológica	10
3.2	Contextualização Ecológica	13
3.3	Hipóteses de Trabalho	16
3.4	Objetivo Geral	16
3.5	Objetivos Específicos	16
4	METODOLOGIA	17
4.1	Os Modelos	17
4.2	Os dados	19
4.3	Ajuste dos Modelos	22
4.4	Inferência Bayesiana	22
4.4.1	Dados Simulados	22
4.4.2	Dados do Ganso Loon - Estudo de caso I	23
4.4.3	Análise de Resíduo - Qualidade do Ajuste	23
4.4.4	Dados da Toninha - Estudo de caso II	24
5	RESULTADOS	25
5.1	Dados Simulados	25
5.2	Loon - Estudo de Caso I	27
5.3	Toninha - Estudo de Caso II	27
6	CONCLUSÃO	31
6.1	Dados Simulados	31
6.2	Estudo de Caso I	32
6.3	Estudo de Caso II	32
7	CONSIDERAÇÕES FINAIS	33

8	REFERÊNCIAS	34
9	ANEXOS	39
9.1	Anexo I	39
9.1.1	Bayesian Inference in Line Transects with Double Count Sampling and Imperfect on-Line Detection	39
9.2	Anexo II	40
9.2.1	Viés de Percepção em Estimação da Abundância de Toninhas na Baía da Babitonga, Brasil	40

1 RESUMO

Estimar abundância é primordial no estudo e gerenciamento de populações selvagens. Um dos métodos amplamente utilizado com essa finalidade é o *Distance Sampling* (DS). Em sua formulação convencional, DS tem entre suas prerrogativas a detecção de todos os indivíduos sobre a linha percorrida. No ambiente marinho, a violação dessa prerrogativa é comum e decorre da dificuldade de avistar os indivíduos quando submersos. Nesse caso o não atendimento à premissa causará subestimativas na abundância, fenômeno conhecido por viés de percepção. Ao propormos a estruturação hierárquica dos modelos DS com a incorporação de dois observadores simultâneos e independentes e uma abordagem inferencial bayesiana, procuramos incluir a estimativa do viés de percepção simultaneamente com abundância, para assim evitar esta limitação dos modelos convencionais. Os resultados mostraram que a incorporação da informação do segundo observador permite estimar o viés de percepção para cada um dos observadores, corrige a estimativa da abundância e reduz o desvio padrão desta estimativa em relação aos modelos que ignoraram a dupla observação. Implementamos e testamos uma extensão que permite identificar a máxima probabilidade de detecção a distâncias positivas. Finalmente, aplicamos os modelos hierárquico a uma população de toninhas (*Pontoporia blainvillei*) residente da Baía de Babitonga, SC. Usamos uma distância de truncamento de 87m utilizada em campo mas também efetuamos ajuste com os dados completos, sem truncamento. Para dados truncados, o melhor modelo estima a máxima probabilidade de detecção (MPD) sobre a derrota (distância = 0). Para dados não-truncados o melhor modelo estima a MPD a distâncias maiores que zero. Os melhores ajustes (menor DIC) indicam curvas MPDs distintas para cada observador tanto para dados truncados quanto não-truncados. O ajuste para dados não-truncados evitou a eliminação de 26% dos dados e indica que a distância média ideal para truncamento é de 123m, superior àquela utilizada em campo (87m). A densidade de grupos foi estimada em $1.219 N_g km^{-2}$ (ICr95%: 0.739 a 2.075). A contribuição deste trabalho ao contexto oceanográfico diz respeito ao fato da modelagem bayesiana em código BUGS permitir uma fácil compreensão dos modelos e flexibilização, que outras propostas não contemplam. E a obtenção de distribuições posteriores da distância de máxima probabilidade de detecção, taxa de recaptura e viés de percepção.

2 ABSTRACT

Abundance estimation is essential in the study and management of wildlife populations. Distance sampling (DS) is a method largely used for this purpose. In its conventional formulation, DS has amongst its assumptions the detection of all individuals on the track line. In the marine environment this assumption is almost always unrealistic due to the difficulty to detect submersed individuals. The failing of this assumption causes underestimation of abundance, a phenomenon known as perception bias. By proposing a hierarchical DS model with the inclusion of two simultaneous and independent observers, and an inferential bayesian approach, we expect to estimate perception bias together with abundance in order to avoid the limitation of perfect on-line detection imposed by conventional DS. Our results show that incorporating a second observer allows for the estimation of observer-specific perception bias, corrects abundance estimates and reduces standard deviation of this estimate in comparison to models that ignore information on double observers. We implement and test a model extension that allows for maximum detection probabilities at positive distances. Finally, we apply hierarchical DS models to a franciscana (*Pontoporia blainvillei*) population resident in Baía de Babitonga, SC. We truncated distances at 87m as recommended by field work researchers but used non-truncated data as well, for comparison. For truncated data the best fitted model estimates the maximum detection probability (MDP) to be located on the transect line (distance = 0). For non-truncated data, the best model estimates MDPs at positive distances. The best fitted models (lowest DIC) indicate distinct detection functions for the observers. The fit with non-truncated data avoided exclusion of 26% of data and estimates average distance for truncation at 123m, larger than *ad hoc* distance of 87m used at field. Group density was estimated to be $1.219 N_g km^{-2}$ (ICr95%: 0.739 a 2.075). The contribution of this work within the oceanographic context is the fact that bayesian modeling in BUGS code allows for easy understanding of the models and flexibility, that other proposals do not address. Posterior distribution of distance of maximum detectability, recapture rate, and perception bias are observed.

3 INTRODUÇÃO

3.1 Contextualização Metodológica

A obtenção de uma estimativa de abundância precisa, a partir da qual se possa elaborar planos de manejo, requer um trabalho de campo com logística difícil e financeiramente dispendiosa, cujos resultados nem sempre são satisfatórios devido ao nível de dificuldades na obtenção dos dados ([Kery & Royle, 2010]). No ambiente marinho, a extensa área de distribuição da espécie ([Siciliano, 1994]; [Crespo et al., 1998]) e a dificuldade de observação dos indivíduos submersos por longo tempo ([Dolphin, 1987]; [Janik & Thompson, 1996]; [Beasley & Jeffersom, 2002]; [Crespo, 2009]), são exemplos dessa dificuldade. Além do mais, a acurácia das estimativas está fortemente influenciadas pelos fatores ambientais (clima, ambiente, luz); pelo observador (fadiga, experiência) e pelo animal em si (camuflagem, tamanho, comportamento) ([Marsh & Sinclair, 1989]).

A elaboração de um plano amostral que seja logística e economicamente viável e estatisticamente adequado geralmente apresenta algumas dificuldades. Tais dificuldades levaram a substituição da inferência tradicional (sample-based) por uma abordagem alternativa fundamentada prioritariamente em modelos estatísticos (model-based) ([Borchers et al., 2002]; [Royle & Dorazio, 2008]). Em estudos ecológicos, a segunda abordagem vem se estabelecendo como uma alternativa vantajosa porque muitas vezes não se consegue dispor de planos amostrais ideais e tampouco de amostras suficientemente grandes exigidas nos procedimentos convencionais baseados na estrutura da amostragem. Em contraste, na inferência baseada em modelo, todo dado observado é representado por uma variável aleatória seguindo alguma distribuição de probabilidade descrita a partir de suposições que consideram a realidade dos processos ecológicos e observacionais subjacentes ([Borchers et al., 2002]; [Laake & Borchers, 2004]). São exemplos da abordagem baseada em modelos os métodos de estimativa de abundância como marcação-recaptura, (MR) ([Borchers et al., 2002]; [McKenzie et al., 2006]) e amostragem por transecção linear, *distance sampling*, (DS) ([Borchers et al., 2002]; [Laake & Borchers, 2004]), desenvolvidos com o propósito de determinar abundância absoluta ou densidades populacionais, inicialmente para espécies terrestres e posteriormente adaptados também para o meio aquático ([Laake, 1999]; [Andriolo et al., 2010]; [Dalla Rosa et al., 2012]).

Entretanto, a violação de alguns pressupostos intrínsecos a esses métodos caracterizam eventuais limitações ao seu uso. Um exemplo que será relevante nesta tese é o fenômeno denominado “detecção

imperfeita” em modelos DS. No contexto dos métodos DS, [Marsh & Sinclair, 1989] caracterizaram as imperfeições no processo de registro visual de animais ou grupos como “viés de visibilidade”, subdividindo-o em dois componentes. O “viés de percepção”, referente à detecção imperfeita de animais disponíveis na área varrida durante o processo de amostragem; e o “viés de disponibilidade” referente à ausência do animal na área varrida. Na formulação convencional do método DS é estabelecida a premissa de que são detectados todos os animais diretamente sobre a linha de transecção. Essa premissa é particularmente questionável no estudo de mamíferos marinhos, dado sua eventual permanência sob a superfície durante longo tempo, podendo não ser detectados embora se encontrem sob a linha de transecção percorrida no momento do levantamento de campo ([Dolphin, 1987]; [Janik & Thompson, 1996]; [Beasley & Jeffersom, 2002]).

Uma abordagem estatística nova que incorpora o viés de percepção no processo inferencial consiste na utilização de modelos hierárquicos ([Marsh & Sinclair, 1989]; [Jonsen et al., 2003]; [Clark, 2005]; [Royle & Dorazio, 2008]). Esta abordagem já tem fornecido soluções simples e eficientes para o viés causado pela heterogeneidade na probabilidade de captura em modelos de MR ([Royle & Dorazio, 2008]). Adotá-la no contexto do DS proporcionará novas perspectivas, ao incluir explicitamente o viés de percepção na formulação do modelo para estimação de abundância.

Em termos estatísticos, os modelos hierárquicos permitem decompor a função de verossimilhança reescrevendo-a em vários submodelos (ou componentes) que descrevem aspectos específicos dos processos ecológicos e observacionais, privilegiando a compreensão e aumentando a versatilidade. O primeiro componente chamado modelo do processo (MP) descreve a dinâmica dos processos ecológicos subjacentes com base nas hipóteses formuladas sobre seu funcionamento. O segundo componente chamado modelo da observação (MO) contém a descrição probabilística do mecanismo que produz o dado observado. Cada um desses componentes (MP e MO) podem ser decompostos em vários submodelos, estendendo a estrutura hierárquica indefinidamente ([Berliner, 1996]). O viés de percepção pertence ao MO uma vez que está associado ao protocolo de coleta de dados, enquanto o viés de disponibilidade integra o MP e não será abordado nesta tese.

Visando a redução do viés de percepção e consequente correção nas estimativas de abundância, a estrutura convencional do DS ([Borchers et al., 2006]; [Royle & Dorazio, 2008]) tem sido estendida de várias maneiras com o intuito de: (i) reduzir a incerteza ([Quang & Becker, 1996]; [Quang & Becker, 1997]; [Quang & Becker, 1998]; [Bailey et al., 2007]; [Conn et al., 2012]; [Laake, 1999]), e (ii) incorporar o efeito de covariáveis ambientais na probabilidade de detecção ([Quang & Becker, 1996]; [Quang & Becker, 1998]; [Durban et al., 2005]; [King & Brooks, 2008]).

Neste sentido, [Laake, 1999] demonstrou o aumento da precisão, conjugando transecção linear e dupla observação numa estrutura de verossimilhança, relatando a heterogeneidade da probabilidade de detecção. [Quang & Becker, 1996] e 1997 fizeram o mesmo, porém usando covariáveis e uma curva de detecção diferente da meia-Normal. Os mesmos autores em 1998, implementam a função logit como função de ligação. [Durban et al., 2005] usando a abordagem bayesiana, demonstraram a dependência entre diversos modelos (hipóteses) às diferentes estimativas de abundância. O método pelo qual estimaram a abundância foi MR com uma função log-linear. [Bailey et al., 2007] estimam simultaneamente a ocupação e a probabilidade de detecção por meio de uma estrutura de verossimilhança, fazendo uso do pacote fechado GENPRES. [King & Brooks, 2008] comparam oito modelos (hipóteses) usando uma abordagem bayesiana para verificar a heterogeneidade da incerteza entre eles. Fazendo uso do pacote fechado RJMCMC, a estrutura hierárquica selecionava modelos e parâmetros. [Eguchi & Gerrodette, 2009] por meio da abordagem bayesiana e clássica, compararam as estimativas de abundância feitas por meio do método DS convencional. Já [Conn et al., 2012] estimaram a abundância por meio de uma estrutura hierárquica, conjugando transecção linear e dupla observação. Usaram uma função probit e covariáveis, porém não estimaram a probabilidade de detecção. No primeiro momento, estimaram a partir de uma população simulada e em seguida com dados publicados. Também usaram um pacote fechado, o RJMCMC. [Oedekoven et al., 2014] usaram um GLMM e a função log-linear para analisar o desempenho da estimativa da abundância feita pelo método de DS. Demonstraram, que o uso de covariáveis afeta a probabilidade de detecção. Aqui também fizeram uso do pacote fechado RJMCMC na seleção entre meia-Normal ou *Razard rate* como função de detecção.

A análise bayesiana vem se popularizando rapidamente no estudo de populações de animais selvagens, pois permite uma avaliação das incertezas que envolvem a abundância e os parâmetros relacionados ([Karunamuni & Quinn, 1995]; [Berliner, 1996]; [Wade, 2000]; [Durban et al., 2005]; [McCarthy, 2007]; [King & Brooks, 2008]). Em contraste com a estatística convencional (frequentista), a bayesiana define probabilidade como sendo a métrica adequada para quantificar todo tipo de incerteza devido à informação apenas parcial de qualquer proposição ([Jaynes, 2003]). Portanto, uma estimativa de abundância populacional, resultará em uma distribuição (posterior) de probabilidade. A abordagem bayesiana também é capaz de integrar formalmente toda informação relevante extra-dados (denominada distribuição a priori) com os dados efetivamente observados (formalizados na função de verossimilhança) para, em conjunto, produzir o estado atualizado do conhecimento (distribuição posterior) por meio do Teorema de Bayes ([Kinas & Andrade, 2010]).

Modelos hierárquicos com uma abordagem bayesiana fornecem uma proposta muito flexível e mais realista para descrever dados coletados em campo nas pesquisas em ecologia ([Clark, 2005]; [Schofield & Barcker, 2010]). O método de simulação de Monte Carlo via Cadeias de Markov (MCMC - *Markov Chain Monte Carlo*) é, por sua vez, uma ferramenta extremamente útil para obter distribuições posteriores em altas dimensões com relativa facilidade ([Martin & Quinn, 2006]; [McCarthy, 2007]). O surjimento de *software* específico implementado no aplicativo R ([R Core Team, 2014]) tornou a inferência bayesiana e a utilização de modelos hierárquicos acessíveis a comunidade científica em geral.

Estes modelos hierárquicos com abordagem bayesiana para DS convencional ($g(0) = 1$) foram sistematizados por [Royle & Dorazio, 2008] (Chapter 7). E com o intuito de simular a distribuição posterior conjunta para todos os parâmetros desconhecidos usando MCMC, eles formularam estes modelos dentro da estrutura de dados aumentados (*data augmentation*) fornecendo uma estrutura, que em princípio, permite ajustar modelos com várias formas de generalizações ([Royle & Dorazio, 2008], p.181). Segundo ainda [Royle & Dorazio, 2008], seus estudos conceituais mostraram que o erro de observação pode levar a um $g(0) < 1$, causando confusão com o viés de percepção.

3.2 Contextualização Ecológica

Por outro lado, populações de mamíferos marinhos estão sujeitas a ameaças diretas e indiretas. Entre as ameaças diretas destacam-se a captura incidental nas redes de pesca, colisões com embarcações e, em menor escala, a caça. Entre as pressões indiretas destacam-se a degradação ambiental tanto via perda de habitat quanto o esgotamento de recursos; proliferação de algas; introdução de espécies exóticas; potenciais patógenos; poluição industrial ou agrícola; descarte doméstico; eutrofização; poluição sonora e escombros marinhos. A longo prazo, também vale mencionar o aquecimento global dos oceanos como consequência das mudanças climáticas ([Harwood, 2001]; [Learmonth et al., 2006]).

Ações de efeito direto associadas às alterações ambientais que afetam a capacidade de suporte, podem levar uma espécie a migrar para áreas mais favoráveis causando mudanças na sua distribuição e abundância, o que afetará a estrutura das comunidades ecológicas em que se inserem. O sucesso da adaptação às novas condições dependerá ainda da suscetibilidade a doenças e contaminantes e do sucesso reprodutivo. Caso contrário, levará ao aumento no risco de extinção ([Learmonth et al., 2006]).

Conhecer o tamanho de uma população ajuda a compreender o tamanho da área de uso, a dinâmica territorial e populacional, a interação entre comunidades. Além de ajudar na avaliação do risco de extinção ([Jonsen et al., 2003]). Portanto, apesar de não ser uma tarefa fácil para a pesquisa ecológica e gestão dos recursos naturais ([Senar & Conroy, 2004]; [O’Connell et al., 2006]; [Newman et al., 2009]; [Williams & Thomas, 2009]), estimar o tamanho de uma população tem implicações na avaliação do status de conservação e no monitoramento à suscetibilidade das ações de conservação ([Williams, 2001]; [Jonsen et al., 2003]).

Um exemplo é o que ocorre com a toninha (*Pontoporia blainvillei* Gervais and D’Orbigny, 1844) na Baía da Babitonga, que é caracterizada como a última grande formação de manguezal do Hemisfério Sul, sendo o mais importante estuário de Santa Catarina ([IBAMA, 1997]). As áreas de entorno do estuário foram categorizadas pelo Projeto de Conservação e Utilização Sustentável da Diversidade Biológica Brasileira, o PROBIO (2003), como de “Prioridade Extremamente Alta”. Esse estuário é uma região rica em diversidade dando suporte físico como área de criadouro para muitas espécies devido à disponibilidade alimentar e fornecimento de abrigo, promovendo variados movimentos migratórios.

A Baía da Babitonga, apesar de sua importância ambiental, é palco do acelerado processo de ocupação urbana desordenada e exploração de seus recursos. Sofre com a problemática da poluição lançada no estuário, proveniente de esgotos domésticos das cidades de entorno (Joinville, São Francisco do Sul e Araquari), de rejeitos e efluentes industriais (Joinville e Araquari), defensivos agrícolas e fertilizantes oriundos de áreas rurais (Joinville, Araquari e Garuva) ([Perez, 2015]). Sofre ainda com a construção do Porto de Itapoá e a atividade portuária do Porto de São Francisco do Sul. Em termos de Brasil, o porto de São Francisco do Sul é o quinto maior em movimentação de contêineres e o sexto em volume de carga, operando há 55 anos ([Porto de São Francisco (Autarquia), 2015]).

Neste cenário de pleno desenvolvimento (urbano, industrial e agrícola), e em meio à diversidade local, encontra-se a toninha. De hábito predominantemente costeiro ([Danilewicz et al., 2009]), é a única espécie do gênero *Pontoporia* (Família Pontoporiidae, Mammalia: Cetartiodactyla). Endêmica do Oceano Atlântico Sul Ocidental, ocorre desde Itaúnas (18°25’S), no Espírito Santo (Brasil) até o Golfo Nuevo (42°35’S), na Província de Chubut (Argentina) ([Siciliano, 1994]; [Crespo et al., 1998]). A população divide-se em quatro subpopulações onde FMA-I representa a área de manejo da franciscana I, que compreende o litoral do Espírito Santo até o litoral sul do Rio de Janeiro; FMA-II, que compreende o litoral de São Paulo até o sul de Santa Catarina e que inclui

a Baía de Babitonga; FMA-III, compreendendo o litoral do Rio Grande do Sul até o Uruguai e a FMA-IV, compreendendo a faixa litorânea que vai do Rio da Prata até a província de Chubut ([Secchi et al., 2003]). Uma toninha adulta chega a medir entre 121 e 177cm de comprimento ([Crespo, 2009]) e apresenta dimorfismo sexual, sendo as fêmeas maiores que os machos ([Ramos et al., 2002]). A gestação dura em torno de onze meses, e a lactação nove ([Danilewicz, 2003]). Os nascimentos ocorrem predominantemente na primavera e verão (e.g. [Ramos et al., 2000]; [Danilewicz, 2003]). No sistema de acasalamento, o macho permanece junto à fêmea durante e após o seu período reprodutivo ([Wells et al., 2013]).

Devido ao hábito costeiro ([Pinedo et al. 1989]; [Cremer & Simoes-lobes, 2005]; [Danilewicz et al., 2009]), a toninha está em constante ameaça. Tal vulnerabilidade é decorrente das ações antrópicas diretas e indiretas. Atualmente, a espécie é considerada o pequeno cetáceo mais ameaçado de extinção no Oceano Atlântico Sul Ocidental, sendo listada como “em perigo” no “Livro Vermelho da Fauna Brasileira Ameaçada de Extinção” e “vulnerable” na “IUCN Red List of Threatened Species” ([Danilewicz et al., 2008]; [Reeves et al., 2012]).

No presente trabalho propomos desenvolver, testar e aplicar variações de modelos estatísticos hierárquico conjugando os métodos de transecção linear e dupla observação. Estes modelos estimaram simultaneamente a abundância de grupos e a probabilidade de detecção, cujo $g(0) < 1$, ou seja considerando o viés de percepção, porém, sem considerarmos o erro de medição. Fazendo uso da técnica de dados aumentados; do método de simulação de Monte Carlo via Cadeia de Markov (MCMC) e da abordagem bayesiana, obtivemos distribuições de probabilidade posteriores da estimativa de abundância e probabilidade de detecção da toninha, residente na Baía da Babitonga.

3.3 Hipóteses de Trabalho

- 1) Modelos hierárquicos usados para estimar abundância de populações, permitem ampliar DS convencional incorporando o efeito do viés de percepção por meio de submodelos na forma de regressões *logit* ou *probit*.
- 2) A estimativa simultânea da abundância e da probabilidade de detecção via modelos hierarquizados, que conjugam a traseção linear e a dupla observação; dados aumentados e simulação por MCMC, incorporam o viés de detecção.

3.4 Objetivo Geral

Desenvolver modelos estatísticos hierárquicos, integrando DS com MR, para estimar, por meio da abordagem bayesiana, conjuntamente as probabilidades de detecção observador-dependentes e o tamanho da população.

3.5 Objetivos Específicos

- Formular modelos hierárquicos para dados de amostragem por distância (DS), que incorporem a dupla observação (MR) e que prescindam da premissa de detecção perfeita em alguma distância fixada (geralmente sobre a linha - distância igual a zero);
- Avaliar o desempenho desses modelos com estudos simulados;
- Comparar o desempenho desses modelos com abordagens alternativas em dados que indicaram máxima probabilidade de detecção fora da linha de transecção (distância maior que zero) (estudo de caso I, Anexo I 9.1)
- Aplicar esses modelos, estimando a abundância de toninhas pertencentes a subpopulação da FMA-II e residentes na Baía de Babitonga, SC, Brasil (estudo de caso II, Anexo II 9.2).

4 METODOLOGIA

4.1 Os Modelos

Os modelos são apresentados numa estrutura hierárquica e recorrem à técnica de dados aumentados (*data augmentation*, [Royle & Dorazio, 2008], p.181), em que a população real de tamanho desconhecido N está contida numa super-população de tamanho conhecido M , escolhido de modo que seja razoável supor que $M \gg N$.

Considerando apenas a existência dos dados de distância, definimos y_i como uma variável binária, que indica se o i -ésimo animal ou grupo de animais foi detectado. A variável y_i é descrita por uma distribuição de Bernoulli com parâmetro μ_i .

$$y_i \sim \text{Bernoulli}(\mu_i)$$

Este parâmetro μ_i consiste em dois componentes: a probabilidade de detecção, $g(x_i)$ como uma função de distância perpendicular x_i , entre o i -ésimo indivíduo ou grupo de indivíduos e a transeção linear; e a variável binária w_i , indicando se o i -ésimo indivíduo ou grupo de indivíduos na "população aumentada (super-população)" de tamanho conhecido, M é parte da "população real" de tamanho desconhecido, N . Ou seja, $\mu_i = w_i \cdot g(x_i)$, onde a variável w_i tem distribuição Bernoulli com probabilidade $\psi = N \cdot M^{-1}$.

$$w_i \sim \text{Bernoulli}(\psi)$$

A abundância será $N = \sum(w_i)$, que tem distribuição binomial (M, ψ) . Portanto, estimar o valor esperado de N é equivalente a estimar ψ .

Como função de detecção, $g(x)$, $[0 \leq g(x) \leq 1]$ e $[0 \leq x \leq m]$ sendo m a máxima distância perpendicular a ser considerada, utilizamos:

a função inversa à função de ligação logit

$$g(x) = \frac{\exp(\beta_0 + \beta_1 \cdot x)}{1 + \exp(\beta_0 + \beta_1 \cdot x)} ;$$

e a função inversa à função de ligação probit

$$g(x) = \Phi(\beta_0 + \beta_1 \cdot x) \quad ;$$

Sendo Φ a função acumulada de uma distribuição Normal padrão.

Uma forma conveniente de expressar as funções logit e probit é definir $g(x) = p$ e reescrevê-las como função linear dos parâmetros, conforme mostrado na tab. 1.

A probabilidade de detecção sobre a linha de transecção (**on-line detection**) onde $x = 0$, é uma função de β_0 tanto para a função de ligação logit: $g(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$, quanto para a função probit: $g(0) = \Phi(\beta_0)$.

Com os dados de distância apenas, β_0 não pode ser estimado. A estimativa se torna possível com a incorporação da informação de dois observadores independentes e simultâneos, percorrendo a transeção ([Buckland et al., 2010]).

Com a inclusão das informações do duplo observador (DO) ao método hierárquico descrito acima (transeção linear - LT), a variável binária y_i é estendida para y_{ji} para o observador $j = 1, 2$ e a função de detecção torna-se observador-dependente $g_j(x_i)$. Com a incorporação das informações de marcação-recaptura ao i -ésimo indivíduo, este agora, estará associado ao vetor $Z_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i})$ em que apenas um valor é igual a 1 e os demais são iguais a zero. Assim sendo, quando $z_{1i} = 1$ a avistagem foi exclusivamente do observador 1 e o vetor associado é $Z_i = (1, 0, 0, 0)$; quando $z_{2i} = 1$ a avistagem foi exclusivamente do observador 2 e $Z_i = (0, 1, 0, 0)$; quando $z_{3i} = 1$ a avistagem foi feita por ambos e $Z_i = (0, 0, 1, 0)$; finalmente, $z_{4i} = 1$ e $Z_i = (0, 0, 0, 1)$ para animais que nenhum observador avistou.

Considerando que: 1) $g_j(x_i) = p_{ji}$ é a probabilidade de que o j -ésimo observador detecte o animal i ; e 2) que as detecções são independentes entre observadores (i.e. $p_{ji} = p_{ji|(3-j)i}$ com o termo à direita da igualdade denotando a probabilidade condicional de que j detecte i dado que $3 - j$ tenha detectado i , para $j = 1, 2$). Então, as probabilidades dos 4 componentes de Z_i são, respectivamente:

$$\begin{aligned}
Z_i &\sim \text{Multinomial}(1, (\mu_{1i}, \mu_{2i}, \mu_{3i}, \mu_{4i})) \\
\mu_{1i} &= g_1(x_i) \cdot (1 - g_2(x_i)) \cdot w_i \\
\mu_{2i} &= (1 - g_1(x_i)) \cdot g_2(x_i) \cdot w_i \\
\mu_{3i} &= g_1(x_i) \cdot g_2(x_i) \cdot w_i \\
\mu_{4i} &= 1 - (\mu_{1i} + \mu_{2i} + \mu_{3i}) \\
w_i &\sim \text{Bernoulli}(\psi)
\end{aligned}$$

Onde $Z \sim \text{Multinomial}(s, (p_1, \dots, p_k))$ denota uma distribuição multinomial k -dimensional com uma amostra de tamanho s e vetor probabilidade $(p_1, \dots, p_k)$ que satisfaz $\sum p_r = 1$.

As funções logit e probit com ajustes individuais para cada observador ($j = 1, 2$) ficam conforme expressado nos modelos 5 e 6 na Tab. 1.

Table 1: Lista dos modelos ajustados para apenas dados de distância (LT) e para dados complementados com a dupla observação da marcação e recaptura (DOLT). A função de ligação logit é representada pela numeração ímpar e a probit pela numeração par. O termo "single" refere-se a modelos com uma única função de detecção enquanto "double" refere-se a modelos com funções de detecção observador-específico. $g(x_i) = p_i$

	Modelos	Descrição
M1	$\log \left[\frac{p_i}{(1-p_i)} \right] = \beta_0 + \beta_1 \cdot x_i$	<i>single LT – logit</i>
M2	$\Phi^{-1}(p_i) = \beta_0 + \beta_1 \cdot x_i$	<i>single LT – probit</i>
M3	$\log \left[\frac{p_{ji}}{(1-p_{ji})} \right] = \beta_0 + \beta_1 \cdot x_i$	<i>single DOLT – logit</i>
M4	$\Phi^{-1}(p_{ji}) = \beta_0 + \beta_1 \cdot x_i$	<i>single DOLT – probit</i>
M5	$\log \left[\frac{p_{ji}}{(1-p_{ji})} \right] = \beta_{j0} + \beta_{j1} \cdot x_i$	<i>double DOLT – logit</i>
M6	$\Phi^{-1}(p_{ji}) = \beta_{j0} + \beta_{j1} \cdot x_i$	<i>double DOLT – probit</i>
M7	$\log \left[\frac{p_{ji}}{(1-p_{ji})} \right] = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2$	<i>single quad DOLT – logit</i>
M8	$\Phi^{-1}(p_{ji}) = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2$	<i>single quad DOLT – probit</i>
M9	$\log \left[\frac{p_{ji}}{(1-p_{ji})} \right] = \beta_{j0} + \beta_{j1} \cdot x_i + \beta_{j2} \cdot x_i^2$	<i>double quad DOLT – logit</i>
M10	$\Phi^{-1}(p_{ji}) = \beta_{j0} + \beta_{j1} \cdot x_i + \beta_{j2} \cdot x_i^2$	<i>double quad DOLT – probit</i>

4.2 Os dados

Utilizamos três conjuntos de dados. O primeiro foi simulado e usado com o propósito de avaliar a performance da inferência. O segundo é um conjunto de dados reais previamente analisados em [Quang & Becker, 1997]. O terceiro conjunto de dados, são dados inéditos, coletados na baía da Babitonga.

Para o estudo simulado consideramos: 1) uma população de $N = 200$ indivíduos cujas distâncias perpendiculares a linha de transeção x_i , são variáveis aleatórias independentes com distribuição uniforme no intervalo $[0, 1]$. A distância ($x = d$), que é geralmente medida considerando uma distância máxima a , pode sempre ser padronizada tal que $x = d \cdot a^{-1}$ esteja no intervalo $[0, 1]$; 2) curvas de detecção, $g_j(x_i) = k_j \cdot \exp(-x^2 \sigma_j^{-2})$ meia-Normal distintas com $k_1 = 0.8$; $\sigma_1 = 8$, para o observador 1 e $k_2 = 0.5$; $\sigma_2 = 12$ para o observador 2. Estes valores foram obtidos após 142 combinações entre os parâmetros k e σ , de modo que a declividade destas curvas de detecção refletissem o melhor possível o que ocorre em campo ([Borchers et al., 2002]; [Laake & Borchers, 2004]). Para o i -ésimo indivíduo simulado a detecção pelo observador j é modelada como uma variável Bernoulli $y_{ji} \sim \text{Bern}(g_j(x_i))$. Cada indivíduo foi classificado como detectado ($y_{ji} = 1$) ou não-detectado ($y_{ji} = 0$) pelo observador j . Assim sendo, os componentes do vetor Z_i são definidos para o i -ésimo indivíduo conforme segue:

$$\begin{aligned} z_{1i} &= y_{1i} \cdot (1 - y_{2i}) \\ z_{2i} &= (1 - y_{1i}) \cdot y_{2i} \\ z_{3i} &= y_{1i} \cdot y_{2i} \\ z_{4i} &= (1 - y_{1i}) \cdot (1 - y_{2i}) \end{aligned}$$

A amostra é constituída do número total de $n = \sum_{i=1}^N (z_{1i} + z_{2i} + z_{3i})$ indivíduos para os quais são retidas as distâncias x_i e o vetor Z_i .

O procedimento de geração de amostra foi repetido mais 19 vezes, gerando assim 20 réplicas (transecções), (t_r) ; sendo $r = 1, 2, \dots, 20$. As distribuições de frequências das distâncias individuais e o número total de indivíduos detectados nas amostras estão sumarizadas na Fig. 1.

O segundo conjunto de dados refere-se a uma população composta por duas espécies de gansos (*Gavia pacifica* e *Gavia immer*), popularmente chamados de gansos Loon, no Parque Nacional e refúgio da vida silvestre *Yukon Flats*, no Alaska. Foi adaptado de [Quang & Becker, 1997]. Para que trabalhássemos com distâncias individuais, simulamos por meio de uma distribuição Uniforme, distâncias a partir dos intervalos de classe fornecidos pelos autores (Tab. 2). Além das distância, os dados também contêm as informações dos dois observadores, decorrentes da marcação e recaptura. Comparamos nossas inferências às estimativas de abundância reportadas por [Quang & Becker, 1997].

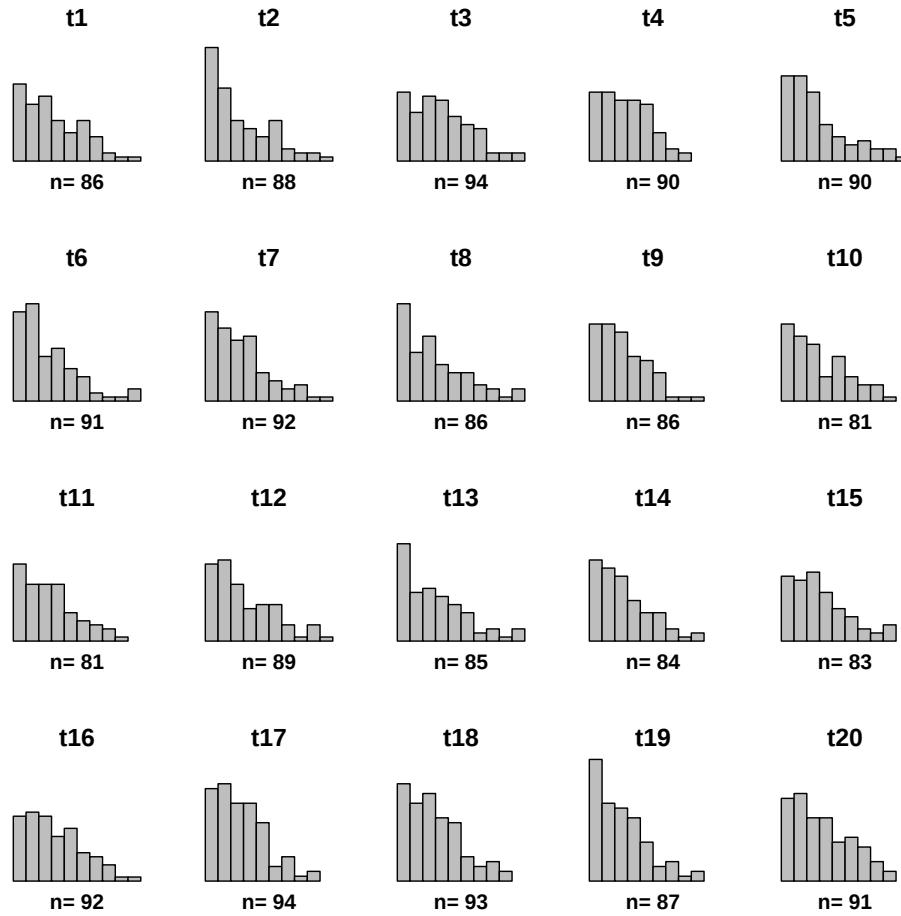


Figure 1: Histograma das distâncias pela frequência. Abaixo de cada histograma está o tamanho n para as 20 amostras replicadas, extraídas da população simulada. Todos os histogramas estão na mesma escala. No eixo x estão as distâncias perpendiculares entre os indivíduos e as transecções, padronizadas entre $[0, 1]$. No eixo y estão as frequências das detecções. Aqui, chamamos a atenção para os variados perfis dos histogramas que foram gerados pela mesma função de detecção, meia-Normal, sempre fazendo uso dos mesmos parâmetros. (Anexo I 9.1).

O terceiro conjunto de dados refere-se ao estoque de toninhas da Baía da Babitonga, foi coletado em 2011 por meio de sobrevoos. A aeronave voou a uma velocidade aproximada de 170-190 km/h e a uma altitude entre 150-152 metros. Quatro observadores com independência auditiva e visual foram utilizados nas mesmas posições durante todo o período de coleta de dados. A informações coletadas foram: local, latitude, longitude, transparência, reflexo no campo de visão, estado do mar ("beaufort"), ângulo de declinação entre o horizonte e o grupo avistado; presença de filhote e tamanho de grupo. Os ângulos foram obtidos por meio de clinômetros. Posteriormente, de posse do ângulo e da altura de voo, foi possível calcular as distâncias perpendiculares entre a transecção e os grupos observados ([Lerczak & Hobbs, 1998]). Os dados eram registrados em gravador digital e GPS sincronizados com os relógios dos quatro observadores ([Perez, 2015]).

Após o estabelecimento das avistagens simultâneas ([Perez, 2015]), elaboramos uma matriz

Table 2: Dados de detecção do Loon publicados por [Quang & Becker, 1997], Tabela 1. Esta tabela foi transcrita de acordo como publicada pelos autores.

Classe de Distância (metros)	5-30	30-60	60-90	90-120	120-150	150-190	190-250
<i>Front passenger only</i>	3	1	2	3	3	5	5
<i>Rear passenger only</i>	4	3	0	3	5	1	2
<i>Both passengers</i>	0	1	2	10	6	6	2

composta das distâncias x_i ; do vetor $Z_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i})$ de comprimento quatro e da variável indicadora w_i , para $i = 1, 2, \dots, M$. O vetor Z_i indica a forma de detecção de um grupo de animais, como já descrito acima no ítem 2.1.

4.3 Ajuste dos Modelos

Os modelos de inferência se distinguem quanto ao uso da informação exclusivamente de distâncias (LT) ou complementadas com dados de dupla observação (Marcação e Recaptura) (DOLT, [Buckland et al., 2010]), além da função de ligação logit ou probit para a transformação das probabilidades de detecção (Tab. 1). Passaremos a nos referir aos modelos pela nomenclatura dada na primeira coluna da tabela 1. No caso dos modelos LT utilizamos o total de detecções n obtido por ambos como sendo apenas um observador, ignorando a informação de qual grupo foi visto pelo primeiro ou pelo segundo observador. Isto permite comparar o impacto sobre a precisão das estimativas causadas pela informação complementar do segundo observador mantendo o mesmo tamanho amostral.

Para os estudos de caso I e II, incluímos nos modelos a possibilidade de um termo quadrático para estimar uma possível máxima probabilidade de detecção a uma distância $x > 0$ (modelos 7 a 10, Tab. 1).

4.4 Inferência Bayesiana

4.4.1 Dados Simulados

Ajustamos os modelos de 1 a 6 (Tab. 1) às 20 réplicas de dados simulados. Usamos prioris não informativas para os parâmetros: $\psi \sim U(0, 1)$; $\beta_{j0} \sim N(0, 1E5)$ e $\beta_{j1} \sim N(0, 1E5)$ para $j = 1, 2$. As distribuições posteriores foram obtidas por meio do método de simulação de Monte Carlo com Cadeia de Markov (MCMC) ([Martin & Quinn, 2006]; [McCarthy, 2007]). Após efetuarmos vários testes preliminares, padronizamos os ajustes trabalhando com três cadeias, um burn-in de 10.000 e

um thinning de 40, visando eliminar a dependência entre os elementos da cadeia e seu subsequente, gerando uma amostra posterior de tamanho 3000. Para obter as estimativas e diagnósticos, usamos as bibliotecas "R2jags" ([Su & Yajima, 015]) "rjags" ([Plummer, 2013]) e "coda" ([Plummer et al., 2010]; [Plummer, 2013]). As bibliotecas "R2jags" e "rjags" têm como algoritmo de desenvolvimento da cadeia o JAGS - Just Another Gibbs Sampler ([Plummer, 2003]; [Plummer, 2012]). Tais bibliotecas foram executadas no aplicativo R (version 3.1.1) ([R Core Team, 2014]).

Para avaliar a eficiência do procedimento inferencial na estimativa do tamanho da população N e dos valores de $g(0)$ e $g_j(0)$ para o observador $j = 1, 2$, utilizamos o erro quadrado médio entre réplicas, calculando a variância entre as médias posteriores somado a diferença quadrática entre o valor médio das médias posteriores e o parâmetro real (raiz quadrada do erro quadrático médio - rmse) ([Bussab & Morettin, 2002]).

Finalmente, com o propósito de comparar a qualidade preditiva dos modelos ajustados para cada uma das 20 réplicas, utilizamos o critério de informação da *deviance* - DIC ([Spiegelhalter et al., 2002]; [Celeux et al., 2006]; [Plummer et al., 2010]). Desta maneira pode-se avaliar se há consistência na seleção do melhor modelo, comparando-se esta escolha entre as 20 réplicas.

4.4.2 Dados do Ganso Loon - Estudo de caso I

Ajustamos os modelos de 7 a 10 (Tab. 1) ao dados do loon ([Quang & Becker, 1997]). Usamos *prioris* conforme o item 4.4.1. As distribuições posteriores foram obtidas por meio do método de Monte Carlo com Cadeia de Markov (mcmc) conforme já descrito em Dados Simulados.

Para diagnosticar a convergência da cadeia, foi utilizado Rhat, n.eff ([Spiegelhalter et al., 2002]; [Celeux et al., 2006]; [Plummer et al., 2010]). Trabalhamos com três cadeias simultâneas, burn-in, thinning e amostra posterior, conforme já descrito em Dados Simulados, item 4.4.1.

4.4.3 Análise de Resíduo - Qualidade do Ajuste

Após os ajustes dos modelos ao dados do ganso loon, procedemos com a análise da qualidade do ajuste. Nosso objetivo é saber se, com os modelos propostos e os parâmetros estimados destes modelos, seríamos capazes de gerar valores preditivos em consonância com os dados amostrais observados ([Gelman & Hill, 2007], p.513). Ou seja, trata-se, por exemplo, de comparar os dados de captura Z , supostamente originários de uma distribuição multinomial, com a distribuição preditiva de dados gerados a partir do modelo hierárquico proposto com as estimativas dos parâmetros pela

distribuição posterior. Maiores detalhes acerca da análise preditiva estão descritas no artigo 1, anexo I.

4.4.4 Dados da Toninha - Estudo de caso II

Para o segundo estudo de caso, das toninhas na Baía da Babitonga, que só ocorreu após as conclusões do estudo de caso I, selecionamos os modelos de M5 a M10, que fizeram uso da informação da distância e do duplo observador (DOTL) (Tab. 1). Com base no DIC (Anexo I 9.1), descartamos os modelos de M1 a M4. Os modelos M1 e M2 fizeram uso apenas da distância (TL). Já os modelos M3 e M4 fizeram uso da informação do duplo observador, usando uma única curva de detecção para ambos (Tab. 1).

As n distâncias $x_i = (d_i - d_{min})(d_{max} - d_{min})^{-1}$ foram usadas para o ajuste dos modelos (Tab. 1) de duas formas distintas: 1) dados não truncados [$d_{min} = 0, d_{max} = 286$]. Aqui não houve restrições à esquerda; e 2) dados truncados [$d_{min} = 87, d_{max} = 286$]. Houve a eliminação de observações com distâncias inferiores a 87m. A faixa de distância [0-87m] representa uma "área de sombra" abaixo da aeronave, que, teoricamente, seria cega ao segundo observador, devido a janela ser plana. A distância de 286m, delimitaria o início de uma faixa com aumento no erro de medição ([Borchers et al., 2002]; [Laake & Borchers, 2004]) e a partir da qual não houveram detecções. Apesar de não haver restrições no uso das distâncias em nossos modelos, incorporamos à análise os dados truncados para comparações.

Os modelos aqui empregados, não fazem uso da taxa de recaptura. Todavia, é possível obtê-la para cada distância perpendicular, a partir das curvas de detecção estimadas. A taxa de recaptura para a distância x_i , ($rr(x_i)$) é definida como "a probabilidade de ambos observadores detectarem um grupo de animais dado que este grupo foi detectado", ou seja:

$$rr(x_i) = Pr(z_{3i} = 1 | [z_{1i} + z_{2i} + z_{3i}] = 1, x_i) = \frac{p_{1i} \cdot p_{2i}}{p_{1i} + p_{2i} - p_{1i} \cdot p_{2i}}$$

Quanto à inferência, procedemos conforme descrito nos itens 4.4.1 e 4.4.2. Na comparação entre o melhor modelo para dados truncados com o melhor modelo para dados não truncados, utilizamos os coeficientes de variação (cv) das distribuições posteriores ([Bussab & Morettin, 2002]).

5 RESULTADOS

5.1 Dados Simulados

Todos os modelos foram ajustados com as funções logit e probit. No entanto, com base no *DIC* não houve uma diferença superior entre uma e outra função. Portanto, vamos nos concentrar na descrição apenas da ligação logit, mantendo nas tabelas e figuras os resultados de ambas para fins de comparação e avaliação por parte do leitor. Resaltamos o fato de que, conforme a distribuição de probabilidade da variável resposta é possível usar logit ou probit para melhor ajuste aos dados.

A estimativa de abundância fazendo uso apenas da informação da distância (M1) tem um desvio padrão posterior médio de 68, que reduz para 31 quando a informação do segundo observador é incluída com ambos os observadores fazem uso apenas de uma única curva de detecção (M3). O desvio padrão reduz ainda mais, para 21, quando curvas de detecção específica para cada observador são consideradas (M5) (ver Anexo I 9.1). Todavia usando a média posterior das 20 réplicas nós observamos (*rb*), que os modelos M1 e M3 são positivamente enviesados (9%) em relação ao valor real enquanto o modelo M5 é negativamente enviesado (12%) em relação ao valor real (Anexo I 9.1). O custo-benefício entre viés e precisão é melhor descrito pela raiz quadrada do erro quadrado médio (*rmse*) calculado a partir das médias posteriores replicadas (Tab. 3). Baseado no menor *rmse* (33,76), o modelo M5 é o melhor modelo entre os três modelos considerados. Finalmente, em relação ao intervalo de credibilidade de 95% (CrI95) o parâmetro real N foi coberto por 16 das 20 réplicas para ambos os modelos DOLT, M3 e M5 (Fig. 2). A cobertura está abaixo do esperado de 19 para 20.

A comparação dos modelos por intermédio do menor *DIC*, consistentemente identificou o modelo M5, com funções de detecção observador-específico, como melhor do que o modelo M3, que assume uma curva única para ambos os observadores (Anexo I 9.1). Esta consistência também é percebida para os modelos que fizeram uso da função probit (M6 melhor do que M4).

Todos os modelos DOLT tiveram um bom desempenho na estimativa do $g(0)$. Para o modelo M5, esta estimativa teve um viés positivo de 5% e de 8% para o primeiro e segundo observadores respectivamente (Anexo I 9.1). Consequentemente, apresentando um menor viés do que para sua estimativa de abundância descrita anteriormente. O modelo M3 que assume uma curva única de detecção, também é bastante eficiente em estimar a média entre os valores reais (0,8 e 0,5) para

ambos os observadores, que é de $g(0) = 0,65$ (Anexo I 9.1). Em relação ao intervalo de credibilidade de 95%, 17 e 19 das 20 réplicas são as taxas de sucesso para o primeiro e segundo observador no modelo M5 (Anexo I 9.1).

Estimativa da Abundância dos Modelos M1 à M6

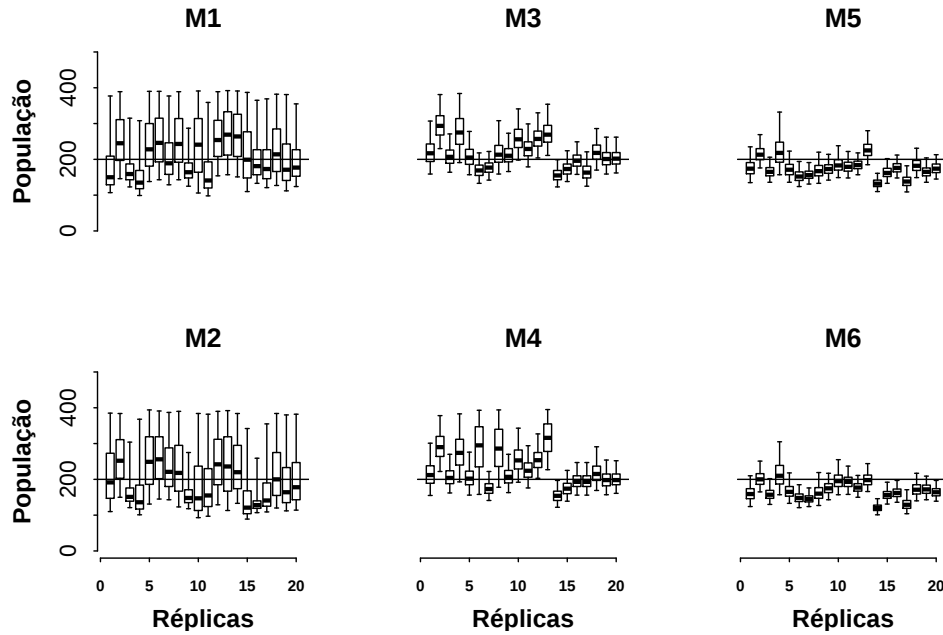


Figure 2: Box-plot-personalizado resumindo a distribuição posterior da abundância (N) para os modelos M1 até M6. CrI50% (*boxes*), CrI95% (*whiskers*), mediana (traço no interior dos *boxes*), parâmetro real do tamanho da população (linha horizontal). (Anexo I 9.1).

Table 3: Resumo das 20 distribuições posteriores da abundância N . Média (mean); desvio padrão entre as médias posteriores (sdpm); raiz quadrada do erro quadrado médio (rmse) e o DIC médio entre as réplicas (DIC_m).

	M1	M2	M3	M4	M5	M6
mean	217.264	206.169	217.003	226.419	176.464	169.809
sdpm	37.134	39.866	39.544	47.167	24.197	23.927
rmse	40.951	40.340	43.044	54.062	33.755	38.523
DIC_m	3185	3924	909	905	808	793

Table 4: Resumo das 20 distribuições posteriores de $g(0)$. Média (mean); variância entre a média das posteriores (vpm); desvio padrão entre as médias posteriores (sdpm); raiz quadrada do erro quadrado médio; média do viés relativo (rb) e o DIC médio entre as réplicas (DIC_m).

	M3	M4	M5		M6	
			g1(0)	g2(0)	g1(0)	g2(0)
mean	0.681	0.651	0.838	0.540	0.834	0.530
sdpm	0.085	0.091	0.073	0.102	0.080	0.099
rmse	0.090	0.091	0.083	0.109	0.087	0.103
rb	0.047	0.001	0.047	0.079	0.042	0.060
DIC_m	909	905	808		793	

5.2 Loon - Estudo de Caso I

Todos os modelos quadráticos (DOLT) listados na Tabela 1 foram ajustados aos dados do Loon. Os modelos com menor DIC são M8 e M10. No entanto, ambos os modelos não apresentam diferenças entre si em relação ao desempenho preditivo, uma vez que seus DIC's diferem apenas em uma unidade (Anexo I - Tab.3). Na verdade, as diferenças na abundância estimada, N são insignificantes uma vez que apresentam estimativas de 156 (CrI95: 128 a 189) e 154 (CrI95: 125-185) para os modelos M8 e M10, respectivamente. No entanto, estas estimativas excedem em cerca de 36% [1-(99/154)] a estimativa de 99 indivíduos obtidos por [Quang & Becker, 1997] (QB). Para mais comparações com os resultados de QB, vamos nos concentrar no modelo M10, que faz uso da função de detecção observador-específico como eles fizeram.

No modelo M10 as probabilidades de detecção para cada observador na distância zero ($g_1(0)$, $g_2(0)$) e as máximas probabilidades de detecção ($g_1(x_0)$, $g_2(x_0)$) são mais elevadas do que as correspondentes estimativas relatadas por [Quang & Becker, 1997] (Anexo I, Tab.5 e Fig. 3 c e d). Quanto aos desvios padrão posteriores para as máximas probabilidades de detecção sobre a transecção, são semelhantes aos erros padrão assintóticos (ASE) apresentados por QB. Já os valores dos desvios padrão posterior para as máximas probabilidades de detecção são muito menores. Eles são cerca de 0.05, enquanto o ASEs estão acima de 0.25 (Tab. 4).

5.3 Toninha - Estudo de Caso II

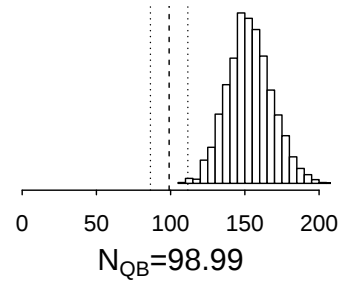
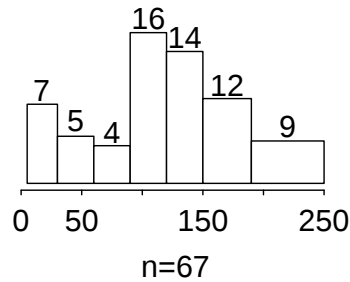
Os dados observados estão sumarizados nas Tabela 5. Dados truncados representam 104 grupos observados e dados não truncados, 140. O truncamento reduziu em 26% o tamanho da amostra.

Modificado de Quang & Becker, 1997

M10

QB'Data

N_{QB}



$g_1(x_0)$

$g_2(x_0)$

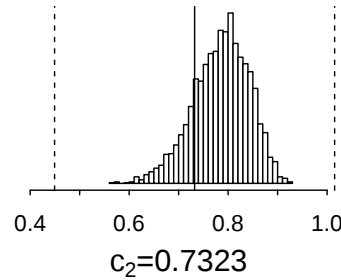
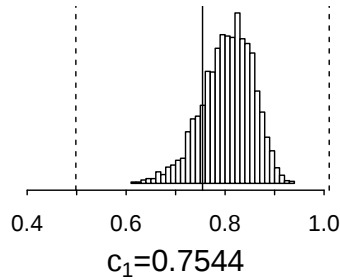


Figure 3: Canto superior esquerdo: histograma dos dados observados por QB; Canto superior direito: distribuição posterior da abundância estimada e a estimativa correspondente, N_{QB} com o intervalo de confiança estimado por QB [linhas verticais]; Canto inferior esquerdo: distribuição posterior da maior probabilidade de detecção para o observador 1 [$g_1(x_0)$] e a estimativa deste parâmetro (c_1) feita por QB [linha vertical]; Canto inferior direito: distribuição posterior da maior probabilidade de detecção para o observador 2 [$g_2(x_0)$] e a estimativa deste parâmetro (c_2) feita por QB [linha vertical]. Todas as distribuições posteriores, referem-se ao modelo M10. (Anexo I 9.1).

É relevante observar, que os dados não truncados não se comportam a semelhança de uma curva half-Normal ou Hazard rate (Fig. 4), ambas comumente usadas no *Distance Sampling* convencional ([Borchers et al., 2002]; [Laake & Borchers, 2004]), juntamente com a premissa de que a máxima probabilidade de detecção se dê à distância zero.

Table 5: **Dados de Campo.** Número de grupos observados pelo observador 1, (Obs1); pelo observador 2, (Obs2); por ambos os observadores, (Ambos); taxa de recaptura total, (rr); frequência (de 1 a 6) e média (mgs) do tamanho de grupo para dados truncados e dados não truncados.

Dado	Obs1	Obs2	Ambos	rr	1	2	3	4	5	6	mgs
Truncado	48	46	10	0.0962	36	33	26	5	3	1	2.13
não Truncado	80	49	11	0.0786	49	43	37	6	4	1	2.11

Para dados truncados, o melhor modelo foi o M6 (M2 no Anexo II 9.2), que apresentou o menor

Curva de Detectabilidade

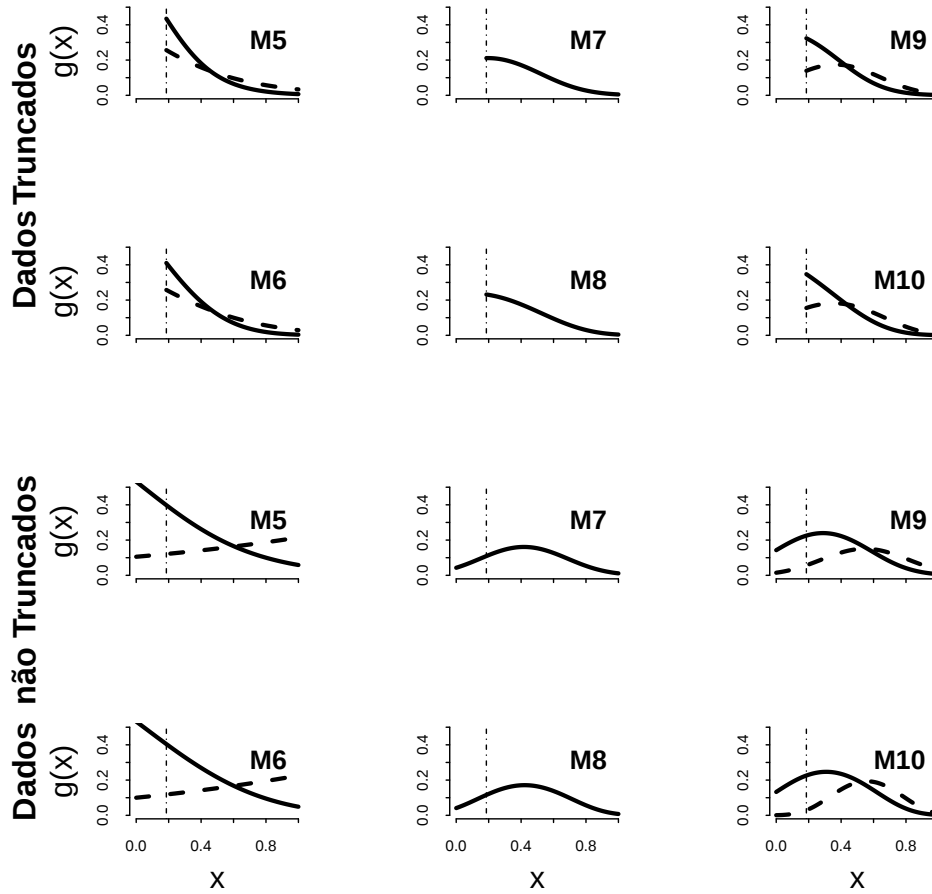


Figure 4: Curva de detectabilidade estimada pelos modelos M1 até M6. A linha contínua representa o observador 1 ou curva única para ambos, e a linha tracejada representa o observador 2. A linha vertical representa o truncamento (87m). No eixo x estão as distâncias padronizada $[0,1]$ para cada grupo observado e y as probabilidades de detecção $(g(x))$ (Anexo II 9.2).

DIC (Tab.3 - Anexo II). Devido à ausência do termo quadrático em M6, a maior probabilidade de detecção se encontra sobre a linha de transecção, porém com valores diferentes para cada observador. Com o truncamento dos dados a linha de transecção está a 87 m da derrota da aeronave (Fig. 4). Quanto ao viés de percepção, estimado por esse modelo, é de 0.586 $(1 - g_1(0))$, para o observador 1 e 0.737 $(1 - g_2(0))$, para o observador 2 (Tab.3 - Anexo II 9.2).

Já para dados não truncados, o melhor modelo foi o M10 (M6 no Anexo II 9.2), que inclui o termo quadrático (Tab.3; Fig.3 - Anexo II 9.2). Por consequência, ele estima as máximas probabilidades de detecção $(g_1(x_0); g_2(x_0))$ a distâncias maiores que zero. A distância correspondente à média de $g_1(x_0)$ e $g_2(x_0)$ é de 122,48 m (CrI95: 100,13 - 136,96). O viés de percepção estimado pelo modelo M10 foi de 0,745 $(1 - g_1(x_0))$ para o observador 1 e 0,801 $(1 - g_2(x_0))$ para o observador 2 a uma distância média perpendicular de 122,48 m da linha de transecção.

A densidade estimada para o melhor modelo com dados truncados (M6, M2 no Anexo II 9.2)

Estimativa da Curva da Taxa de Recaptura

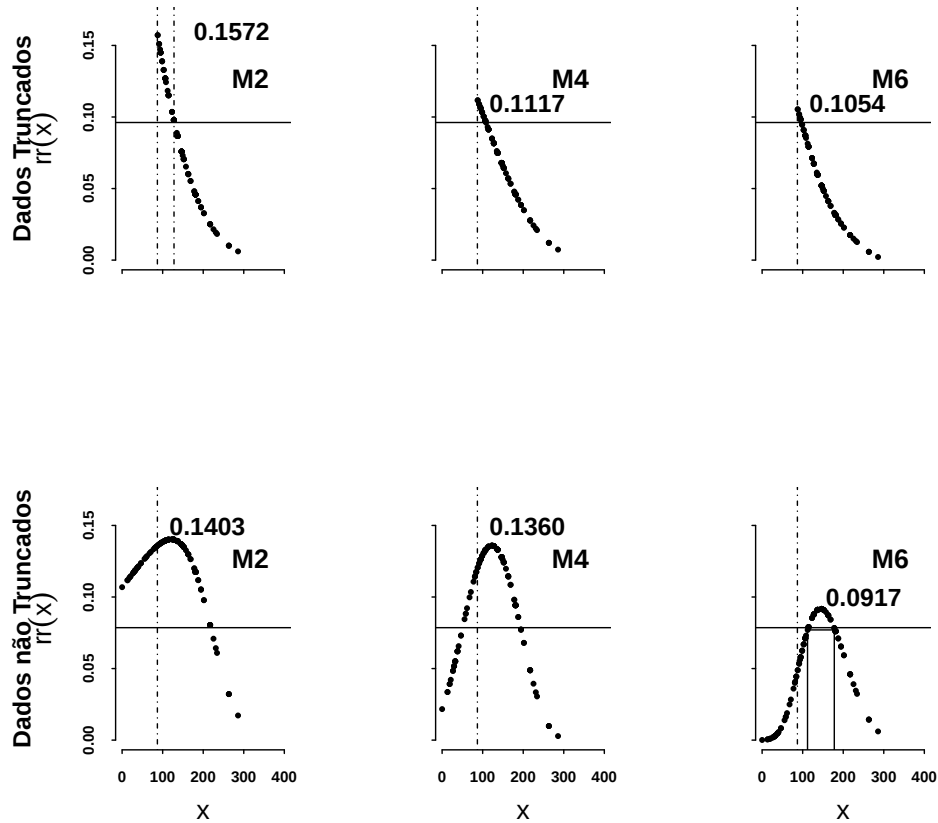


Figure 5: **Estimativa da Curva da Taxa de Recaptura.** Frequência de grupos observados pela distância (metros) para dados truncados e dados não truncados. A linha tracejada vertical representa o truncamento em 87 m e a linha contínua horizontal representa a taxa de captura obtida com os dados de campo. Para dados truncados $rr = 0,0962$ e para dados não truncados, $rr = 0,0786$. O valor no ponto máximo de cada curva representa a taxa de recaptura máxima para cada modelo. O retângulo vertical que aparece no modelo M6, representa a área de maior probabilidade de detecção de animais (entre 112 m e 178 m), onde se encontra a distância média ideal para o truncamento, estimada em 123 m. (Anexo II 9.2).

resultou em 0,950 grupos por km^2 (ICr95: 0,595 a 1,610). Já para o melhor modelo com dados não truncados (M10, M6 no Anexo II 9.2) a estimativa foi 1,219 (ICr95: 0,739 a 2,075). A pequena diferença entre as estimativas de densidade e a semelhança dos cv dos dois modelos não permitiu identificar um dos dois modelos como sendo claramente superior ao outro (Tab.4 - Anexo II 9.2).

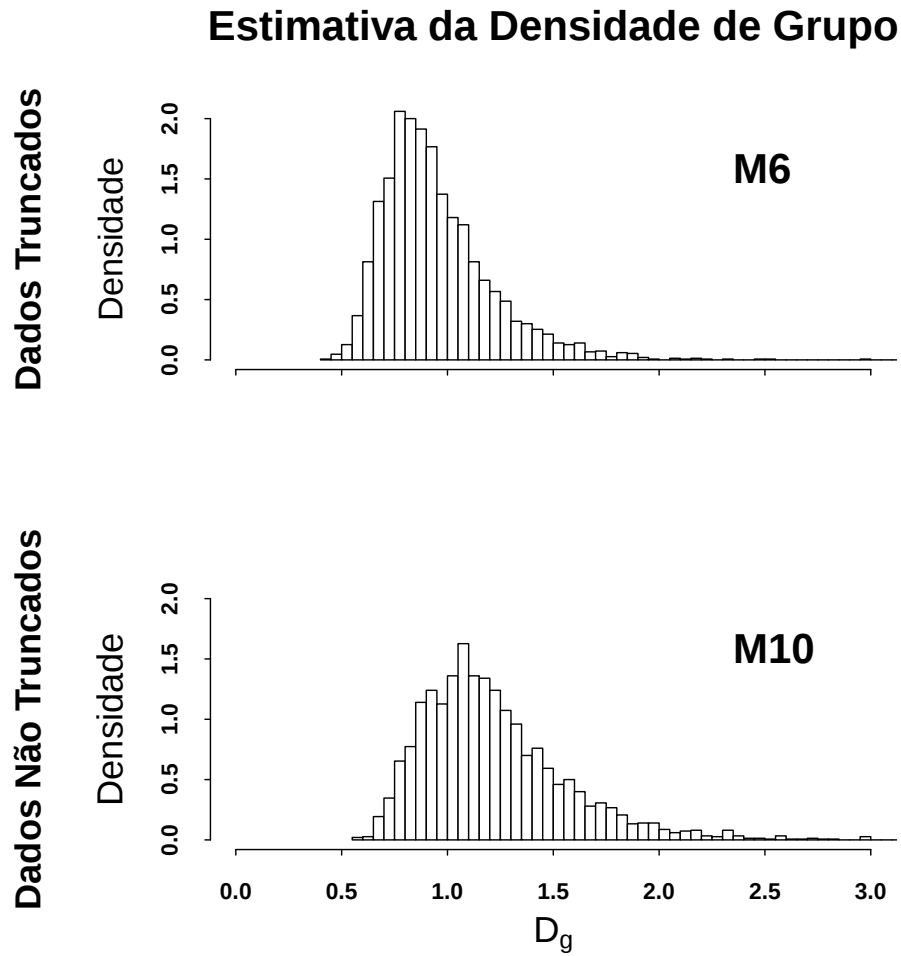


Figure 6: **Estimativa da densidade de grupos.** Figura superior representa dados truncados. Figura inferior, dados não truncados. Histograma da distribuição posterior da estimativa de densidade para toninhas na Baía da Babitonga. (Anexo II 9.2).

6 CONCLUSÃO

6.1 Dados Simulados

O estudo de simulação mostrou que o modelo hierárquico proposto, que incorpora dados de distância e dupla observação e sua implementação com o aumento de dados é uma forma viável e relativamente fácil para modelar dados de distância com probabilidade de detecção imperfeita. O estudo ainda mostrou que, apesar da não detecção de animais, houve uma redução da variabilidade, melhorando a estimativa. De forma geral, não conseguimos concluir acerca do viés apresentado na estimativa de abundância. talvez uma explicação do fato, seja o aumento da precisão penalizando a estimativa de abundância.

Dados de distância com dupla observação quando modelados por uma estrutura hierárquica com regressão logit ou probit em função da distância são capazes, numa abordagem bayesiana, de estimar o viés de percepção observador-dependente mesmo quando os dados são gerados por outro modelo.

6.2 Estudo de Caso I

O estudo da estimativa de abundância do ganso loon mostrou que os modelos estimam uma detecção máxima fora da linha de transecção com a inclusão de um termo quadrático. Com esta extensão, a perda de dados devido à truncagem pode ser evitada. Isto pode representar uma vantagem importante para as populações esquivas, onde as amostras são difíceis de serem obtidas e cada dado contém informações valiosas.

A inclusão de um termo quadrático nas regressões logit ou probit permitem identificar a distância de máxima detecção que não esteja sobre a linha de transecção (distância > 0), tornando desnecessário a premissa convencional de que seja sobre a linha.

6.3 Estudo de Caso II

O ajuste dos melhores modelos indica curvas de detecção distintas para cada observador tanto para dados truncados quanto não truncados. Apesar de tal fato apenas ser uma constatação, consequência de nossos resultados, pode estar indicando a precisão dos modelos que considerem curvas de detecção observador-dependente. A partir do ajuste efetuado sem truncamento, verificou-se que a distância média ideal para truncamento (estimada em 123 m) é superior àquela que foi utilizada de modo *ad hoc* (87 m).

Finalmente, a densidade estimada em número de grupos por km^2 , ($N_g \cdot \text{km}^{-2}$) é 1.219 com um intervalo de 95% de credibilidade, estendendo-se de 0.739 a 2.075 (Fig. 6).

A distância de máxima probabilidade de detecção é observador-dependente, sendo menor para o observador 1, que usou a janela bolha, e maior para o observador 2, que usou a janela plana.

A flexibilidade do modelo hierárquicos propostos permite estimar densidade da população sem que haja necessidade de qualquer tipo de truncamento dos dados. O que evita o descarte de informações, obtidas muitas vezes, por intermédio de logística onerosa.

7 CONSIDERAÇÕES FINAIS

Não encontramos trabalhos, que abordassem o tema abundância de mamíferos marinhos; fizesse uso dos dados de distância de forma integral e cuja distribuição de frequência não correspondesse a uma *half-Normal* ou *Hazard rate*. Nossa intenção era comparar os resultados obtidos com as estimativas feitas pelos modelos aqui propostos. Provavelmente, porque não haja na literatura trabalhos publicados com esse perfil. Isso demonstra um dos aspectos do caráter inédito de nosso trabalho.

Não conseguimos implementar, como a técnica dos dados aumentados, covariáveis, cujos parâmetros da distribuição de probabilidade fossem independentes dos dados coletados em campo. Trabalhos futuros, considerando a relevância do tema, "dados aumentados", podem propor outra abordagem para a incorporação de covariáveis ambientais.

A contribuição que trazemos para o contexto oceanográfico diz respeito a simplificação. Neste aspecto, nosso trabalho permitiu uma fácil compreensão e flexibilização dos modelos, que outras propostas não contemplam. Possibilitando estimar distância de máxima probabilidade de detecção, taxa de recaptura e viés de percepção. Além de evitar a eliminação de dados, resultando no uso integral da informação oriunda de campanhas oceanográficas de logística financeiramente dispendiosas.

8 REFERÊNCIAS

References

- [Andriolo et al., 2010] Andriolo, A., Kinas, P.G., Engel, M.H., Martins, C.C.A., Rufino, A.M. 2010. Humpback whale within the brazilian breeding ground: distribution and population size estimate. *Endangered Species Research* 11:233-243. doi:10.3354/esr00282.
- [Bailey et al., 2007] Bailey, L.L., Hines, J.N.J., MacKenzie, D. 2007. Sampling design trade-offs in occupancy studies with imperfect detection: examples and software. *Ecological Applications* 17(1):281-290.
- [Beasley & Jeffersom, 2002] Beasley, I., Jefferson, T.A. 2002. Surface and dive times of finless porpoises in Hong Kong's coastal waters *Neophocaena phocaenoides*. *Raffles Bulletin of Zoology Supplement* 10:125-129.
- [Becker & Quang, 2009] Becker, E., Quang, P. 2009. A Gamma-Shaped Detection Function for Line-Transect Surveys With Mark-Recapture and Covariate Data. *Journal of Agricultural, Biological and Environmental Statistics* 14(2):207-223. doi:10.1198/jabes.2009.0013.
- [Berliner, 1996] Berliner, L. 1996. "Hierarchical Bayesian time series models", *Maximum Entropy and Bayesian Methods* 15-22.
- [Borchers et al., 2002] Borchers, D.L., Bukland, S., Zucchini, W. 2002. Estimating animal abundance. Great Britain: Springer. 314p.
- [Borchers et al., 2006] Borchers, D.L., Laake, J.S., Southwell, C., Paxton, C. 2006. Accommodating Unmodeled Heterogeneity in Double-Observer Distance Sampling Surveys. *Biometrics* 62:372-378. doi:10.1111/j.1541-0420.2005.00493.x.
- [Buckland et al., 2010] Buckland, S.T., Laake, J.L., Borchers, D.L. 2010. Double-Observer Line Transect Methods: Levels of Independence. *Biometrics* 66:169-177. doi:10.1111/j.1541-0420.2009.01239.x. <http://digitalcommons.unl.edu/usdeptcommercepub/199>. (Accessado 10/10/2015).
- [Bussab & Morettin, 2002] Bussab, W.O., Morettin, P.A. 2002. Estatística Básica. São Paulo: Saraiva. ISBN978-85-02-08177-2.
- [Celeux et al., 2006] Celeux, G., Forbes, F.R.C., Titterton, D. 2006. Deviance Information Criteria for Missing Data Models. *International Society for Bayesian Analysis* 4:651-674.
- [Clark, 2005] Clark, J. 2005. Why environmental scientists are becoming Bayesians?. *Ecology Letters* 8:2-14. doi:10.1111/j.1461-0248.2004.00702.x.
- [Cremer & Simoes-lobes, 2005] Cremer, M.J., Simões-Lopes, P.C. 2005. The occurrence of *Pontoporia blainvillei* (Gervais and D'Orbigny) (Cetacea, Pontoporiidae) in a estuarine area in southern Brazil. *Revista Brasileira de Zoologia* 22(3):717-723.
- [Crespo, 2009] Crespo, E.A. 2009. Franciscana dolphin *Pontoporia blainvillei*. In: Perrin WF, Würsig B. Thewissen JGM. *Encyclopedia of Marine Mammals*. (eds). 2.ed. San Diego. CA: Academic Press.466-469.
- [Crespo et al., 1998] Crespo, E.A., Harris, G., Gonzáles, R. 1998. Group size and distributional range of the franciscana, *Pontoporia blainvillei*. *Marine Mammal Science* 14(4):845-849.
- [Conn et al., 2012] Conn, P.B., Laake, J., Johnson, D. 2012. A Hierarchical Modeling Framework for Multiple Observer Transect Surveys. *PLoS ONE* 7(8):e42294. doi:10.1371/journal.pone.0042294.
- [Dalla Rosa et al., 2012] Dalla Rosa, L., Ford, J.K.B., Trites, A.W. 2012. Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters. *Continental Shelf Research* (36): 89-104. doi:10.1016/j.csr.2012.01.017.2012.

- [Danilewicz, 2003] Danilewicz, D. 2003. Reproduction of female franciscana (*Pontoporia blainvillei*) in Rio Grande do Sul, southern Brazil. *Latin American Journal of Aquatic Mammals* 2(2):67-78.
- [Danilewicz et al., 2008] Danilewicz, D., Secchi, E.R., Di Benedetto, A.P.M. 2008. *Pontoporia blainvillei* (Gervais and d'Orbigny, 1844). In: Livro Vermelho da Fauna Brasileira Ameaçada de Extinção. Machado ABM, Drummond GM, Paglia AP. (eds). Brasília: MMA v.II ,p.814-815.
- [Danilewicz et al., 2009] Danilewicz, D., Secchi, E.R., Ott, P.H., Moreno, I.B., Bassoi, M., Borges-Martins, M. 2009. Habitat use patterns of franciscana dolphins (*Pontoporia blainvillei*) off southern Brazil in relation to waters depth. *Journal of Marine Biological Association of the United Kingdom* 89(5):943-949.
- [Dolphin, 1987] Dolphin, W.F. 1987. Ventilation and dive patterns of Humpback Whale, *Megaptera novaeangliae*, on their Alaskan feeding grounds. *Canadian Journal Zoology* 65:83-90.
- [Durban et al., 2005] Durban, J.W., Elston, D.A., Ellifrit, D.K., Dickson, E., Hammond, P.S., Thompson, P.M. 2005. Multisite mark-recapture for cetaceans: population estimates with Bayesian model averaging. *Marine Mammal Science* 21(1):80-92.
- [Eguchi & Gerrodette, 2009] Eguchi, T., Gerrodette, T. 2009. A Bayesian approach to line-transect analysis for estimating abundance. *Ecological Modelling* 220:1620-1630. doi:10.1016/j.ecolmodel.2009.04.011.
- [Fewter et al., 2009] Fewter, R.M., Buckland, S.T., Burnham, K.P., Borchers, D.L., Jupp, P.E., Laake, J.L., Thomas, L. 2009. Estimating the Encounter Rate Variance in Distance Sampling. *Biometrics* 65:225-236. doi:10.1111/j.1541-0420.2008.01018.x.
- [Gelman & Hill, 2007] Gelman, A., Hill, J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 5.ed. New York: Cambridge University Press 645pp. ISBN-13978-0-511-26878-6 eBook (EBL).
- [Guilherme-silveira & Kinas, 2016] Guilherme-Silveira, F.R., Kinas, P.G. 2016. Bayesian inference in line transects with double count sampling and imperfect on-line detection. *Revista Brasileira de Biometria*. No Prelo
- [Harwood, 2001] Harwood, J. 2001. Marine mammals and their environment in the twenty-first century. *Journal of Mammal* 82(3):630-640.
- [IBAMA, 1997] Instituto Brasileiro do Meio Ambiente e Recursos Renováveis. 1997. *Relação das Espécies sob Maior Pressão Antrópica*. In: Mamíferos Aquáticos do Brasil: Plano de Ação II. Instituto Brasileiro de Meio Ambiente e Recursos Renováveis (Eds). Brasília. p.25-27.
- [Janik & Thompson, 1996] Janik, V.M., Thompson, P.M. 1996. Changes in surfacing patterns of bottlenose dolphins in response to boat traffic *Tursiops truncatus*. *Marine Mammal Science* 12(4):597-602.
- [Jaynes, 2003] Jaynes, E.T. 2003. *Probability Theory - The logic of science*. Cambridge: Cambridge University Press. 727p.
- [Jonsen et al., 2003] Jonsen, I.D., Myers, R.A., Flemming, J.M. 2003. Meta-analysis of animal movement using state-space models. *Ecology* 84(11):3055-3063
- [Karunamuni & Quinn, 1995] Karunamuni, R., Quinn, I.T. 1995. Bayesian Estimation of Animal Abundance for Line Transect Sampling. *Biometrics* 51(4):1325-1337. <http://www.jstor.org/stable/2533263>.
- [Kery & Royle, 2010] Kery, M., Royle, J.A. 2010. Hierarchical modelling and estimation of abundance and population trends in metapopulation designs. *Journal of Animal Ecology* 79, p.453-461. doi:10.1111/j.1365-2656.2009.01632.x.
- [King & Brooks, 2008] King, R., Brooks, S.P. 2008. On the Bayesian Estimation of a Closed Population Size in the Presence of Heterogeneity and Model Uncertainty. *Biometrics* 64: 816-824. doi:10.1111/j.1541-0420.2007.00938.x.

- [Kinas & Andrade, 2010] Kinas, P.G., Andrade, H. 2010. Introdução a análise bayesiana (com R). 1.ed. Porto Alegre: MaisQnada editora 240p.
- [Laake, 1999] Laake, J. 1999. Distance sampling with independent observers: Reducing bias from heterogeneity by weakening the conditional independence assumption. Marine Mammal Survey and Assessment Methods. Garner et al. (eds). Balkema, Rotterdam: Brookfield. 290p. ISBN99058090434.
- [Laake & Borchers, 2004] Laake, J.L., Borchers, D.L. 2004. Methods for incomplete detection at distance zero, In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. Advanced distance sampling: estimating abundance of biological populations. Norfolk, Great Britain: Oxford University Press. 416pp. ISBN0198507836 (hbk).
- [Learmonth et al., 2006] Learmonth, J.A., MacLeod, C.D., Santos, M.B., Pierce, G.J., Crick, H.Q.P., Robinson, R.A. 2006. Potential effects of climate change on marine mammals. In: Oceanography and Marine Biology: An annual review 44:431-464.
- [Lerczak & Hobbs, 1998] Lerczak, J.A., Hobbs, R.C. 1998. Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammals survey. Marine Mammal Science 14(3):590-599.
- [Marsh & Sinclair, 1989] Marsh, H.A., Sinclair, D.F. 1989. An Experimental Evaluation of Dugong and Sea Turtle Aerial Survey Techniques. Australian Wildlife Research 16:639-650.
- [Martin & Quinn, 2006] Martin, A., Quinn, K. 2006. Applied bayesian inference in R using MCMC-pack. R News 6(1):2-7.
- [McCarthy, 2007] McCarthy, M. 2007. Bayesian Methods for Ecology. 1.ed. New York: Cambridge University Press. ISBN-13978-0-521-61559-4. 305p.
- [McKenzie et al., 2006] MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E. 2006. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier Inc. 342pp. ISBN13:978-0-12-088766-8.
- [Newman et al., 2009] Newman, K.B., Fernandez, C., Thomas, L., Buckland, S.T. 2009. Monte Carlo Inference for State-Space Models of Wild Animal Populations. Biometrics 65:572-583. doi:10.1111/j.1541-0420.2008.01073.x.
- [O'Connell et al., 2006] O'Connell Jr, A.F., Talancy, N.W., Bailey, L.L., Sauer, J.R., Cook, R., Gilbert, A.T. 2006. Estimating site occupancy and detection probability parameters for meso and large mammals in a coastal ecosystem. Journal of Wildlife Management 70(6):1625-1633.
- [Oedekoven et al., 2014] Oedekoven, C.S., Buckland, S.T., Mackenzie, M.L., King, R., Evans, K.O., Burger Jr, L.W. 2014. Bayesian Methods for Hierarchical Distance Sampling Models. International Biometric Society. doi:10.1007/s13253-014-0167-0.
- [Perez, 2015] Perez, F.S. 2015. Fatores que afetam a detectabilidade da toninha (*Pontoporia blainvillei*) em estudos de estimativas populacionais. [Discertação de mestrado] Universidade Federal de Juiz de Fora. 80pp.
- [Pinedo et al. 1989] Pinedo, M.C., Praderi, R., Brownell Jr, R.L. 1989. Review of the biology and status of the franciscana, *Pontoporia blainvillei*. In: Biology and Conservation of the River Dolphins. Perrin WF, Brownell Jr RL, Zhou K, Liu J. (eds). Hong Kong. IUCN. p.46-51.
- [Plummer, 2003] Plummer, M. 2003. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. ISSN 1609-395X. Vienna, Austria. Available: <http://www.ci.tuwien.ac.at/Conferences/DSC-2003>. p.1-10.
- [Plummer et al., 2010] Plummer, M., Best, N., Cowles, K., Vines, K. 2010. CODA: Convergence diagnosis and output analysis for MCMC. R News 6(1):7-11.

- [Plummer, 2012] Plummer, M. 2012. JAGS Version 3.3.0. Disponível: <http://people.math.aau.dk/~kbb/undervisning/Bayes13/sorenh/doc>.
- [Plummer, 2013] Plummer, M. 2013. Package "rjags": Bayesian graphical models using MCMC. Disponível: <http://cran-r.c3sl.ufpr.br/web/packages/rjags/rjags.pdf>.
- [Porto de São Francisco (Autarquia), 2015] Porto de São Francisco. 2015. Porto de São Francisco (Autarquia). Governo do Estado de Santa Catarina. disponível em: <http://www.apsfs.sc.gov.br/> (Acessado em: 02/07/2015).
- [Quang & Becker, 1996] Quang, P., Becker, E. 1996. Line Transect Sampling Under Varying Conditions with Application to Aerial Surveys. *Ecological Society of America* 77(4):1297-1302. Disponível: <http://www.jstor.org/stable/2265601>.
- [Quang & Becker, 1997] Quang, P., Becker, E. 1997. Combining Line Transect and Double Count Sampling Techniques for Aerial Surveys. *Journal of Agricultural, Biological and Environmental Statistics* 2(2):230-242. Disponível: <http://www.jstor.org/stable/1400405>.
- [Quang & Becker, 1998] Quang, P., Becker, E. 1998. Aerial survey sampling of contour transects using double-count and covariate data. In: *Marine Mammal Survey and Assessment Methods*, 25-27 February 1998. Seattle Washington USA. Proceedings of the Symposium on Surveys, Status and Trends of Marine Mammal Populations. Washington: Seattle. p1-11.
- [R Core Team, 2014] R Core Team. 2014. R: A Language and Environment for Statistical Computing. R. Foundation for Statistical Computing. Vienna, Austria. Disponível: <http://www.R-project.org/>.
- [Ramos et al., 2000] Ramos, R.M.A., Di Benedetto, A.P.M., Lima, N.R.W. 2000. Growth parameter of *Pontoporia blainvillei* and *Sotalia fluviatilis* (Cetacea) in northern Rio de Janeiro, Brazil. *Aquatic Mammals* 26(1):65-75.
- [Ramos et al., 2002] Ramos, R.M.A., Di Benedetto, A.P.M., Siciliano, S., Santos, M.C.O., Zerbini, A.N., Bertozzi, C., Vicente, A.F.C., Zampiroli, E., Alvarenga, F.S., Lima, N.R.W. 2002. Morphology of the franciscana (*Pontoporia blainvillei*) off southeastern Brazil: Sexual dimorphism, growth and geographic variation. *Latin American Journal of Aquatic Mammals* 1(1):129-144.
- [Reeves et al., 2012] Reeves, R.R., Dalebout, M., Jefferson, T.A., Karkzmariski, L., Laidre, K.L., O'Corry-Crowe, G., Rojas-Bracho, L., Secchi, E.R., Slooten, E., Smith, B.D., Wang, J.Y., Zerbini, A.N., Zhou, K. 2012. *Pontoporia blainvillei*. IUCN Red List of Threatened Species: IUCN.
- [Royle, 2008] Royle, J.A. 2008. Hierarchical Modeling of Cluster Size in Wildlife. Surveys *Journal of Agricultural, Biological, and Environmental Statistics* 13(1):23-36. doi:10.1198/108571108X273188.
- [Royle & Dorazio, 2008] Royle, J.A., Dorazio, R. 2008. Hierarchical modeling and inference in ecology, the analysis of data from populations, metapopulations and communities. 1.ed. USA: Elsevier. 444p.
- [Secchi et al., 2003] Secchi, E.R., Danilewicz, D., Ott, P.H. 2003. Applying the phylogeographic concept to identify franciscana dolphin stocks: implications to meet management objectives. *Journal Cetacean Research Management* 5(1):61-68.
- [Senar & Conroy, 2004] Senar, J.C., Conroy, M.J. 2004. Multi-state analysis of the impacts of avian pox on a population of Serins (*Serinus serinus*): the importance of estimating recapture rates. *Animal Biodiversity and Conservation* 27(1):133-146.
- [Siciliano, 1994] Siciliano, S. 1994. Review of small cetaceans and fishery interactions in coastal waters of Brazil. *International Whaling Commission (Special Issue 5)*. p.241-250.
- [Spiegelhalter et al., 2002] Spiegelhalter, D.J., Best, N., Carlin, B., van der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64:583-640.
- [Schofield & Barcker, 2010] Schofield, M., Barcker, R. 2010. Data augmentation and reversible jump MCMC for multinomial index problems. Disponível: <http://arxiv.org/abs/1009.3507v1>.

- [Su & Yajima, 015] Su, Y.S., Yajima, M. 2015. Pachege "R2jags": A Package for Running 'JAGS' from R. Disponível: <http://cran-r.c3sl.ufpr.br/web/packages/R2jags/R2jags.pdf>.
- [Wade, 2000] Wade, P. 2000. Bayesian methods in conservation biology. *Conservation Biology* 14(5):1308-1316.
- [Wells et al., 2013] Wells, R.S., Bordino, P., Douglas, D.C. 2013. Patterns of social association in the franciscana, *Pontoporia blainvillei*. *Marine Mammal Science* 29(4):E520–E528. ISSN 08240469.
- [Williams, 2001] Williams, B.K. 2001. Uncertainty, learning, and the optimal management of wildlife. *Environmental and Ecological Statistics* 8:269-288.
- [Williams & Thomas, 2009] Williams, R., Thomas, L. 2009. Cost-effective abundance estimation of rare animals: Testing performance of small-boat surveys for killer whale in British Columbia. *Biological Conservation* 142:1542–1547. doi:10.1016/j.biocon.2008.12.028.

9 ANEXOS

9.1 Anexo I

9.1.1 Bayesian Inference in Line Transects with Double Count Sampling and Imperfect on-Line Detection

Conhecer a abundância absoluta de populações animais é primordial para seu manejo e sua conservação. Porém, estimativas de abundância, que ignoram a detecção imperfeita dos indivíduos presentes nas áreas, resultam em subestimativas. Modelos hierárquicos com uma abordagem bayesiana, que fazem uso de dados de distâncias com a informação do segundo observador permitem estimar simultaneamente abundância e detectabilidade. Propomos uma alternativa de implementação simples, usando dados aumentados e simulação de Monte Carlo com Cadeia de Markov (mcmc). As probabilidades de detecção são modeladas por regressões logit e probit em função das distâncias aos indivíduos detectados. Validamos nossos modelos com amostras simuladas de uma população fictícia de tamanho conhecido e com funções de detectabilidade distintas. Implementamos novos recursos de diagnósticos para *goodness-of-fit* dos modelos aos dados. A complementação dos dados de distância com dados de um segundo observador, aumentou a precisão da estimativa de abundância em 29,6% com relação ao uso exclusivamente dos dados de distância. O melhor modelo, M5, estima corretamente os valores $g(0)$, porém a abundância é subestimada em 12% considerando-se a média da distribuição como estimativa pontual. Também aplicamos o modelo a uma população de gansos Loon descrita e analisada em Quang and Becker (1997). Nossa estimativa de 154 loons é maior que a estimativa de 99 indivíduos reportada em QB. O diagnóstico de *goodness-of-fit*, no entanto, indicam que o modelo é adequado. O estudo simulado sugere que a população seja ainda maior. Modelos hierárquicos integrando amostras de distância com dados de marcação e recaptura permitem estimar simultaneamente abundância e curvas de detecção. A modelagem das probabilidades de detecção por regressão logit ou probit, permite flexibilidade para ajuste de curvas não-convencionais com potencial para inclusão de outras covariáveis. Apesar da estimativa de curva de detectabilidade para ambos os observadores não ser o objetivo mais relevante do trabalho, os modelos propostos lidam com a problemática do viés de percepção de tal forma a fornecer uma estimativa de abundância com bastante precisão.

9.2 Anexo II

9.2.1 Viés de Percepção em Estimação da Abundância de Toninhas na Baía da Babitonga, Brasil

O objetivo desse trabalho foi conjugar o método de amostragem de distância com o método de marcação e recaptura numa estrutura hierárquica. Estimou-se conjuntamente densidade de toninhas e o viés de percepção. Para dados truncados, (DT), o melhor modelo estima a máxima probabilidade de detecção, (MPD) sobre a derrota (distância = 0). Para dados não truncados (DnT) o melhor modelo, que inclui um termo quadrático, estima as MDPs a distâncias maiores que zero. Os melhores ajustes indicam curvas MDPs distintas para cada observador tanto para DT quanto DnT. O ajuste para DnT evitou a eliminação de 26% dos dados. Verificou-se ainda, que a distância média ideal para truncamento é em 123 m. A densidade foi estimada em $1.219 \text{ } N_g km^{-2}$ (ICr95%: 0.739 a 2.075). A flexibilidade do modelo hierarquicos propostos permite estimar densidade da população sem que haja necessidade de qualquer tipo de truncamento dos dados. O que evita o descarte de informações, obtidas muitas vezes, por intermédio de logística onerosa.

BAYESIAN INFERENCE IN LINE TRANSECTS WITH DOUBLE COUNT SAMPLING AND IMPERFECT ON-LINE DETECTION

Fernando Roberto GUILHERME-SILVEIRA¹
Paul Gerhard KINAS²

■ **ABSTRACT:** *For the management and conservation of wild animal populations it is fundamental to know its abundance. However, if imperfect detection, a very common phenomenon in field counts, is ignored, abundance will be underestimated. We show that Bayesian hierarchical models for double observer distance sampling data are capable of simultaneously estimating abundance and detection probabilities and propose a simple model where detection probabilities are modeled as logit or probit regressions of distance-to-line and give its implementation in BUGS code. With a simulation study we verify empirically that double observer information increases the precision in abundance estimates by about 30% when compared with estimates from distance data only. We further verify that the model is capable to correctly estimate observer-specific detection probability, but underestimates abundance by 12% on average. We also apply an extension of these models to a population of loon (QUANG and BECKER, 1997; URL:<http://www.jstor.org/stable/1400405.1997>). Our estimate of 154 (posterior mean) was much higher than the estimated 99 individuals reported by QB although other model parameters are similar. Some new model-specific goodness-of-fit diagnostics are proposed and applied.*

■ **KEYWORDS:** *animals distribution; ecological modeling; state process; observation model.*

¹Universidade Federal do Rio Grande-FURG, Graduate Program in Biological Oceanography, Oceanography Institute, Postal code: 96203-900, Rio Grande, Rio Grande do Sul, Brazil. E-mail: guilhermesilveira.fr@gmail.com

²Universidade Federal do Rio Grande-FURG, Institute of Mathematics, Statistics and Physics, Postal code: 96203-900, Rio Grande, Rio Grande do Sul, Brazil. E-mail: paulkinas@furg.br

1 Introduction

The absolute size of an animal population is important to evaluate its conservation status and to monitor the sustainability of management actions. However, the estimation of abundance is often a difficult and expensive task, particularly in the marine environment. This is due to the extensive area of distribution and the elusiveness of individuals or groups who remain submersed. Marsh and Sinclair (1989) propose two sources of bias in abundance estimates, denoting them as availability and perception biases. Availability bias refers to the uncertain presence of the animal within the searched area at time of survey, while perception bias is caused by imperfect detection of animals or groups which are available.

Transect line distance sampling analysis (distance sampling, for short) is a commonly used method of estimating density and abundance for a variety of marine mammal populations, usually for the purpose of management and conservation (BORCHERS, BUCKLAND and ZUCCHINI, 2002; DALLA ROSA, FORD and TRITES, 2012). Data are collected by an observer travelling along prespecified routes (the transect line) while recording the perpendicular distances to the detected animals. If the animals occur in groups or clusters, such as flocks of birds or schools of dolphins, the number of animals in each group is recorded along with the distance to the groups centroid. Since some animals are missed (perception bias), approaches are needed to correct for it. This is achieved by modeling the probability of detection as a known function $g(\cdot)$ of perpendicular distance x , usually assuming that higher distances associate with lower detection probabilities. In order to have identifiable parameters, this function needs to assume perfect detection at some known distance. Conventional distance sampling assumes perfect detection on the transect line, $g(0) = 1$. If this assumption is violated (i.e. $g(0) < 1$), we say there is imperfect on-line detection and denote $1 - g(0)$ as perception bias. This feature is common for marine mammals and if unaccounted for will cause a downward bias in abundance estimates (BORCHERS, 1999; BORCHERS, BUCKLAND and ZUCCHINI, 2002).

Extensions of distance sampling to incorporate imperfect detection on the transect line as another parameter and guarantee parameter identifiability rely on double observer distance sampling with mark-recapture data between them (KARUNAMI and QUINN, 1995; LAAKE, 1999). These finding have been confirmed and extended by many other studies of various authors (QUANG and BECKER, 1996, 1997, 1998; BAILEY et al., 2007; CONN, LAAKE and JOHNSON, 2012; EGUCHI and GERRODETTE, 2009; ROYLE and DORAZIO, 2008).

Bayesian analysis has become increasingly popular in statistical inference of wildlife population abundance and related parameter (KARUNAMUNI and QUINN, 1995; BERLINER, 1996; WADE, 2000; DURBAN et al., 2005; MCCARTHY, 2007; KING and BROOKS, 2008). Unlike the orthodox, more restrictive, definition of frequentist probability, a Bayesian probability is a much broader metric capable to quantify any kind of uncertainty caused by incomplete

information (JAYNES, 2003). Hence, a Bayesian population abundance estimate is given in the form of a (posterior) probability distribution. The Bayesian approach is further capable to combine available extra-data information (prior distributions) with new observed data (likelihood) to produce an updated state of information (posterior distributions) by way of Bayes theorem (KINAS and ANDRADE, 2010).

Hierarchical Bayesian models offer a flexible and realistic approach to ecological research (CLARK, 2005; SCHOFIELD and BARKER, 2010). Markov Chain Monte Carlo (MCMC) is a handy tool to obtain simulated high-dimensional posterior distributions with relative ease (MARTIN and QUINN, 2006; McCARTHY, 2007). Its implementation in BUGS code (LUNN et al., 2009; RESNIK and HARDISTY, 2010) with the use of specialized software (e.g. JAGS) (PLUMMER, 2003; PLUMMER, 2012) and related R libraries (rjags) (PLUMMER, 2013) have made Bayesian inference and hierarchical model fitting more accessible to applied scientist in general.

Bayesian hierarchical distance sampling models were sistematized by Royle and Dorazio, (2008, Chapter 7). In order to simulate the joint posterior distribution for all unknown parameters using MCMC, they formulate the model within a data augmentation framework providing a flexible structure to fit very general models (ROYLE and DORAZIO, 2008, P.181).

The aim of this paper is to formulate simple Bayesian hierarchical models implemented in BUGS code to estimate population abundance without assuming perfect detection on the transect line. Double observer distance sampling data are the basic requirements. Firstly, with the help of a simulated study we examine the reliability of the proposed hierarchical model to estimates abundance and imperfect on-line detection probability. Secondly, we explore further flexibilities of hierarchical model formulation with a case study on double observer distance data for a loon population in Alaska (QUANG and BECKER, 1997) in which we allow for maximum detection probability to occur off the transect line and evaluate the goodness-of-fit for these models with some novel diagnostic tools.

2 Method

2.1 The Models

The models are formulated within a hierarchical structure and implemented with data augmentation (ROYLE and DORAZIO, 2008, P.181). The idea behind data augmentation is that the actual population of unknown size N is contained in a super-population of given size M assumed much larger than N .

Considering just distance data, we define a binary variable y_i which indicates whether the i -th subject has been detected ($y_i = 1$) or not ($y_i = 0$) for $i = 1, \dots, M$. The variable y_i has therefore a Bernoulli distribution with parameter μ_i . We express this distribution as

$$y_i \sim \text{Bernoulli}(\mu_i)$$

For each subject the parameter μ_i has two components: the detection probability $g(x_i)$, a function of the perpendicular distance to the transect line x_i , and the binary variable w_i , indicating whether the i -th subject, in the "augmented population" of known size M is part of the "real population" of unknown size N . For all detected animals $w_i = 1$ while it is unknown (or latent) for the remaining $M - n$ elements and has to be estimated. Hence, μ_i is defined as $\mu_i = w_i \cdot g(x_i)$ and w_i are independent Bernoulli random variables with success probability $\psi = N \cdot M^{-1}$ and therefore expressed as

$$w_i \sim \text{Bernoulli}(\psi)$$

Consequently it results that the sum $S_w = \sum(w_i)$, has a Binomial distribution $\text{Bin}(M, \psi)$. Hence, estimating N is equivalent to estimate ψ since $E(S_w) = M \cdot \psi = N$.

For perpendicular distances x [$0 \leq x \leq m$], where m is some fixed maximum perpendicular distance, we define the distance-dependent detection probability $g(x)$ with the inverse logit link function

$$g(x) = \frac{e^{(\beta_0 + \beta_1 \cdot x)}}{1 + e^{(\beta_0 + \beta_1 \cdot x)}}$$

and the inverse probit link function

$$g(x) = \Phi(\beta_0 + \beta_1 \cdot x)$$

where Φ is the standard Normal distribution function.

The detection probability on the transect line (i.e. $g(0)$) becomes a function of β_0 only for both link functions. Hence, by estimating β_0 we are estimating a known function of the perception bias $1 - g(0)$. When only distance data are available, β_0 is confounded with abundance and cannot be estimated individually. One way to circumvent this difficulty is the incorporation of mark-recapture information for two observers who simultaneously search the transect line and collect distance data individually (BUCKLAND, LAAKE and BORCHERS, 2010).

With the inclusion of the double observers (DO) information, the binary variable y_i extends to y_{ji} for observer j where $j = 1, 2$ and the detection probability becomes also observer-dependent $g_j(x_i)$. Variable y_{ji} is now associated with the indicator vector $Z_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i})$, defined such that just one element is equal to one and all others are zeros. Thus, by $z_{1i} = 1$ we mean that the detection is just by observer 1 (i.e. $y_{1i} = 1$ and $y_{2i} = 0$) and the associated vector is $Z_i = (1, 0, 0, 0)$; when $z_{2i} = 1$ the detection is just by observer 2 and $Z_i = (0, 1, 0, 0)$; when $z_{3i} = 1$, the detection is by both observers and $Z_i = (0, 0, 1, 0)$; finally, $z_{4i} = 1$ and $Z_i = (0, 0, 0, 1)$ for all $M - n$ undetected subjects.

We consider that detections are independent between observers (i.e. $g_j(x_i) = g_{j|(3-j)}(x_i)$ with the right-hand side denoting the conditional probability that

observer j detects subject i given that the other observer also detects it). Thus, the probabilities of the four components of Z_i are easily calculated and the Binomial model for y_i is extended to a Multinomial model for Z_i with the following parameter structure:

$$\begin{aligned} Z_i &\sim \text{Multinomial}(1, (\mu_{1i}, \mu_{2i}, \mu_{3i}, \mu_{4i})) \\ \mu_{1i} &= g_1(x_i) \cdot (1 - g_2(x_i)) \cdot w_i \\ \mu_{2i} &= (1 - g_1(x_i)) \cdot g_2(x_i) \cdot w_i \\ \mu_{3i} &= g_1(x_i) \cdot g_2(x_i) \cdot w_i \\ \mu_{4i} &= 1 - (\mu_{1i} + \mu_{2i} + \mu_{3i}) \\ w_i &\sim \text{Bernoulli}(\psi) \end{aligned}$$

where $\text{Multinomial}(s, (p_1, \dots, p_k))$ denotes a k -dimensional multinomial distribution with sample size s and probability vector $(p_1, \dots, p_k)$ satisfying $\sum p_r = 1$.

Finally, based on the n sampled detections with known distances x_i and the $M - n$ undetected augmented data the estimated population abundance becomes

$$N = \sum_{i=1}^M w_i = n + \sum_{i=n+1}^M w_i$$

For the $M - n$ undetected subjects with missing perpendicular distances x_i , these distances are assumed uniformly distributed in the interval $(0, m)$ and imputed.

2.2 The Data

For the simulation study, we consider a virtual population of size $N = 200$ subjects, whose perpendicular distances to the transect line x , are independent random variable with uniform distribution in the interval $[0, 1]$. Any distance d , effectively measured in the field until some maximum fixed distance m , can always be standardized such that $x = d \cdot m^{-1}$ is in the range $[0, 1]$. Hence, there is no loss of generality by using this distribution. To generate the data we also define half-Normal detection functions $g_j(x_i) = k_j \cdot \exp(-x^2 \sigma_j^{-2})$ with parameters $k_1 = 0.8$; $\sigma_1 = 8$, for observer one and $k_2 = 0.5$; $\sigma_2 = 12$ for observer two. The detection of the i -th individual by observer j is modeled with a Bernoulli distribution $y_{ji} \sim \text{Bern}(g_j(x_i))$, where each individual i is classified as detected ($y_{ji} = 1$) or undetected ($y_{ji} = 0$) by observer j . The components of the vector Z_i for $i = 1, \dots, N$ are defined for the i -th individual as follows:

$$\begin{aligned}
z_{1i} &= y_{1i} \cdot (1 - y_{2i}) \\
z_{2i} &= (1 - y_{1i}) \cdot y_{2i} \\
z_{3i} &= y_{1i} \cdot y_{2i} \\
z_{4i} &= (1 - y_{1i}) \cdot (1 - y_{2i})
\end{aligned}$$

A simulated sample consists in the distances x_i and associated vectors (z_{1i}, z_{2i}, z_{3i}) for a total $n = \sum_{i=1}^N (z_{1i} + z_{2i} + z_{3i})$ detected subjects. The full sampling procedure is replicated 20 times, each identified as transect t_r , where $r = 1, 2, \dots, 20$. The observed perpendicular distances and the total number of detected subjects are summarized and displayed in Fig. 1.

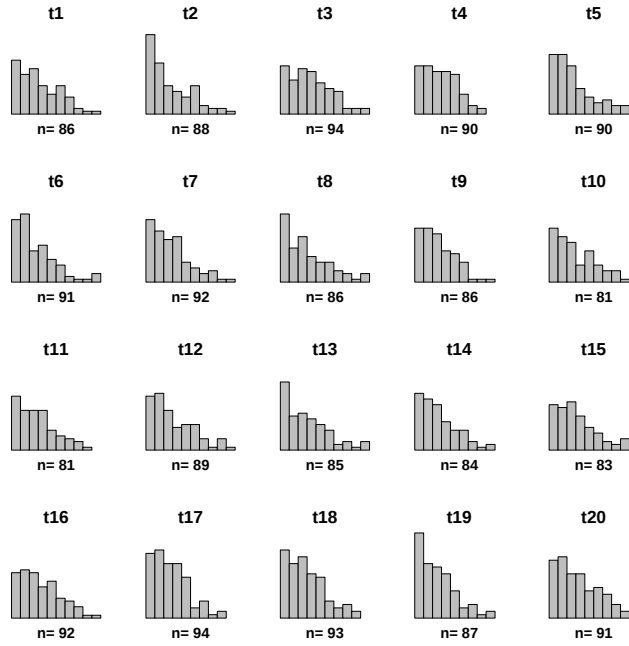


Figure 1 - Histogram of individual distances and total sample size n for 20 replicated samples extracted from the simulated population. All the frequencies (y -coordinates) are on the same scale and distances (x -coordinates) are on the interval $[0, 1]$

The second data set is adapted from Quang and Becker (1997) and consists of a mixed population with two species of loons (*Gavia pacifica* and *Gavia immer*) inhabiting the Yukon Flats National Park, Alaska. Since double count distance data were reported only in interval classes (Table 1), we simulated within each class, the individual distances with uniform distributions.

Table 1 - Loon detection data reproduced from Quang and Becker (1997), Table 01

Distance class	5-30	30-60	60-90	90-120	120-150	150-190	190-250
Front passenger only	3	1	2	3	3	5	5
Rear passenger only	4	3	0	3	5	1	2
Both passengers	0	1	2	10	6	6	2

2.3 Fitted Model

Models used for inference are distinct regarding data type and link function. In some models, data are only perpendicular distances to the line transect (LT); in others, double observer mark-recapture data are also included (DOLT). The link functions are logit or probit. From now on we will just use model numbers in Table 2 when referring to them. From now on we will use the shorthand 'DOLT-logit' when referring to the model fitted to double-observer distance data fitted with logit link; applying the obvious changes when referring to all other situations. In the case of LT models M1 and M2 we use the total number n of subjects detected by at least one observer and simply ignore capture-recapture information. This allows for the evaluation of changes in precision of abundance estimates obtained for the same sample size and distance data but including double observer data as well (M3 and M4).

Table 2 - List of models fitted to distance data only (LT) and complemented by double observer mark-recapture data (DOLT). The logit link is used in odd-numbered and the probit link in even-numbered models. The term *single* refers to models with a single detection function while *obs* refers to models with observer-specific detection functions. The term *quad* indicates the inclusion of a quadratic term ($p_{ji} = g_j(x_i)$).

	Model	Description
M1	$\log\left[\frac{p_i}{(1-p_i)}\right] = \beta_0 + \beta_1 \cdot x_i$	<i>single LT – logit</i>
M2	$\Phi^{-1}(p_i) = \beta_0 + \beta_1 \cdot x_i$	<i>single LT – probit</i>
M3	$\log\left[\frac{p_{ji}}{(1-p_{ji})}\right] = \beta_0 + \beta_1 \cdot x_i$	<i>single DOLT – logit</i>
M4	$\Phi^{-1}(p_{ji}) = \beta_0 + \beta_1 \cdot x_i$	<i>single DOLT – probit</i>
M5	$\log\left[\frac{p_{ji}}{(1-p_{ji})}\right] = \beta_{j0} + \beta_{j1} \cdot x_i$	<i>double DOLT – logit</i>
M6	$\Phi^{-1}(p_{ji}) = \beta_{j0} + \beta_{j1} \cdot x_i$	<i>double DOLT – probit</i>
M7	$\log\left[\frac{p_{ji}}{(1-p_{ji})}\right] = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2$	<i>single quad DOLT – logit</i>
M8	$\Phi^{-1}(p_{ji}) = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2$	<i>single quad DOLT – probit</i>
M9	$\log\left[\frac{p_{ji}}{(1-p_{ji})}\right] = \beta_{j0} + \beta_{j1} \cdot x_i + \beta_{j2} \cdot x_i^2$	<i>double quad DOLT – logit</i>
M10	$\Phi^{-1}(p_{ji}) = \beta_{j0} + \beta_{j1} \cdot x_i + \beta_{j2} \cdot x_i^2$	<i>double quad DOLT – probit</i>

To fit the loon data, we further include a quadratic term of x to allow for the possibility of a maximum detection probability at some distance off the transect line

($x > 0$). To select the most parsimonious model we use the deviance information criterion (DIC), with smaller DIC meaning a better fit (SPIEGELHALTER et al., 2002). To facilitate later reference just by model number, all fitted models are listed in Table 2.

2.4 Bayesian Inference - simulated data

Models 1 to 6 (Table 2) are fitted to each of the 20 replicated transects. We define vague marginal priors for all parameters: $\psi \sim U(0, 1)$; $\beta_{j0} \sim N(0, 1E5)$ and $\beta_{j1} \sim N(0, 1E5)$ to $j = 1, 2$. Posterior distributions are obtained by Markov Chain Monte Carlo (mcmc) simulations (MARTIN and QUINN, 2006; McCARTHY, 2007) with the libraries R2jags (SU and YAJIMA, 2015) and rjags (PLUMMER, 2013), which run JAGS (PLUMMER, 2003; PLUMMER, 2012) from within R (version 3.1.1) (R CORE TEAM, 2014).

We evaluate convergence with the diagnostic tools Rhat and n.eff provided by R2jags, in combination with the standard diagnostics provided in the 'coda' package (PLUMMER, 2010; PLUMMER, 2013). After some preliminary testing we have fixed the fitting procedure that achieved satisfactory convergence diagnostics, running three chains with a burn-in of 10.000 steps and a thinning of 40, generating a posterior sample of size 3000.

To assess inferential efficiency in estimating population size N and detection probability at maximum detection $g_j(0)$ we use the relative bias (rb) defined as the difference between the posterior mean and the true value divided by the true value. As a summary over the twenty replicates we further calculate the average rb and also the root mean squared error ($rmse$) among posterior means; where $rmse$ is defined as the variance of the posterior mean plus the squared difference between the average posterior mean and the true parameter.

2.5 Bayesian Inference - loon data

Models 3 to 10 (Table 2) are fitted to the loon data with vague prior distributions for all model parameter: $\psi \sim U(0, 1)$; $\beta_{j0} \sim N(0, 1E5)$, $\beta_{j1} \sim N(0, 1E5)$ e $\beta_{j2} \sim N(0, 1E5)$. The posterior distributions are obtained as previously, with three chains, a burn-in of 10.000 steps and a thinning of 40, generating a posterior sample of size 3000. Preliminary testing indicated good convergence diagnostics when running mcmc with these options.

Models with an additional quadratic term β_2 (models 7 and 8) or the observer-specific extension β_{j2} (models 9 and 10) have the maximum detection probability at some positive perpendicular distance $x_0 > 0$. For some fixed values β_1 and β_2 this distance is:

$$x_0 = -\frac{\beta_1}{2\beta_2}$$

Therefore, in a quadratic model, $g(x_0)$ replaces $g(0)$ as the distance with maximum detection probability. Within the Bayesian framework, it is easy to obtain a posterior distribution for this probability, since it is a known function of the uncertain parameters β_0, β_1 and β_2 (or their j -indexed equivalents) for which joint posterior distribution is readily available. For the inverse-logit and inverse-probit link functions, these maximum detection probabilities are, respectively

$$\begin{aligned} g(x_0) &= (1 + e^{-(\beta_0 - \frac{\beta_1^2}{4\beta_2})})^{-1} \\ g(x_0) &= \Phi(\beta_0 - \frac{\beta_1^2}{4\beta_2}) \end{aligned}$$

For later reference, we mention that these maximum detection probabilities $g_j(x_0)$ are equivalent to the parameters c_j for observer $j = 1, 2$ in Quang and Becker (1997) defined in another very different model formulation.

2.6 Goodness-of-fit - loon data

The rational to evaluate the quality of model fit is to check whether, with a proposed model and the resulting parameter estimates, we are able to generate predictive distributions consistent with the observed sample data (GELMAN and HILL, 2007, p.513). That is, to compare the observed total number of detections by the front passenger only ($n_{1o} = 22$), by the rear passenger only ($n_{2o} = 18$) and by both passengers ($n_{bo} = 27$), with the three-dimensional predictive distribution induced by the posterior distribution.

Based on the given perpendicular distances x_i ($i = 1, \dots, M$) in the augmented population of size M and given the posterior sample of size 3000 for the parameter vector $\theta_{(s)}$ ($s = 1, \dots, 3000$), which for the most general models (M9 and M10) is $\theta_{(s)} = (\psi_{(s)}, \beta_{j0(s)}, \beta_{j1(s)}, \beta_{j2(s)})$ ($j = 1, 2$) we repeat the steps describe next for each s .

Generate M random quantities $w_{i(s)}$ from the Bernoulli distribution with parameter $\psi_{(s)}$. For each of the M subjects calculate the probability vector $\mu_{i(s)} = (\mu_{1i(s)}, \mu_{2i(s)}, \mu_{3i(s)}, \mu_{4i(s)})$ and simulate $Z_{i(s)} = (z_{1i(s)}, z_{2i(s)}, z_{3i(s)}, z_{4i(s)})$ from a multinomial distribution of size one and parameter $\mu_{i(s)}$. Obtain the sums $n_{r(s)} = \sum_{i=1}^M z_{ri(s)}$ for $r = 1, 2, 3$ which are the predicted numbers of subjects detected by observer one only ($n_{1(s)}$), observer two only ($n_{2(s)}$) and by both ($n_{3(s)}$). Hence, the predicted total sample size is $n_{(s)} = n_{1(s)} + n_{2(s)} + n_{3(s)}$.

The procedure outlined in the previous paragraph, is repeated for all 3000 simulated posterior parameter vectors $\theta_{(s)}$ to generate a predictive distributions to be confronted with observed mark-recapture data. A model with adequate fit is expected to display predictive distributions in line with actual observations.

Finally, two further model checks are used: (i) the empirical cumulative distribution function (ecdf) of observed distances is compared to ecdf-s build from predicted observed distances simulated for posterior parameter vectors $\theta_{(s)}$; (ii) the

observed sample size is compared to sample sizes of the posterior predicted samples,

3 Results

3.1 Simulated Data

All models were fitted with the logit and the probit links. However, based on *DIC* no link provides a uniformly superior fit. Therefore, we focus the description below only on the logit link, but retain in Tables and Figures the results for both.

Abundance estimation with distance data only (M1) has an average posterior standard deviation of 68, which reduces to an average of 31 when double observer data are included and a single detection function is assumed (M3) and reduces still further to 21 when observer-specific detection functions are assumed (M5)(Table 3a). However, using the posterior means of the 20 replicas we observe (*rb*) that models M1 and M3 are positively biased by 9% of the true value while model M5 is biased negatively by about 12% (Table 3). A compromise between bias and precision is best described by the root mean square error (*rmse*) calculated over the replicated posterior means (Table 4). Based on the smallest *rmse* (33.76) model M5 is the best among these three models. Finally, regarding the coverage provided by the posterior 95% credibility intervals (CrI95), the true parameter N was covered by 16 out of 20 replicas for both DOLT models M3 and M5 (Fig. 2). This coverage is below the 19 out of 20 as announced by the credibility interval.

Model comparisons with *DIC* consistently identified the model with observer-specific detection function M5 as better (i.e. lower *DIC*) than the model that assumes a single detection curve M3 (Table 3). This consistency also holds for models with probit link (i.e., M6 is better than M4).

In all DOLT models the estimates of $g(0)$ have a good performance. For the observer-specific model M5, these estimates have a positive bias of 5% and 8% for observers one and two, respectively (Table 3). Hence, they are less biased than estimates of abundance described previously. Model M3, that assumes a single detection function is also quite able to estimate the average between the true values (0.8 and 0.5) for both observers, which is $g(0) = 0.65$ (Table 3). Regarding coverage of the 95% credibility intervals, 17 and 18 out of 20 are the success rates for observers one and two in model M5 (Fig. 3).

3.2 Case Study

All linear and quadratic *DOLT* models listed in Table 2 were fitted to the loon data; summaries are in Table 5. The models with smallest *DIC* are M8 and M10 which include a quadratic term for observed distances and uses the probit link. However, both models are indistinguishable regarding predictive performance, since *DICs* differ only by one unit. In fact, differences in abundance estimates N are negligible since they have posterior mean 156 (CrI95: 128 to 189) and 154 (CrI95:

Abundance estimate to modelos front M1 to M6

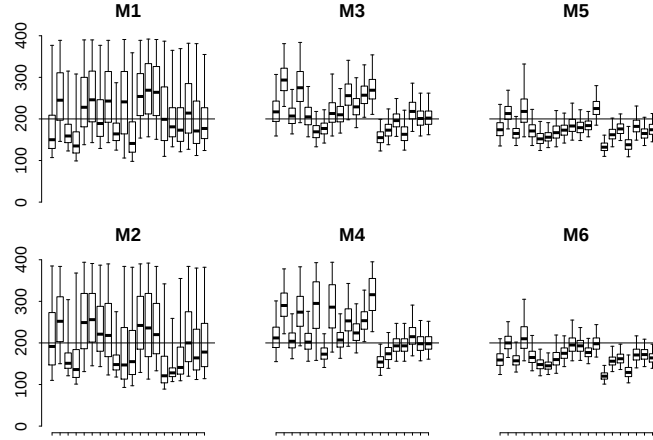


Figure 2 - Box-plot-type posterior distribution summaries of abundance (N) for models M1 to M6. CrI50 (boxes), CrI95 (whiskers), median (inbox horizontal line), true parameter (horizontal line)

125 to 185) for models M8 and M10, respectively. Nevertheless, these estimates exceed in about 35% the estimated 99 individuals obtained by Quang and Becker (QB). For further comparisons with results by QB, we focus on model M10 which retains observer-specific detection functions as they did.

In model M10 the observer-specific detection probabilities at distance zero ($g_1(0)$, $g_2(0)$) and the detection probabilities at the distances of maximum detection probabilities ($g_1(x_{10})$, $g_2(x_{20})$) are higher than the corresponding estimates reported by QB (Table 4; 5; Fig. 4c and 4d). However, while posterior standard deviations in the former estimates are similar to the asymptotic standard errors (ASE) presented by QB, our values are much smaller for the latter where posterior standard deviation are close to 0.05 while the ASEs are above 0.25 (Table 4).

Goodness-of-fit diagnostics suggest that model M10 is less capable of adequately predicting the observed number of mark-recapture data than the simpler model M6 (Fig. 5 and 6). The graphic displays suggest that M10 predicts larger sample sizes of detected loon than the 67 individuals that were actually observed. Only with regard to the predicted distribution of distances (Fig. 6) M10 seems to be slightly superior. This results are surprising because M10 has a much smaller

Table 3 -
 (a) Posterior means of abundance, (N); standard deviation, (sd); relative bias, (rb) and DIC.
 (b) Posterior means of $g(0)$; standard deviation (sd); relative bias (rb) for M1 to M4; and $g_1(0)$ and $g_2(0)$ for M5 and M6.

a	M1				M2				M3				M4				M5				M6			
	N	sd	rb	DIC	N	sd	rb	DIC	N	sd	rb	DIC	N	sd	rb	DIC	N	sd	rb	DIC	N	sd	rb	DIC
t1	180	74	-0.10	4226	214	81	0.07	4623	220	38	0.10	821	216	38	0.08	839	177	25	-0.12	665	161	22	-0.20	656
t2	255	71	0.28	2696	258	67	0.29	2454	296	39	0.48	1113	293	40	0.47	1180	216	24	0.08	1005	202	22	0.01	953
t3	172	45	-0.14	2376	165	45	-0.17	2550	209	28	0.05	1013	207	28	0.04	1023	167	17	-0.17	837	159	18	-0.20	948
t4	153	53	-0.23	3480	165	72	-0.17	5477	279	51	0.40	1194	278	50	0.39	1156	224	44	0.12	1303	216	39	0.08	1184
t5	243	74	0.21	3314	254	78	0.27	3384	208	31	0.04	992	206	31	0.03	984	173	23	-0.13	927	168	21	-0.16	892
t6	255	73	0.27	3051	261	72	0.31	2831	170	22	-0.15	772	289	69	0.44	738	154	18	-0.23	749	149	16	-0.25	704
t7	210	68	0.05	3520	238	70	0.19	3259	178	20	-0.11	771	175	20	-0.12	787	158	15	-0.21	708	146	14	-0.27	679
t8	253	74	0.27	2813	234	80	0.17	3745	219	38	0.09	1151	288	63	0.44	996	170	23	-0.15	908	164	25	-0.18	1021
t9	175	41	-0.13	1719	160	40	-0.20	2030	213	27	0.06	888	209	27	0.05	878	174	19	-0.13	816	176	19	-0.12	800
t10	242	86	0.21	3820	183	88	-0.09	6060	260	38	0.30	1150	257	38	0.29	1192	186	23	-0.07	1045	198	25	-0.01	1063
t11	168	70	-0.16	3895	188	84	-0.06	4854	231	31	0.16	1005	227	30	0.14	1001	181	19	-0.10	863	194	20	-0.03	825
t12	260	65	0.30	2220	250	77	0.25	3453	259	32	0.30	815	256	32	0.28	827	185	16	-0.08	623	178	16	-0.11	635
t13	273	70	0.36	2178	243	85	0.22	3935	273	36	0.36	990	316	48	0.58	922	227	24	0.13	882	200	20	-0.00	862
t14	267	71	0.34	2242	236	74	0.18	3026	157	20	-0.22	608	155	19	-0.23	614	133	13	-0.33	492	121	11	-0.40	483
t15	217	82	0.09	4014	149	67	-0.26	5007	175	22	-0.13	784	176	22	-0.12	804	164	18	-0.18	708	157	16	-0.21	650
t16	201	61	0.00	2981	138	36	-0.31	2220	198	23	-0.01	782	195	23	-0.02	799	177	17	-0.12	658	164	16	-0.18	665
t17	195	67	-0.02	3781	170	67	-0.15	5202	166	25	-0.17	720	164	26	-0.18	766	140	19	-0.30	673	131	17	-0.34	653
t18	228	74	0.14	4000	220	79	0.10	4838	221	31	0.11	941	219	31	0.10	973	184	22	-0.08	857	173	20	-0.13	839
t19	198	76	-0.01	4148	193	75	-0.04	4646	204	26	0.02	798	200	25	0.00	784	166	17	-0.17	670	173	17	-0.13	630
t20	197	61	-0.01	3230	203	77	0.02	4893	205	25	0.02	877	201	23	0.00	838	175	18	-0.12	771	165	15	-0.18	709
mean	217	68	0.09	3185	206	71	0.03	3924	217	31	0.09	909	226	34	0.13	905	176	21	-0.12	808	170	19	-0.15	793

b	M3			M4			M5			M6		
	g(0)	sd	rb	g(0)	sd	rb	g1(0)	sd	rb	g1(0)	sd	rb
t1	0.62	0.08	-0.046	0.59	0.08	-0.095	0.88	0.06	0.098	0.42	0.08	-0.160
t2	0.60	0.07	-0.071	0.57	0.07	-0.116	0.84	0.06	0.046	0.45	0.08	-0.105
t3	0.66	0.07	0.019	0.65	0.07	-0.002	0.83	0.07	0.038	0.50	0.09	-0.004
t4	0.46	0.09	-0.298	0.44	0.08	-0.326	0.66	0.11	-0.172	0.30	0.08	-0.397
t5	0.64	0.07	-0.021	0.62	0.08	-0.046	0.74	0.09	-0.079	0.51	0.09	0.029
t6	0.73	0.06	0.127	0.53	0.48	-0.182	0.78	0.07	-0.027	0.68	0.08	0.361
t7	0.76	0.06	0.163	0.75	0.06	0.147	0.91	0.05	0.133	0.60	0.08	0.209
t8	0.56	0.08	-0.144	0.54	0.40	-0.177	0.84	0.08	0.046	0.38	0.08	-0.238
t9	0.71	0.06	0.090	0.70	0.06	0.078	0.77	0.07	-0.032	0.63	0.08	0.257
t10	0.59	0.07	-0.088	0.57	0.07	-0.123	0.77	0.08	-0.041	0.41	0.08	-0.180
t11	0.67	0.07	0.030	0.65	0.07	0.003	0.85	0.06	0.064	0.51	0.08	0.022
t12	0.78	0.06	0.192	0.74	0.06	0.144	0.92	0.04	0.151	0.62	0.08	0.243
t13	0.70	0.06	0.081	0.62	0.38	-0.046	0.78	0.07	-0.030	0.63	0.08	0.254
t14	0.80	0.05	0.233	0.79	0.06	0.216	0.96	0.03	0.196	0.63	0.08	0.253
t15	0.73	0.07	0.124	0.72	0.07	0.109	0.90	0.05	0.121	0.59	0.09	0.181
t16	0.77	0.06	0.182	0.75	0.06	0.159	0.93	0.03	0.165	0.60	0.08	0.200
t17	0.69	0.08	0.057	0.67	0.08	0.031	0.83	0.08	0.039	0.57	0.10	0.137
t18	0.67	0.07	0.031	0.65	0.07	0.004	0.82	0.07	0.020	0.54	0.09	0.071
t19	0.75	0.06	0.153	0.73	0.06	0.123	0.89	0.05	0.108	0.63	0.08	0.259
t20	0.74	0.06	0.136	0.73	0.06	0.118	0.88	0.05	0.102	0.59	0.08	0.189
mean	0.68	0.07	0.047	0.65	0.12	0.001	0.84	0.06	0.047	0.54	0.08	0.079

Table 4 - Summaries over 20 replicated simulations of the posterior mean for abundance N : average (mean); standard deviation among posterior means (sdpm); root mean square error (rmse) and average DIC among replicates (DIC_m).

	M1	M2	M3	M4	M5	M6
mean	217.264	206.169	217.003	226.419	176.464	169.809
sdpm	37.134	39.866	39.544	47.167	24.197	23.927
rmse	40.951	40.340	43.044	54.062	33.755	38.523
DIC_m	3185	3924	909	905	808	793

DIC than M6 and therefore one would expect to observe features in support of a better fit.

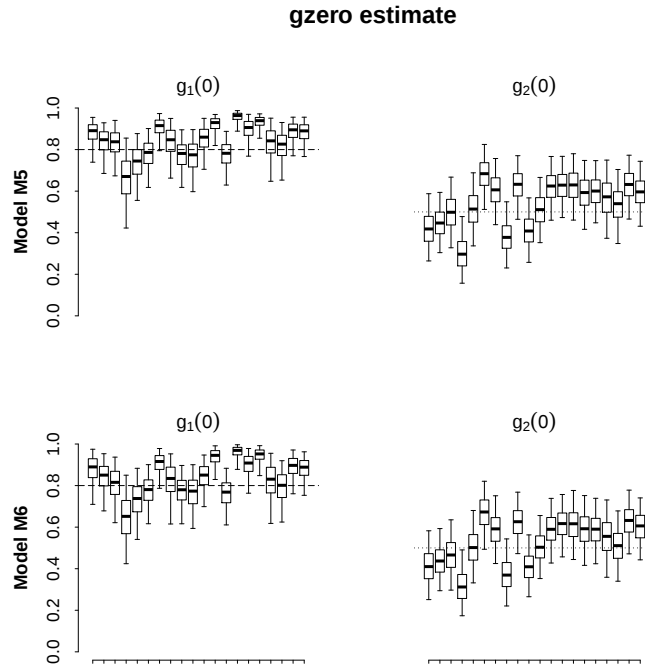


Figure 3 - Box-plot-type posterior distribution summaries of detection probability at distance zero for observer 1 ($g_1(0)$) and observer 2 ($g_2(0)$) for simulated data fitted to models M5 and M6. CrI50 (boxes), CrI95 (whiskers), median (inbox horizontal line), true parameter (dotted horizontal line)

Table 5 - Posterior mean and sd for abundance (N); detection probability at distance zero for single detection function or for observer 1, ($g_1(0)$) and for observer 2, ($g_2(0)$); maximum detection probability at distance x_0 for single detection function or for observer 1, ($g_1(x_0)$) and for observer 2, ($g_2(x_0)$). Last line (QB) reproduces maximum likelihood estimates and asymptotic standard error from Quang and Becker (1997).

	N	sd	$g_1(0)$	sd	$g_2(0)$	sd	$g_1(x_0)$	sd	$g_2(x_0)$	sd	DIC
M3	84	7.38	0.440	0.124	-	-	-	-	-	-	477
M4	85	7.72	0.442	0.122	-	-	-	-	-	-	498
M5	82	6.84	0.396	0.091	0.531	0.096	-	-	-	-	453
M6	83	6.85	0.338	0.136	0.541	0.148	-	-	-	-	454
M7	144	14.11	0.224	0.056	-	-	0.852	0.040	-	-	278
M8	156	15.58	0.124	0.056	-	-	0.902	0.041	-	-	251
M9	141	14.19	0.202	0.068	0.302	0.103	0.876	0.052	0.932	0.052	287
M10	154	15.12	0.095	0.058	0.193	0.094	0.879	0.054	0.924	0.056	252
QB	99	6.34	0.110	0.080	0.162	0.089	0.754	0.256	0.732	0.283	-

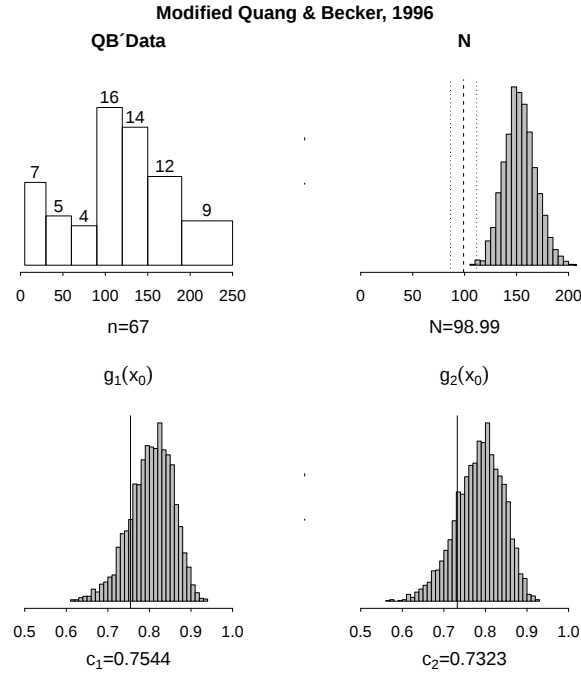


Figure 4 - Top left: histogram of observed data by QB; Top right: posterior distribution for abundance N and the corresponding estimate with confidence range by QB [vertical lines]; Bottom left: posterior distribution of highest detection probability for observer 1 [$g_1(x_0)$] and the estimate of this parameter (c_1) by QB [vertical line]; Bottom right: posterior distribution of highest detection probability for observer 2 [$g_2(x_0)$] and the estimate of this parameter (c_2) by QB [vertical line]. All posterior distributions refer to model M10.

4 Discussion

4.1 Simulated Data

Our simulations confirm similar findings already reported by other authors that double observer mark-recapture data, when combined with perpendicular distances from the detected subject to the transect line, enables the estimation of $g(0)$ and increases the accuracy in abundance estimates (BORCHERS, 1999; KING, 2014). Hence, the double-observer data allowed for successful observer-specific estimation of the imperfect on-line detection probability $g(0)$ (Table 3b, Fig. 3). The literature on distance sampling uses both terms $g(0)$ and perception bias ($1 - g(0)$) and some care is necessary when comparisons are intended. For instance, on-line detection probability of $g(0) = 0.8$ translates to a perception bias of 20%. Of course, perception bias is zero and can be ignored when $g(0) = 1$ as in standard distance sampling.

Furthermore, in standard distance sampling maximum detection is presumed

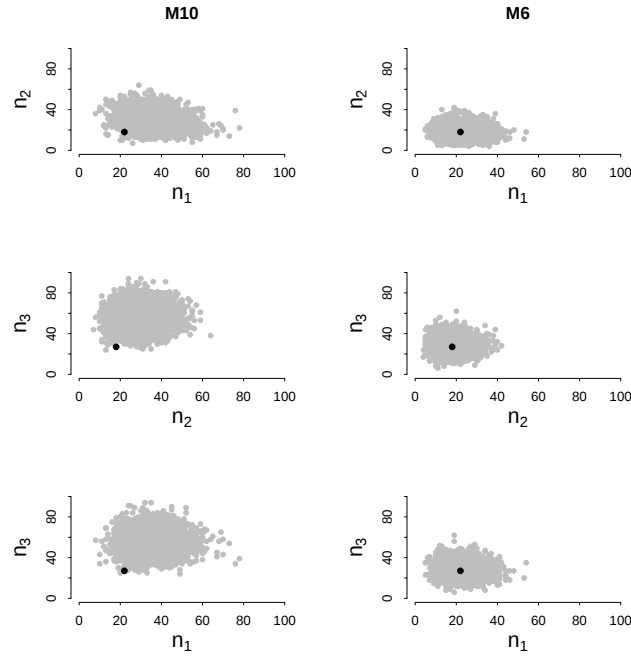


Figure 5 - Predictive multivariate distribution under models M10 (left column) and M6 (right column) [grey dots]. n_1 - number of individuals detected by observer one only; n_2 - by observer two only; n_3 - by both observers. The observed numbers (n_{1o}, n_{2o}, n_{3o}) are displayed with black dots.

to occur on the transect line ($x = 0$) and model with the shape of half-normal or hazard-rate detection functions (BORCHERS, BUKLAND and ZUCCHINI, 2002; FEWSTER et al., 2009; EDEKOVEN et al., 2013). But this maximum will be found at some positive distance $x_0 > 0$ when the detection function is assumed Weibull (QUANG and BECKER, 1997) or gamma (BECKER and QUANG, 2009). In contrast, we allow for an off transect line maximum detection simply by adding a quadratic term into de logit and probit regressions of distance. The important point in all this being that, with distance data only, parameter identifiability of detection functions in all these models are only possible when $g(x_0) = 1$ is assumed at some fixed $x_0 \geq 0$ (Z AHL, 1989).

In the simulation study the use of a logit or probit regression to model detection as a function of distance proved succesful although these were not the models that actually generated the data. With the chosen modeling strategy we took a quite distinct approach from most of the distance sampling literature (e.g. BORCHERS, 1999; BECKER and QUANG, 2009). To our knowledge, only Conn, Laake and Johnson (2013) used a somewhat similar modeling approach with a multivariate probit transformed detection function. However they used a much more complex

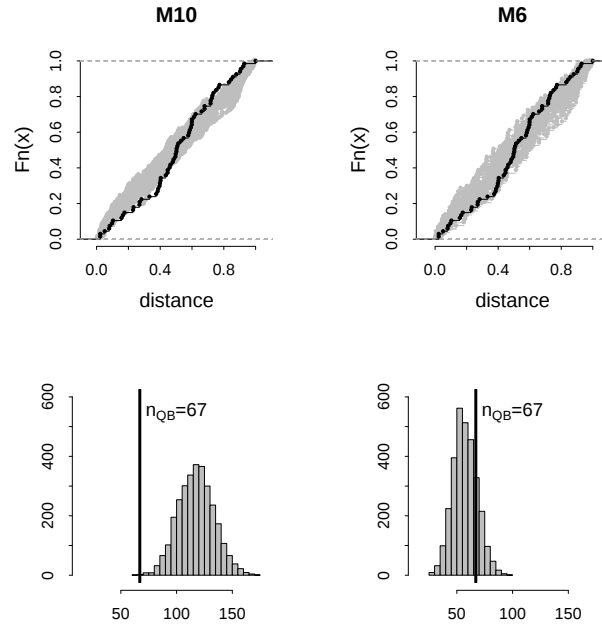


Figure 6 - Top row: A random sample of 16 empirical cumulative distribution functions (ecdf) [grey] of predicted distances and the observed ecdf [black]; Bottom row: predictive distribution of total number of detected loons [grey histogram] and observed number (67) [black vertical line]. Left column: M10 and Right column: M6.

reversible jump mcmc algorithm to obtain posterior distributions.

4.2 Case Study

The inclusion of a quadratic term into the logit and probit regressions of detection probability by distance, resulted in a substantial reduction in DIC dropping from a smallest value of 454 for the best linear-term-only model M6 to 252 when a quadratic term was added (M10) (Table 5). This is not surprising since the data suggest a mode at some intermediate distance away from zero (Fig. 4a). The posterior distributions obtained with model M10, estimate these modes at 140m (CrI95 123.2m to 163.5m) and 118m (CrI95 102.7m to 135.7m) for both observers. Using a completely distinct model structure and maximum likelihood inference, QB estimates them at 135.05m and 122.82m respectively. In conventional distance sampling, the requirement of maximum detection on the transect line forces all distances smaller than the mode to be discarded and the remaining distances rescaled; a procedure usually known as data truncation (LAAKE, 1999; BORCHERS, BUCKLAND and ZUCCHINI, 2002; ANDRIOLO et al., 2010). This

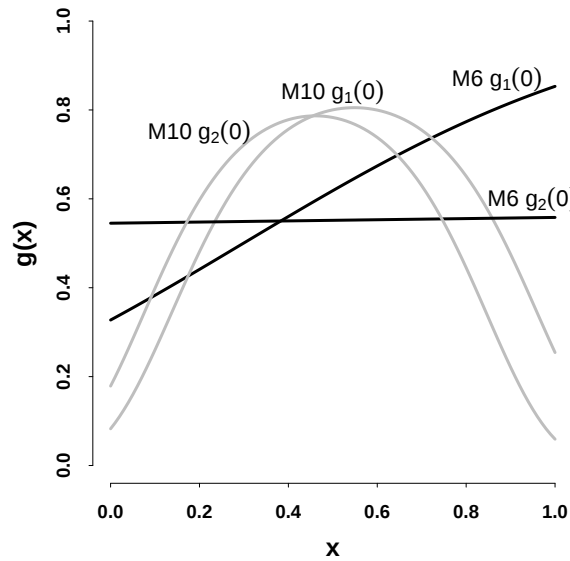


Figure 7 - Estimated detection functions for both observers under models M10 [grey] and M6 [black], when the probit regression parameters are fixed at the posterior means.

is unfortunate because it requires the exclusion of data that are often hard and expensive to collect.

Although the estimated distances of maximum detection x_0 show good agreement with QB, the estimated detection probabilities at these distances are lower in QB when compared to ours, although within acceptable ranges since both are covered by our posterior 95% credibility intervals (Table 4, Fig. 4c and 4d). In contrast, abundance estimates cannot be reconciled as our posterior mean of 154 (CrI95 128 to 185) is much higher than the estimated 99 individuals reported by QB. It is surprising that, while QB infer a perception bias around 25% that is higher than our inferred value around 10%, their abundance estimate is smaller. This fact suggests that, with the inclusion of a quadratic term into our (probit- or logit-) regression models, the relationship between abundance and perception bias becomes less obvious than naive intuition would indicate. It is also to be noted that in all quadratic models the probit regressions (M8 and M10) outperform their logit equivalents (M7 and M9, respectively).

In the goodness-of-fit analysis we confronted the most complete quadratic model M10 to the most complete linear model M6. Based on posterior predictions, the visual examination of both Figures suggest that the observed number of

detections was lower than model M10 would have predicted, being in better agreement with M6 (Fig. 5 and 6). These findings contradict the ranking obtained by DIC, which might have been affected by missing distances induced by data augmentation (CELEUX et al, 2006). Finally, in Fig. 7 we see that the detection probabilities estimated by M10 are reasonable in light of the data, while those for M6 are much less so and might be difficult to justify in practice. We think these aspects of model goodness-of-fit diagnostics need further investigation as it is unclear at this time how to reconcile these surprising results of lower DIC associated to worse lack-of-fit.

Finally, the simulation study has shown that the proposed model and its implementation with data augmentation in BUGS code, is a workable and comparatively easy way to model distance data with imperfect maximum detection probability. Furthermore, the case study has shown another versatility of this model by allowing for maximum detection off the transect line by simply including a quadratic term into the model. With this simple model extension, data loss due to truncation can be avoided. This can represent a critical advantage for elusive populations where samples are hard to get and each datum contains valuable information.

Acknowledgments

This work was developed as part of graduate studies in Biological Oceanography at FURG (Universidade Federal do Rio Grande) with a scholarship provided to the first author by Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). The authors thank an anonymous reviewer for comments that help to improve the manuscript.

References

- 1 ANDRIOLO, A.; KINAS, P.G.; ENGEL, M.H.; MARTINS, C.C.A.; RUFINO, A.M. *Humpback whale within the brazilian breeding ground: distribution and population size estimate*. Endangered Species Research, 11:233-243, 2010. doi 10.3354/esr00282.
- 2 BAILEY, L. L.; HINES, J. N. J.; MacKENZIE, D. *Sampling design trade-offs in occupancy studies with imperfect detection: examples and software*. Ecological Applications, 17(1):281-290, 2007.
- 3 BECKER, E.; QUANG, P. *A Gamma-Shaped Detection Function for Line-Transect Surveys With Mark-Recapture and Covariate Data*. Journal of Agricultural, Biological, and Environmental Statistics, 14(2):207-223, 2009. doi: 10.1198/jabes.2009.0013.
- 4 BERLINER L. Hierarchical Bayesian time series models, *Proceeding of the XVth Workshop on Maximum Entropy and Bayesian Methods*.1996. Accessible at <http://> (Accessed: 03/10/2015)
- 5 BOCHERS, D. L. *Composite mark-recapture line transect surveys*. Marine Mammal Survey and Assessment Methods. Garner et al. (eds). Balkema, Rotterdam: Brookfield, 1999. 290p. ISBN 990 5809 043 4.
- 6 BORCHERS, D. L.; BUCKLAND, S.; ZUCCHINI, W. *Estimating animal abundance*. Great Britain: Springer, 2002. 314p.

- 7 BUCKLAND, S. T.; LAAKE, J. L. and BORCHERS, D. L. *Double-Observer Line Transect Methods: Levels of Independence*. Biometrics 66:169–177, 2010.
doi: 10.1111/j.1541-0420.2009.01239.x.
URL <http://digitalcommons.unl.edu/usdeptcommercepub/199>. (Accessed 10/10/2015).
- 8 CELEUX, G.; FORBES, F. R. C.; TITTERINGTON, D. *Deviance Information Criteria for Missing Data Models*. International Society for Bayesian Analysis, 4:651-674, 2006.
- 9 CLARK, J. *Why environmental scientists are becoming Bayesians?*. Ecology Letters, 8:2-14, 2005.
doi: 10.1111/j.1461-0248.2004.00702.x.
- 10 CONN, P. B.; LAAKE, J.; JOHNSON, D. *A Hierarchical Modeling Framework for Multiple Observer Transect Surveys*. PLoS ONE, 7(8), 2012: e42294. doi:10.1371/journal.pone.0042294.
- 11 DALLA ROSA, L.; FORD, J.K.B.; TRITES, A. W. *Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters*. Continental Shelf Research, (36): 89–104, 2012.
doi:10.1016/j.csr.2012.01.017.
- 12 DURBAN, J. W.; ELSTON, D. A.; ELLIFRIT, D. K.; DICKSON, E.; HAMMOND, P. S.; THOMPSON, P. M. *Multisite mark-recapture for cetaceans: population estimates with Bayesian model averaging*. Marine Mammal Science, 21(1):80-92., 2005.
- 13 EGUCHI, T.; GERRODETTE, T. *A Bayesian approach to line-transect analysis for estimating abundance*. Ecological Modelling, 220:1620-1630, 2009
doi:10.1016/j.ecolmodel.2009.04.011.
- 14 FEWSTER, R. M.; BUCKLAND, S. T.; BURNHAM, K. P.; BORCHERS, D. L.; JUPP, P. E.; LAAKE, J. L.; THOMAS, L. *Estimating the Encounter Rate Variance in Distance Sampling*. Biometrics, 65:225–236, 2009.
doi: 10.1111/j.1541-0420.2008.01018.x.
- 15 GELMAN, A.; HILL, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 5.ed. New York: Cambridge University Press, ISBN-13 978-0-511-26878-6 eBook (EBL)., 2007. 645.p
- 16 JAYNES, E. T. *Probability Theory - The logic of science*. Cambridge: Cambridge University Press, 2003. 727p.
- 17 KARUNAMUNI, R.; QUINN, I. T. *Bayesian Estimation of Animal Abundance for Line Transect Sampling*. Biometrics, 51(4):1325-1337, 1995.
URL <http://www.jstor.org/stable/2533263>. (Accessed: 08/01/2015).
- 18 KINAS, P. G.; ANDRADE, H. *Introdução a análise bayesiana (com R)*. 1.ed. Porto Alegre: MaisQnada editora, 2010. 240p.
- 19 KING, R.; BROOKS, S. P. *On the Bayesian Estimation of a Closed Population Size in the Presence of Heterogeneity and Model Uncertainty*. Biometrics, 64: 816–824, 2008. doi: 10.1111/j.1541-0420.2007.00938.x.
- 20 KING, R. *Statistical Ecology*. Annual Review of Statistics and Its Application, 1:401-426, 2014.
doi: 10.1146/annurev-statistics-022513-115633.
- 21 LAAKE, J. *Distance sampling with independent observers: Reducing bias from heterogeneity by weakening the conditional independence assumption*. Marine Mammal Survey and Assessment Methods. Garner et al. (eds). Balkema, Rotterdam: Brookfield, 1999. 290p.
ISBN 990 5809 043 4.
- 22 LUNN, D.; SPIEGELHALTER, D.; THOMAS, A.; BEST, N. *The BUGS project: Evolution, critique and future directions*. 2009. Published online in Wiley InterScience (urlprefixwww.interscience.wiley.com).
doi: 10.1002/sim.3680. (Accessed: 04/06/2015).
- 23 MARSH, H. A.; SINCLAIR, D.F. *An Experimental Evaluation of Dugong and Sea Turtle Aerial Survey Techniques*. Australian Wildlife Research, 16:639-650, 1989.
- 24 MARTIN, A.; QUINN, K. *Applied bayesian inference in R using MCMC-pack*. R News, 6(1):2-7, 2006.

- 25 MCCARTHY, M. *Bayesian Methods for Ecology*. 1.ed. New York: Cambridge University Press, ISBN-13 978-0-521-61559-4., 2007. 305p.
- 26 OEDEKOVEN, C. S.; BUCKLAND, S. T.; MACKENZIE, M. L.; KING, R.; EVANS, K. O.; BURGER, Jr L. W. *Bayesian Methods for Hierarchical Distance Sampling Models*. Journal of Agricultural, Biological, and Environmental Statistics, 2014. doi: 10.1007/s13253-014-0167-0.
- 27 PLUMMER, M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, ISSN 1609-395X*. Vienna, Austria. p.1-10. 2003.
URL <http://www.ci.tuwien.ac.at/Conferences/DSC-2003>. (accessed:10/11/2014).
- 28 PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. *CODA: Convergence diagnosis and output analysis for MCMC*. R News, 6(1):7-11, 2006.
- 29 PLUMMER, M. *JAGS Version 3.3.0*. 2012.
URL <http://people.math.aau.dk/~kbb/undervisning/Bayes13/sorenh/doc>. (Accessed: 05/07/2013).
- 30 PLUMMER, M. *Package "rjags": Bayesian graphical models using MCMC*. 2013.
URL <http://cran-r.c3sl.ufpr.br/web/packages/rjags/rjags.pdf>. (Accessed: 20/01/2014).
- 31 QUANG, P.; BECKER, E. *Line Transect Sampling Under Varying Conditions with Application to Aerial Surveys*. Ecological Society of America, 77(4):1297-1302, 1996.
URL <http://www.jstor.org/stable/2265601>.
- 32 QUANG, P.; BECKER, E. *Combining Line Transect and Double Count Sampling Techniques for Aerial Surveys*. Journal of Agricultural, Biological and Environmental Statistics, 2(2):230-242, 1997.
URL <http://www.jstor.org/stable/1400405>. (Accessed: 18/10/2012).
- 33 QUANG, P.; BECKER, E. Aerial survey sampling of contour transects using double-count and covariate data. In: Marine Mammal Survey and Assessment Methods, 25-27 February 1998, Seattle, Washington, USA. *Proceedings of the Symposium on Surveys, Status and Trends of Marine Mammal Populations* Washington: Seattle, 1998. p.1-11.
- 34 R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria, 2014.
URL <http://www.R-project.org/>. (Accessed:22/01/2014)
- 35 RESNIK, P.; HARDISTY, E. *Gibbs sampling for the uninitiated*. CS-TR-4956, UMIACS-TR-2010-04, LAMP-TR-153, 2010.
URL <http://www.umiacs.umd.edu/~resnik/pubs/LAMP-TR-153.pdf>.
- 36 ROYLE, J.; DORAZIO, R. *Hierarchical modeling and inference in ecology, the analysis of data from populations, metapopulations and communities*. 1.ed. USA: Elsevier, 2008. 444p.
- 37 SCHOFIELD, M.; BARCKER, R. *Data augmentation and reversible jump MCMC for multinomial index problems*. 2010.
URL <http://arxiv.org/abs/1009.3507v1>.
- 38 SPIEGELHALTER, D. J.; BEST, N.; CARLIN, B.; van der LINDE, A. *Bayesian measures of model complexity and fit*. Journal of the Royal Statistical Society, Series B, 64:583-640, 2002.
- 39 SU, Y-S. and YAJIMA, M. *Package "R2jags": A Package for Running 'JAGS' from R*. 2015.
URL <http://cran-r.c3sl.ufpr.br/web/packages/R2jags/R2jags.pdf>.
- 40 WADE, P. *Bayesian methods in conservation biology*. Conservation Biology, 14(5):1308-1316, 2000.
- 41 ZAHL, S. *Line Transect Sampling with Unknown Probability of Detection Along the Transect*. Biometric, 45:453-470, 1989.

Received in 0x.0x.20xx.

Approved after revised in 0x.0x.20xx.

Viés de Percepção na Estimativa da Abundância de Toninhas na Baía da Babitonga, Brasil

Fernando Roberto Guilherme-Silveira^{1*}, Federico Sucunza Perez³, Alexandre N. Zerbini³, Daniel Danilewicz³, Artur Andriolo³, Paul Gerhard Kinas^{2*},

1 Instituto de Oceanografia/Programa de Pós-Graduação em Oceanografia Biológica/Universidade Federal do Rio Grande-FURG, Rio Grande, RS, Brazil

2 Instituto de Matemática, Estatística e Física/Universidade Federal do Rio Grande-FURG, Rio Grande, RS, Brazil

3 Universidade Federal de Juiz de Fora - UFJF, Juiz de Fora, Minas Gerais, Brazil

Resumo

Implementamos o método de amostragem de distância conjugando o método de transecção linear com o método de marcação e recaptura numa estrutura hierárquica com abordagem bayesiana. Nosso objetivo foi estimar conjuntamente densidade de toninhas na Baía da Babitonga e o viés de percepção. Para dados truncados, (DT), o melhor modelo estima a máxima probabilidade de detecção, (MPD) sobre a derrota (distância = 0). Para dados não truncados (DnT) o melhor modelo, que inclui um termo quadrático, estima as MDPs à distâncias maiores que zero. Os melhores ajustes indicam curvas MDPs distintas para cada observador tanto para DT quanto DnT. O ajuste para DnT evitou a eliminação de 26% dos dados e verificou que a distância média ideal para truncamento (123 m) é superior àquela utilizada em campo (87 m). A densidade foi $1.219 N_g km^{-2}$ (ICr95%: 0.739 a 2.075).

Abstract

We implemented into the same linear transect sampling framework Distance Sampling with Mark and Recapture methods. Our objective was to jointly estimate density of Fransciscana dophin at Babitonga Bay and the perception bias through a hierarchical statistical structure and bayesian approach. For truncated data (TD), the best model estimates the maximum detection probability (MDP) on the transect (distance = 0). For not truncated data, (nTD), the best model, which includes a quadratic term, estimates the MDPs for distances greater than zero. The best fit indicated distinct MDP curves by observer for both TD as nTD. The fitting for nTD avoided eliminating 26% of data and found that the ideal average distance for truncation (123 m) was greater than the distance used in the field (87 m). The density was $1.219 N_g km^{-2}$ (ICr95%: 0.739 a 2.075).

Introdução

O estuário da Baía da Babitonga, Sul do Brasil (Fig. 1), é a última grande formação de manguezal do Hemisfério Sul. Região de rica diversidade, áreas de criadouro, disponibilidade alimentar, abrigo e migração (Ibama, 1997). Apesar de sua relevância ambiental é palco do acelerado processo de ocupação desordenada e exploração de seus recursos. Sofre com a problemática da poluição proveniente do esgotos domésticos, rejeitos e efluentes industriais, defensivos agrícolas, fertilizantes e atividade portuária. Neste cenário e em meio a diversidade local, encontra-se a Franciscana (*Pontoporia blainvillei* - Gervais & d'Orbigny, 1844). É a única espécie da família *Pontoporiidae*, endêmica do Oceano Atlântico Sul Ocidental, cuja gestação dura em torno de onze meses, e a lactação nove (Danilewicz, 2003; Wells *et al.* 2013). Devido ao hábito predominantemente costeiro (Cremer & Simoes-Lope, 2005; Danilewicz *et al.* 2009; Pinedo *et al.* 1989), está em constante ameaça. Considerada o pequeno cetáceo mais ameaçado de extinção no Oceano Atlântico Sul Ocidental, sendo listada como “em perigo” no “Livro Vermelho da Fauna Brasileira Ameaçada de Extinção” e “vulnerable” na “IUCN Red List of Threatened Species” (Danilewicz *et al.* 2008; Reeves *et al.* 2012). Deste modo, é de extrema importância estimar o tamanho desta população para fins de gerenciamento e conservação.

Entretanto, estimar abundância de populações selvagens é uma tarefa de logística difícil e financeiramente dispendiosa, com resultados nem sempre satisfatórios, dado a incerteza na detecção e erros de medição (Kery & Royle, 2010; Royle, 2008). No ambiente marinho, o problema se agrava devido à distribuição da espécie em sua área de uso (Crespo *et al.* 1998; Siciliano, 1994) e o tempo de submersão (Beasley & Jeffersom, 2002; Crespo, 2009; Dolphin, 1987; Janik & Thompson, 1996). A capacidade de detecção está fortemente influenciadas por fatores ambientais (clima, ambiente, luz); pelo observador (fadiga, experiência) e pelo animal em si (camuflagem, tamanho, comportamento) (Marsh & Sinclair, 1989). Tudo isso leva a uma probabilidade positiva de não detecção dos animais, fenômeno denominado de “viés de visibilidade”. Marsh and Sinclair (1989) subdividiram este fenômenos em dois componentes: “viés de percepção”, referente à detecção imperfeita dos animais disponíveis; e o “viés de disponibilidade” referente à ausência dele no momento da amostragem.

Propostas convencionais para minimizar o viés de percepção, vinculam o método "Distance Sampling" (Borchers *et al.* 2002; Laake & Borchers, 2004) e de Marcação e Recaptura (Borchers *et al.* 2002; McKenzie *et al.* 2006) em uma só estrutura. Isto se dá, na prática, com a implementação de um segundo observador simultâneo e independente no procedimento de coleta de dados de distância

(Bailey *et al.* 2007; Conn *et al.* 2012; Quang & Becker, 1996; 1997; 1998).

Diferentemente de implementarmos a transecção linear com dupla observação em uma só verossimilhança, de forma convencional, optamos pela hierárquização do modelo de modo a separarmos em submodelos distintos a variabilidade ecológica, inerente ao meio, da variabilidade/incerteza oriunda do erro de medição, facilitando a correção do viés de percepção. Através do método de simulação de Monte Carlo com Cadeia de Markov (MCMC) (Martin & Quinn, 2006; McCarthy, 2007) e da técnica dos dados aumentados (Royle & Dorazio, 2008), estruturamos modelos hierárquico-bayesianos, que facilmente abordam o problema da correção do viés, bem como possibilita a incorporação de informações oriundas do ambiente, que auxilia na redução desse viés com aumento da precisão.

Portanto, ao sobrevoarem a Baía de Babitonga, dois observadores registrariam o número e a distância perpendicular dos grupos de toninhas até a linha percorrida pela aeronave. O nosso objetivo é estimar a abundância, corrigindo o viés de percepção.

Materiais e Métodos

Coleta de dados

Os dados foram coletados durante o período de 2011 por meio de sobrevoos na Baía da Babitonga (Fig. 1). A aeronave usada foi um "Aerocommander" 500B com asa alta; duas janelas bolhas na lateral, à frente e duas janelas planas na lateral atrás. Durante as transecções a aeronave voou a uma velocidade aproximada de 170-190 km/h e entre 150-152 metros de altura. Quatro observadores foram utilizados nas mesmas posições durante todo o período de coleta de dados. A independência auditiva foi obtida por meio do uso de auriculares, evitando que qualquer observador ouvisse as informações registradas pelos outros. A independência visual foi obtida por meio de anteparo entre os observadores da frente e os de trás. Os dados ambientais foram coletados por cada observador, de forma independente um do outro, no início de cada transecção ou havendo mudanças significativa das condições ambientais. As informações coletadas foram: latitude, longitude, transparência, reflexo no campo de visão, estado do mar ("beaufort"), ângulo de declinação entre o horizonte e o grupo avistado; presença de filhote e tamanho de grupo. Os ângulos foram obtidos por meio de clinômetros "Suunto" PM-5. Posteriormente, de posse do ângulo e da altura de voo, foi possível calcular as distâncias perpendiculares entre a transecção e os grupos observados (Lerczak & Hobbs, 1998). Os dados eram registrados em gravador digital e GPS sincronizados com os relógios dos quatro

observadores.

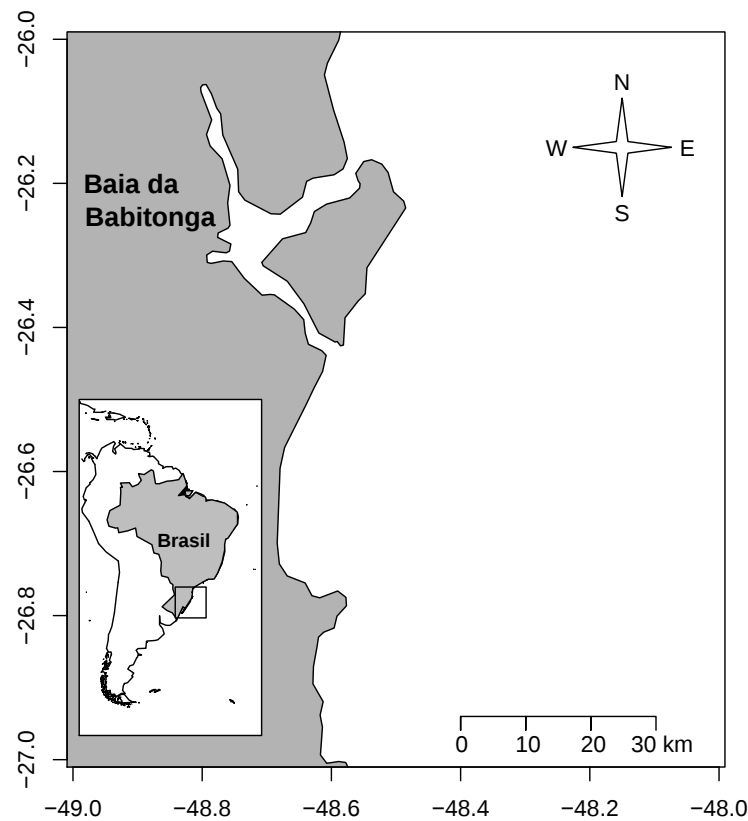


Figure 1: **Baía de Babitonga**. A área de estudo foi percorrida por uma aeronave em transecções previamente estabelecidas.

Análise de dados

Supondo independência entre os quatro observadores, agrupamos as informações dos dois observadores da frente, que fizeram uso das janelas bolhas e chamamos de primeiro observador. Agrupamos ainda as informações dos dois observadores de traz, que fizeram uso das janelas planas e chamamos de segundo observador. Adotou-se os seguintes critérios para a obtenção dos registros que identificassem os grupos detectados simultaneamente: 1) intervalo de tempo entre as detecções inferior ou igual a quinze segundos; 2) diferença entre ângulos de declinação registrados inferior ou igual a dez graus e 3) diferença nas estimativas de tamanho de grupo inferior ou igual a dois indivíduos (Perez, 2015). Assim, com o atendimento simultâneo dos três critérios, o grupo foi classificado como detectado por ambos os observadores.

Após o estabelecimento das detecções simultâneas, elaboramos uma matriz, cuja i -ésima linha é composta da distância d_i ; do vetor $Z_i = (z_{1i}, z_{2i}, z_{3i}, z_{4i})$ de comprimento quatro, que indica a forma de detecção de um grupo de animais, informando se o grupo i , foi detectado apenas pelo primeiro observador, $Z_i = (1, 0, 0, 0)$; apenas pelo segundo observador, $Z_i = (0, 1, 0, 0)$; por ambos, $Z_i = (0, 0, 1, 0)$ ou não detectado, $Z_i = (0, 0, 0, 1)$. A última coluna, z_4 está relacionada com a técnica dos dados aumentados e não se refere a dados efetivamente observados ao longo do estudo (Royle & Dorazio, 2008). O vetor Z_i tem uma distribuição de probabilidade Multinomial com vetor de parâmetros, $[\mu_{1i}, \mu_{2i}, \mu_{3i} \text{ e } \mu_{4i}]$. Na técnica dos dados aumentados, a amostra de n linhas correspondente aos grupos de animais detectados é aumentada com outras $M_g - n$ linhas correspondente a grupos não detectados e aos quais se atribui o vetor $Z_i = (0, 0, 0, 1)$. O valor M_g é fixado de modo que seja bem maior que o tamanho (desconhecido) de grupos N_g . Se $\psi = N_g \cdot M_g^{-1}$ denota a fração de M_g representada por N_g , então estimar N_g é equivalente a estimar ψ . Finalmente define-se a variável indicadora w_i tal que $w_i = 1$ para todas os N_g grupos e $w_i = 0$ para os demais $M_g - N_g$ grupos. Para maiores detalhes sobre a técnica de dados aumentados no contexto da estimativa de abundância, consulte Royle & Dorazio (2008). Logo, o modelo de observação multinomial (Royle & Dorazio, 2008, p.188) estruturado de forma hierárquica terá a seguinte forma.

$$\begin{aligned}
 Z_i &\sim \text{Multinomial}(1, [\mu_{1i}, \mu_{2i}, \mu_{3i}, \mu_{4i}]) \\
 \mu_{1i} &= p_{1i} \cdot (1 - p_{2i}) \cdot w_i \\
 \mu_{2i} &= (1 - p_{1i}) \cdot p_{2i} \cdot w_i \\
 \mu_{3i} &= p_{1i} \cdot p_{2i} \cdot w_i \\
 \mu_{4i} &= 1 - (\mu_{1i} + \mu_{2i} + \mu_{3i}) \\
 w_i &\sim \text{Bernoulli}(\psi)
 \end{aligned}$$

Quanto a p_{1i} e p_{2i} , estas são respectivamente as probabilidades de detecção do primeiro e do segundo observador associada à distância perpendicular padronizada x_i e ao i -ésimo indivíduo. Se d_0 denota a menor distância (usualmente $d_{\min} = 0$) e $d_{\max}$ a distância fixada como limite superior, então $x_i = \frac{d_i - d_{\min}}{d_{\max} - d_{\min}}$ padroniza a distância d_i ao intervalo $[0, 1]$. Ambos os conjuntos de valores, p_{1i} e p_{2i} , guardam uma relação com as respectivas distâncias perpendiculares, x_i , utilizando para isso das regressões logit e probit conforme detalhado nos modelos M1 e M2 na Tabela 1. Os modelos

M3 e M4, embora apresentem uma única curva de detecção para os dois observadores, incluem um termo quadrático que permite estimar a máxima detecção a uma distância positiva $x_0 > 0$. Finalmente os modelos M5 e M6 estendem a estrutura quadrática, permitindo curvas de detecção distintas para cada observador.

Table 1: Lista dos Modelos. Os números ímpares indicam os modelos, cuja função de ligação é logit e os números pares, probit. Rescrevemos com p_i a probabilidade de detecção, $g(x_i) = p_i$ apenas para melhor visualização do modelo. Enquanto x_i representa a covariável distância perpendicular do animal à linha de transecção.

	Modelo
M1	$\log \left[\frac{p_{ji}}{(1-p_{ji})} \right] = \beta_{j0} + \beta_{j1} \cdot x_i$
M2	$\Phi^{-1}(p_{ji}) = \beta_{j0} + \beta_{j1} \cdot x_i$
M3	$\log \left[\frac{p_{ji}}{(1-p_{ji})} \right] = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2$
M4	$\Phi^{-1}(p_{ji}) = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2$
M5	$\log \left[\frac{p_{ji}}{(1-p_{ji})} \right] = \beta_{j0} + \beta_{j1} \cdot x_i + \beta_{j2} \cdot x_i^2$
M6	$\Phi^{-1}(p_{ji}) = \beta_{j0} + \beta_{j1} \cdot x_i + \beta_{j2} \cdot x_i^2$

As n distâncias $x_i = (d_i - d_{min})(d_{max} - d_{min})^{-1}$ foram usadas para o ajuste dos modelos (Tab. 1) de duas formas distintas: 1) dados não truncados [$d_{min} = 0, d_{max} = 286$]. Aqui não houve restrições à esquerda; e 2) dados truncados [$d_{min} = 87, d_{max} = 286$]. Houve a eliminação de observações com distâncias inferiores a 87m. A faixa de distância [0-87m] representa uma "área de sombra" abaixo da aeronave, que, teoricamente, seria cega ao segundo observador, devido a janela ser plana. A distância de 286m, delimitaria o início de uma faixa com aumento no erro de medição (Borchers et al., 2002; Laake & Borchers, 2004) e a partir da qual não houveram detecções. Apesar de não haver restrições no uso das distâncias, incorporamos à análise os dados truncados para comparações.

A taxa de recaptura ("recapture rate - rr") é definida como probabilidade de ambos observadores detectarem um grupo de animais dado que este grupo foi detectado. Isto para cada distância observada x_i expressa como $rr(x_i)$ é função das probabilidades individuais de detecção.

$$rr(x_i) = Pr(z_{3i} = 1 | [z_{1i} + z_{2i} + z_{3i}] = 1, x_i) = \frac{p_{1i} \cdot p_{2i}}{p_{1i} + p_{2i} - p_{1i} \cdot p_{2i}}$$

Análise Bayesiana

A fim de manter uma estrutura matematicamente conveniente para a obtenção das posteriores e devido à inexistência de informações externas aos dados que fossem imprescindível ao modelo, foram

usadas *prioris* vagas e não informativas respectivamente para $\beta_{jk} \sim N(0, 1E5)$, onde ($j = 1, 2$ e $k = 0, 1, 2$) e $\psi \sim U(0, 1)$. As $M_g - n$ distâncias não observadas de d_i são consideradas como dados não observados ("missing values") no modelo para inferência e modelados convencionalmente com distribuições uniforme ($d_i \sim U(d_{min}, d_{max})$).

As distribuições posteriores foram obtidas por meio do método de simulação de Monte Carlo com Cadeia de Markov (MCMC) (Martin & Quinn, 2006; McCarthy, 2007). Usamos três cadeias, um "burn-in" de 10.000 passos e um "thinning" de 40, gerando uma amostra posterior de tamanho 3000. Diagnosticamos o nível de convergência das cadeias, usando a biblioteca "coda" (Plummer *et al.* 2010; Plummer, 2013). Para efetuar as simulações via MCMC, usamos as bibliotecas "R2jags" (Su & Yajima, 2015) e "rjags" (Plummer, 2013), que tem como algoritmo de desenvolvimento da cadeia o "JAGS - Just Another Gibbs Sampling" (Plummer, 2003; 2012). Tais bibliotecas foram executadas no aplicativo R (versão 3.1.1) ("R core team", 2014).

A seleção do melhor modelo (Tab. 1) para dados truncados e para dados não truncados, foi feita por meio do critério de DIC ("Deviance Information Criteria"), sendo o menor valor de DIC correspondente ao modelo com melhor ajuste (Spiegelhalter *et al.* 2002).

A comparação de modelos para dados truncados e não truncados, via DIC, não é possível devido tratar-se de amostras de tamanhos diferentes. Neste caso utilizou-se o coeficiente de variação (cv) das distribuições posteriores do desvio padrão e média das estimativas de abundância e probabilidade de detecção para efeito comparativo.

Para a estimativa da densidade, usamos a terça parte da área da Baía da Babitonga 1.567,00 km^2 (<http://www.ecobabitonga.org>, acesso em 15/01/2016).

Resultados

Os dados observados estão sumarizados na Tabela 2, bem como na Figura 2. Dados truncados representam 104 indivíduos observados e dados não truncados, 140. O truncamento reduziu em 26% o tamanho da amostra. É relevante observar (Fig. 2) que os dados não truncados não se comportam a semelhança de uma curva "half-Normal" ou "Hazard rate", ambas comumente usadas no "Distance Sampling" convencional (Borchers *et al.* 2002; Laake & Borchers, 2004), juntamente com a premissa de que a máxima probabilidade de detecção se dê à distância zero.

As estimativas posteriores para os parâmetros das curvas de detecção estão sumarizadas na Tabela 3 e ilustradas na Fig. 3.

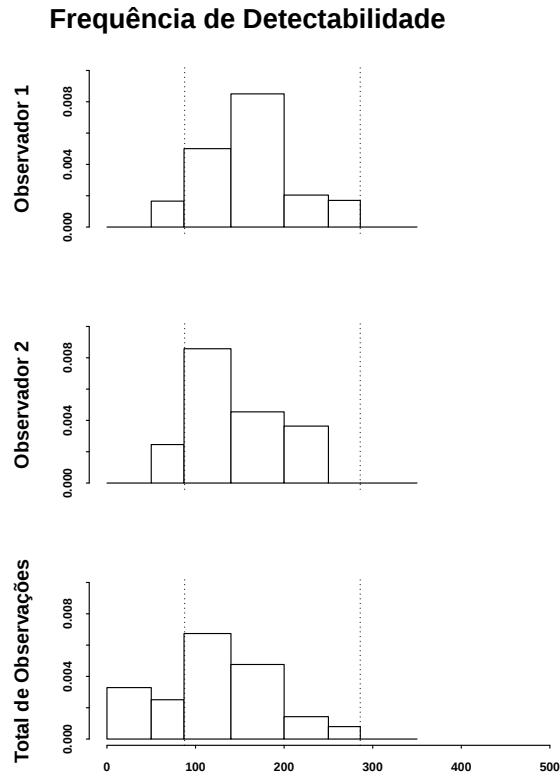


Figure 2: **Frequência de detectabilidade.** Frequência dos grupos observados pela distância (metros) para dados truncados (entre as linhas verticais tracejadas [87,286 m]) e dados não truncados.

Table 2: Número de grupos observados, para dados truncados e não truncados, pelo observador 1, (Obs1); observador 2, (Obs2); ambos os observadores (ambos); taxa de recaptura total (rr); frequência de tamanho de grupo, (de 1 até 6) e a média do tamanho de grupos, (mtg).

Dados	Obs1	Obs2	Ambos	rr	1	2	3	4	5	6	mtg
Truncados	48	46	10	0.0962	36	33	26	5	3	1	2.13
não Truncados	80	49	11	0.0786	49	43	37	6	4	1	2.11

Para dados truncados, o melhor modelo foi o M2 que apresentou o menor DIC (Tab. 3 a). Devido à ausência do termo quadrático em M2, a maior probabilidade de detecção se encontra sobre a linha de transecção, porém com valores diferentes para cada observador (Tab. 3 a). Com o truncamento dos dados a linha de transecção está a 87 m da derrota da aeronave. Quanto ao viés de percepção, estimado por esse modelo, é de 0.586 ($1 - g_1(0)$), para o observador 1 e 0.737 ($1 - g_2(0)$), para o observador 2.

Já para dados não truncados, o melhor modelo foi o M6, que inclui o termo quadrático (Tab. 3 b). Por consequência, ele estima as máximas probabilidades de detecção ($g_1(x_0)$; $g_2(x_0)$) a distâncias maiores que zero, ou seja, à distâncias de 85.2 m e 159.7 m para os observadores 1 e 2, respecti-

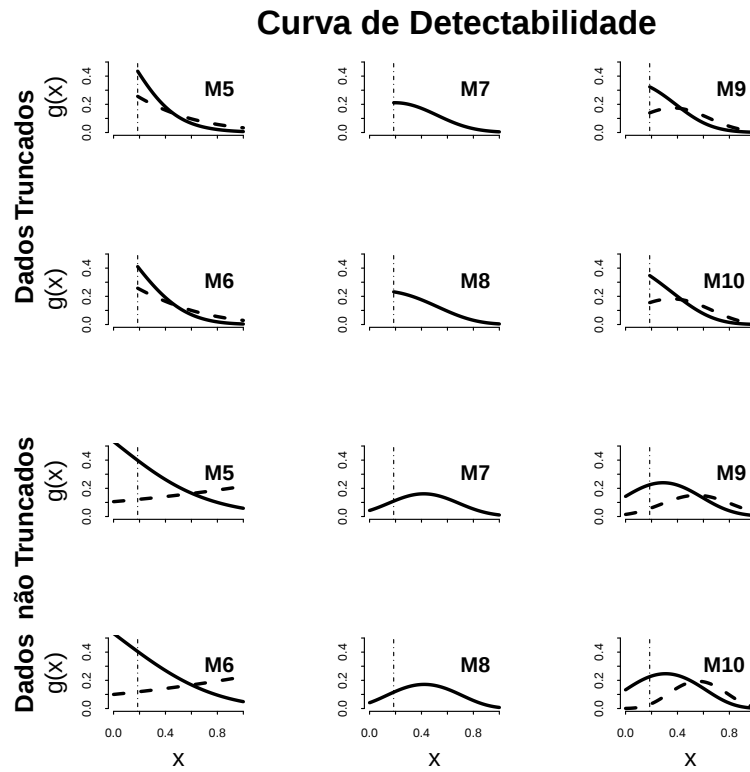


Figure 3: **Curva de detectabilidade estimada pelos modelos M1 até M6.** A linha contínua representa o observador 1 ou curva única para ambos, e a linha tracejada representa o observador 2. A linha vertical representa o truncamento (87m). No eixo x estão as distâncias mensurada para cada grupo observado e y as probabilidades de detecção ($g(x)$).

vamente (Tab. 3 b). O viés de percepção estimado pelo modelo M6 é de 0.745 ($1 - g_1(x_0)$) para o observador 1 e 0.801 ($1 - g_2(x_0)$).

A densidade estimada para o melhor modelo com dados truncados (M2) resultou em 0.950 grupos por km^2 (ICr95: 0.595 a 1.610). Em comparação, para o melhor modelo com dados não truncados (M6) a estimativa foi 1.219 (ICr95: 0.739 a 2.075). A diferença entre as estimativas de densidade resultou pequena e os desvios padrão dos dois modelos também são semelhantes não permitindo identificar um dos dois modelos como sendo claramente superior ao outro.

Discussão

Convencionalmente, o viés de percepção é definido como sendo a probabilidade de não-deteccção sobre a linha de transecção; isto é, $1 - g(0)$. Uma definição mais geral, e que se torna necessária aqui, é considerar o viés de percepção como sendo a probabilidade de não-deteccção na distância em que ocorre a máxima detecção e que não é necessariamente sobre a linha de transecção. Portanto, no modelo M6 o viés de percepção de cada observador está associado ao valor $1 - g(x_0)$ com $x_0 > 0$,

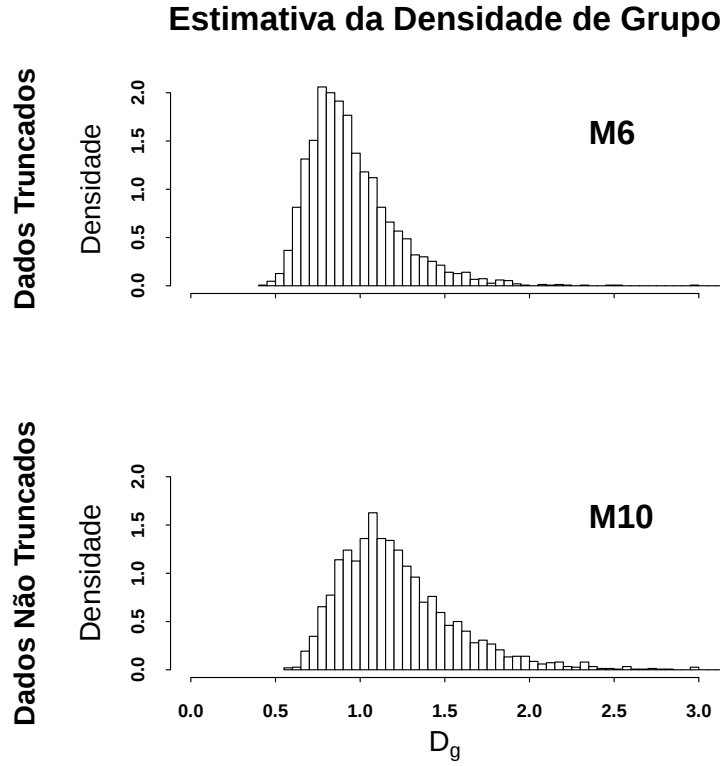


Figure 4: **Estimativa da densidade de grupos.** Figura superior representa dados truncados. Figura inferior, dados não truncados. Histograma da distribuição posterior da estimativa de densidade para toninhas na Baía da Babitonga.

sendo este o valor que devemos considerar quando confrontamos o viés de percepção entre modelos com formas distintas como são os modelos M2 e M6.

Os modelos sem termo quadrático têm a sua máxima probabilidade de detecção sobre a linha de transecção. Portanto, não surpreende que, para dados truncados o modelo com menor DIC tenha sido o modelo M2 sem termo quadrático já que o truncamento tem exatamente este propósito: garantir que a máxima detecção seja sobre a linha de transecção. Este truncamento se deu em $d_{min} = 87 \text{ m}$ sobre o qual se define o $g(0)$ neste caso. Além disso, a escolha de M2 indicou que os dois observadores tinham curvas de detecção distintas.

Modelos com termo quadrático são propícios quando se espera que a máxima probabilidade de detecção ocorra fora da linha de transecção. De fato, ao se utilizar dados sem truncamento ($d_{min} = 0$), o modelo selecionado foi o M6 que também indicou curvas distintas para cada observador. A grande diferença entre as máximas probabilidades de detecção para ambos ($g_1(x_0)$ e $g_2(x_0)$ - Tab. 3 b) pode ser decorrência das janelas bolha e plana respectivamente. Entretanto, uma estimativa da distância média de máxima probabilidade de detecção ($[g_1(x_0) + g_2(x_0)](2^{-1})$) é $d = 123 \text{ m}$. Este valor é superior à distância de truncamento que havia sido utilizado em campo e sugere que a

Table 3: Média e desvio padrão posteriores das probabilidades de detecção à distância zero para funções de detecção única ou para o observador 1, ($g_1(0)$) e para o observador 2, ($g_2(0)$); máxima probabilidade de detecção e a distância x_0 para função de detecção única ou para observador 1, (x_{01}) e para observador 2, (x_{02}). Todos os modelos foram ajustados com dados truncados, (a) e dados não truncados, (b).

a Dados Truncados													
Modelos	g1(0)	sd	g2(0)	sd	g1(x0)	sd	x01	sd	g2(x0)	sd	x02	sd	DIC
M1	0.439	0.114	0.264	0.079	-	-	-	-	-	-	-	-	2888
M2	0.414	0.106	0.263	0.076	-	-	-	-	-	-	-	-	2657
M3	0.221	0.077	-	-	0.263	0.166	65.054	980.625	-	-	-	-	3101
M4	0.238	0.078	-	-	0.296	0.208	-110.260	6162.845	-	-	-	-	2865
M5	0.334	0.117	0.150	0.065	0.400	0.270	130.272	11552.010	0.200	0.103	116.516	346.800	3009
M6	0.354	0.120	0.165	0.071	0.391	0.315	148.865	8659.387	0.210	0.116	138.658	4245.372	2807
b Dados não Truncados													
Modelos	g1(0)	sd	g2(0)	sd	g1(x0)	sd	x01	sd	g2(x0)	sd	x02	sd	DIC
M1	0.528	0.144	0.112	0.043	-	-	-	-	-	-	-	-	4727
M2	0.529	0.141	0.106	0.044	-	-	-	-	-	-	-	-	4713
M3	0.046	0.019	-	-	0.167	0.047	119.137	7.699	-	-	-	-	5330
M4	0.045	0.021	-	-	0.176	0.048	120.257	7.523	-	-	-	-	4887
M5	0.154	0.068	0.016	0.007	0.250	0.071	80.279	18.666	0.157	0.044	157.177	11.418	4827
M6	0.142	0.063	0.001	0.002	0.255	0.068	85.219	17.492	0.199	0.050	159.749	7.345	4556

distância de truncamento seja em torno de 123 *m*.

Ao não truncar os dados evitou-se descartar dados simplesmente por restrições impostas pelo modelo como ocorre com o uso convencional de distance sampling. No caso em estudo, o truncamento representou a eliminação de 26% dos dados o que representa informação relevante sendo perdida. Na estimativa de densidade de grupos, no entanto, o modelo M6 diferiu pouco do modelo M2. (Tab. 5 e Fig. 4). Porém ainda assim é possível observar a menor variabilidade do modelo M6 em relação ao M2 como consequência do não truncamento. A possível explicação para esta diferença inexpressiva está no fato de que, embora sem truncamento houvessem mais dados, a inclusão do componente quadrático aumentou a complexidade do modelo exigindo a estimação de seis ao invés de quatro parâmetros.

A taxa de recaptura global (rr) é simplesmente a proporção de avistagens que se caracteriza como sendo recapturas. Aqui, esta taxa passa a ser distância-específica ($rr(x)$). A Fig. 5 apresenta essas taxas de recaptura estimadas para todos os modelos. Merecem destaque duas situações: modelo M2 para dados truncados e modelo M6 para dados não-truncados. No primeiro caso (M2) verificam-se taxas de recaptura acima da média entre 87 m e 128 m com declínio a medida que a

Table 4: Média e desvio padrão posteriores, coeficiente de variação e quantis (0.05, 0.5 e 0.975) para a estimativa de densidade por modelo. Dados truncados e não truncados.

	Dados Truncados						Dados não Truncados					
Modelos	média	dp	cv	2.5%	50%	97.5%	médias	dp	cv	2.5%	50%	97.5%
M1	0.958	0.293	0.306	0.592	0.900	1.693	0.770	0.212	0.275	0.492	0.728	1.304
M2	0.950	0.267	0.281	0.595	0.896	1.610	0.770	0.212	0.275	0.492	0.728	1.304
M3	1.094	0.340	0.311	0.632	1.028	1.951	1.507	0.486	0.323	0.860	1.413	2.759
M4	1.050	0.306	0.291	0.637	0.990	1.817	1.423	0.422	0.297	0.833	1.344	2.493
M5	1.018	0.308	0.303	0.620	0.951	1.773	1.247	0.365	0.293	0.741	1.179	2.173
M6	0.966	0.282	0.291	0.595	0.909	1.658	1.219	0.349	0.286	0.739	1.154	2.075

distância aumenta. No segundo caso (M6), taxas de recaptura acima da média se concentram numa faixa que vai de 112 m a 178 m. Neste segundo caso (M6), Caracterizando uma região intermediária como sendo de maior efetividade de detecção.

Finalmente, destacamos que uma limitação do modelo em seu estágio atual de desenvolvimento está na impossibilidade de incluir outras variáveis preditoras, sejam ambientais ou espaciais na função de detecção. Isto decorre da necessidade de utilizar a técnica de dados aumentados que exigiria conhecer estas covariáveis para grupos de animais não detectados. No caso, as distâncias também são quantidades desconhecidas para grupos não-detectados e, portanto, incluídas como se fossem dados não-observados. Mas embora desconhecidas, estas distâncias tem distribuição supostamente uniforme entre d_{min} e d_{max} . Com isso o problema da ausência desta informação é facilmente contornado. Para outras variáveis preditoras este tipo de modelo obviamente não existe; daí a dificuldade. Formas de obter as distribuições posteriores sem recorrer a técnica de dados aumentados terão que ser desenvolvidas para poder incorporar outras variáveis preditoras. Este é um desafio em aberto.

Conclusão

O ajuste dos melhores modelos indica curvas de detecção distintas para cada observador tanto para dados truncados quanto não truncados. A partir do ajuste efetuado sem truncamento, verificou-se que a distância ideal para truncamento (estimada em 123 m) é superior àquela que foi utilizada de modo *ad hoc* (87 m). A densidade estimada em número de grupos por km^2 , ($N_g \cdot \text{km}^{-2}$) é 1.219 com um intervalo de 95% de credibilidade, estendendo-se de 0.739 a 2.075. A flexibilidade do

Estimativa da Curva da Taxa de Recaptura

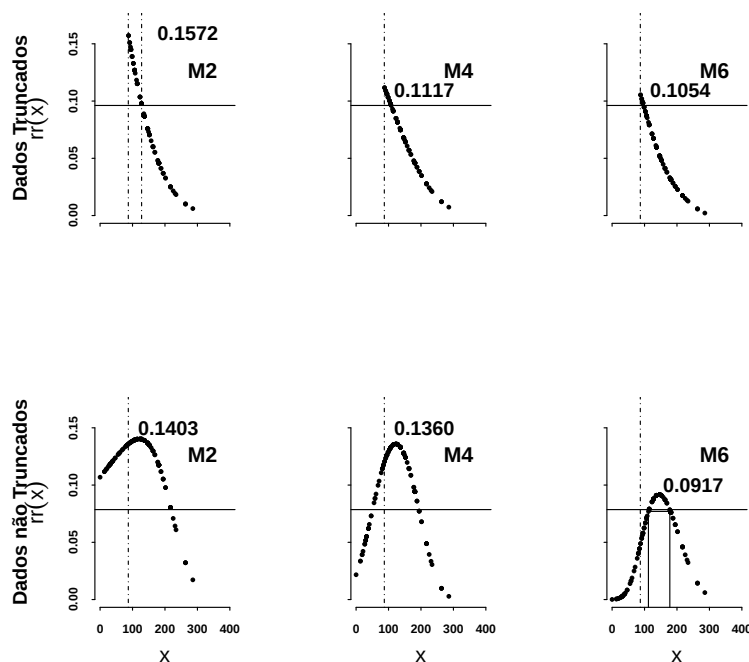


Figure 5: **Estimativa da Curva da Taxa de Recaptura.** Frequência de grupos observados por distância (metros) para dados truncados e dados não truncados. A linha tracejada vertical representa o truncamento em 87 m e a linha contínua horizontal representa a taxa de captura obtida com os dados de campo. Para dados truncados $rr = 0,0962$ e para dados não truncados, $rr = 0,0786$. O valor no ponto máximo de cada curva representa a taxa de recaptura máxima para cada modelo. O retângulo vertical que aparece no modelo M6, representa a área de maior probabilidade de detecção de animais (entre 112 m e 178 m), onde se encontra a distância média ideal para o truncamento, estimada em 123 m.

modelo hierarquicos propostos permite estimar densidade da população sem que haja necessidade de qualquer tipo de truncamento dos dados. O que evita o descarte de informações, obtidas muitas vezes, por intermédio de logística onerosa.

References

- [1] Andriolo, A., Kinas, P.G., Engel, M.H., Martins, C.C.A. & Rufino, A.M. 2010. Humpback whale within the brazilian breeding ground: distribution and population size estimate. *Endangered Species Research* 11:233-243. doi 10.3354/esr00282.
- [2] Bailey, L.L., Hines, J.N.J. & MacKenzie, D. 2007. Sampling design trade-offs in occupancy studies with imperfect detection: examples and software. *Ecological Applications* 17(1):281-290.
- [3] Beasley, I. & Jefferson, T.A. 2002. Surface and dive times of finless porpoises in Hong Kong's coastal waters *Neophocaena phocaenoides*. *Raffles Bulletin of Zoology Supplement* 10:125-129.
- [4] Becker, E. & Quang P. 2009. A Gamma-Shaped Detection Function for Line-Transect Surveys With Mark-Recapture and Covariate Data. *Journal of Agricultural, Biological and Environmental Statistics* 14(2):207-223. doi: 10.1198/jabes.2009.0013.
- [5] Borchers, D.L., Bukland, S. & Zucchini, W. 2002. Estimating animal abundance. Great Britain: Springer. 314p.

- [6] Borchers, D.L., Laake, J.S., Southwell, C. & Paxton, C. 2006. Accommodating Unmodeled Heterogeneity in Double-Observer Distance Sampling Surveys. *Biometrics* 62:372–378. doi: 10.1111/j.1541-0420.2005.00493.x.
- [7] Clack, J. 2005. Why environmental scientists are becoming Bayesians?. *Ecology Letters* 8:2-14. doi: 10.1111/j.1461-0248.2004.00702.x.
- [8] Cremer, M.J. & Simões-Lopes, P.C. 2005. The occurrence of *Pontoporia blainvillei* (Gervais and D'Orbigny) (Cetacea, *Pontoporiidae*) in a estuarine area in southern Brazil. *Revista Brasileira de Zoologia* 22(3):717-723.
- [9] Crespo, E.A., 2009. Franciscana dolphin *Pontoporia blainvillei*. In: Perrin, W.F., Würsig, B. & Thewissen, J.G.M. *Encyclopedia of Marine Mammals*. (eds). 2.ed. San Diego, CA: Academic Press p.466-469.
- [10] Crespo, E.A., Harris, G. & Gonzáles, R. 1998. Group size and distributional range of the franciscana, *Pontoporia blainvillei*. *Marine Mammal Science* 14(4):845-849.
- [11] Conn, P.B., Laake, J. & Johnson, D., 2012. A Hierarchical Modeling Framework for Multiple Observer Transect Surveys. *PLoS ONE* 7(8):e42294. doi:10.1371/journal.pone.0042294.
- [12] Dalla Rosa, L., Ford, J.K.B. & Trites, A.W. 2012. Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters. *Continental Shelf Research* (36): 89–104. doi:10.1016/j.csr.2012.01.017.
- [13] Danilewicz, D. 2003. Reproduction of female franciscana (*Pontoporia blainvillei*) in Rio Grande do Sul, southern Brazil. *Latin American Journal of Aquatic Mammals* 2(2):67-78.
- [14] Danilewicz, D., Secchi, E.R. & Di Benedetto, A.P.M. 2008. *Pontoporia blainvillei* (Gervais and d'Orbigny, 1844). In: *Livro Vermelho da Fauna Brasileira Ameaçada de Extinção*. Machado, A.B.M., Drummond, G.M. & Paglia, A.P. (eds). Brasília: MMA. v.II ,p.814-815.
- [15] Danilewicz, D., Secchi, E.R., Ott, P.H., Moreno, I.B., Bassoi, M. & Borges-Martins, M. 2009. Habitat use patterns of franciscana dolphins (*Pontoporia blainvillei*) off southern Brazil in relation to waters depth. *Journal of Marine Biological Association of the United Kingdom* 89(5):943-949.
- [16] Dolphin, W.F. 1987. Ventilation and dive patterns of Humpback Whale, *Megaptera novaeangliae*, on their Alaskan feeding grounds. *Canadian Journal Zoology* 65:83-90.
- [17] Durban, J.W., Elston, D.A., Ellifrit, D.K., Dickson, E., Hammond, P.S. & Thompson, P.M. 2005. Multisite mark-recapture for cetaceans: population estimates with Bayesian model averaging. *Marine Mammal Science* 21(1):80-92.
- [18] Gelman, A. & Hill, J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 5.ed. New York: Cambridge University Press, ISBN-13 978-0-511-26878-6 eBook (EBL). 645.p
- [19] Guilherme-Silveira, F.R. & Kinas, P.G. 2016. Bayesian inference in line transects with double count sampling and imperfect on-line detection. *Revista Brasileira de Biometria*. No Prelo
- [20] Harwood, J. 2001. Marine mammals and their environment in the twenty-first century. *Journal of Mammal* 82(3):630-640.
- [21] Instituto Brasileiro do Meio Ambiente e Recursos Renováveis. 1997. Relação das Espécies sob Maior Pressão Antrópica. In: *Mamíferos Aquáticos do Brasil: Plano de Ação II*. Instituto Brasileiro de Meio Ambiente e Recursos Renováveis (Eds). Brasília. p.25-27.
- [22] Janik, V.M. & Thompson, P.M. 1996. Changes in surfacing patterns of bottlenose dolphins in response to boat traffic *Tursiops truncatus*. *Marine Mammal Science* 12(4):597-602.
- [23] Jonsen, I.D., Myers, R.A. & Flemming, J.M. 2003. Meta-analysis of animal movement using state-space models. *Ecology* 84(11):3055-3063.

- [24] Kery, M. & Royle, J.A. 2010. Hierarchical modelling and estimation of abundance and population trends in metapopulation designs. *Journal of Animal Ecology* 79:453–461. doi: 10.1111/j.1365-2656.2009.01632.x
- [25] King, R. & Brooks, S.P. 2008. On the Bayesian Estimation of a Closed Population Size in the Presence of Heterogeneity and Model Uncertainty. *Biometrics* 64:816–824. doi: 10.1111/j.1541-0420.2007.00938.x.
- [26] Kinas, P.G. & Andrade, H. 2010. Introdução a análise bayesiana (com R). 1.ed. Porto Alegre: MaisQnada editora. 240p.
- [27] Laake, J. 1999. Distance sampling with independent observers: Reducing bias from heterogeneity by weakening the conditional independence assumption. *Marine Mammal Survey and Assesment Methods*. Garner et al. (eds). Balkema, Rotterdam: Brookfield. 290p. ISBN 990 5809 043 4.
- [28] Laake, J.L. & Borchers, D.L. 2004. Methods for incomplete detection at distance zero, In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. *Advanced distance sampling: estimating abundance of biological populations*. Norfolk, Breat Britain: Oxford University Press. 416pp. ISBN 0 19 850783 6 (hbk).
- [29] Learmonth, J.A., MacLeod, C.D., Santos, M.B., Pierce, G.J., Crick, H.Q.P. & Robinson, R.A. 2006. Pontencial effects of climate change on marine mammals. *In: Oceanography and Marine Biology: An annual review* 44:431-464.
- [30] Lerczak, J.A. & Hobbs, R.C. 1998. Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammals survey. *Marine Mammal Science* 14(3):590-599.
- [31] Marsh, H.A. & Sinclair, D.F. 1989. An Experimental Evaluation of Dugong and Sea Turtle Aerial Survey Techniques. *Australian Wildlife Research* 16:639-650.
- [32] Martin, A. & Quinn, K. 2006. Applied bayesian inference in R using MCMC-pack. *R Newes* 6(1):2-7.
- [33] McCarthy, M. 2007. *Bayesian Methods for Ecology*. 1.ed. New York: Cambridge University Press. ISBN-13 978-0-521-61559-4 305p.
- [34] MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. 2006. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier Inc. 342pp. ISBN 13: 978-0-12-088766-8.
- [35] Newman, K.B., Fernandez, C., Thomas, L. & Buckland, S.T. 2009. Monte Carlo Inference for State-Space Models of Wild Animal Populations. *Biometrics* 65:572–583. doi: 10.1111/j.1541-0420.2008.01073.x.
- [36] O’Connell Jr, A.F., Talancy, N.W., Bailey, L.L., Sauer, J.R., Cook, R. & Gilbert, A.T. 2006. Estimating site occupancy and detection probability parameters for meso and large mammals in a costal ecosystem. *Journal of Wildlife Manangement* 70(6):1625-1633.
- [37] Perez, F.S. 2015. Fatores que afetam a detectabilidade da toninha (*Pontoporia blainvillei*) em estudos de estimativas populacionais. *Disertação 80f*. Universidade Federal de Juiz de Fora. Juiz de Fora.
- [38] Pinedo, M.C., Praderi, R. & Brownell Jr, R.L. 1989. Review of the biology and status of the franciscana, *Pontoporia blainvillei*. *In: Biology and Conservation of the River Dolphins*. Perrin, W.F., Brownell Jr, R.L., Zhou, K. & Liu, J. (eds). Hong Kong. IUCN. p.46-51.
- [39] Plummer, M. 2003. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. ISSN 1609-395X. Vienna, Austria. Available: <http://www.ci.tuwien.ac.at/Conferences/DSC-2003>. p.1-10.
- [40] Plummer, M., Best, N., Cowles, K. & Vines, K. 2010. CODA: Convergence diagnosis and output analysis for MCMC. *R Newes* 6(1):7-11.
- [41] Plummer, M. 2012. JAGS Version 3.3.0. 2012. Available: <http://people.math.aau.dk/~kbb/undervisning/Bayes13/sorenh/doc>.

- [42] Plummer, M. 2013. Package "rjags": Bayesian graphical models using MCMC. Available: [http://cran-r.c3sl.ufpr.br/web/packages/rjags/rjags.pdf](http://cran.r.c3sl.ufpr.br/web/packages/rjags/rjags.pdf).
- [43] Quang, P. & Becker, E. 1996. Line Transect Sampling Under Varying Conditions with Application to Aerial Surveys. *Ecological Society of America* 77(4):1297-1302. Available: <http://www.jstor.org/stable/2265601>.
- [44] Quang, P. & Becker, E. 1997. Combining Line Transect and Double Count Sampling Techniques for Aerial Surveys. *Journal of Agricultural, Biological and Environmental Statistics* 2(2):230-242. Available: <http://www.jstor.org/stable/1400405>.
- [45] Quang, P. & Becker, E. 1998. Aerial survey sampling of contour transects using double-count and covariate data. *In: Marine Mammal Survey and Assessment Methods*, 25-27 February 1998. Seattle Washington USA. Proceedings of the Symposium on Surveys, Status and Trends of Marine Mammal Populations. Washington: Seattle. p1-11.
- [46] R Core Team. 2014. R: A Language and Environment for Statistical Computing. R. Foundation for Statistical Computing. Vienna, Austria. Available: <http://www.R-project.org/>.
- [47] Ramos, R.M.A., Di Benedetto, A.P.M. & Lima, N.R.W. 2000. Growth parameter of *Pontoporia blainvillei* and *Sotalia fluviatilis* (Cetacea) in northern Rio de Janeiro, Brazil. *Aquatic Mammals* 26(1):65-75.
- [48] Ramos, R.M.A., Di Benedetto, A.P.M., Siciliano, S., Santos, M.C.O., Zerbini, A.N., Bertozzi, C., Vicente, A.F.C., Zampiroli, E., Alvarenga, F.S. & Lima, N.R.W. 2002. Morphology of the franciscana (*Pontoporia blainvillei*) off southeastern Brazil: Sexual dimorphism, growth and geographic variation. *Latin American Journal of Aquatic Mammals* 1(1):129-144.
- [49] Reeves, R.R., Dalebout, M., Jefferson, T.A., Karkzmarski, L., Laidre, K.L., O'Corry-Crowe, G., Rojas-Bracho, L., Secchi, E.R., Slooten, E., Smith, B.D., Wang, J.Y., Zerbini, A.N. & Zhou, K. 2012. *Pontoporia blainvillei*. IUCN Red List of Threatened Species: IUCN.
- [50] Royle, J.A. 2008. Hierarchical Modeling of Cluster Size in Wildlife. *Surveys Journal of Agricultural, Biological, and Environmental Statistics* 13(1):23-36. doi: 10.1198/108571108X273188.
- [51] Royle, J.A. & Dorazio, R. 2008. Hierarchical modeling and inference in ecology, the analysis of data from populations, metapopulations and communities. 1.ed. USA: Elsevier. 444p.
- [52] Senar, J.C. & Conroy, M.J. 2004. Multi-state analysis of the impacts of avian pox on a population of Serins (*Serinus serinus*): the importance of estimating recapture rates. *Animal Biodiversity and Conservation*. 2004; 27(1):133-146.
- [53] Siciliano, S. 1994. Review of small cetaceans and fishery interactions in coastal waters of Brazil. *International Whaling Commission* (Special Issue 5). p.241-250.
- [54] Spiegelhalter, D.J., Best, N., Carlin, B. & van der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64:583-640.
- [55] Su, Y.S. & Yajima, M. 2015. Package "R2jags": A Package for Running 'JAGS' from R. Available: <http://cran-r.c3sl.ufpr.br/web/packages/R2jags/R2jags.pdf>.
- [56] Wells, R.S., Bordino, P. & Douglas, D.C. 2013. Patterns of social association in the franciscana, *Pontoporia blainvillei*. *Marine Mammal Science* 29(4):E520-E528. ISSN 08240469.
- [57] Williams, B.K. 2001. Uncertainty, learning, and the optimal management of wildlife. *Environmental and Ecological Statistics* 8:269-288.
- [58] Williams, R. & Thomas, L. 2009. Cost-effective abundance estimation of rare animals: Testing performance of small-boat surveys for killer whale in British Columbia. *Biological Conservation* 142:1542-1547. doi:10.1016/j.biocon.2008.12.028.