

UNIVERSIDADE FEDERAL DO RIO GRANDE
PÓS-GRADUAÇÃO EM OCEANOGRAFIA BIOLÓGICA

**MONITORAMENTO PESQUEIRO
INFERENCIAL: ANÁLISES POR
DELINEAMENTO AMOSTRAL E POR
MODELOS BAYESIANOS**

LAURA VILLWOCK DE MIRANDA

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

Orientador: Prof. Dr. Paul Gerhard Kinas

RIO GRANDE
Agosto, 2017

“In God we trust, all others must bring data.”

W. EDWARDS DEMING

AGRADECIMENTOS

Ao prof. Dr. Paul Gerhard Kinas pela oportunidade de trabalharmos juntos, pela orientação atenta, criteriosa e prestativa, pelo apoio e tempo dedicados em todas as fases deste trabalho, pela paciência, por fazer todas as minhas dúvidas parecerem difíceis e importantes e por suas lições admiráveis de objetividade. OBRIGADA POR TUDO, KINAS!

Ao pessoal do NAAE-IMEF/FURG (de agora e de antes): Raquel, Juliano, Marcos, Hans, Bruno, Carlos, Inácio, Ana Carolina, Liana, Rayd, Aline, Marie, Fernandinho, Fernandão, Baila... pelo companheirismo e amizade, por aturarem e entenderem minha necessidade de silêncio, pelas discussões e ajuda com a tese e com o R e pelas boas conversas regadas a muito café. Às Luluzinhas, meu agradecimento especial. A amizade de vocês foi um dos maiores ganhos deste doutorado! Obrigada pela companhia, apoio, amizade incondicionais e pelas muitas risadas!

Ao Programa de Pós-Graduação em Oceanografia Biológica e à Universidade Federal do Rio Grande que proporcionaram toda a infraestrutura necessária para que eu pudesse realizar este doutorado.

Aos membros da banca examinadora: Prof. Dr. Humber Agrelli de Andrade (DEPAq-UFRPE), Prof. Dr. Luiz Felipe Cestari Dumont, Prof. Dr. Gonzalo Velasco Canziani, Prof. Dr. João Paes Vieira sobrinho e Prof. Dr. Maurício Garcia de Camargo (IO-FURG), que gentilmente aceitaram esta tarefa e que muito contribuíram para o aperfeiçoamento deste trabalho e da minha formação como pesquisadora.

Ao professor e amigo Maurício Camargo por me introduzir ao mundo do R (meses antes do início deste doutorado) e por estar sempre disponível para ajudar e me ensinar novos “truques” que facilitaram muito a minha vida! Por isso, também agradeço ao Rodrigo Sant’Ana, que mesmo à distância me ajudou a “descascar alguns pepinos” no R.

Aos técnicos do IBGE, autores da metodologia “Estatística da Pesca - Pesca Embarcada”, MSc Guilherme Guimarães Moreira e MSc Aristides Pereira Lima-Green pela ajuda indispensável na fase inicial desta tese.

Aos pesquisadores e à equipe técnica do Programa de Monitoramento de Atividade Pesqueira do Instituto de Pesca-APTA-SAA/SP pelos melhores dados pesqueiros que eu poderia ter e por entenderem eu nem sempre estar presente. Obrigada Rafael Cabrera Namora por toda a ajuda! Agradeço, em especial, aos monitores Alexandre, Diogo, Sandro e Thaís e a todos os agentes de campo do PMAP-NPDLN-Litoral Norte. Obrigada por segurarem a barra sem a minha presença constante na coordenação e por me pouparem de muitas atividades que seriam minha responsabilidade para que eu pudesse ter mais tempo para me dedicar à tese!! Por tudo isso, eu também agradeço ao PqC Marcus Henrique Carneiro, mas também pela amizade, apoio, discussões e opiniões, por estar disponível sempre que precisei... Marculino, obrigada! Também sou grata aos diretores dos NPDLN, CAPTAPM e de Departamento do Instituto de Pesca que permitiram e agilizaram todos os trâmites necessários para os meus inúmeros afastamentos de curta duração. Sem isso não haveria tese...

Agradeço, e vou passar uma vida agradecendo (e ainda não será suficiente!), ao meu pai, à minha mãe e à minha irmã (Ronaldo, Vera e Leticia)!! O amor, apoio, compreensão, incentivo e carinho inesgotáveis de vocês foram e serão sempre essenciais na minha vida!! Vocês são tudo para mim!! Ao meu cunhado Henrique que trouxe para a nossa família a alegria da Princesa Sofia e, de cara, já me fez ser tia (te adoro, Pinduca!!) A ele e a mana, pois me deram o meu melhor presente: meu afilhado Gabriel! OBRIGADA POR ACREDITAREM EM MIM, MAIS DO QUE EU MESMA! OBRIGADA POR SEREM A MINHA FAMÍLIA, POR ME ENSINAREM A DAR VALOR ÀS COISAS CERTAS E POR FAZEREM DE MIM TUDO O QUE EU SOU!! AMO MUITO VOCÊS!!

Ao meu marido, amigo, parceiro, Alexandre! Não consigo nem listar tudo que eu precisaria te agradecer!! Sem ti, sem teu apoio, compreensão, cumplicidade, paciência, amor, ouvidos e abraços para me ouvir e me acalmar nada disso teria sido possível! Te amo tamanho céu! De montão, pra chuchu, pra mais de “milha náutica! (metro é pouco!!) CONTIGO VOU ATÉ O INFINITO, E MUITO MAIS ALÉM!! OBRIGADA, MEU AMOR!!

ÍNDICE

RESUMO.....	1
ABSTRACT.....	3
INTRODUÇÃO	5
OBJETIVOS	17
METODOLOGIA GERAL.....	18
1. Fonte das informações pesqueiras	18
2. Monitoramento pesqueiro inferencial por delineamento amostral	19
2.1. Cadastro dos portos pesqueiros	19
2.2. Elaboração do plano amostral baseado na captura descarregada	20
2.3. Expansão e inferência estatística	21
2.4. Obtenção dos custos de monitoramento pesqueiro	22
3. Monitoramento pesqueiro inferencial por modelos Bayesianos.....	24
3.1. Análise dos dados	25
3.2. Estrutura geral dos modelos inferenciais.....	26
3.3. Dados de entrada	27
3.4. Distribuições de probabilidade prioris.....	28
3.5. Definição dos modelos inferenciais.....	29
3.6. Modelos preditivos	30
SÍNTESE DOS RESULTADOS	32
CONSIDERAÇÕES FINAIS	36
LITERATURA CITADA	44
ANEXO I: Survey Sampling for Fisheries Monitoring in Brazil: Implementation and Analysis	50
ANEXO II: Fisheries Monitoring: Is Bayesian Inference for Finite Population Sampling an Alternative?	65

RESUMO

Informações pesqueiras podem ser coletadas por censo ou por amostragem. Quando obtidas por amostragem, recorrem à inferência estatística por delineamento amostral ou, alternativamente, baseada em modelos probabilísticos paramétricos. O objetivo geral deste trabalho foi analisar os resultados do monitoramento pesqueiro inferencial por delineamento amostral (ANEXO I – artigo publicado) e por modelos Bayesianos (ANEXO II – artigo submetido) em relação à coleta censitária de dados pesqueiros. Para isso, informações censitárias de todas as descargas pesqueiras realizadas em 2011 ao longo da costa de São Paulo foram utilizadas para simular um monitoramento pesqueiro inferencial e a obtenção de amostras probabilísticas de acordo com a metodologia proposta pelo IBGE. Os resultados obtidos pela inferência baseada em delineamento amostral demonstraram que quanto maior é o nível de detalhamento necessário, menos precisas e mais enviesadas tornam-se as estimativas. Quando comparada com as capturas totais censitárias, a ordem de importância das categorias de pescado baseada nas capturas estimadas alterou-se ligeiramente em algumas posições após o quinto lugar. A redução do custo do monitoramento pesqueiro obtida com a aplicação de amostragem foi mínima em comparação à coleta censitária dos dados (máxima de 15,4%) e provavelmente não compensa a perda de qualidade das informações pesqueiras. No monitoramento pesqueiro inferencial por modelos Bayesianos, foi utilizada a mesma amostra de portos pesqueiros, previamente escolhida para minimizar o erro quadrático médio das estimativas de captura. Os resultados desta etapa do trabalho demonstraram que o modelo se ajustou muito bem aos dados de vários municípios e categorias de pescado do litoral paulista, fornecendo boas estimativas de captura total descarregada (Y) e do número total de descargas realizadas (D), porém não para todos os casos. Para as categorias de pescado com sazonalidade marcada, os

melhores resultados foram obtidos com o maior detalhamento dos modelos e a inclusão de informações mensais. A inclusão do número de descargas por porto pesqueiro como covariável conhecida não melhorou os resultados como seria o esperado. Uma das razões apontada para a falta de ajuste dos modelos aos dados de pesca em alguns municípios ou categorias de pescado foi a estratificação dos portos pesqueiro, definida por municípios e originalmente adotada neste trabalho. Diante disto, uma nova proposta de desenho amostral, respeitando a relação linear entre o $\ln(Y)$ e o $\ln(D)$ de todos os portos pesqueiros do litoral paulista, foi, então, apresentada e discutida. A nova proposta de modelo inferencial Bayesiano, baseada nesta relação linear, forneceu bons resultados, embora preliminares, para os cinco principais municípios pesqueiros do estado, com expressivo aumento da eficiência amostral.

Palavras-chave: pesca, monitoramento pesqueiro, amostragem, inferência, custos de monitoramento, litoral de São Paulo.

ABSTRACT

Fishery information can be obtained by census or by sampling methods. When sampling methods are used, inference will be carried out with design-based or model-based approaches. The general objective of this work was to analyze the results of a design-based inferential fisheries monitoring (APPENDIX I - published) and a new Bayesian model-based alternative (APPENDIX II - submitted article) in relation to the census data collection. For this, census fishing data about the landings carried out along the São Paulo coast during 2011 were used to simulate an inferential fishing monitoring and the obtaining of probabilistic samples according to the methodology proposed by IBGE. The results obtained by the design-based approach demonstrated that the higher the level of required detail, the less accurate and more biased the catch estimates become. When compared to the total census catches, the order of importance of fish categories based on estimated catches changed slightly in some positions after the fifth place. The reduction of the fishing monitoring costs obtained with the adoption of the sampling methodology was minimal when compared with the costs of collecting census data (15.4% at most) and might not compensate for the loss of accuracy in the final results. For the model-based approach (inferential Bayesian model), the same sample of fishing ports, previously chosen to minimize the mean square error of the catch estimates, was used. The results demonstrated that the Bayesian model had a very good data fit for several municipalities and fish categories of the coast of São Paulo State, providing mostly good estimates of the total landed catch (Y) and the total number of landings (D); but not for all cases. For fish categories with marked seasonality, the best results were obtained by breaking down the information by month and include these into the models. Assuming the total number of landings per fishing port as a known covariate did not improve the results as expected. One of the reasons pointed out for the lack of

adjustment of the models to the fishing data in some municipalities or fish categories was the stratification of the fishing ports, defined by municipalities and originally adopted in this work. In view of this, a new proposal of sampling design, respecting the linear relationship between $\ln(Y)$ and $\ln(D)$ for all fishing ports of the São Paulo coast, was, then, presented and discussed as a possible alternative. These new sampling criterion seems promising as it provided good preliminary results under the Bayesian inferential model for the five most important fishery municipalities of the São Paulo State, with an expressive increase in sample efficiency.

Key words: fishery, fishing monitoring, sampling, inference, monitoring costs, São Paulo coast.

INTRODUÇÃO

Informações de captura e esforço de pesca, obtidas somente a partir das pescarias comerciais podem resultar em avaliações equivocadas dos estoques pesqueiros (Hilborn & Branch 2013, Kotaro et al. 2015). Contudo, são as informações mais básicas, baratas, de fácil coleta e, muitas vezes, as únicas disponíveis para o manejo das pescarias (Cadima 2003, Maunder & Punt 2004, Pauly 2013).

A Organização das Nações Unidas para Agricultura e Alimentação (FAO) reúne, atualiza e publica informações sobre a estatística pesqueira e aquicultura mundial a cada dois anos (FAO 2016). São informações obtidas com a adoção de metodologias distintas, geralmente dependentes da pesca comercial, e repassadas à FAO por diversos países. Em países desenvolvidos, é comum haver investimentos no monitoramento e controle da atividade pesqueira e os dados coletados são, com frequência, complementados por informações provenientes de pesquisas científicas antes de estes serem compilados em um banco mundial de dados pesqueiros. Porém, esta não é a realidade da maior parte dos países onde os dados são coletados e repassados diretamente à FAO, sem nenhuma complementação ou validação científica (Gelpke et al. 2013). Muitas vezes, as informações pesqueiras são incompletas ou incorretas, com qualidade variável e que, frequentemente, ignoram as descargas da pesca de subsistência, da pesca ilegal e da pesca esportiva (Cressey 2015). A baixa qualidade, até mesmo a inexistência, das informações pesqueiras, bem como a falta de investimento no monitoramento pesqueiro são uma realidade bastante comum em muitas regiões do mundo, principalmente para países subdesenvolvidos e em desenvolvimento (Pauly 2013).

Em geral, informações sobre a estatística pesqueira são compiladas com periodicidade máxima anual (FAO 2016, Ávila-da-Silva et al. 2017). Garantir a

obtenção ininterrupta e a qualidade destas informações anualmente é fundamental para a construção de séries temporais utilizadas na avaliação do estado de exploração dos recursos pesqueiros e para subsidiar políticas públicas eficazes, de forma a promover a manutenção destes recursos, da biodiversidade e da pesca como atividade econômica (Hilborn & Walters 1992, Cadima 2003). Por esta razão, tentativas de compilar a informação disponível para a reconstrução das capturas (Freire et al. 2015, Pauly & Zeller 2016) e métodos de inferência para lidar com dados faltantes (Cadima et al. 2005, Gelman et al. 2014) tornam-se cada vez mais importantes.

O Brasil está entre os países com dificuldades em manter uma coleta continuada de dados pesqueiros. Desde o princípio do desenvolvimento da pesca extrativa nacional, os resultados das estatísticas de pesca no Brasil foram muito influenciados pelos arranjos institucionais e políticos (Dias-Neto 2010, 2011, Lima-Green & Moreira 2012). Por esta razão, períodos de interrupção na coleta de dados em diferentes regiões e oscilações na eficiência de coleta e processamento foram bastante comuns ao longo do tempo.

O início do desenvolvimento da atividade pesqueira no país foi marcado pela criação da SUDEPE - Superintendência do Desenvolvimento da Pesca - em 1962. Desde esta época, a obtenção de informações pesqueiras já foi atribuição de diferentes instituições (Serviço de Estatística da Produção do Ministério da Agricultura até 1967; SUDEPE de 1967 a 1989; IBAMA/MMA de 1989 a 1998; IBAMA/MMA, DPA/MAPA e SEAP/PR de 1998 a 2009, MPA e IBAMA/MMA 2009 a 2015, MAPA 2016 e MDIC, a partir de 2017), e executada com a adoção de distintas metodologias.

Apesar de todas estas trocas de responsabilidades, a atribuição formal de fazer o levantamento de informações e produzir as estatísticas do setor pesqueiro do Brasil é, desde o final da década de 70, do Instituto Brasileiro de Geografia e Estatística – IBGE.

No entanto, desde a extinção da SUDEPE, no final da década de 80, o IBGE participou apenas dos fechamentos anuais das estatísticas de pesca. Desde o início da década de 90, o IBAMA assumiu a tarefa de levantamento das informações pesqueiras e, junto com outras instituições regionais, adotaram o EstatPesca (Aragão & Martins 2006) em apenas alguns estados do Brasil como a metodologia para o monitoramento pesqueiro (Lima-Green & Moreira 2012). Esta metodologia foi baseada em amostras fixas de embarcações pesqueiras, o que exigia a manutenção de um cadastro permanentemente atualizado de embarcações em operação. Isto se mostrou inviável, principalmente para a pesca de pequena escala, onde a venda e a troca dos nomes e das características das embarcações são frequentes. Além disso, a escolha das embarcações que comporiam a amostra acabava sendo feita de forma intencional e, portanto, com levantamentos amostrais não probabilísticos, podendo trazer vícios ao processo de amostragem (Isaac et al. 2008). Para reduzir as estimativas viciadas, o número de amostras deveria ser maior, aumentando os custos do processo amostral. Isaac et al. (2008) observaram uma superestimação da captura (kg) obtidos pelo método de amostragem EstatPesca no litoral do estado do Pará e concluíram que 70% da frota do litoral paraense deveria ser amostrada para que o erro das estimativas fosse aceitável.

Em 2009, foi formalizado o Acordo de Cooperação Técnica entre IBGE e Ministério da Pesca e Aquicultura (MPA) para reunir os esforços de coordenação das atividades do SINPESQ (Sistema Nacional de Informações da Pesca e Aquicultura, criado em 1995) e de produção de estatísticas da pesca. Este acordo estabeleceu como atribuição do IBGE fornecer orientação metodológica e assistência técnica ao MPA no processo de construção, implantação e manutenção do SINPESQ, no que diz respeito à produção de estatísticas das atividades pesqueira (Lima-Green & Moreira 2012).

Diante disto e após uma avaliação técnica do EstatPesca e de outras metodologias de levantamento de dados de pesca, uma nova metodologia por amostragem para o monitoramento pesqueiro foi proposta por Lima-Green & Moreira (2012), técnicos IBGE. Esta metodologia buscou solucionar os problemas detectados no EstatPesca definindo como unidade amostral as descargas pesqueiras ao invés das embarcações. Seu uso ganhou expressão no Brasil, uma vez que o governo federal, embora com muitas dificuldades de manter a estatística pesqueira, incentivou e financiou a sua adoção nos programas de monitoramento pesqueiro.

No final de 2015, o MPA foi extinto e todas as suas atribuições foram transferidas para a Secretaria de Aquicultura e Pesca ligada ao Ministério da Agricultura, Pecuária e Abastecimento (SAP/MAPA). Em março de 2017, a SAP foi novamente transferida, agora para o Ministério do Desenvolvimento, Indústria, Comércio Exterior e Serviços (MDIC).

Atualmente, o sistema nacional de coleta e processamento de dados pesqueiros colapsou. Apenas alguns estados, entre eles São Paulo, Paraná e Santa Catarina, mantêm o monitoramento da atividade pesqueira marinha através de programas específicos, criados para atender as demandas do licenciamento ambiental dos empreendimentos de petróleo e gás da Petrobras na Bacia de Santos. Desde 2012, o Rio Grande do Sul também manteve a coleta de dados por meio do "Projeto de Estatísticas de Desembarque Pesqueiro no Estuário da Lagoa dos Patos e Região Oceânica Adjacente", financiado pelo MPA. O projeto, no entanto, tem futuro incerto, pois atualmente já opera apenas parcialmente (somente frotas industrial e semi-industrial) e não há perspectivas para novos aportes financeiros que possam viabilizar sua continuidade em 2018.

O monitoramento pesqueiro no estado de São Paulo é uma exceção no Brasil e os

primeiros registros pesqueiros datam de 1944, feitos pela então Secretaria de Agricultura, Indústria e Comércio. Desde 1969, o Instituto de Pesca, vinculado à Secretaria de Agricultura e Abastecimento (SAA/SP) é o órgão responsável pela coleta, armazenamento, processamento e disponibilização de informações sobre a produção pesqueira marinha descarregada no estado de São Paulo. A coleta de dados pesqueiros é realizada com metodologia censitária e de forma ininterrupta desde então (Mendonça & Miranda 2008, Ávila-da-Silva et al. 2017).

O monitoramento pesqueiro pode ser realizado com a adoção de diferentes metodologias e estas podem ser estruturadas, basicamente, de duas maneiras: monitoramento por amostragem (Caddy & Basigos 1985, Aragão & Martins 2006, Lima-Green & Moreira 2012) ou monitoramento censitário (FAO 1999, Mendonça & Miranda 2008, Ávila-da-Silva et al. 2017).

Comumente, a coleta censitária é indicada para populações pequenas, onde o custo de coleta é baixo ou quando as consequências de se tomar uma decisão errada são graves e os erros amostrais grandes (Bolfarine & Bussab 2005). O monitoramento pesqueiro marinho dos estados de São Paulo e Paraná vem sendo executado através de metodologia censitária pelo Instituto de Pesca. A experiência desta instituição no monitoramento pesqueiro e a adoção de estratégias de redução de custos tornaram esta metodologia economicamente viável.

Por outro lado, considerar apenas parte de uma população por meio de amostragem pode ter várias motivações destacando-se populações muito grandes e restrições financeiras ou de tempo. Por outro lado, pesquisas por amostragem podem ter uma maior flexibilidade e abrangência quanto ao tipo e à quantidade de informações necessárias (FAO 1999, Bolfarine & Bussab 2005). Lidar com menos informação também pode reduzir o tempo de consolidação e, frequentemente, o custo de coleta e

processamento (Cochran 1977). O monitoramento pesqueiro em Santa Catarina e no Rio Grande do Sul, com algumas adaptações, é realizado com a metodologia por amostragem proposta por Lima-Green & Moreira (2012).

Para dados obtidos a partir de processos de amostragem, as inferências podem ser feitas de duas maneiras: com base no delineamento amostral ou a partir de um modelo probabilístico para os dados. Na abordagem clássica, baseada no delineamento amostral, toda a inferência é fundamentada na variabilidade induzida pelo plano amostral, sem considerar como as variáveis se comportam (Little 2004, Bussab & Morettin 2012). Em geral, produz inferências acuradas sem a necessidade de assumir um modelo probabilístico para os dados, porém, para isso precisa de amostras grandes. Na inferência baseada em modelo, toda a variabilidade nos dados está explicitada no modelo probabilístico. Os parâmetros do modelo, estimados a partir da amostra, são então usados para inferir sobre os dados não observados na população e, assim, após preencher as lacunas desses “dados faltantes”, obter os totais de uma população finita (Little 2004). Sua eficiência depende de um modelo bem ajustado aos dados para produzir boas inferências.

Para ilustrar a diferença entre estas duas abordagens, considera-se o seguinte exemplo: supondo que a captura total descarregada em toneladas (Y) foi coletada a partir uma amostra aleatória de descargas pesqueiras e é preciso conhecer a probabilidade de haver captura total descarregada superior a 5 t, $P(Y > 5)$. Na inferência baseada no delineamento amostral esta probabilidade só será conhecida se forem observadas descargas com captura maior do que 5 t; caso contrário, a inferência será zero. Já na inferência baseada em modelo, a captura pode ser descrita por uma distribuição log-normal. Seus parâmetros são estimados a partir da amostra e a

probabilidade é calculada pela cauda superior da distribuição, independentemente de haver descargas acima de 5 t na amostra.

A inferência baseada em modelo pode seguir dois caminhos: modelos de superpopulação ou por verossimilhança (abordagem clássica) (Thompson 1988 e Dorfman et al. 2000 *apud* Little 2004) e modelos Bayesianos (Little 2004, McCarthy 2007, Kinas & Andrade 2010, Gelman et al. 2014, Kruschke 2015). Os modelos de superpopulação assumem que os valores da população (composta por unidades amostrais finitas, algumas contidas na amostra e outras não) são uma amostra aleatória de uma “superpopulação” descrita por um modelo probabilístico. Seus parâmetros são fixos e estimados a partir da função de verossimilhança de uma amostra, necessariamente, grande. Em modelos Bayesianos, os dados observados são igualmente resumidos na função de verossimilhança, complementada com distribuições de probabilidade para os parâmetros do modelo. Toda a informação prévia existente sobre os parâmetros do modelo pode ser incorporada na forma de distribuições *prioris* informativas. Assim, a distribuição posterior dos parâmetros, através do Teorema de Bayes, reúne as informações atuais e prévias e contém toda a informação necessária para realizar a inferência (Kinas & Andrade 2010). Em muitas situações, o uso de distribuições *prioris* informativas é uma vantagem, o que permite que a abordagem Bayesiana lide bem com pequenas amostras. Com amostras grandes e distribuições *prioris* não informativas, modelos Bayesianos e por verossimilhança obtêm resultados muito similares (Little 2004).

Na prática, problemas ligados aos dados incompletos surgem frequentemente. A expressão “dados faltantes” pode referir-se àqueles não-intencionalmente faltantes (*e.g.* não resposta, dados censurados ou truncados) e aos intencionalmente faltantes, consequência de processos amostrais, onde somente parte dos dados é observada

(Gelman et al. 2014). Os modelos de dados faltantes são uma maneira de completar as unidades amostrais de uma população finita que não foram observadas e podem ter uma abordagem frequentista ou Bayesiana. No entanto, a inferência frequentista em populações finitas não é o foco deste trabalho. Logo, a formalização teórica apresentada a seguir refere-se à inferência Bayesiana em populações finitas e é uma adaptação de Gelman et al. (2014) e Zangeneh & Little (2015).

Quando a coleta de dados assume aleatoriedade, o modelo do desenho amostral I e seus parâmetros φ podem ser ignoráveis (Rubin 1976). Para facilitar a notação, a ignorabilidade do desenho amostral será suposta válida e adotada a partir de agora. Para lidar com os dados intencionalmente faltantes é conveniente dividir a análise em duas partes:

- (1) inferência em superpopulação $p(\theta|x, y_{obs})$ (ou modelo inferencial), que estima as distribuições posteriores dos parâmetros (θ) condicionadas aos dados observados (y_{obs}) e às possíveis covariáveis (x);
- (2) inferência em população finita $p(y_{mis}|x, y_{obs})$ (ou modelo preditivo) que estima a distribuição posterior preditiva para os dados faltantes (y_{mis}) condicionadas, aos dados observados, às possíveis covariáveis.

Na abordagem Bayesiana, as simulações da distribuição posterior preditiva são denominadas de *imputação múltipla* e consiste em duas etapas: primeiro obter o conjunto θ dos parâmetros do modelo a partir da sua distribuição posterior $p(\theta|x, y_{obs})$; e depois simular y_{mis} a partir do modelo para os dados $p(y_{mis}|x, \theta)$. Em suma:

$$p(y_{mis}|x, y_{obs}) = \int p(\theta|x, y_{obs}) * p(y_{mis}|x, \theta) d\theta$$

A inferência Bayesiana em populações finitas (baseada em modelo) supera a inferência baseada em delineamento amostral quando o modelo está corretamente especificado, pois possui uma estrutura formal para combinar informações de diversas fontes. O uso de informações de anos anteriores (em forma de distribuições *prioris*) quando disponíveis e pertinentes e a ausência de argumentos assintóticos permitem obter boas inferências mesmo diante de pequenas amostras. Além disso, modelos inferenciais Bayesianos são capazes de lidar com desenhos amostrais complexos de forma natural e simples e permitem a inclusão relativamente fácil de covariáveis (Little 2004, Kéry 2010, Hall 2011).

Todas estas características levam a acreditar que a abordagem Bayesiana é uma ideia atrativa para lidar com a inferência de dados pesqueiros que são altamente variáveis e muitas vezes escassos. Portos pesqueiros de diferentes tamanhos, tanto em quantidade de captura descarregada quanto em número de descargas realizadas e uma grande diversidade de espécies-alvo, de períodos de safra e de artes de pesca utilizadas, entre outras características, são uma realidade comum da atividade pesqueira no Brasil (Isaac et al. 2000, 2008, Mendonça & Miranda 2008). Tamanha variação e diversidade dificilmente poderão ser captadas unicamente por um delineamento amostral. Além disso, o número de embarcações e pescadores e o tipo e a dinâmica da frota de um porto pesqueiro, assim como a época do ano e o período de safra dos recursos pesqueiros alvos das pescarias podem ser covariáveis potencialmente úteis e facilmente incorporadas em um modelo inferencial Bayesiano para a atividade pesqueira.

Diversos autores discutiram métodos de obtenção de dados pesqueiros e a sua aplicação em análises de dinâmica populacional e avaliação de estoques pesqueiros (*e.g.* Gulland 1966, Caddy & Basigos 1985, Sparre & Venema 1997, FAO 1999, Sparre 2000, Cadima et al. 2005). O que estas metodologias apresentam em comum é a

abordagem estatística frequentista, seja via plano amostral ou censitário. A motivação para a realização do presente trabalho surgiu diante de: (1) o histórico problemático do monitoramento pesqueiro no Brasil; (2) o uso de diferentes metodologias de coleta ao longo da costa e a necessidade de que as informações obtidas sejam compatíveis, confiáveis e possam ser consolidadas sem prejuízos às avaliações pesqueiras; (3) a existência e a disponibilidade das informações pesqueiras censitárias do litoral de São Paulo através do banco de dados ProPesqWEB; e (4) o fato de que a inferência Bayesiana para o monitoramento pesqueiro é uma abordagem inédita no contexto da estatística pesqueira até este momento.

Neste trabalho, são apresentados estudos que utilizaram as informações pesqueiras censitárias do estado de São Paulo para simular amostras probabilísticas e o monitoramento pesqueiro inferencial no litoral paulista. Os resultados obtidos pela inferência baseada tanto em delineamento amostral quanto em modelos Bayesianos foram analisados e comparados à coleta censitária de dados pesqueiros. O trabalho está dividido em dois capítulos, apresentados no formato de artigos científicos para publicação. Os resultados são apresentados resumidamente no item “Síntese dos Resultados” (em português) e integralmente nos anexos desta Tese (em inglês). O capítulo 1, intitulado “Survey Sampling for Fisheries Monitoring in Brazil: Implementation and Analysis” (publicado), adotou o desenho amostral e aplicou a inferência por delineamento amostral para o monitoramento pesqueiro, propostos pelo IBGE (Lima-Green & Moreira 2012). Além da qualidade das estimativas, os custos para realizar o monitoramento pesqueiro na costa de São Paulo também foram considerados para avaliar as perdas e ganhos da metodologia por amostragem quando comparada à coleta censitária dos dados. O capítulo 2, intitulado “Fisheries monitoring: is Bayesian inference for finite population sampling an alternative?” (submetido) utilizou o mesmo

desenho amostral e a amostra de portos pesqueiros do capítulo 1. As informações dos portos pesqueiros observados e faltantes nesta amostra foram utilizadas com o objetivo de formular um modelo inferencial Bayesiano para o monitoramento pesqueiro.

HIPÓTESES

Hipótese 1: O monitoramento pesqueiro inferencial baseado em delineamento amostral gera estimativas acuradas, juntamente com uma importante redução de custos quando comparado com a obtenção de informações pesqueiras censitárias.

Pressupostos da hipótese 1: As estimativas acuradas e confiáveis podem ser produzidas com um custo bem menor e os dados podem ser obtidos e consolidados mais rapidamente se métodos por amostragem forem adotados (Cochran 1977, Bolfarine & Bussab 2005).

Hipótese 2: O modelo inferencial bayesiano para o monitoramento pesqueiro lida melhor com a alta variabilidade dos dados pesqueiros e com situações de poucas informações quando comparado com a inferência baseada em delineamento amostral.

Pressupostos da hipótese 2: A inferência Bayesiana supera a inferência baseada em delineamento amostral quando o modelo está corretamente especificado, pois combina formalmente diferentes tipos de informações (atuais e prévias) e permite que as incertezas, em qualquer momento, sejam revistas à medida que novas informações surgem. O uso de distribuições *prioris* informativas é uma vantagem que permite que a abordagem Bayesiana lide bem com pequenas amostras e com dados variáveis (McCarthy 2007, Kéry 2010, Kinas & Andrade 2010, Gelman et al. 2014, Kruschke 2015).

OBJETIVOS

Objetivo Geral

Analisar os resultados do monitoramento pesqueiro inferencial por delineamento amostral e por modelos Bayesianos em relação à coleta censitária de dados pesqueiros.

Objetivos Específicos

1. Validar a metodologia por amostragem para o monitoramento pesqueiro proposta pelo IBGE (Lima-Green & Moreira 2012), comparando e avaliando as diferenças entre os resultados obtidos com a amostragem em relação à coleta censitária de dados pesqueiros do litoral de São Paulo (ANEXO I);
2. Avaliar as diferenças nas estimativas do custo econômico do monitoramento pesqueiro no litoral de São Paulo simulando a coleta por amostragem das informações pesqueiras comparado ao custo real da coleta censitária dos dados (ANEXO I).
3. Formular e validar o modelo inferencial Bayesiano para o monitoramento pesqueiro, utilizando as informações censitárias do monitoramento da pesca marinha do litoral de São Paulo (ANEXO II);
4. Utilizar informações pesqueiras prévias na construção de distribuições de probabilidade *prioris* informativas e avaliar como isto influencia nas estimativas do modelo inferencial de monitoramento pesqueiro (ANEXO II);

METODOLOGIA GERAL

1. Fonte das informações pesqueiras

As informações censitárias de todas as descargas pesqueiras realizadas na costa de São Paulo ao longo de um ano foram utilizadas para simular amostras probabilísticas e o monitoramento pesqueiro inferencial no litoral paulista. Dados pesqueiros de São Paulo foram utilizados, já que neste estado há uma série temporal ininterrupta desde 1969. O ano escolhido para análises foi 2011, pois no início deste trabalho este ano era o mais recente da série temporal. E a escolha de apenas um ano de dados pesqueiros se justifica uma vez que a metodologia para o monitoramento pesqueiro proposta pelo IBGE (Lima-Green & Moreira 2012), e que está sendo validada neste trabalho, foi elaborada para estimar totais anuais. No entanto, a escolha do ano específico de 2011 foi arbitrária, pois as análises deste trabalho poderiam ser realizadas com informações pesqueiras de qualquer ano ou de outras regiões. Os modelos inferenciais Bayesianos aqui desenvolvidos são suficientemente genéricos para serem adotados em diferentes anos ou regiões.

As informações pesqueiras utilizadas foram obtidas através do Programa de Monitoramento da Atividade Pesqueira (PMAP), coordenado por pesquisadores do Instituto de Pesca da Secretaria de Agricultura e Abastecimento do Estado de São Paulo. Em março de 2008, o PMAP passou a ser utilizado para avaliar a influência sobre a atividade pesqueira dos empreendimentos de petróleo e gás da Petrobras na Bacia de Santos. O programa segue o método censitário para o acompanhamento das descargas de pescado (FAO 1999, Mendonça & Miranda 2008, Ávila-da-Silva et al. 2017) e, atualmente, monitora 205 pontos de descarga de pescados em 15 municípios do litoral de São Paulo. Para preservar a confidencialidade das informações, o município de Santos, que possui apenas um porto pesqueiro, junto com Guarujá são considerados

um único município na consolidação dos dados. Para obtenção dos dados pesqueiros, os agentes de campo realizam entrevistas estruturadas com os mestres das embarcações e pescadores na ocasião das descargas pesqueiras. As informações são complementadas pelo resgate dos registros dos pescadores sobre suas operações pesqueiras diárias (auto-registro), pelos mapas de bordo e registros fornecidos pelas empresas de pesca. O armazenamento, processamento, análise e disponibilização das informações pesqueiras são realizadas através do Sistema Gerenciador ProPesq® (Ávila-da-Silva et al. 1999), atualmente operando em plataforma web, denominado ProPesqWEB (<http://www.propesq.pesca.sp.gov.br>). As variáveis disponíveis neste banco de dados envolvem informações sobre captura e esforço de pesca das descargas realizadas, entre elas nome da embarcação, data e local da descarga, duração da viagem, dias de pesca, aparelho de pesca utilizado e área de pesca, além de informações sobre a captura descarregada por categoria de pescado. Entre as informações de esforço pesqueiro estão o número e a duração média das operações de pesca, número e características dos petrechos de pesca utilizados, entre outras variáveis.

2. Monitoramento pesqueiro inferencial por delineamento amostral

A organização e estruturação das informações pesqueiras censitárias do estado de São Paulo e do desenho para compor a amostra de descargas foram discutidas durante o workshop realizado na FURG, em abril de 2014, com a presença dos técnicos do IBGE, autores da metodologia “Estatística da Pesca - Pesca Embarcada” que está sendo validada (Lima-Green & Moreira 2012).

2.1. Cadastro dos portos pesqueiros

As informações sobre todas as descargas pesqueiras realizadas no litoral de São Paulo durante o ano 2011 foram extraídas do banco de dados do Instituto de Pesca.

Nesse ano, foram registradas descargas pesqueiras em 227 pontos, porém alguns com descargas eventuais e outros localizados bastante próximos entre si.

Para definir a população de interesse e finalizar o cadastro de portos pesqueiros de 2011, foram adotados os seguintes critérios: (1) excluir os portos com menos de 40 descargas no ano (menos que uma descarga semanal); (2) reincluir todos os portos com, ao menos, uma descarga superior a 500 t descarregadas; (3) reunir em um único porto todos aqueles que, por questões de logística, possuem nomes distintos no banco de dados, mas, na prática, fazem parte de um único entreposto pesqueiro; (4) excluir do cadastro os portos que se encontravam desativados. Após cumprir estas etapas, finalizou-se o cadastro de portos pesqueiros de 2011 com um total de 133 portos localizados ao longo de todo o litoral de São Paulo (Figura 1 - ANEXO I).

2.2. Elaboração do plano amostral baseado na captura descarregada

A metodologia IBGE baseia-se na elaboração de um plano amostral complexo, com amostragem estratificada e por conglomerados, para obter estimativas de capturas totais descarregadas e de seus erros amostrais associados, sobretudo o coeficiente de variação das estimativas. Segundo Lima-Green & Moreira (2012), os portos pesqueiros devem ser previamente divididos em estratos de acordo com sua importância, principalmente, quanto à quantidade de pescado descarregado. Um primeiro estrato, chamado de estrato gerencial ou estrato certo, é composto por portos selecionados arbitrariamente, sem sorteio, por sua maior importância. Nos demais estratos, chamados estratos amostrados, uma seleção de amostras aleatórias simples, sem reposição, dos portos é realizada. Em cada um dos portos pode ser realizada uma amostra estratificada das descargas pesqueiras, com seleção sistemática ou censitária das mesmas. No entanto, este estudo assumiu um plano amostral estratificado e por conglomerado em

um estágio, *i.e.*, as informações de todas as descargas realizadas nos portos amostrados foram consideradas nas análises.

Especialistas do Instituto de Pesca, ligados diretamente à Estatística Pesqueira, realizaram a estratificação portos pesqueiros de cada município paulista. As regras para esta estratificação foram: (1) separar os pontos de descarga de cada município em estratos gerencial e amostrado, de acordo com a quantidade de captura descarregada de cada um; (2) separar um máximo de três portos no estrato gerencial; (3) subdividir os demais portos em quantos estratos julgassem necessários para que cada um ficasse homogêneo (*i.e.* pequeno, médio ou grande); e (4) cada estrato amostrado composto por, no mínimo, três portos. Os municípios de Itanhaém, São Vicente e Bertioga possuem três ou menos portos e, por isso, todos compuseram o estrato gerencial, sem a necessidade de estratificação pelos especialistas nestes municípios. Quaisquer portos com grande especificidade, principalmente em relação ao conjunto de espécies descarregadas, quando identificados, foram transferidos para o estrato gerencial. Isto porque, mesmo com quantidade de captura descarregada semelhante aos demais pontos de seu estrato, sua especificidade poderia causar distorções na posterior expansão da amostra.

2.3. Expansão e inferência estatística

As estimativas constantes nesta etapa do trabalho foram obtidas com a utilização do software R (R Core Team 2017) sendo utilizado o pacote *Sampling* (Tillé & Matei 2016) para a seleção das amostras e o pacote *Survey* (Lumley 2014) para os cálculos das estimativas. Em relação ao tamanho das amostras, dois portos foram sorteados nos estratos com até seis portos e três portos nos demais estratos amostrados. Assim, um total de 77 portos compôs a amostra, 36 nos estratos gerenciais e 41 nos estratos amostrados.

A rotina R foi elaborada de forma que fossem realizadas 100 simulações contendo as etapas de seleção de amostras e cálculos de expansão para cada um dos 14 municípios paulistas. Para cada uma das 100 simulações, a estimativa de captura total por município (Y) e do erro padrão associado (SE), o coeficiente de variação (CV), a raiz quadrada do erro quadrático médio (RMSE), o viés percentual em relação à captura total realmente descarregada (%B) e o efeito de delineamento (Deff) foram calculados. Além disso, o custo anual para o monitoramento pesqueiro, obtido pela soma dos custos individuais dos portos amostrados também foi obtido para cada simulação. As principais fórmulas utilizadas estão descritas na Tabela 1 - ANEXO I. O pacote *Survey* (Lumley 2014) estima o erro padrão como a raiz quadrada da variância estimada de Horwitz-Thompson do total populacional e foi calculada utilizando o método de *Ultimate Cluster*. Comparando os valores obtidos entre as 100 simulações de cada município, os resultados de dois planos amostrais foram considerados: plano amostral 1 – amostra com menor RMSE (PA1); e plano amostral 2 – amostra com o menor custo econômico de monitoramento (PA2). É importante esclarecer que sempre que se fizer referência à acurácia ou a uma estimativa acurada ao longo deste texto, o compromisso entre a variância e o quadrado do viés da estimativa será considerado (menor RMSE).

A estimativa final da captura total por município e do erro padrão associado foram obtidos por média simples entre as 100 estimativas e o coeficiente de variação foi calculado a partir do erro padrão. A cobertura do intervalo de confiança ($1-\alpha=0,95$) (IC95) foi obtida pela contagem das amostras (simulações) nas quais o IC95 englobou o real valor da captura total descarregada.

2.4. Obtenção dos custos de monitoramento pesqueiro

Informações reais obtidas pelo PMAP foram utilizadas para: a obtenção do custo de monitoramento de cada porto pesqueiro, do custo total dos municípios obtido através

da soma dos custos dos portos que compuseram a amostra e também para selecionar a melhor amostra de acordo com o critério PA2.

No PMAP, o monitoramento de portos distantes é realizado através de saídas de campo regulares que utilizam veículos institucionais e particulares. A combinação de portos monitorados por saída é variável, por diversas razões que muitas vezes fogem ao controle do agente de campo. Para facilitar o cálculo, assumiu-se que o monitoramento destes portos é realizado individualmente, com uma única saída de campo para monitorar cada porto pesqueiro. O custo com combustível (em litros/mês) por porto foi obtido baseado na distância percorrida (ida e volta), na frequência semanal de monitoramento e no consumo de combustível (l/km) de acordo com o veículo utilizado (carro, motocicleta ou barco). Todos os outros custos também foram utilizados para obter o custo de cada funcionário e incluem: Pessoal (salários e benefícios como planos de saúde e odontológico); Transporte para coleta (locação de veículos, aquisição de embarcações e motocicletas, motores de popa, combustível, manutenção de barcos e motocicletas); Diárias e hospedagem; Operacional (material de consumo e uniformes); e Serviços (construção do banco de dados, manutenção de computadores, gráfica).

Dependendo do número e do conjunto de portos monitorados, o número de monitores (supervisores), agentes de campo e digitadores varia. Visando um menor custo, o número de agentes de campo em PA2, comparado com PA1, foi reduzido. O número de digitadores foi baseado no total de horas necessárias para incluir todas as descargas reportadas nos portos amostrados, considerando a capacidade de inclusão no banco de dados de 20 descargas por hora por digitador. O custo total com a digitação por porto considerou o custo de inclusão no banco de dados por descarga (custo total com digitadores dividido pelo número total de descargas monitoradas, posteriormente multiplicado pelo número de descargas do porto). O custo total individual dos demais

funcionários foi dividido pelo número de portos sob sua responsabilidade de monitoramento. Despesas relacionadas com a coordenação e gestão do PMAP, apesar de serem custos fixos, foram consideradas e divididas igualmente pelo conjunto de portos de cada amostra. Os cálculos foram baseados nos maiores salários de cada cargo e não incluíram taxas administrativas.

3. Monitoramento pesqueiro inferencial por modelos Bayesianos

Para fins de comparação, a amostra simulada com dados pesqueiros de 2011, que em Miranda et al. (2016) resultou nas estimativas de captura com o menor erro quadrático médio, foi também utilizada para a aplicação dos modelos inferenciais Bayesianos elaborados.

Foram elaboradas cinco modelos inferenciais Bayesianos para o ajuste da amostra simulada de dados pesqueiros. O **Modelo 1** foi elaborado para estimar, prioritariamente, o número de descargas realizadas e a captura total descarregada por município do estado de São Paulo. O número de descargas e a captura por categoria de pescado foram fixados pela proporção de ocorrência de cada categoria de pescado nas descargas totais dos estratos amostrados. E o **Modelo 2**, formulado para primeiro estimar o número total de descargas e a captura por categoria de pescado do estado de São Paulo e depois, através de somatórios, obter as estimativas dos totais para os municípios paulistas.

Inicialmente, duas versões do modelo 2 foram elaboradas: **Modelo 2A**, que se baseou em um modelo de presença e ausência das categorias de pescado nas descargas pesqueiras, combinado com o modelo da captura total descarregada destas mesmas categorias; e **Modelo 2B**, que possui a mesma lógica do modelo 2A, porém considerou o número de descargas dos portos pesqueiros como uma covariável do modelo, sendo conhecida em todos os portos pesqueiros. A partir dos modelos 2A e 2B, foram

elaboradas as versões *Modelo 2Am* e *Modelo 2Bm*, cujas estruturas foram análogas, porém com dados de entrada e inferências detalhadas por mês, na tentativa de melhor absorver as sazonalidades das categorias de pescado.

3.1. Análise dos dados

A estimativa de cada parâmetro foi obtida com a utilização de métodos de Monte Carlo via processo estocástico de Cadeias de Markov (MCMC) (Gelman et al. 2014). Três cadeias MCMC foram simuladas com 20.000 iterações cada. O intervalo de retenção na amostra foi de uma iteração a cada cinco (salto) e as 5.000 iterações iniciais foram ignoradas (aquecimento), o que resultou em uma amostra posterior com 12.000 valores para cada parâmetro dos modelos. As estimativas do número de descargas pesqueiras realizadas e da captura total descarregada foram expressas como sumários das distribuições posteriores preditivas. A convergência das cadeias foi avaliada através do diagnóstico de Gelman-Rubin (Brooks & Gelman 1998), acompanhando os valores de $\hat{R}$, onde valores ≤ 1.1 sugerem a convergência. Todas as análises desta etapa do trabalho foram realizadas através do software R (R Core Team 2017) e do JAGS (Plummer 2016), ligado ao R através do pacote *R2jags* (Su & Yajima 2015).

A simbologia e as principais equações utilizadas na formulação dos modelos inferenciais e preditivos para o monitoramento da atividade pesqueira no Brasil estão apresentadas na Tabela 1 - ANEXO II. A sequência hierárquica das relações entre os dados observados, parâmetros estimados pelos modelos e estimativas da captura e do número de descargas foi descrita a partir de grafos acíclicos dirigidos (*Directed Acyclic Graphs* - DAG), onde as constantes conhecidas e os dados observados são representados por retângulos. As elipses representam variáveis aleatórias ou parâmetros latentes, ambos com uma distribuição de probabilidade associada. As setas pontilhadas

representam dependência estocástica enquanto que as setas sólidas representam relações determinísticas conhecidas.

3.2. Estrutura geral dos modelos inferenciais

No JAGS, as distribuições Normal e log-Normal são parametrizadas pela média μ e pela precisão τ , definida como o inverso da variância ($\tau = \sigma^{-2}$). No restante do texto, distribuições parametrizadas com τ serão sempre indicadas com um asterisco, seguindo o critério adotado em Kinas & Andrade (2010).

Para obter as estimativas por município, dentro de um estrato amostrado h , o número total de descargas realizadas em um porto pesqueiro (D_{hp}) foi descrito por uma distribuição de Poisson cujo parâmetro η_h representa o número médio de descargas dos portos no estrato. De fato, D_{hp} foi descrita por sua generalização em forma de binomial negativa (Poisson-Gamma), onde a variância sempre excede a média (Kinas & Andrade 2010). Esta foi uma medida adotada para lidar com a alta variabilidade entre portos e mês, mesmo que dentro do mesmo estrato.

A captura total descarregada de um porto pesqueiro Y_{hp} condicionado a D_{hp} correspondente foi dada pelo somatório das capturas por descarga C_{hpd} . Se as capturas têm média (m) e a variância (v), então assumimos distribuição log-Normal $C_{hpd} \sim LN^*(\mu, \tau)$, cujos os parâmetros são obtidos a partir de m e v , conforme equações 1 e 2 (Tabela 1 - ANEXO II). Se $Y_{hp} = \sum_{d=1}^{D_{hp}} C_{hpd}$ denota a captura total para D_{hp} descargas, independentes e identicamente distribuídas, então aproximamos a sua distribuição por $Y_{hp} \sim LN^*(\mu_D, \tau_D)$, sendo μ_D e τ_D definidos conforme as equações 3 e 4 (Tabela 1 - ANEXO II). Definiu-se $P(Y_{hp} = 0) = 1$ sempre que $D_{hp} = 0$.

Para as estimativas por categoria de pescado j no estado de São Paulo, dentro de cada estrato h , o número de descargas realizadas em um porto pesqueiro em (h, j)

(D_{hpj}) , dado o número total de descargas de um porto pesqueiro (D_{hp}) , foi estimado por um modelo de presença e ausência. Assim, D_{hp} tornou-se um dos parâmetros da distribuição Binomial que descreve D_{hpj} , sendo a proporção de cada categoria de pescado em relação ao número total de descargas no estrato $(\rho_{(D)_{hj}})$ o segundo parâmetro desta distribuição, isto é, $D_{hpj} \sim \text{Bin}(D_{hp}, \rho_{(D)_{hj}})$. Em $\rho_{(D)_{hj}}$, a inclusão do índice D é necessária para diferenciá-la da proporção de cada categoria de pescado em relação a captura total descarregada no estrato $(\rho_{(Y)_{hj}})$, que será utilizada adiante.

A captura total descarregada de um porto pesqueiro, condicionado ao total de descargas realizadas no mesmo por categoria de pescado (Y_{hpj}) foi descrita por uma log-Normal, cujos parâmetros, μ_{Dj} e τ_{Dj} , foram calculados de maneira análoga ao já descrito nas equações 3 e 4 (Tabela 1 - ANEXO II), porém para cada categoria de pescado analisada. Aqui, também se assumiu $P(Y_{hpj} = 0) = 1$ quando $D_{hpj} = 0$.

3.3. Dados de entrada

As informações obtidas a partir da amostragem e disponíveis para o ajuste do modelo 1 foram: vetores com o número de descargas realizadas $(D_{ih}^{(obs)})$ e com a captura total descarregada em toneladas $(Y_{ih}^{(obs)})$ nos portos monitorados em (i, h) para o ano 2011. A cada $Y_{ih}^{(obs)}$ foi somada a constante $K = (0.01 * (D_{ih}^{(obs)} + 1))$ para evitar os zeros nos dados de entrada. Esta constante foi posteriormente subtraída na predição das capturas dos portos não observados. O número de portos amostrados m_h também foi incluído como dados necessários ao ajuste do modelo. Informações pesqueiras de 2010 foram utilizadas para a determinação das distribuições *prioris* dos parâmetros e serão detalhadas abaixo.

As informações fornecidas como dados dos modelos 2A e 2Am incluíram um vetor com o número de descargas realizadas em 2011 nos portos observados ($D_{hp}^{(obs)}$) e vetores com o número de descargas e a captura total descarregada por categoria de pescado, porto e estrato amostrado de 2011 ($D_{hpj}^{(obs)}$ e $Y_{hpj}^{(obs)}$, respectivamente). Aqui, $Y_{hpj}^{(obs)}$ foi igualmente somada à constante K . O número de portos amostrados m_h e o total de categorias de pescado ($n.sp$) também foram incluídos como dados necessários ao ajuste do modelo.

Para os modelos 2B e 2Bm, os dados de entrada foram os mesmos que nos modelos 2A e 2Am. No entanto, D_{hp} tornou-se um vetor com valores conhecidos para todos os M_h portos. Os valores dos portos observados (1 a m_h) foram utilizados no modelo inferencial e os ($M_h - m_h$) portos não observados no modelo preditivo.

3.4. Distribuições de probabilidade *prioris*

No modelo inferencial 1, o conjunto dos parâmetros (θ_{ih}) para os quais obtivemos uma amostra posterior foi composto por (η_{ih} , m_{ih} e v_{ih}). Foram utilizadas distribuições *prioris* informativas (Hall 2011), baseadas nas informações pesqueiras de São Paulo em 2010, conforme apresentadas na Tabela 2 - ANEXO II. Os portos pesqueiros ativos em 2010 foram estratificados sob os mesmos critérios utilizados para 2011. Somente os dados dos portos observados na amostra simulada foram utilizados para obter os parâmetros das distribuições *prioris*.

Os valores de α_{ih}^{10} e β_{ih}^{10} foram obtidos para descrever uma distribuição gama com moda igual ao número médio de descargas (η_{ih}) em (i, h) para 2010 e assumindo uma média duas vezes maior do que a moda. Os valores de m_{min}^{10} e m_{max}^{10} foram obtidos como $\pm 30\%$ da captura média individual (m_{ih}^{10}) em (i, h) de 2010, porém garantindo o intervalo mínimo (0.01, 0.03) quando m_{ih}^{10} fosse igual ou muito próximo

de zero e em torno do próprio m_{ih}^{10} nos demais casos. Já os valores de v_{min}^{10} e v_{max}^{10} foram pensados para garantir o limite inferior do intervalo da variância entre as capturas individuais (v_{ih}^{10}) em (i, h) de 2010 igual a zero e v_{max}^{10} em torno de v_{ih}^{10} e maior que zero.

No modelo inferencial 2 (versões 2A e 2Am), θ_h foi composto por $(\eta_h, m_{hj}, v_{hj} \text{ e } \rho_{(D)_{hj}})$. Também foram utilizadas distribuições *prioris* informativas, baseadas nas informações pesqueiras de São Paulo em 2010, conforme apresentadas na Tabela 2 - ANEXO II. Os valores de m_{min}^{10} e m_{max}^{10} foram calculados com os mesmos critérios do modelo 1, porém em relação à captura média individual (m_{hj}^{10}) em (h, j) de 2010, assim como v_{min}^{10} e v_{max}^{10} . A distribuição *priori* de $\rho_{(D)_{hj}}$ foi definida como uma distribuição Uniforme em $[0,1]$. Nas versões dos modelos 2B e 2Bm, η_h não fez parte de θ_h , uma vez que D_{hp} passou a ser uma informação conhecida e não precisou mais ser estimada por uma distribuição de Poisson.

3.5. Definição dos modelos inferenciais

As diferentes versões dos modelos inferenciais seguiram a estrutura geral já descrita e foram formalmente definidos conforme a Tabela 3 - ANEXO II.

O DAG apresentado na Figura 1 - ANEXO II define a estrutura do modelo inferencial 1 e as sequências hierárquicas das relações entre os dados observados e os parâmetros do modelo em (i, h, p) . O DAG da Figura 2 - ANEXO II faz o mesmo para os modelos inferenciais 2A, 2Am, 2B e 2Bm, porém com os parâmetros do modelo em (h, p, j) . Os modelos inferenciais 2B e 2Bm foram praticamente os mesmos que os dos modelos 2A e 2Am, porém sem a necessidade de modelar η_h e D_{hp} (Figura 2 – ANEXO II, desconsiderando a parte em azul).

3.6. Modelos preditivos

A imputação múltipla das informações dos portos não observados foi realizada em conformidade com o modelo, com base nos parâmetros estimados sob a forma de distribuições posteriores.

A estrutura dos modelos preditivos e as relações entre os parâmetros estimados e as estimativas de captura total descarregada e do número de descargas por município e categoria de pescado foram descritas pelos DAGs apresentado na Figura 3 – ANEXO II (modelo 1) e na Figura 4 - ANEXO II (modelos 2A, 2Am, 2B e 2Bm). Cada valor simulado (componente da distribuição posterior dos parâmetros) passou a ser uma constante que fornece, dentro da sequência hierárquica, parâmetros das distribuições de probabilidade do que será estimado, resultando nas distribuições (posteriores) preditivas. A sequência dos cálculos realizados está detalhada a seguir.

No modelo 1, considerando (i, h) e para cada valor da amostra posterior dos parâmetros estimados, $M_h - m_h$ valores de $\hat{Y}_{ihp}^{(miss)}$ e $\hat{D}_{ihp}^{(miss)}$ foram simulados. Vale ressaltar que $\hat{Y}_{ihp}^{(miss)}$ foi simulado sempre que $\hat{D}_{ihp}^{(miss)} > 0$, caso contrário foi igual a zero. Uma vez isto feito, obteve-se a captura total descarregada dos portos observados no estrato h ($Y_{ih}^{(obs)}$) e as seguintes distribuições preditivas de interesse (Tabela 1 – ANEXO II): captura dos portos não observados do estrato h ($\hat{Y}_{ih}^{(miss)}$) e a captura total do estrato h ($\hat{Y}_{ih}$) (Equação 5), a captura total descarregada no município i $\hat{Y}_i$ (Equação 6), além da captura total descarregada no estado de SP ($\hat{Y}_{SP}$) (Equação 7). Também foram obtidos o número de descargas realizadas nos portos observados no estrato h ($D_{ih}^{(obs)}$) e as distribuições preditivas do número de descargas dos portos não observados do estrato h ($\hat{D}_{ih}^{(miss)}$), além do total de descargas do estrato h ($\hat{D}_{ih}$) (Equação 9). As estimativas do número descargas para o município i ($\hat{D}_i$) e para o

estado de SP ($\widehat{D}_{SP}$) foram obtidas de forma semelhante ao descrito nas equações 6 e 7, porém utilizando os valores de $\widehat{D}$. A captura total descarregada e o número total de descargas por categoria de pescado no estado de São Paulo ($\widehat{Y}_j$ e $\widehat{D}_j$) (Equações 8 e 10) foram obtidos considerando as proporções de cada categoria de pescado em relação à captura total descarregada ($\rho_{(Y)_{hj}}$) e ao número total de descargas ($\rho_{(D)_{hj}}$) dos portos observados em cada estrato amostrado h .

Nos modelos 2A e 2Am, a predição da captura total descarregada e do número de descargas realizadas por categoria de pescado no estado de São Paulo foi feita de maneira similar ao modelo 1, com as adaptações pertinentes. Os modelos preditivos 2B e 2Bm foram praticamente iguais aos 2A e 2Am. As mesmas equações já mencionadas foram utilizadas, porém sem a necessidade de estimar $\widehat{D}_{hp}^{(miss)}$, uma vez que D_{hp} tornou-se uma covariável com todos os valores conhecidos.

Finalmente, pôde-se obter a captura total descarregada por categoria de pescado do estrato h ($\widehat{Y}_{hj}$) (Equação 11), além da captura total por categoria de pescado no estado de São Paulo ($\widehat{Y}_j$) (Equação 12). As mesmas estimativas foram feitas para número total de descargas realizadas nos portos: $\widehat{D}_{hj}$ (equação 14) e $\widehat{D}_j$, obtida da mesma forma ao descrito na equação 12. Nestes modelos, $\widehat{Y}_i$ e $\widehat{D}_i$ foram obtidos de acordo com as equações 13 e de forma análoga ao descrito na equação 6, respectivamente. É preciso esclarecer que $\widehat{D}_i$ não pode ser obtido a partir da soma das descargas de todas as categorias de pescado, pois pode haver mais de uma categoria em uma mesma descarga, o que faz com que $\sum_{j=1}^{n.sp} \widehat{D}_{hj} > \widehat{D}_h$. As estimativas do número de descargas ($\widehat{D}_{SP}$) e da captura total descarregada para o estado de São Paulo ($\widehat{Y}_{SP}$) foram obtidas pelas mesmas equações descritas no modelo preditivo 1.

SÍNTESE DOS RESULTADOS

CAPÍTULO 1

Survey Sampling for Fisheries Monitoring in Brazil: Implementation and Analysis

L.V. Miranda, P. G. Kinas, G. G. Moreira, R. C. Namora & M.H. Carneiro

Artigo publicado em Brazilian Journal of Oceanography em dezembro de 2016

Versão completa disponível do ANEXO I

As informações censitárias de todas as descargas pesqueiras realizadas em 2011 ao longo da costa de São Paulo foram utilizadas com o objetivo de aplicar a metodologia por amostragem para o monitoramento pesqueiro proposta pelo IBGE (Lima-Green & Moreira 2012) e, assim, comparar e avaliar as diferenças entre os resultados obtidos com a amostragem em relação à coleta censitária de dados pesqueiros. Estas informações pesqueiras foram obtidas através do Programa de Monitoramento da Atividade Pesqueira (PMAP), coordenado por pesquisadores do Instituto de Pesca da Secretaria de Agricultura e Abastecimento do Estado de São Paulo. Os resultados demonstraram que a captura média descarregada, obtida entre as 100 simulações de amostragem, é um bom estimador da captura total para a maior parte dos municípios, pois se mostrou não viciada e como uma alta precisão (exceto para Caraguatatuba). A cobertura do intervalo de confiança ($1-\alpha=0,95$) não abrangeu o verdadeiro valor das capturas em 95% das amostras como deveria, exceto para Ilha Comprida onde a cobertura foi completa (100%), em razão de haver uma discordância com a distribuição Gaussiana utilizada para construir esses intervalos. Quando os resultados da expansão

de cada uma das 100 amostras simuladas foram comparados, ficou claro que o conjunto de portos que fornece o menor RMSE é diferente daquele que resulta no menor custo de monitoramento. A captura estimada por mês e por categoria de pescado e por ambos (domínios não considerados no desenho amostral) demonstraram que quanto maior é o nível de detalhamento necessário nos resultados, menos precisas e mais enviesadas tornam-se as estimativas de captura. Quando comparada com as capturas reais para 2011, a ordem de importância das categorias de pescado baseada nas capturas estimadas alterou-se ligeiramente em algumas posições após o quinto lugar. O pior resultado foi obtido para a ostra (*Cassostrea brasiliiana*), com estimativas sobrestimadas e com muita variabilidade. Isto foi consequência de problemas na estratificação que só foram identificados após a obtenção dos resultados, o que demonstrou a importância de se ter um bom conhecimento prévio dos portos pesqueiros antes de definir a estratificação dos mesmos. Uma pequena redução de custos foi observada com o uso da amostragem (77 portos pesqueiros em 2011) em relação à metodologia censitária (133 portos em 2011) adotada no monitoramento pesqueiro de São Paulo. O custo de monitorar o conjunto de portos da amostra de menor RMSE foi 11,2% menor do que o custo de monitorar todas as descargas de todos os portos paulistas. Esta porcentagem de redução foi de 15,4% quando o menor custo de monitoramento foi levado em consideração na escolha da amostra. Os resultados deste estudo demonstraram que estes custos são diretamente proporcionais ao nível de detalhes e a qualidade dos dados necessários.

CAPÍTULO 2

Fisheries monitoring: is Bayesian inference for finite population sampling an alternative?

Laura Villwock de Miranda & Paul Gerhard Kinas

Artigo submetido em Scientia Marina em julho de 2017

Versão completa disponível do ANEXO II

Cinco modelos inferenciais Bayesianos para o monitoramento pesqueiro de captura descarregada (Y) e do número de descargas (D) para 11 municípios do estado de São Paulo e 20 categorias de pescado foram propostos. Todas as informações pesqueiras utilizadas foram obtidas através do Programa de Monitoramento da Atividade Pesqueira (PMAP), coordenado por pesquisadores do Instituto de Pesca da Secretaria de Agricultura e Abastecimento do Estado de São Paulo. A mesma amostra probabilística de portos pesqueiros, determinada previamente em Miranda et al. (2016) para minimização do erro quadrático médio, foi adotada nas inferências deste trabalho. O modelo que descreve totais de descargas e de captura total descarregada por município resultou nas melhores estimativas de Y para a maioria dos municípios paulistas. O modelo que descreve sequencialmente o número total de descargas e de captura por categoria de pescado foi mais bem ajustado para a maioria das categorias de pescado não sazonais. Para aquelas com sazonalidade marcada, os melhores resultados foram obtidos com a inclusão de informações mensais nos modelos. A inclusão do número de descargas como covariável conhecida não trouxe um melhor ajuste do modelo aos dados como era esperado. A estratificação dos portos pesqueiros adotada no desenho

amostral foi uma das razões para os ajustes inadequados em alguns municípios ou categorias de pescado. Finalmente foi apresentada e discutida uma nova proposta de desenho amostral que respeitasse explicitamente a relação linear entre o $\ln(Y)$ e $\ln(D)$, presente nos modelos inferenciais Bayesianos aqui usados. Embora preliminares, resultados encorajadores da nova proposta foram obtidos para os cinco principais municípios pesqueiros do estado, com expressivo aumento da eficiência amostral.

CONSIDERAÇÕES FINAIS

Monitoramento pesqueiro inferencial baseado em delineamento amostral

Quanto maior for o nível de detalhamento necessário nos resultados, menos precisas e mais enviesadas serão as estimativas de captura total descarregada.

Em uma metodologia baseada em delineamento amostral, se for necessário ter boas estimativas da captura total por município, o que é importante para o estado de São Paulo, esta variável deve ser considerada explicitamente no plano amostral (Bolfarine & Bussab 2005, Lumley 2010). Se a variável município não fosse incluída no plano amostral, haveria o risco de não ter algum município representado na amostra, já que nesta abordagem, para um estrato não observado na amostra não pode ser efetuada inferência. Por outro lado, ao se inferir sobre variáveis que não foram consideradas no plano amostral, como categoria de pescado, a informação contida na amostra ficou dividida entre estratos e com menor representatividade, resultando em estimativas menos acuradas. Detalhar dados ainda mais, por artes de pesca ou tipo de frota, seria apenas piorar a acurácia das estimativas.

O conjunto de portos que fornece o menor erro quadrático médio é diferente daquele que resulta no menor custo de monitoramento.

Entre as 100 amostras simuladas, nenhuma foi considerada ótima na forma como Cochran (1977) a define, na qual a alocação ótima é obtida quando, para um dado tamanho de amostra, a amostra resulta nas estimativas de maior precisão com o menor custo de monitoramento. Aumentar o tamanho das amostras visando ter estimativas mais acuradas aumentaria os custos. Por outro lado, diminuir o tamanho destas amostras para reduzir custos poderia significar ter que deixar alguns estratos de fora e, conseqüentemente, perder qualidade nos resultados inferenciais. Os resultados deste

trabalho demonstraram que, para o litoral de São Paulo, realizar o monitoramento pesqueiro inferencial significa ter que escolher entre: estimativas mais acuradas ou baixo custo de monitoramento.

A redução de custo do monitoramento pesqueiro obtida com a aplicação de amostragem foi mínima quando comparado aos dados obtidos de forma censitária.

Mesmo com uma redução de 42,1% no número de portos monitorados, a redução nos custos do monitoramento por amostragem foi muito pequena e há algumas explicações para que isto tenha ocorrido. O menor número de portos monitorados não necessariamente reduz a distância a ser percorrida para monitorá-los. No monitoramento pesqueiro do litoral de São Paulo, o Instituto de Pesca adota estratégias para reduzir o custo de monitoramento, tais como o auto-registro e o uso de agentes de campo volantes (em constante deslocamento para monitorar mais de um porto, muitas vezes através do resgate de informações e do auto-registro). Assim, em municípios onde a costa é extensa, *e.g.* Ubatuba e São Sebastião, quando portos nos extremos da área são amostrados, a distância a ser percorrida, os custos com o agente de campo, assim como com o uso do automóvel e com combustível serão os mesmos necessários para monitorar todos os portos ao longo do percurso.

Além disso, os custos fixos e a infraestrutura necessária para realizar o monitoramento pesqueiro pouco se alteram entre as coletas censitárias e por amostragem. O custo com pessoal é, certamente, o principal custo do monitoramento pesqueiro e, embora tenha havido redução no número de funcionários, esta não foi suficiente para diminuir o custo final de forma expressiva. Com a adoção do plano amostral, ainda é preciso ter agentes de campo exclusivos para monitorar todas as

descargas dos portos pesqueiros alocados nos estratos gerenciais dos diferentes municípios e, uma vez que a maior parte das descargas continua sendo monitorada, a redução no número de monitores e digitadores também não foi expressiva.

A análise do custo de monitoramento demonstrou que, para o litoral de São Paulo, a redução de custos obtida com a aplicação da amostragem não compensa a perda de qualidade das informações pesqueiras quando comparada à coleta de dados censitária.

Monitoramento pesqueiro inferencial por modelos Bayesianos

O maior detalhamento nos dados de entrada do modelo resultou em estimativas mais acuradas para algumas categorias de pescado.

A acurácia das estimativas de captura total descarregada para algumas categorias de pescado obtidas através de modelos inferenciais foi maior quando informações mensais foram fornecidas como dados de entrada para o modelo. Este foi o caso de algumas categorias de pescado com sazonalidade pesqueira marcada e/ou com períodos de parada da pesca estabelecidos pela legislação vigente. Por outro lado, passar a lidar com informações mensais por categoria de pescado e porto pesqueiro acentuou um problema comum em dados de pesca, principalmente nas categorias de pescado menos frequentes nas descargas: a presença de muitos zeros.

Difícilmente um único modelo será bem especificado e fornecerá boas estimativas para todas as situações. Uma das vantagens da abordagem Bayesiana hierárquica é que diferentes versões de modelos podem ser elaboradas para lidar com situações específicas. Toda a estrutura do modelo é explícita, o que permite a inclusão relativamente fácil de covariáveis e de outras modificações a qualquer momento. O mesmo vale para outras informações relevantes que podem ser consideradas no modelo

através do uso de distribuições *prioris* informativas. Desta forma, pode-se elaborar um modelo exclusivo para obter estimativas para as categorias de pescado com sazonalidade marcada e outro, sem informações mensais e com menos zeros nos dados de entrada e, portanto, mais estável para aquelas sem sazonalidade. Esta mesma estratégia pode ser estendida para as estimativas por aparelho de pesca que também apresentam períodos de parada de pesca ao longo do ano.

A inclusão do número de descargas como covariável conhecida não trouxe um melhor ajuste do modelo aos dados como era esperado.

Todas as versões do modelo inferencial Bayesiano foram estruturadas para descrever sequencialmente o número total de descargas (D_{hp}) e a captura total descarregada (Y_{hp}) de um porto pesqueiro dentro de um estrato amostrado, assumindo uma relação diretamente proporcional entre estas variáveis. Entretanto, os estratos criados por município para a aplicação da metodologia IBGE, e adotados nos modelos inferenciais Bayesianos com o objetivo de comparar os resultados das duas abordagens, foram compostos por portos pesqueiros relativamente homogêneos em relação à captura total descarregada, mas não necessariamente em relação ao número de descargas realizadas. A escolha deste desenho amostral foi a principal razão para que as estimativas do número de descargas fossem menos acuradas tanto na inferência baseada em delineamento amostral quanto nos modelos inferenciais Bayesianos.

A falta de homogeneidade no número de descargas em conjunto com o pequeno número de portos pesqueiros na maioria dos estratos amostrados fez com que a relação linear entre D_{hp} e Y_{hp} não fosse observada em muitos destes estratos. Sendo assim, mesmo considerando D_{hp} uma covariável com valores conhecidos para todos os portos pesqueiros (observados ou não na amostra), estimativas de captura total descarregada

por município e por categoria de pescado não foram expressivamente melhores daquelas obtidas com D_{hp} sendo considerada uma variável aleatória estimada pelo modelo nos portos não amostrados.

A adoção do desenho amostral recomendado na metodologia por amostragem proposta pelo IBGE não foi adequada para a inferência baseada em modelo.

Para utilizar a inferência baseada em modelo na atividade pesqueira deve-se pensar em uma estratificação totalmente diferente da que foi adotada originalmente neste trabalho. Incluir os municípios na estratificação foi ruim para os resultados tanto da inferência por delineamento amostral quanto por modelos Bayesianos, pois resultou em estratos muito pequenos e pouco homogêneos em relação à captura total descarregada (Y) e, principalmente, ao número total de descargas (D). O que distingue as duas abordagens é que, ao contrário da inferência por delineamento amostral, a inferência baseada em modelo não precisa ter portos pesqueiros amostrados em todos os municípios para obter estimativas de captura e número de descargas por município. Os parâmetros do modelo são estimados a partir da amostra e usados para obter as estimativas de todos os portos pesqueiros não amostrados. Posteriormente, os valores dos portos são agregados de acordo com o município ao qual cada porto pertence.

As propriedades e vantagens da abordagem Bayesiana baseada em modelo são válidas quando o modelo é bem especificado para os dados observados (Zangeneh & Little 2015). Em modelos inferenciais Bayesianos, o desenho amostral deve ser planejado para proporcionar um bom ajuste dos dados observados ao modelo em questão. Por isso, pensou-se em uma nova maneira de estratificar os portos pesqueiros que respeitasse a relação linear entre o $\ln(Y)$ e o $\ln(D)$ de todos os portos pesqueiros do litoral paulista. Os *outliers* desta relação linear foram identificados e alocados um

estrato gerencial (18 portos) e todos os demais portos (115) em estratos amostrados. Esta nova versão do desenho amostral e também do modelo inferencial é apenas uma primeira ideia, simplificada, mas que já apresentou uma maior eficiência amostral e resultados promissores para os principais municípios pesqueiros do litoral de São Paulo.

Notadamente, a pesquisa por amostragem para o monitoramento pesqueiro pode ser útil quando recursos financeiros são limitados, mas, principalmente, em regiões onde a atividade pesqueira for mais homogênea. Em regiões como o litoral de São Paulo, onde a pesca é altamente diversificada (Mendonça & Miranda 2008, Carneiro et al. 2013), o monitoramento pesqueiro inferencial, se não for bem planejado e executado para lidar com esta grande variabilidade, pode comprometer a qualidade e confiabilidade das informações pesqueiras e, portanto, a gestão dos recursos pesqueiros.

Neste trabalho ficou bastante claro que um desenho amostral, por si só, dificilmente captará toda a variabilidade da atividade pesqueira em uma amostra que seja, ao mesmo tempo, representativa e de baixo custo de obtenção. Por outro lado, características que podem causar esta variabilidade tanto na captura total descarregada quanto no número total de descargas, como o número de embarcações e pescadores entre outras informações conhecidas para todos os portos pesqueiros, podem ser covariáveis úteis para o melhor ajuste do modelo inferencial Bayesiano para a atividade pesqueira. A possibilidade de incluir informações prévias e diferentes covariáveis na estrutura do modelo é uma vantagem importante da inferência Bayesiana, permitindo que esta abordagem lide melhor com pequenas amostras e informações altamente variáveis quando comparada com a abordagem frequentista (Little 2004, Kéry 2010).

Sabidamente, a avaliação dos estoques e a gestão dos recursos pesqueiros são dependentes de séries temporais ininterruptas, com dados pesqueiros confiáveis e com

boa qualidade. A escolha da metodologia para a coleta destes dados pesqueiros depende muito dos objetivos do monitoramento pesqueiro, das atribuições institucionais de quem o realiza, da dinâmica pesqueira da região monitorada, além da disponibilidade de pessoal e recursos financeiros, entre outros fatores (FAO, 1999). Portanto, não é correto falar-se em método certo ou errado, mas sim naquele que é mais adequado para a realidade da atividade pesqueira de cada região. Acima de tudo, é preciso priorizar métodos de coleta que tragam qualidade para as informações sobre as espécies, sejam estas importantes pela sua captura total descarregada ou por critérios ecológicos, econômicos, sociais ou culturais. Independente da metodologia a ser adotada e seja qual for seu custo de aplicação, é um fato que: o investimento no monitoramento pesqueiro será sempre menor do que os custos econômicos, sociais e ambientais de não se ter informações de qualidade para permitir a gestão da atividade pesqueira baseada em evidências.

Depois de todas as tentativas de ajustar as diferentes versões aos dados, acredita-se que a inferência Bayesiana em populações finitas pode ser uma boa alternativa para o monitoramento inferencial da atividade pesqueira. Diante de um modelo bem especificado para os dados pesqueiros observados, é possível obter estimativas acuradas a partir de amostras menores e, conseqüentemente, com um menor custo para o monitoramento pesqueiro.

Os resultados obtidos até aqui foram o primeiro passo para a estruturação de um modelo inferencial Bayesiano para o monitoramento pesqueiro e, certamente, os próximos desafios ainda serão grandes e numerosos. As próximas etapas deverão incluir o desenvolvimento de mecanismos e procedimentos de análise de dados que permitam: (1) detectar covariáveis apropriadas e disponíveis no momento da elaboração da estratégia amostral e da formulação do modelo; (2) construir distribuições de

probabilidade *prioris* informativas que melhorem o ajuste e a estabilidade do modelo inferencial Bayesiano para a predição aos portos pesqueiros não amostrados; (3) elaborar desenhos amostrais mais adequados para o modelo proposto; e, por fim, (4) usar de maneira eficiente a inferência Bayesiana em populações finitas no monitoramento pesqueiro por amostragem.

LITERATURA CITADA

- ARAGÃO, JAN & S MARTINS. 2006. Censo Estrutural da Pesca – Coleta de Dados e Estimação de Desembarques de Pescado – IBAMA, Brasília/DF. 180p.
- ÁVILA-DA-SILVA, AO, MH CARNEIRO, & L FAGUNDES. 1999. Sistema gerenciador de banco de dados de controle estatístico de produção pesqueira marinha – ProPesq. In: Anais do XI Congresso Brasileiro de Engenharia de Pesca e I Congresso Latinoamericano de Engenharia de Pesca. 2:824-832. Recife/PE (17-21/01/1999).
- ÁVILA-DA-SILVA, AO, MH CARNEIRO, JT MENDONÇA, GCC BASTOS, LV MIRANDA, WR RIBEIRO & S SANTOS. 2017. Produção Pesqueira Marinha e Estuarina do Estado de São Paulo – Dezembro de 2016. *Inf Pesq de São Paulo*. 80: 1-4. Disponível em <http://www.propesq.pesca.sp.gov.br/10/conteudo>. Acesso em : 12/06/2017.
- BOLFARINE, H & WO BUSSAB. 2005. Elementos de Amostragem, 1ª Ed. Editora Blucher, São Paulo/SP. 274p.
- BROOKS, SP & A GELMAN. 1998. Alternative methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7:434-455.
- BUSSAB, WO & PA MORETTIN. 2012. Estatística Básica. 7ª Ed. Editora Saraiva, São Paulo/SP. 540p.
- CADDY, JF & GP BAZIGOS. 1985. Practical guidelines for statistical monitoring of fisheries in manpower limited situations. *FAO Fisheries Technical Paper*. Nº 257. FAO, Roma. 86p.
- CADIMA, EL. 2003. Fish stock assessment manual. *FAO Fisheries Technical Paper*. Nº 393. FAO, Roma. 161p.

- CADIMA, EL, AM CAMELO, M AFONSO-DIAS, P CONTE DE BARROS, MO TANDSTAD & JI DE LEIVA-MORENO. 2005. Sampling methods applied to fisheries science: a manual. *FAO Fisheries Technical Paper*. Nº 434. FAO, Roma, 88p.
- CARNEIRO, MH, JA KOLLING, AO ÁVILA-DA-SILVA, JT MENDONÇA, RC NAMORA, LV MIRANDA. 2013. Pesca nas Áreas de Proteção Ambiental Marinha do estado de São Paulo, Brasil, entre agosto de 2008 e julho de 2009. *Inf Pesq de São Paulo*. 36: 1-34. Disponível em: <http://www.propesq.pesca.sp.gov.br/10/conteudo>. Acesso em: 11/07/2017.
- COCHRAN, WG. 1977. Sampling Techniques. Wiley: New York. 428p.
- CRESSEY, D. 2015. Eyes on the ocean. *Nature*. 519: 280-282.
- DIAS-NETO, J. 2010. Pesca no Brasil e seus aspectos institucionais – um registro para o futuro. *Revista CEPSUL - Biodiversidade e Conservação Marinha*, 1 (1): 66-80.
- DIAS-NETO, J. 2011. Números e Baionetas – A Nova Estatística da Produção Pesqueira do Brasil. Erro Estatístico ou Equívoco Político?. *Pesca & Mar - Informativo SAPERJ* (março/abril). Rio de Janeiro/RJ. 132: 31-34.
- FAO. 1999. Guidelines for the routine collection of capture fishery data. *FAO Fisheries Technical Paper*. Nº 382. FAO, Roma. 113p.
- FAO. 2016. The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all. Roma. 200p.
- FREIRE, KMF, JAN ARAGÃO, ARR ARAÚJO, G VELASCO-CANZIANI, MH CARNEIRO, FDS GONÇALVES, KA KEUNECKE, JT MENDONCA, PS MORO, FS MOTTA, G OLAVO, PR PEZZUTO, RF SANT'ANA, RA SANTOS, I TRINDADE-SANTOS, JA VASCONCELOS, M VIANNA & E DIVOVICH.

2015. Reconstruction of catch statistics for Brazilian marine waters (1950-2010). *Fisheries Centre Research Reports*. 23: 3-30.
- GELMAN, A, JB CARLIN, HS STERN, DB DUNSON, A VEHTARI & DB RUBIN. 2014. Bayesian Data Analysis. 3ª Ed. Texts in Statistical Science Series. CHAPMAN & HALL/CRC PRESS. Boca Raton, Florida, USA. 639p.
- GELPKE, N, A BEHNAM & M VISBECK. 2013. World Ocean Review 2013: Living with the oceans. 2: The future of fish—the fisheries of the future. *Hamburg: Maribus GmbH*. 145p. Disponível em: http://worldoceanreview.com/wp-content/downloads/wor2/WOR2_english.pdf. Acesso em 02/03/2016.
- GULLAND, JA. 1966. Manual of Sampling and Statistical Methods for Fisheries Biology – Part 1 – Sampling Methods. *FAO Manual in Fishery Science*, Nº 3. FAO, Roma. 87p.
- HALL, B. 2011. Bayesian inference. Disponível em: <https://cran.r-project.org/web/packages/LaplacesDemon/vignettes/BayesianInference.pdf>. Acesso em: 02/03/2016.
- HILBORN, R & T BRANCH. 2013. Does catch reflect abundance? No, it is misleading. *Nature*. 494: 303-306.
- HILBORN, R. & CJ WALTERS. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. New York: Chapman and Hall. 570p.
- ISSAC, VJ, ML RUFFINO & P MELLO. 2000. Considerações sobre o Método de Amostragem para a Coleta de Dados sobre Captura e Esforço Pesqueiro no Médio Amazonas. *IBAMA – Coleção Meio Ambiente. Série Estudos Pesca*. 22: 175-199.
- ISSAC, VJ, RV ESPÍRITO SANTO & JLG NUNES. 2008. A estatística pesqueira no litoral do Pará: resultados divergentes. *Pan-American Journal of Aquatic Sciences*, 3(3): 205-213.

- KÉRY, M. 2010. Introduction to WinBUGS for Ecologists: a Bayesian approach to regression, ANOVA, mixed models and related analyses. 1ª Ed. Academic Press. Burlington/USA. 302p.
- KINAS, PG & HA ANDRADE. 2010. Introdução à Análise Bayesiana (com R). Editora maisQnada. Porto Alegre/RS. 240p.
- KOTARO, O, AE PUNT & R HILBORN. 2015. How do marine closures affect the analysis of catch and effort data? *Can. J. Fish. Aquat. Sci.* 72: 1177–1190.
- KRUSCHKE, JK. 2015. Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. 2ª Ed. Elsevier. London/England. 759p.
- LIMA-GREEN, AP & GG MOREIRA. 2012. Metodologia Estatística da Pesca – Pesca Embarcada. Textos para Discussão – Diretoria de Pesquisas (IBGE), Rio de Janeiro/RJ. 40: 1-52.
- LITTLE, RJ. 2004. To model or not to model? Competing modes of inference for finite population sampling. *J. Amer. Statist. Assoc.* 99(466): 546-556.
- LUMLEY, T. 2014. Survey: analysis of complex survey samples. R package version 3.30.
- MAUNDER, MN & AE PUNT. 2004. Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*. 70: 141–159.
- MCCARTHY, MA. 2007. Bayesian methods for ecology. Cambridge University Press. Cambridge/England. 296p.
- MENDONÇA, JT & LV MIRANDA. 2008. Estatística pesqueira do litoral sul do estado de São Paulo: subsídios para gestão compartilhada. *Pan-American Journal of Aquatic Sciences*, 3(3): 152-173.

- MIRANDA, LV, PG KINAS, GG MOREIRA, RC NAMORA & MH CARNEIRO. 2016. Survey sampling for fisheries monitoring in Brazil: implementation and analysis. *Braz. J. Oceanogr.* 64(4): 401-414.
- PAULY, D. 2013. Does catch reflect abundance? Yes, it is a crucial signal. *Nature*. 494: 303-306.
- PAULY, D & D ZELLER. 2016. Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nature communications*. 7:10244.
- PLUMMER, M. 2016. JAGS Version 4.2.0 user manual. Disponível em: <https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/>. Acesso em 07/07/2016.
- R CORE TEAM 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Disponível em: <http://www.R-project.org>. Acesso em: 14/03/2016.
- RUBIN, DB. 1976. Inference and missing data. *Biometrika*. 581-592.
- SPARRE, PJ. 2000. Manual on sample-based data collection for fisheries assessment - Examples from Viet Nam. *FAO Fisheries Technical Paper*, Nº 398. FAO, Roma. 171p.
- SPARRE, PJ & SC VENEMA. 1997. Introdução à avaliação de mananciais de peixes tropicais. Parte 1: manual. *FAO Documento Técnico sobre as pescas*, Nº 306 (1), Revisão 2. FAO, Roma. 404p.
- SU, YS. & M YAJIMA. 2015. R2jags: Using R to run 'JAGS'. R package version 0.5-7. Disponível em: <https://cran.r-project.org/web/packages/R2jags/index.html>. Acesso em: 03/09/2015.

TILLÉ, Y. & A MATEI. 2016. Sampling: Survey Sampling. R package version 2.8.

Disponível em: <http://CRAN.R-project.org/package=sampling>. Acesso em: 05/01/2017

ZANGENEH, SZ & RJ LITTLE. 2015. Bayesian inference for the finite population total from a heteroscedastic probability proportional to size sample. *J. Surv. Stat. Methodol.* 3(2): 162-192.

ANEXO I: Survey Sampling for Fisheries Monitoring in Brazil:

Implementation and Analysis

*Laura Villwock de Miranda, Paul Gerhard Kinas, Guilherme Guimarães Moreira,
Rafael Cabrera Namora, Marcus Henrique Carneiro*

Artigo publicado em dezembro de 2016

BRAZILIAN JOURNAL OF OCEANOGRAPHY, 64(4):401-414

Em arquivo PDF anexado ao email

Survey sampling for fisheries monitoring in Brazil: implementation and analysis

Laura Villwock de Miranda^{1,5}, Paul Gerhard Kinas^{1,2}, Guilherme Guimarães Moreira^{3#}, Rafael Cabrera Namora⁴, Marcus Henrique Carneiro⁵*

¹ Programa de Pós-Graduação em Oceanografia Biológica, Universidade Federal do Rio Grande (FURG).

(Av. Itália, km 8, Carreiros, 96203-900, Rio Grande, RS, Brazil).

² Instituto de Matemática, Estatística e Física (IMEF/FURG).

(Av. Itália, km 8, Carreiros, 96203-900, Rio Grande, RS, Brazil).

³ Instituto Brasileiro de Geografia e Estatística (IBGE).

(Rua Oliveira, 523, 4th floor, Cruzeiro, 31310-150, Belo Horizonte, MG, Brazil).

⁴ Fundação de Desenvolvimento da Pesquisa do Agronegócio (FUNDEPAG)

(Rua Dona Germaine Burchard, 409, Água Branca, 05002-062, São Paulo, SP, Brazil).

⁵ Instituto de Pesca de São Paulo (IP/APTA-SP)

(Rua Joaquim Lauro de Monte Claro Neto, 2275, Itaguá, 11680-000, Ubatuba, SP, Brazil).

*Corresponding author: lauravillwockmiranda@gmail.com

ABSTRACT

Census of fishing data about the landings carried out along the São Paulo coast during 2011 was used to evaluate and compare the survey sampling for fisheries monitoring, expecting reliable results along with an important cost reduction. Estimates of total catch for the São Paulo State as a whole and by municipality were relatively accurate (high precision and low bias). Estimated catch by month, by fish categories and both (factors not considered in the sampling design) demonstrated that, as the level of required detail increased, the catch estimates became more biased and less precise. However, when comparing to the 2011 true catches, the order of importance of fish categories based on estimated catches changed slightly in some positions after the fifth place. There was a minor cost reduction due to the sampling in comparison with the census methodology currently in use (15.4% at most). The results demonstrated that fisheries monitoring costs are directly proportional to the required level of details and data quality.

Descriptors: Fishing activity, Fishing landings, Sampling design, Inference, Monitoring costs.

RESUMO

Informações sobre as descargas pesqueiras realizadas em 2011 ao longo da costa de São Paulo foram utilizadas com o objetivo de avaliar e comparar os métodos de amostragem em campanhas voltadas para o monitoramento pesqueiro. Espera-se com isto um conjunto de dados consistentes, além de uma importante redução de custos. As estimativas da captura total para o estado de São Paulo e por municípios foram relativamente acuradas (alta precisão e baixo viés). A captura estimada por mês, por categoria de pescado e por ambos (domínios não considerados no desenho amostral) demonstraram que quanto maior é o nível de detalhamento menos precisas e mais enviesadas tornam-se as estimativas de captura. Quando comparada com as capturas reais para 2011, a ordem de importância das categorias de pescado baseada nas capturas estimadas alterou-se ligeiramente em algumas posições após o quinto lugar. Houve uma pequena redução de custos devido à amostragem em comparação com a metodologia censitária atualmente em uso no estado de São Paulo (máxima de 15,4%). Os resultados demonstraram que os custos do monitoramento pesqueiro são diretamente proporcionais ao nível de detalhamento e à qualidade dos dados requeridos.

Descritores: Atividade pesqueira, Descargas pesqueiras, Desenho amostral, Inferência, Custos de monitoramento.

[#] IBGE is exempt from any responsibility for the opinions, information, data and concepts presented in this report which are exclusive responsibility of the authors.

<http://dx.doi.org/10.1590/S1679-87592016132706404>

INTRODUCTION

Catch and fishing effort are the most basic information that can be obtained about any fishing activity. To guarantee that at least these data are reliably collected and maintained over time is crucial to formulate effective fisheries policies and management plans (HILBORN; WALTERS, 1992; CADIMA, 2003).

Monitoring and obtaining fishing information can be performed in two forms: by sampling surveys (CADDY; BAZIGOS, 1985; ARAGÃO; MARTINS, 2006; LIMA-GREEN; MOREIRA, 2012) or by census (FAO, 1999; MENDONÇA; MIRANDA, 2008; ÁVILA-DA-SILVA et al., 2015). In general, a census is recommended when the population is small, sampling errors are large, information is cheap to obtain or the cost in making the wrong decisions is high. Sampling techniques must be used when the population is very large and/or the cost (concerning money and time) to obtain information is high (CADDY; BAZIGOS, 1985; BOLFARINE; BUSSAB, 2005).

When the available data constitute only a portion of a population (collected by sampling), then there are two ways of dealing with the inferences: (1) based on a sampling plan specially designed by a finite population with a controlled random selection procedure where all probabilities involved can be known (design-based); and (2) based on observational research (model-based), where there is no control over the sampling plan and the specification of a model plays a fundamental role to connect the observed data to the parameters of the population (COCHRAN, 1977; BUSSAB; MORETTIN, 2012). Basically, in a model-based approach, data are assumed to have been generated from a random process specified by a probability model so that conclusions can be generalized to other situations where the same process operates, while design-based inference cannot be generalized to other populations which were not sampled (LUMLEY, 2010).

The design-based approach is usually applied to the analysis of complex survey samples and, up to now, widely adopted by fisheries monitoring methodologies (FAO, 1999). Estimates of total catch, their variance and any other population quantities are obtained based on the Horvitz-Thompson estimator (HORVITZ; THOMPSON, 1952). This is an unbiased estimator of population total applicable to any sampling design with or without replacement, from a finite population, when unequal but known selection probabilities are used. The estimation

procedure weighs each selected unit by the inverse of its overall selection probability and known nonzero pairwise probabilities are required for unbiased variance estimation (LUMLEY, 2010).

Historically, fisheries monitoring in Brazil has certainly been influenced by different political and institutional arrangements made along the development of national extractive fishery (DIAS-NETO, 2010, 2011; LIMA-GREEN; MOREIRA, 2012). The adoption of different methodologies for different fisheries or for the same fishery in time has been common, with periods of interruption in data collection in different regions along the Brazilian coast.

In some States of Brazil, the EstatPesca (ARAGÃO; MARTINS, 2006) was the most adopted sampling methodology for fisheries monitoring since the nineties (LIMA-GREEN; MOREIRA, 2012). This methodology was based on the follow-up of fixed samples of fishing vessels, which required a permanently updated registry of all vessels in operation. This proved impracticable, mainly for small-scale fisheries, where sales and changes in the names and in the characteristics of the vessels are very frequent. It was also usual that vessel sampling was intentionally motivated by logistics considerations and not conducted as a probabilistic sampling survey and, therefore, subjected to bias (ISAAC et al., 2008). In order to reduce biased estimates, more samples should be taken, increasing the costs of the sampling process. ISAAC et al. (2008) observed a catch overestimation when EstatPesca was applied to the fisheries monitoring of the Pará State (Northern Brazil) and concluded that at least 70% of the fleet should be sampled to place the error of the estimates at acceptable levels.

A new sampling methodology for fisheries monitoring has been proposed by LIMA-GREEN and MOREIRA (2012), technicians from the official Brazilian Institute of Statistics, the “Instituto Brasileiro de Geografia e Estatística” (IBGE). This methodology aimed to avoid the weakness detected in EstatPesca by defining fishing landings as sample unit instead of fishing vessels. Its use has been gaining recognition in Brazil, since the government has been supporting and encouraging its adoption for new programs for fisheries monitoring. However, this is a sampling design method that can present some limitations for use in survey sampling of fishing activity, by its great diversity, variable characteristics and different strategies. The use of different combinations of more than one fishing gear on a same trip, several fishing seasons and many

target species, besides fishing landings with small catches spread over large extensions of the coast are some of the characteristics described for small-scale fishing in Brazil (ISAAC et al., 2000, 2008; MENDONÇA; MIRANDA, 2008). Furthermore, when the IBGE methodology is applied with a two stages sampling, the total numbers of landings carried out in a fishing landing place must be known. This information can be very difficult to be obtained and, of course, it is not known in advance. Many factors can influence the dynamics of fishing landings, such as the size of the landing facility, the number of fishing vessels and fishermen at this facility, the fishing seasons, the type of fishing fleet and the fishing gears used by this fleet.

Fisheries monitoring of the São Paulo coast is, however, an exception in Brazil and its first records of fishing information dates back to 1944. Since its creation in 1969, the Fisheries Institute of the Department of Agriculture and Food Supply of São Paulo State has been the institution responsible for the collection, storage, processing and disclosure of census data (FAO, 1999) about the marine fisheries production landed along the São Paulo coast (MENDONÇA; MIRANDA, 2008; ÁVILA-DA-SILVA et al., 2015).

Realistic and good quality data, where the true total population is known, are required to evaluate and compare survey sampling methods (LUMLEY, 2010). In this paper, the complete fishing data of São Paulo State collected during 2011 were used to simulate probability samples following the sampling design described by LIMA-GREEN and MOREIRA (2012) and to compare the results of these simulations to the true total landed catches. In addition to the quality of estimates, the costs to perform fisheries monitoring on the São Paulo coast were also considered in order to evaluate losses and gains of the sampling methodology when compared to census data collection. The hypothesis of this study is that the survey sampling method applied to fisheries monitoring of the São Paulo coast will generate reliable results along with an important cost reduction when compared to the census data collection.

MATERIAL AND METHODS

OBTAINING FISHING INFORMATION

Fishing landings census data collected on the São Paulo coast during 2011 were used to apply the sampling methodology for fisheries monitoring proposed by

IBGE (LIMA-GREEN; MOREIRA, 2012). These data were obtained through the Fishing Activity Monitoring Program (PMAP), coordinated by fisheries scientists from the Fisheries Institute of the Department of Agriculture and Food Supply of São Paulo State.

In March 2008, PMAP began to be used aiming to evaluate the impact on fishing activity by oil and gas exploration activities by Petrobras in the Santos Basin. The PMAP applies the census methodology to collect fisheries statistics (FAO, 1999; MENDONÇA; MIRANDA, 2008; ÁVILA-DA-SILVA et al., 2015), and currently monitors 196 fishing ports and landing places (just “ports” in the remaining text) in 15 municipalities included in the area of influence of the oil and gas exploration in Santos Basin. The municipality of Santos has only one port and was considered a single municipality together with the neighboring city of Guarujá to preserve the confidentiality of information. In order to obtain information on catch and fishing effort, field agents perform structured interviews with fishermen on the occasion of landing. This information is complemented with retrieved fishermen’s records about their daily fishing operations (self-registration), in logbooks and with records provided by fishing enterprises. The storage, processing, analysis and provision of fishery statistics are carried out by the System Manager ProPesq® (ÁVILA-DA-SILVA et al., 1999), currently operating in a web platform, called ProPesqWEB (<http://www.propesq.pesca.sp.gov.br>).

APPLYING SAMPLING METHODOLOGY TO FISHERIES MONITORING

The organization and structuring of fishing landings census data of the State of São Paulo and the sampling design to extract fishing landings from it were defined during a Workshop with technicians of IBGE, who are the authors of the methodology being validated (LIMA-GREEN; MOREIRA, 2012).

THE REGISTRY OF PORTS

Information about all reported fishing landings (just “landings” in the remaining text) carried out on São Paulo coast during 2011 was extracted from the database of the Fisheries Institute, comprising 227 ports, although some of them with very few landings or located too close to one another. In order to define the population of interest and conclude this register, the following criteria were adopted: (1) Remove the ports with fewer than 40 landings per year

(fewer than one weekly landing); (2) Reconsider all ports with at least one landing greater than 500 t; (3) Gather in a single port all places which, for logistics, have distinct names in the database, but in practice could be part of only one port; (4) Remove from the registry, the ports that were deactivated in 2013. After accomplishing these steps, the register was finalized with a total of 133 ports located along the entire coast of São Paulo State (Figure 1).

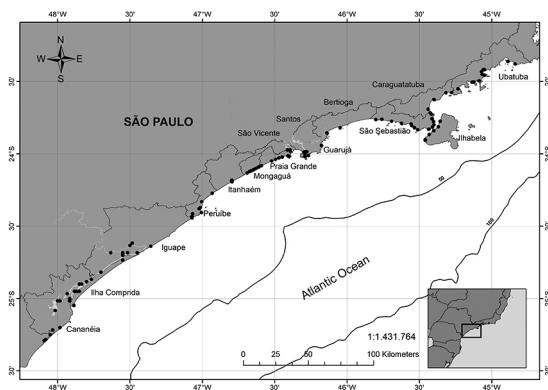


Figure 1: Location of fishing ports and landing places monitored in 2011 on the coast of São Paulo, Brazil included in the register (population of interest) for the analysis.

SAMPLING DESIGN BASED ON CATCH INFORMATION

The IBGE methodology is based on a complex sampling design, composed of stratification of the ports and conglomeration of landings within ports for the calculation of total catch estimates and their associated sampling errors or coefficients of variation. This study assumed a single conglomerate sampling design, i.e., the information of all landings carried out at sampled ports was considered in the analysis.

According to LIMA-GREEN and MOREIRA (2012), the ports should be previously divided into strata regarding their importance. One stratum called census stratum was composed of ports selected arbitrarily by their importance, according to historical total landed catches. For all others, called sampled strata, simple random samples (without replacement) of ports were used.

Experts from the Fisheries Institute of São Paulo made the stratification of ports for each municipality into strata. The rules for stratification were: (1) to separate a maximum of three ports into census stratum; (2) to subdivide the remaining ports into as many strata as deemed necessary to achieve approximate homogeneous landings (i.e. small, medium or large); and (3) each sampled stratum should be composed of at least three ports. The municipalities

of Itanhaém, São Vicente and Bertioga have three or fewer ports and therefore all of them comprised the census stratum. Any port with unmatched specificity in landings, when identified, was transferred to the census stratum because it could cause distortions in later sample expansion.

EXPANSION AND STATISTICAL INFERENCE

All estimates of this study were obtained through software R (R CORE TEAM, 2015) using package Sampling (TILLÉ; MATEI, 2015) for the port sampling and package Survey (LUMLEY, 2014) for the estimate calculations. The sample sizes were defined to be two ports, randomly chosen, in strata with up to six ports and three ports for all others. Therefore, a total of 77 ports composed the sample, 36 in the census strata and 41 in the sampled strata. Main equations used in this analysis are specified in Table 1.

This study performed 100 simulations, each containing the sample selection of ports and total catch estimation for each one of the 14 municipalities of São Paulo. For each of the 100 simulations, the estimated total landed catch by municipality ($\bar{Y}_i^{(k)}$ – Equation 10) and associated standard error ($SE_i^{(k)}$ – Equation 3), the coefficient of variation ($CV_i^{(k)}$ – Equation 4), the square root of the mean squared error ($RMSE_i^{(k)}$ – Equation 17), the percentage bias ($\%B_i^{(k)}$ – Equation 5) with respect to the true total landed catch of 2011 (Y_i – Equation 7), and the design effect ($Deff_i^{(k)}$ – Equation 18) were calculated. Furthermore, the annual economic cost of fisheries monitoring, obtained as a sum of costs for sampled ports, was also obtained in each simulation. The R package Survey (LUMLEY, 2014) estimates standard errors (SE) as the square root of the Horwitz-Thompson estimated variance of the total population, and was implemented with the option 'ultimate cluster' method (Equation 13). The best sample allocation of ports within each municipality was chosen from the 100 simulations under two criteria: Sampling Plan 1 (SP1) - the sample with the lowest RMSE; Sampling Plan 2 (SP2) - the sample with the lowest economic cost. It is important to clarify that whenever referring to accuracy or to an accurate estimate in the remaining text, the compromise between the variance and the squared bias of the estimate will be considered (low RMSE).

The final estimate of the total catch by municipality ($\bar{Y}_i^{(k)}$ – Equation 19) was obtained by a simple average of the 100 simulated estimates and the coefficient of variation (CV_i) was calculated from its standard error

Table 1. Main equations used to obtain estimated total catch and its measures of variability and bias by municipalities and total of São Paulo State. A – Symbology; i – Municipality; p – Port; h – Stratum; H_i – Total number of sampled strata in municipality i ; M_{ih} – Total number of ports in sampled stratum h and municipality i ; m_{ih} – Total number of sampled ports in sampled stratum h and municipality i ; k – Simulation; Y – True total landed catch; $\widehat{Y}$ – Estimated total landed catch.

General equations		Description
Eq. 1	$\widehat{T}_x = \sum_{a=1}^n \frac{1}{\pi_a} X_a$ where $\pi_a = \frac{m}{M}$	Horvitz-Thompson estimator $\widehat{T}_x$ for the population total T_x of the variable X , based on a sample of size n , with π_a probability that a particular unit is sampled.
Eq. 2	$\widehat{V}(\widehat{T}_x) = \sum_{a,b} \left(\frac{X_a X_b}{\pi_{a,b}} - \frac{X_a}{\pi_a} * \frac{X_b}{\pi_b} \right)$	Variance estimate of Horvitz-Thompson estimator $\widehat{T}_x$ where π_a is the probability that unit a is sampled and $\pi_{a,b}$ is the pairwise sampling probability that both units, a and b , are sampled.
Eq. 3	$\widehat{SE}(\widehat{T}_x) = \sqrt{\widehat{var}(\widehat{T}_x)}$	Standard Error
Eq. 4	$\widehat{CV}(\widehat{T}_x) = \frac{\widehat{SE}(\widehat{T}_x)}{\widehat{T}_x} * 100$	Coefficient of Variation
Eq. 5	$\%B = \left(\frac{\widehat{T}_x - T_x}{T_x} \right) * 100$	Bias expressed as percentage of true total landed catch.
Specific Equations		Description
True values of total landed catch		
Eq. 6	$Y_{ih} = \sum_{p=1}^{M_{ih}} Y_{ihp}$	True total landed catch in sampled stratum h and municipality i , where Y_{ihp} is the true total landed catch of each port for the stratum and municipality.
Eq. 7	$Y_i = \sum_{h=1}^{H_i} Y_{ih}$	True total landed catch for municipality i .
Eq. 8	$Y_{SP} = \sum_{p=1}^{14} Y_i$	True total landed catch for São Paulo State as a whole.
Specific Equations		Description
Estimates for each simulation (k)		
Estimated values of total catch		
Eq. 9	$\widehat{Y}_{ih}^{(k)} = \sum_{p=1}^{m_{ih}} \frac{M_{ih}}{m_{ih}} Y_{ihp}^{(k)}$	Estimated total landed catch in sampled stratum h and municipality i , where the sum is over the sampled ports within (i, h) .
Eq. 10	$\widehat{Y}_i^{(k)} = \sum_{h=1}^{H_i} \widehat{Y}_{ih}^{(k)}$	Estimated total landed catch for municipality i .
Eq. 11	$\widehat{Y}_{SP}^{(k)} = \sum_{i=1}^{14} \widehat{Y}_i^{(k)}$	Estimated total landed catch for São Paulo State as a whole.
Estimated Variances		
Eq. 12	$\widehat{\bar{Y}}_{ih}^{(k)} = \frac{\sum_{p=1}^{M_{ih}} \widehat{Y}_{ih}^{(k)}}{M_{ih}}$	Mean of estimated total landed catch in sampled stratum h and municipality i .
Eq. 13	$\widehat{V}(\widehat{Y}_{ih}^{(k)}) = \sum_{h=1}^{H_i} \left[\frac{m_{ih}}{1 - m_{ih}} \sum_{p=1}^{m_{ih}} \left(Y_{ihp}^{(k)} - \widehat{\bar{Y}}_{ih}^{(k)} \right)^2 \right]$	Simplified and conservative estimated variance of the estimates in sampled stratum h and municipality i obtained through the Ultimate cluster method.
Eq. 14	$\widehat{V}(\widehat{Y}_i^{(k)}) = \sum_{h=1}^{H_i} \widehat{V}(\widehat{Y}_{ih}^{(k)})$	Estimated variance of the estimates for municipality i .

Eq. 15	$\widehat{V}(\widehat{Y}_{SP}^{(k)}) = \sum_{i=1}^{14} \widehat{V}(\widehat{Y}_i^{(k)})$	Estimated variance of the estimates for São Paulo State as a whole.
<i>Other definitions</i>		
Eq. 16	$B_i^{(k)} = (\widehat{Y}_i^{(k)} - Y_i)$	Bias of the estimates for municipality i .
Eq. 17	$RMSE_i^{(k)} = \sqrt{\widehat{V}(\widehat{Y}_i^{(k)}) + (B_i^{(k)})^2}$	Squared root of the mean squared error of the estimates for municipality i .
Eq. 18	$Deff_i^{(k)} = \frac{\widehat{V}(\widehat{Y}_i^{(k)})}{\widehat{V}_{srs}(\widehat{Y}_i^{(k)})}$	Design Effect of the estimates for municipality i , where $\widehat{V}_{srs}(\widehat{Y}_i^{(k)})$ is the estimated variance of a simple random sample of landings of the same size.
Estimates with all 100 simulations		
Eq. 19	$\widehat{\bar{Y}} = \frac{\sum_{k=1}^{100} \widehat{Y}_i^{(k)}}{100}$	Mean of estimated total landed catch obtained as a simple average of 100 simulated estimates of the total catch for municipality i .
Eq. 20	$\widehat{V}(\widehat{\bar{Y}}) = \frac{\sum_{k=1}^{100} (\widehat{Y}_i^{(k)} - \widehat{\bar{Y}})^2}{100 - 1}$	Variance of the mean of estimated total landed catch of 100 simulations for municipality i .

(SE_{*i*} - Square root of $\widehat{V}(\widehat{\bar{Y}}_i)$ - Equation 20). The coverage of the confidence interval (1- α =0.95) (CI95) was obtained by counting the samples (simulations) for which the true landed catch (Y_i) was encompassed by the CI95.

OBTAINING FISHERIES MONITORING COSTS

Information gathered by the PMAP was used to obtain: the cost for monitoring each port, the total cost for each sample by adding over the costs of its set of ports and also to select the best sample according to SP2.

In PMAP, distant ports are monitored through regular field trips using institutional or private vehicles. Although the combination of ports monitored per trip can vary for different reasons, to simplify, the cost calculation assumed individual trips to each port. The cost of fuel (in liters/month) per port was obtained based on the distance traveled (round trip), weekly frequency of monitoring and fuel consumption (l/km) in accordance with the type of vehicle (car, motorcycle or boat). Other costs include wages, equipment, supply and maintenance, food and lodging and services like database and computers maintenance, printing and telephony and were all used to obtain the cost of each employee.

Depending on the number and set of ports to be monitored, the number of monitors (supervisors), field agents and typists varies. Aiming at a lower cost, the number of field agents in SP2, compared to SP1, was reduced. The number of typists was based on the total number of hours required to include all landings from

the set of sampled ports, which considered the number of reported landings and the ability to include 20 landings per hour into the database. Typing cost by port considered the inclusion cost by landing into the database (total cost with typists divided by the total number of landings times the number of landing per port). The total cost of each remaining employee was divided by the attended ports under his/her responsibility. Expenses related to the coordination and management of the PMAP, despite being overhead costs, were also considered and equally divided between ports that comprised each sampled set. The calculations were based on the highest wage for each position and did not include the administration fees.

RESULTS

The estimated total catch per municipality was obtained together with the associated coefficient of variation, percentage bias and CI95 coverage (Table 2). For municipalities with few ports (Bertioga, Itanhaém and São Vicente), all ports were allocated in the census stratum and the dispersion measures were therefore equal to zero. Estimates of the total landed catch in all remaining municipalities had low bias, but the maximum CV among samples was obtained in Caraguatatuba (12.0%). The coverage of the CI95 has not encompassed the true value of catches in 95% of samples as it should, except for Ilha Comprida where the coverage was complete (100%).

For each municipality, among the 100 sets of sampled ports, the set resulting in the smallest RMSE (SP1)

was chosen (Table 3). The estimates with the highest variability (CV) were obtained in the municipalities of Caraguatatuba, Iguape, Ilha Comprida, Mongaguá, Praia Grande and São Sebastião, which also presented values of design effect (Deff) greater than one, indicating that the stratification made between ports for these municipalities did not improve in comparison to simple random sampling. Caraguatatuba had estimates with the highest values of CV and bias. The best results were found for Ubatuba.

In Table 4, similar inference results are shown for the selected set of ports for each municipality providing the lowest monitoring cost (SP2). Catch estimates were less accurate than those found for SP1, which is demonstrated by the higher CV value and RMSE for virtually all municipalities (except for Caraguatatuba and Ilha Comprida that had equal values).

The results that will be presented from now on, and in more detail, refer only to the set of ports by municipality resulting in the smallest RMSE (SP1). After obtaining the estimates of total catch per municipality, some other domain estimations (month and fish categories) that were not considered in the sampling design, were made.

Monthly catch estimates for São Paulo State as a whole and associated dispersion measures are displayed in Table 5. Despite the relatively accurate estimates, the worst results were observed in February (largest CV) and November (largest bias).

Fish categories landed on the São Paulo coast over 2011 (identified at minor taxonomic rank possible during data collection) were also considered as domain to compute the catch estimates (Table 6). The 20 most important fish categories (in relation to landed catch) have been listed. All remaining categories have been lumped together in “Others”. Incidental landed catches have also been included in the analysis. Considering CV and bias, *Crassostrea brasiliana* and *Opisthonema oglinum* were the species with the worst and the best estimates of catch, respectively. In relation to the estimated number of trips, *C. brasiliana* had the worst estimates again while *Menticirrhus* spp had the best results. The CV of estimated number of trips by fish category tended to be greater than the CV of estimated catch. With this information, the estimated LPUE (reported Landed catch Per Unit of Effort, $\text{kg} \cdot \text{trip}^{-1}$) calculated for 2011 was compared to the true LPUE and few differences were observed (Table 6). When comparing estimated and actual catches for 2011, it was found that the order of importance of fish categories in landings of São Paulo has remained the same for the first four categories and showed a slight difference in some of the remaining positions (Figure 2). Only *Macrodon atricauda*, *Octopus vulgaris*, *Mugil liza* and *Doryteuthis* spp were ranked worse than actual order by two or more positions while *C. brasiliana* was the only category really badly ranked, with eight positions higher than actual order.

Table 2. Results obtained by simple average between 100 estimates of the total catch (tons) by municipality for São Paulo coast over 2011. Y_i – True total landed catch (tons); $\widehat{Y}_i$ – Estimates of total landed catch (tons); CV_i – Coefficient of variation between averages; $\%B_i$ – Bias expressed as percentage of catch 2011.

Municipality	Y_i	$\widehat{Y}_i$	CV_i	$\%B_i$	Coverage CI95
Bertioga	216.13	216.13	0.00	0.00	
Cananéia	3289.96	3274.14	3.85	-0.48	81
Caraguatatuba	139.58	139.54	12.01	-0.03	68
Iguape	1078.18	1064.39	6.98	-1.28	77
Ilha Comprida	56.97	56.60	4.89	-0.65	100
Ilhabela	702.15	703.85	2.19	0.24	56
Itanhaém	45.47	45.47	0.00	0.00	
Mongaguá	53.94	53.88	4.96	-0.10	67
Peruíbe	147.66	147.44	3.26	-0.15	66
Praia Grande	81.00	80.37	6.04	-0.78	86
Santos/Guarujá	11423.33	11500.55	5.43	0.68	90
São Sebastião	556.93	554.27	9.20	-0.48	71
São Vicente	105.93	105.93	0.00	0.00	
Ubatuba	2064.56	2066.23	0.98	0.08	81
São Paulo State	19961.80	20008.79	3.22	0.24	

Table 3: Results of the Sampling Plan 1 (SP1) – the lowest square root of the mean squared error (RMSE) – by municipality for São Paulo coast over 2011. Y_i – True total landed catch (tons); $\widehat{Y}_i^{(k)}$ – Estimates of total landed catch (tons); $CV_i^{(k)}$ – coefficient of variation within each municipality; $\%B_i^{(k)}$ – Bias expressed as percentage of catch 2011; $Deff_i^{(k)}$ – design effect; Annual cost in December 2015 – rounded and expressed in USD (R\$ 3.70 in Brazilian currency).

Municipality	$Y_i^{(k)}$	$\widehat{Y}_i^{(k)}$	$CV_i^{(k)}$	$\%B_i^{(k)}$	$RMSE_i^{(k)}$	$Deff_i^{(k)}$	Annual Cost
Bertioga	216.13	216.13	0.00	0.00	0.00	0.00	31200.00
Cananéia	3289.96	3287.71	1.82	-0.07	59.88	0.33	120050.00
Caraguatatuba	139.58	156.52	11.14	12.13	24.31	104.88	53675.00
Iguape	1078.18	1112.83	1.57	3.21	38.80	3.71	115820.00
Ilha Comprida	56.97	60.29	3.27	5.83	3.86	24.41	53195.00
Ilhabela	702.15	702.01	1.49	-0.02	10.44	0.31	43975.00
Itanhaém	45.47	45.47	0.00	0.00	0.00	0.00	35340.00
Mongaguá	53.94	55.02	5.00	2.01	2.95	12.27	52215.00
Peruíbe	147.66	150.49	0.91	1.92	3.14	0.26	70830.00
Praia Grande	81.00	80.96	4.21	-0.05	3.41	7.74	47340.00
Santos/Guarujá	11423.33	11186.90	3.60	-2.07	467.13	0.21	113570.00
São Sebastião	556.93	535.32	5.23	-3.88	35.38	2.48	120920.00
São Vicente	105.93	105.93	0.00	0.00	0.00	0.00	37695.00
Ubatuba	2064.56	2057.11	0.10	-0.36	7.70	0.00	130010.00
São Paulo State	19961.80	19752.69	2.07	-1.05	459.50		1025835.00

Table 4. Results of the Sampling Plan 2 (SP2) – the lowest economic cost – by municipality for São Paulo coast over 2011. Y_i – True total landed catch (tons); $\widehat{Y}_i^{(k)}$ – Estimates of total landed catch (tons); $CV_i^{(k)}$ – coefficient of variation within each municipality; $\%B_i^{(k)}$ – Bias expressed as percentage of catch 2011; $Deff_i^{(k)}$ – design effect; Annual cost in December 2015 – rounded and expressed in USD (R\$ 3.70 in Brazilian currency).

Municipality	$Y_i^{(k)}$	$\widehat{Y}_i^{(k)}$	$CV_i^{(k)}$	$\%B_i^{(k)}$	$RMSE_i^{(k)}$	$Deff_i^{(k)}$	Annual Cost
Bertioga	216.13	216.13	0.00	0.00	0.00	0.00	31335.00
Cananéia	3289.96	3274.10	6.45	-0.48	211.64	2.91	116720.00
Caraguatatuba	139.58	116.02	5.13	-16.89	24.31	23.77	49575.00
Iguape	1078.18	1013.83	7.34	-5.97	98.39	48.88	103820.00
Ilha Comprida	56.97	60.29	3.27	5.83	3.86	24.41	46940.00
Ilhabela	702.15	712.45	3.22	1.47	25.16	2.18	56750.00
Itanhaém	45.47	45.47	0.00	0.00	0.00	0.00	35425.00
Mongaguá	53.94	51.00	2.40	-5.44	3.18	2.07	52205.00
Peruíbe	147.66	142.92	4.01	-3.22	7.44	6.96	71275.00
Praia Grande	81.00	71.93	3.19	-11.19	9.35	2.55	47235.00
Santos/Guarujá	11423.33	10953.47	4.06	-4.11	646.91	0.14	97090.00
São Sebastião	556.93	489.36	2.97	-12.13	69.11	0.49	109510.00
São Vicente	105.93	105.93	0.00	0.00	0.00	0.00	38000.00
Ubatuba	2064.56	2043.50	0.51	-1.02	23.51	0.01	121240.00
São Paulo State	19961.80	19296.40	2.58	-3.33	831.55		977120.00

To the extent that the level of detail increased, more biased and less precise the estimates have become. This can be seen in Table 7 where monthly estimated catches by fish categories over 2011 and the associated CV are shown.

With two domains being considered at the same time for catch estimates (month and fish categories), CV of some of these combinations were much greater than when only one of these domains was considered (Tables 5 and 6) as,

Table 5. Results of the Sampling Plan 1 (SP1) – the lowest square root of the mean squared error (RMSE) – by month for São Paulo coast over 2011. Catches are expressed in tons; *CV* - coefficient of variation within each month; % *B* – Bias expressed as percentage of catch 2011.

Month	Catch 2011	Estimated catch	CV	% B
January	1216.95	1170.96	2.99	-3.78
February	1332.73	1355.35	5.69	1.70
March	1208.75	1279.73	2.91	5.87
April	1299.85	1390.45	2.53	6.97
May	1217.87	1265.60	4.25	3.92
June	1547.84	1498.77	1.88	-3.17
July	1300.29	1286.30	3.86	-1.08
August	2233.64	2194.97	2.18	-1.73
September	1830.74	1810.24	1.88	-1.12
October	3414.64	3379.20	1.74	-1.04
November	1700.62	1579.03	5.00	-7.15
December	1657.89	1542.09	3.43	-6.98
Total	19961.80	19752.69	2.07	-1.05

for example, observed for estimated catch of *Cynoscion jamaicensis* in May (Table 7). In Figure 3, monthly estimated catches for four of the 20 main fish categories landed in São Paulo State are presented, each displaying a different situation: estimates with some variability and bias (*Micropogonias furnieri*), estimates with relatively large bias (*M. atricauda*), accurate estimates (*O. oglinum*), and estimates with very large variability and bias (*C. brasiliensis*).

In order to collect data from all 83137 landings carried out in 133 ports on the São Paulo coast over 2011 (census) 32 field agents, five monitors and four typists would be required. The scenarios SP1 (lowest RMSE) and SP2 (lowest fisheries monitoring cost) were compared to the census methodology and, in both, 77 ports were sampled, reducing the number of monitored ports by 42.1% compared to the census methodology. With this reduction, the same five monitors and three typists would be required. Considering only SP1, 27 field agents would be required since a 35.4% reduction in number of monitored landings were observed, decreasing the fisheries monitoring costs by 11.2%. To monitor the ports of SP2, 24 field agents would be required with a 41.5% reduction in number of monitored landings and a 15.4% decrease in fisheries monitoring costs.

DISCUSSION

The fishing activity, especially the small-scale fishery, represents a seasonal, diversified and dynamic activity (CADIMA et al., 2005; ISAAC et al., 2008; MENDONÇA;

MIRANDA, 2008). These characteristics along with the need for accurate information lead to the adoption of complex survey plans for fisheries monitoring, such as the sampling methodology for fisheries monitoring proposed by IBGE (LIMA-GREEN; MOREIRA, 2012).

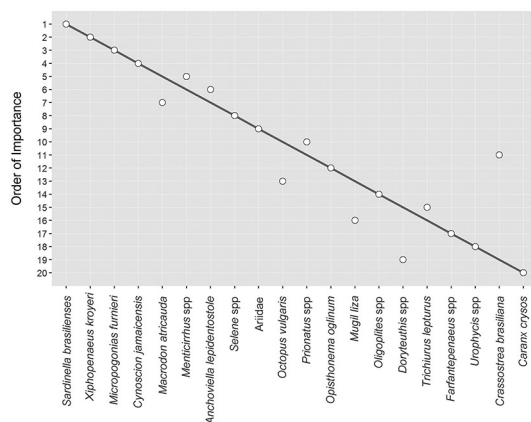
The accuracy of the total landed catch estimated through this methodology could only be judged because the true (population) values of landed catch in all municipalities of the São Paulo State are known. Therefore, a sampling distribution could be obtained by applying the same sampling procedure repeatedly (COCHRAN, 1977). The results demonstrated that the mean landed catch is a good estimator of the total landed catch for most municipalities since it was unbiased and had high precision. The low coverage of CI95 was attributed to the non-conformity of the Gaussian distribution used to build these intervals. The small number of possible sets of sampled ports (conglomerates) compromises the use of IC95 to evaluate the precision and the reliability of the estimates (BUSSAB; MORETTIN, 2012).

Comparing the results of each of the 100 simulated samples, it was clear that the set of sampled ports that provides the lowest RMSE is different from the set that provides the lowest monitoring cost. Hence, neither of the criteria for allocation of ports (SP1 and SP2) is optimum in the sense of COCHRAN (1977) sees it, in which the optimum allocation is achieved when, for a given sample size, the sample provides the most precision of the estimates at the lowest cost. Larger samples would increase costs while further reduction would leave some strata out

Table 6. Results of the Sampling Plan 1 (SP1) – the lowest square root of the mean squared error (RMSE) – by fish category for São Paulo coast over 2011. Catches are expressed in tons; *CV* - coefficient of variation within each fish category; % *B* – Bias expressed as percentage of catch 2011. Landed catch per unit of effort (LPUE) expressed in ton*trip⁻¹.

Minor taxonomic rank	Catch 2011	Estimated catch	CV	% B	Number of trips 2011	Estimated number of trips	CV	% B	LPUE 2011	Estimated LPUE
<i>Sardinella brasilienses</i>	4105.09	4150.42	3.38	1.10	205	211	7.52	2.93	20.025	19.670
<i>Xiphopenaeus kroyeri</i>	2759.11	2399.59	5.31	-13.03	17618	14608	18.25	-17.08	0.157	0.164
<i>Micropogonias furnieri</i>	2102.17	2183.93	6.12	3.89	11311	10184	5.17	-9.96	0.186	0.214
<i>Cynoscion jamaicensis</i>	1036.34	1060.41	7.62	2.32	1003	1015	4.26	1.20	1.033	1.045
<i>Macrodon atricauda</i>	705.18	634.20	2.15	-10.07	8277	6136	7.74	-25.87	0.085	0.103
<i>Menticirrhus spp</i>	646.99	674.74	6.98	4.29	3357	3342	3.23	-0.45	0.193	0.202
<i>Anchoviella lepidontostole</i>	612.01	669.00	6.29	9.31	11630	11709	13.52	0.68	0.053	0.057
<i>Selene spp</i>	480.40	482.36	2.44	0.41	849	759	19.24	-10.60	0.566	0.636
Ariidae	454.35	475.63	16.44	4.68	10835	10083	14.49	-6.94	0.042	0.047
<i>Octopus vulgaris</i>	355.26	316.69	3.66	-10.85	722	668	4.22	-7.48	0.492	0.474
<i>Prionotus spp</i>	352.43	375.29	10.78	6.49	600	644	8.08	7.33	0.587	0.583
<i>Opisthonema oglinum</i>	343.43	342.78	0.40	-0.19	888	753	9.31	-15.20	0.387	0.455
<i>Mugil liza</i>	319.68	237.95	14.89	-25.57	7053	6432	12.36	-8.80	0.045	0.037
<i>Oligoplites spp</i>	290.77	285.05	3.16	-1.97	3110	2802	5.70	-9.90	0.093	0.102
<i>Doryteuthis spp</i>	256.73	208.11	4.34	-18.94	2546	2422	3.46	-4.87	0.101	0.086
<i>Trichiurus lepturus</i>	255.64	271.10	9.03	6.05	3865	4144	6.01	7.22	0.066	0.065
<i>Farfantepenaeus spp</i>	249.70	227.30	14.89	-8.97	872	848	4.59	-2.75	0.286	0.268
<i>Urophycis spp</i>	193.85	217.03	13.63	11.96	259	267	6.62	3.09	0.748	0.813
<i>Crassostrea brasiliana</i>	173.82	355.81	34.24	104.70	5354	10591	43.82	97.81	0.032	0.034
<i>Caranx crysos</i>	170.84	158.05	5.39	-7.48	744	787	24.76	5.78	0.230	0.201
Landed incidental catch *	882.45	834.60	6.88	-5.42	10661	8733	2.10	-18.08	0.083	0.096
Others	3215.56	3192.65	4.63	-0.71	39721	37467	9.06	-5.67	0.081	0.085
Total	19961.80	19752.69	2.07	-1.05	83137	80202	7.21	-3.53	0.240	0.246

* Fish of small size and/or low or no commercial value, however landed and marketed, composing the category called “Mistura” in fisheries statistics of the São Paulo State.

**Figure 2.** Order of importance in terms of catches of the 20 main species landed over 2011 on the São Paulo coast. Abscissa with true order for 2011 and ordinate with estimated order; matches between is a point on grey line.

of the sample. Thus, since there is no marked difference in costs between SP1 and SP2, SP1 was chosen as the most appropriate allocation.

The common features of municipalities with the worst result (total landed catch estimates with low precision, large bias or high values of *Deff*) include few clustered ports within sampled strata with heterogeneous ports, as observed in Caraguatatuba and Ilha Comprida. To improve the precision of the estimated catch, the larger ports and possible outliers may be relocated from the sampled strata to census strata and/or larger samples must be taken, but both measures imply an increase in the monitoring costs. Applying this solution for Caraguatatuba and Ilha Comprida means having all their ports monitored, resulting in a lower reduction in the total fisheries monitoring cost

Table 7. Results of the Sampling Plan 1 (SP1) – the lowest square root of the mean squared error (RMSE) – by month and fish category for São Paulo coast over 2011. Catches are expressed in tons; CV - coefficient of variation within each month and fish category.

Minor taxonomic rank	January		February		March		April		May		June	
	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV
<i>Sardinella brasilienses</i>	0.00	109.54	36.91	0.61	387.29	5.79	349.84	17.50	249.82	0.00	108.28	0.00
<i>Xiphopenaeus kroyeri</i>	265.09	6.96	202.53	5.40	6.50	0.00	0.00	0.00	0.27	22.92	402.81	7.35
<i>Micropogonias furnieri</i>	55.44	13.80	76.62	10.22	125.51	9.08	223.67	3.15	238.83	13.35	229.86	0.53
<i>Cynoscion jamaicensis</i>	54.61	14.56	82.58	14.88	83.55	5.70	115.58	7.49	52.98	33.52	89.59	3.87
<i>Menticirrhus spp</i>	62.30	6.71	55.96	14.49	43.59	9.14	60.07	8.95	32.55	22.20	33.97	0.54
<i>Anchoviella lepidontostole</i>	78.68	9.26	150.63	15.43	10.05	21.42	1.98	37.04	0.77	55.12	0.83	53.43
<i>Macrodon atricauda</i>	52.08	10.41	60.54	1.06	30.68	0.65	44.23	0.31	56.30	0.29	51.61	4.82
<i>Selene spp</i>	6.65	7.14	3.95	34.19	109.37	0.18	54.46	30.08	15.03	8.59	27.18	0.07
<i>Ariidae</i>	11.80	2.90	16.97	12.99	10.33	17.08	23.27	13.80	21.44	19.12	23.96	25.05
<i>Prionotus spp</i>	13.15	4.81	13.45	25.76	10.97	18.56	16.69	25.94	9.07	23.36	36.07	6.55
<i>Crassostrea brasiliiana</i>	0.81	6.20	6.95	36.34	35.55	27.12	37.09	35.17	40.09	34.16	33.91	47.36
<i>Opisthonema oglinum</i>	0.03	4.53	0.02	0.00	0.03	10.14	61.42	0.05	195.15	0.01	26.70	0.00
<i>Octopus vulgaris</i>	10.58	2.58	24.79	0.52	32.49	0.19	23.86	0.01	20.91	0.00	27.07	8.96
<i>Oligoplites spp</i>	54.71	2.57	28.11	1.09	11.54	8.38	19.70	3.18	14.48	11.06	13.03	11.52
<i>Trichiurus lepturus</i>	34.91	6.17	29.89	24.79	17.51	14.33	34.22	5.57	23.17	17.60	18.33	5.04
<i>Mugil liza</i>	12.27	20.62	5.43	19.59	6.06	16.77	7.98	17.94	12.51	15.32	43.95	19.35
<i>Farfantepenaeus spp</i>	4.05	2.35	8.80	12.21	2.36	0.00	0.00	0.00	0.00	0.00	25.35	8.47
<i>Urophycis spp</i>	18.58	16.48	25.84	11.85	29.03	30.73	1.62	0.00	0.83	58.74	4.10	17.94
<i>Doryteuthis spp</i>	44.59	15.94	24.67	2.95	30.12	1.65	44.26	3.93	18.99	3.81	6.90	10.20
<i>Caranx crysos</i>	11.08	1.85	111.94	6.62	16.75	18.68	3.89	19.91	2.66	3.77	2.63	23.33
Landed incidental catch *	55.64	5.46	68.11	8.49	50.79	7.09	56.88	6.82	51.09	3.09	96.29	3.27
Others	323.91	5.88	320.68	8.40	229.68	8.24	0.00	5.42	208.66	6.50	196.36	3.88
Total	1170.96	2.99	1355.35	5.69	1279.73	2.91	1390.45	2.53	1265.60	4.25	1498.77	1.88
Minor taxonomic rank	July		August		September		October		November		December	
	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV	Est. Catch	CV
<i>Sardinella brasilienses</i>	0.00	0.00	895.04	5.47	469.43	0.00	1523.79	3.70	129.09	0.03	0.93	92.58
<i>Xiphopenaeus kroyeri</i>	275.47	7.21	186.48	5.30	266.11	6.78	285.82	6.50	273.84	8.09	234.66	9.63
<i>Micropogonias furnieri</i>	179.24	2.05	268.33	15.97	262.83	1.78	267.38	15.35	130.22	5.89	125.99	1.87
<i>Cynoscion jamaicensis</i>	51.15	1.86	94.59	14.65	91.42	13.45	141.65	3.67	82.67	11.79	120.02	3.25
<i>Menticirrhus spp</i>	39.50	8.84	39.74	6.54	54.11	3.47	106.30	6.89	79.79	4.35	66.88	5.82
<i>Anchoviella lepidontostole</i>	0.63	48.14	0.71	66.99	9.00	30.51	137.08	9.58	169.89	9.40	108.75	15.63
<i>Macrodon atricauda</i>	56.44	3.50	54.54	2.79	31.00	12.17	81.91	2.78	50.06	3.67	64.83	4.59
<i>Selene spp</i>	29.53	0.65	3.39	10.40	81.28	0.11	144.51	1.20	4.56	23.91	2.47	1.31
<i>Ariidae</i>	17.84	14.87	28.18	34.41	21.12	20.75	58.93	15.35	118.10	25.57	123.70	29.89
<i>Prionotus spp</i>	58.58	17.88	51.71	2.76	41.48	4.38	54.38	13.54	39.51	22.01	30.23	9.26
<i>Crassostrea brasiliiana</i>	39.72	37.43	44.16	36.62	37.05	35.77	41.40	32.24	29.72	39.59	9.37	76.70
<i>Opisthonema oglinum</i>	29.95	0.18	21.44	3.22	2.01	15.58	4.30	22.39	0.92	0.35	0.82	1.48
<i>Octopus vulgaris</i>	18.21	4.46	27.13	0.16	48.22	7.64	30.85	10.40	21.52	10.24	31.06	6.73
<i>Oligoplites spp</i>	13.96	14.47	10.17	12.75	5.15	19.62	5.23	18.89	19.84	9.85	89.15	1.27
<i>Trichiurus lepturus</i>	16.78	7.78	16.71	13.56	11.94	10.25	28.17	25.25	17.60	13.49	21.87	3.28

<i>Mugil liza</i>	58.11	10.33	22.47	15.66	44.02	31.75	13.37	23.07	7.67	24.29	4.11	9.97
<i>Farfantepenaeus spp</i>	44.73	28.22	24.23	11.62	22.68	14.17	43.36	18.14	28.56	21.45	23.20	12.72
<i>Urophycis spp</i>	14.74	27.13	35.30	1.72	14.24	0.00	9.78	43.97	24.68	18.83	38.29	26.22
<i>Doryteuthis spp</i>	2.75	43.82	0.80	0.00	0.95	3.63	2.82	73.75	10.64	14.06	20.61	14.57
<i>Caranx crysos</i>	0.59	12.06	0.61	28.57	0.70	26.02	0.50	39.86	2.95	31.93	3.76	13.57
Landed incidental catch *	73.71	18.75	73.42	8.50	65.90	12.80	105.94	12.40	39.16	4.20	97.68	8.50
Others	264.66	9.40	295.84	3.06	229.62	4.58	291.73	6.47	298.05	9.69	323.69	7.01
Total	1286.30	3.86	2194.97	2.18	1810.24	1.88	3379.20	1.74	1579.03	5.00	1542.09	3.43

* Fish of small size and/or low or no commercial value, however landed and marketed, composing the category called “Mistura” in fisheries statistics of the São Paulo State.

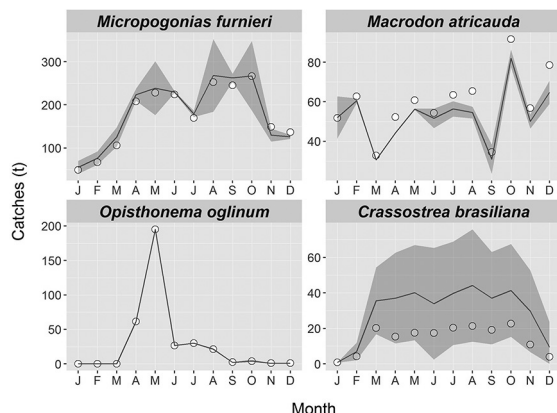


Figure 3. True total landed catches of 2011 (open circle), estimated total landed catches (line), both in tons, and confidence interval of 95% (shaded area) by month to four of the 20 main species landed over 2011 on the São Paulo coast.

of the São Paulo coast with sampling design compared to the census (10.9% in SP1 and 13.8% in SP2).

The more detail is needed and the more variables are involved, the less accurate the estimates become. According to BOLFARINE and BUSSAB (2005), accurate estimates result from considering them explicitly when developing the sampling design, as it is the case herein for the estimated total catch by municipality. Although the variable month has not been considered in the sampling design, estimated monthly total catch for São Paulo State was also a good result since there was sufficient information by month in the sample. In contrast, after breaking down these data by fish category, a much lower accuracy was obtained, particularly for some fish categories such as *C. brasiliiana*, *M. Lisa*, *Farfantepenaeus* spp and *Doryteuthis* spp, all of them economically important for the São Paulo State. Detailing data even further, by fishing gear or fleet type, would only make matters worse.

However, the landed catch per landing (LPUE) turned out to be a robust estimate, for most but a very few fish categories. This was facilitated because all landings of the sampled ports were considered. Furthermore, when

both the catch and the number of trips are simultaneously under- or overestimated, this bias tends to be canceled out. However, since this estimated LPUE was based on a sampling plan designed specifically for the São Paulo State, its performance with the consolidated fishery data from other States with different sample designs must be further investigated. The assessment and interpretation of a temporal series of the estimated LPUE may also be a problem, mainly for species with little information in the sample or with less precise estimates, when sample errors are greater than the variations of the LPUE itself.

The results obtained for four fish categories have been selected to illustrate distinct situations that can also occur in other survey samplings for fisheries monitoring. First of them is the *Opisthonema oglinum*, which had very precise and unbiased estimates since the vast majority of the landings and of the landed catch occurred in ports that had been allocated into the census strata. The second situation is described by the estimated monthly catch of *M. furnieri*, which was unbiased but less precise. Many landings covering a wide range of landed catches of this species occurred in most of the sampled ports, which truly represents the situation in the ports of São Paulo. The third situation is described by *M. atricauda*, for which the sampled ports recorded fewer landings and lower catches in comparison to what really happens in all ports. Thus, moderately imprecise and underestimated landed catches were obtained. Finally, *C. brasiliiana* represents the worst that can happen during a sampling process and shows the importance of having good knowledge about the ports before the stratification is defined. An enormous variation in landed catch of this species was recorded based on sampled ports while a very specific port, which almost exclusively has landings of this resource, was also part of the sample. This port had been wrongly allocated into a sampled stratum rather than into the census stratum as it should, causing distortion and overestimation of the landed catch after being expanded over the sampled stratum.

In general, information gathered from well-designed survey samples may have some advantages compared to a complete data collection (census). According to COCHRAN (1977), accurate and reliable estimates can be produced at a much lower cost and data can be obtained and consolidated more quickly applying sampling methods. In addition, survey samplings may have more scope and suppleness regarding the type and amount of information that can be obtained, since only a part of the population is being considered (FAO, 1999). However, these announced advantages were not clearly observed by the sampling design that was applied to monitor the fisheries on the São Paulo coast.

Some issues that may cause concern will be mentioned next. The high-diversified fishing activity affected the accuracy in the estimated catch by fish category. The cost reduction obtained with the sampling was minor and may not compensate the loss of quality of the fishing information compared to the census data. Fishing vessels and fishermen who are not included in the sample cannot be supplied with a proof of activity and of fisheries production, documents required to obtain benefits such as bank loans and fishing licenses. The true fishing area covered by the fleet in operation might be underestimated, since vessels distribution is underrepresented. Finally, the lack of fishing effort measurements that are more appropriate for each fishing fleet makes the fish stock assessment difficult.

The choice of a fishing data collection methodology depends very much on the goals of fisheries monitoring (FAO, 1999). In this study, it was clear that a survey sampling for fisheries monitoring is very useful when financial resources are limited and there is the interest only in a broad picture, without details about the catches. It is understood that it is possible to have more detailed and reliable fishing data increasing the complexity of the sampling design and the costs of the fishing monitoring as well. Or even, to begin with a simple sampling design and, as far as it is feasible, gradually expand to a census methodology (FAO, 1999). In this case, low cost strategies, such as self-registration and mobile field agents, may be adopted. Both strategies have been adopted since the beginning of the Fishing Activity Monitoring Program (PMAP) of the São Paulo coast. However, regardless of the methodology and whatever the cost of a fisheries monitoring program might be, one thing is for sure, it will always be lower than the economic, social and

environmental costs of not having quality data to perform evidence based fisheries management.

ACKNOWLEDGEMENTS

This work is part of the doctoral degree in Biological Oceanography of the first author working under the supervision of the second. We express our thanks to the fisheries scientists and technical team of the Fishing Activity Monitoring Program (PMAP) and to the Fisheries Institute of the Department of Agriculture and Food Supply of São Paulo State. We are also thankful to MSc Aristides Pereira Lima Green (IBGE) for all his contribution.

REFERENCES

- ARAGÃO, J. A. N.; MARTINS, S. Censo Estrutural da Pesca, Coleta de Dados e Estimação de Desembarques de Pescado. Brasília: IBAMA, 2006. 180 p.
- ÁVILA-DA-SILVA, A. O.; CARNEIRO, M. H.; FAGUNDES, L. Sistema gerenciador de banco de dados de controle estatístico de produção pesqueira marinha – ProPesq. In: Anais do XI Congresso Brasileiro de Engenharia de Pesca e I Congresso Latinoamericano de Engenharia de Pesca. Recife, 1999. p. 824-832.
- ÁVILA-DA-SILVA, A. O.; CARNEIRO, M. H.; MENDONÇA, J. T.; BASTOS, G. C. C.; MIRANDA, L. V.; RIBEIRO, W. R.; SANTOS, S. Produção Pesqueira Marinha e Estuarina do Estado de São Paulo - Dezembro de 2014. Inf. Pesq. São Paulo, v. 54, p. 1-4, 2015.
- BOLFARINE, H.; BUSSAB, W.O. Elementos de Amostragem. 1.ed. São Paulo: Edgar Blücher, 2005. 274 p.
- BUSSAB, W. O.; MORETTIN, P. A. Estatística Básica. 7.ed. São Paulo: Saraiva, 2012. 540 p.
- CADDY, J. F.; BAZIGOS, J. P. Practical guidelines for statistical monitoring of fisheries in manpower limited situations. FAO Fisheries Technical Paper. nº. 257. Rome: FAO, 1985. 86 p.
- CADIMA, E. L. Fish stock assessment manual. FAO Fisheries Technical Paper. nº. 393. Rome: FAO, 2003. 161 p.
- CADIMA, E. L.; CAMELO, A. M.; AFONSO-DIAS, M.; CONTE DE BARROS, P.; TANDSTAD, M. O.; DE LEIVA-MORENO, J. I. Sampling methods applied to fisheries science: a manual. FAO Fisheries Technical Paper. nº 434. Rome: FAO, 2005. 88 p.
- COCHRAN, W. G. Sampling Techniques. New York: John Wiley & Sons, 1977. 428 p.
- DIAS-NETO, J. Pesca no Brasil e seus aspectos institucionais - um registro para o futuro. Rev. CEPSUL - Biodivers. Conserv. Mar., v. 1, n. 1, p. 66-80, 2010.
- DIAS-NETO, J. Números e Baionetas – A Nova Estatística da Produção Pesqueira do Brasil. Erro Estatístico ou Equívoco Político? Pesca & Mar - Informativo SAPERJ (março/abril). Rio de Janeiro/RJ. v. 132, p. 31-34, 2011.
- FAO. Guidelines for the routine collection of capture fishery data. FAO Fisheries Technical Paper. nº 382. Rome: FAO, 1999. 113 p.

- HILBORN, R.; WALTERS, C. J. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. New York: Chapman and Hall, 1992. 570 p.
- HORVITZ, D. G.; THOMPSON, D. J. A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.*, v. 47, n. 260, p. 663-685, 1952.
- ISSAC, V. J.; RUFFINO, M. L.; MELLO, P. Considerações sobre o Método de Amostragem para a Coleta de Dados sobre Captura e Esforço Pesqueiro no Médio Amazonas. In: IBAMA. (Org.). Recursos Pesqueiros do Médio Amazonas: Biologia e Estatística Pesqueira. Brasília: Edições IBAMA, 2000. p. 175-199.
- ISSAC, V. J.; ESPÍRITO SANTO, R. V.; NUNES, J. L. G. A estatística pesqueira no litoral do Pará: resultados divergentes. *Pan-Am. J. Aquat. Sci.*, v. 3, n. 3, p. 205-213, 2008.
- LIMA-GREEN, A. P.; MOREIRA, G. G. Metodologia Estatística da Pesca: pesca embarcada. Textos para Discussão. Diretoria de Pesquisas. Rio de Janeiro: IBGE, 2012. p. 1-52.
- LUMLEY, T. Complex surveys: a guide to analysis using R. Hoboken: John Wiley & Sons, 2010. 276 p.
- LUMLEY, T. Survey: analysis of complex survey samples. R package version 3.30. 2014.
- MENDONÇA, J. T.; MIRANDA, L. V. Estatística pesqueira do litoral sul do estado de São Paulo: subsídios para gestão compartilhada. *Pan-Am. J. Aquat. Sci.*, v. 3, n. 3, p. 152-173, 2008.
- R CORE TEAM R. A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2015. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 dez. 2015.
- TILLÉ, Y.; MATEI, A. Sampling: Survey Sampling. R package version 2.7. 2015. Disponível em: <<http://CRAN.R-project.org/package=sampling>>. Acesso em: 10 dez. 2015.

**ANEXO II: Fisheries Monitoring: Is Bayesian Inference for Finite
Population Sampling an Alternative?**

Laura Villwock de Miranda & Paul Gerhard Kinas

Artigo submetido em julho de 2017

SCIENTIA MARINA

Fisheries monitoring: is Bayesian inference for finite population sampling an alternative?

Laura Villwock de Miranda ^{1,2}, Paul Gerhard Kinas ²

1 Instituto de Pesca de São Paulo, Rua Joaquim Lauro de Monte Claro Neto, 2275, Itaguá, 11680-000, Ubatuba, SP, Brazil (LVM) (Corresponding author) E-mail: lauravillwockmiranda@gmail.com. ORCID-iD: <http://orcid.org/0000-0002-1750-8973>

2 Universidade Federal do Rio Grande. Av. Itália, km 8, Carreiros, 96203-900, Rio Grande, RS, Brazil. (PGK) E-mail: pgkinas@gmail.com. ORCID-iD: <http://orcid.org/0000-0002-1294-9518>

Summary: Inferential Bayesian models for fisheries monitoring of the total landed catches (Y) and the number of landings (D) for 11 municipalities in the state of São Paulo and for 20 fish categories were proposed. Inferences were carried out using a probabilistic sample of fishing ports, previously established to minimize mean squared error. The model that sequentially described D and Y per municipality resulted in the best Y estimates for most of the municipalities in the state of São Paulo. The model describing D and Y per fish category was better adjusted to most of the non-seasonal fish categories. For those with marked seasonality, the best results were obtained by including monthly information in the models. The stratification of fishing ports that was carried over from the design-based approach was a possible reason for the lack of adjustment in some municipalities or fish categories. A new proposal of sampling design, respecting the linear relationship between $\ln(Y)$ and $\ln(D)$ assumed by the model for all fishing ports was presented and discussed. Although preliminary, the new proposal obtained good results for the five major fishery municipalities, with an expressive increase in sample efficiency.

Running Title: Bayesian inference for fisheries monitoring by sampling

Keywords: fishing activity, sampling, model-based, inference, Bayesian models

INTRODUCTION

Catch and fishing effort are the most basic information that can be obtained about any fishing activity in order to assess the fishing stocks and to formulate effective fisheries policies and management plans (Hilborn and Walters 1992, Maunder and Punt 2004). However, the low quality, even inexistence, of this information, as well as the lack of investment in fishing monitoring are quite common in many regions around the globe, mainly for underdeveloped and developing countries (Pauly 2013). Attempts to compile available information for catches reconstruction (Freire et al. 2015, Pauly & Zeller 2016) and inference methods to deal with missing data (Cadima et al. 2005, Gelman et al. 2014) have become increasingly important.

Brazil is among the countries with problems in maintaining an uninterrupted and reliable fishery data collection and there are several reasons for this situation. The national fishing statistics have been heavily influenced by institutional and political arrangements over time (Dias-Neto 2010, 2011, Lima-Green and Moreira 2012). Several government institutions were granted the formal assignment to survey and produce statistics of the Brazilian fishing industry. Consequently, periods of interruption in data collection in different regions and oscillations in collection and processing efficiency were quite common over time. Moreover, the use of administrative records such as logbooks and fishing licenses is hampered by the lack of both maintenance and integration of these collection systems. Actually, the national fishing data collection and processing system collapsed due to lack of funding. Some states, including São Paulo, Paraná and Santa Catarina, currently monitor fishing activity using specific programs designed to ensure compliance with environmental licenses of the Petrobras oil and gas activities in the Santos basin. Since 2012, the state of Rio Grande do Sul has also maintained data collection, funded by the federal

government. The project, however, has an uncertain future, as it currently operates only partially (only industrial and semi-industrial fleets) and there are no prospects for new financial contributions that may enable its continuity in 2018.

Fishing information can be obtained using different methodologies which are structured in two basic ways, namely: by sampling survey (Caddy and Basigos 1985, Aragão and Martins 2006, Lima-Green and Moreira 2012) or by census (FAO 1999, Mendonça and Miranda 2008, Ávila-da-Silva et al. 2017). Census is usually indicated for small populations, when the cost of collecting is low or when the consequences of making a wrong decision are severe and sampling errors are large (Bolfarine and Bussab 2005). In contrast, there are several reasons for considering just part of a population (sampling), such as very large populations and financial or time constraints. In general, sampling surveys can, advantageously, be more flexible and comprehensive in terms of the type and amount of information required (FAO 1999, Bolfarine and Bussab 2005). Dealing with less information can also reduce consolidation time and often reduce the cost of collection and processing (Cochran 1977).

Inferences for data obtained from sampling procedures can be approached in two ways: design-based or model-based. In the design-based approach, any variability is prompted by the sampling plan, without considering the behavior of variables (Little 2004, Bussab and Morettin 2012). In general, it produces accurate inferences without assuming a probabilistic data model; however, it also requires large samples. In model-based inference, any variability in the data is explained by the underlying probabilistic model. The model parameters, estimated from the sample, are then used to infer the non-observed data in the population and, after filling the gaps of these missing data, to obtain the totals of a finite population (Little 2004). However, its efficiency depends on a well fitted-model to produce good inferences.

The following example can illustrate one important difference between these two approaches. Suppose that the total landed catch in tons (Y) was collected by a random sample from a population of fishing landings, for which it is necessary to know the probability of obtaining a total landed catch of, say, over 5 t, $P(Y > 5)$. In the design-based inference, this probability can only be estimated when catches greater than 5 t are sampled; otherwise, the inference will be zero. In the model-based inference, the catch can be described by a log-Normal distribution and its parameters estimated from the sample. The probability of interest is obtained by calculating the upper tail of the distribution, regardless of whether any catches in the sample exceed 5 t.

Model-based inference can follow two paths: superpopulation models (Thompson 1988 and Dorfman et al. 2000 apud Little 2004) and Bayesian models (Little 2004, McCarthy 2007, Kinas and Andrade 2010, Gelman et al. 2014, Kruschke 2015). The assumption of superpopulation models is that the values of the population (composed of finite sampling units, some contained in the sample and others not) are a random sample of a "superpopulation" described by a probabilistic model. Its parameters are fixed and estimated from the likelihood function of a sample that is necessarily large. In Bayesian inference, the observed data are also summarized in the likelihood function, complemented with probability distributions for the model parameters. Any previous existing information about the model parameters can be incorporated as informative prior distributions. Thus, the posterior distribution of the parameters, through the Bayes' theorem, gathers current and previous data and contains all necessary information to perform the inference (Kinas and Andrade 2010). In many situations, the use of informative priors is an advantage, allowing the Bayesian approach to deal with small samples appropriately. With large samples and vague prior

distributions, Bayesian and superpopulation models obtain very similar results (Little 2004).

The expression "missing data" may refer to non-intentionally missing data (e.g. non-response, censored or truncated data) and intentionally missing data, which is a consequence of the sampling processes where only part of the data is observed (Gelman et al. 2014). Missing data models are a way of completing the non-observed sampling units of the finite population and may have a frequentist or Bayesian approach. However, the frequentist inference in finite populations is not the focus of this work. Therefore, the theoretical formalization presented below refers to Bayesian inference in finite populations and was adapted from Gelman et al. (2014) and Zangeneh & Little (2015).

When data collection assumes randomization, the sampling process (I) and its parameter φ can be ignorable (Rubin 1976). To facilitate notation, the ignorability of I will be presumed valid and adopted from now on. The intentionally missing data can be dealt conveniently by dividing the analysis into two parts:

- (1) Superpopulation inference $p(\theta|x, y_{obs})$ (or inferential model) that estimates the subsequent distributions of the parameters (θ) conditioned to the observed data (y_{obs}) and the possible covariates (x);
- (2) Finite population inference $p(y_{mis}|x, y_{obs})$ (or predictive model) that estimates posterior predictive distribution for the missing data (y_{mis}) conditioned to the observed data and possible covariates.

In the Bayesian approach, simulations of the posterior predictive distribution are called *Multiple Imputation* and consist of two steps: first, to obtain the set θ of the model parameters from its posterior distribution $p(\theta|x, y_{obs})$; and second, to simulate y_{mis} from the model to the data $p(y_{mis}|x, \theta)$. In short:

$$p(y_{mis}|x, y_{obs}) = \int p(\theta|x, y_{obs}) * p(y_{mis}|x, \theta) d\theta$$

Bayesian (model-based) inference for finite populations overcomes designed-based inference when the model is correctly specified since it has a formal structure to combine information from various sources. The use of information from previous years (in the form of prior distributions), when available and relevant, and the absence of asymptotic arguments, both allow the obtainment of good inferences even with small samples. Furthermore, inferential Bayesian models can easily deal with complex sample designs and simplifies the inclusion of covariates (Little 2004, Kéry 2010, Hall 2011).

All these characteristics lead us to believe that the Bayesian approach is an attractive way of dealing with the inference of fishing data that are highly variable and often scarce. Fishing ports with varied volume of landed catches and number of landings, a great diversity of target species, harvest periods and fishing gear, among other characteristics, are common in the Brazilian fishing activity (Isaac et al. 2000, 2008, Mendonça and Miranda 2008). Such variation and diversity can hardly be captured only by sampling design. Furthermore, the number of fishing vessels and fishermen, the type and dynamics of the fishing fleet, as well as the seasons and harvest period of the fishing resource can be potentially useful covariates that are easily incorporated in a Bayesian inference model for fishing activity.

Since 1969, the Fisheries Institute of the Department of Agriculture and Food Supply of São Paulo State has been responsible for collecting, storing, processing and disclosure of the census information on marine fisheries production landed along the state of São Paulo (Mendonça and Miranda 2008, Ávila-da-Silva et al. 2017). Information on all fishing landings in São Paulo, uninterrupted since at least 1998, is stored in the ProPesq data management system® (Ávila-da-Silva et al. 1999) and currently available in the web platform called ProPesqWEB

(<http://www.propesq.pesca.sp.gov.br>). This database makes the state of São Paulo an exception in fisheries monitoring in Brazil.

In this paper, probabilistic samples based on the sampling design described by Lima-Green and Moreira (2012) were simulated from the fishing census information of the state of São Paulo collected during 2011. Observed fishing port data in one of these samples were used to formulate an inferential Bayesian model for fisheries monitoring, which is an unprecedented approach in fisheries statistics. Once well structured, the model-based approach can be widely adopted to obtain the total landed catch and the number of landings, all of which is essential information for fisheries management (Cadima 2003, Maunder and Punt 2004, Pauly 2013).

Next, two inferential Bayesian models and their versions, structured to handle more specific situations encountered by the fishing activity, will be presented. The challenges faced during the process of building these models, together with the proposed solutions will be the focus of the discussion.

MATERIAL AND METHODS

Data source, sampling design, and sample

In Miranda et al. (2016), census fishing information for the state of São Paulo collected during 2011 were used to simulate probabilistic samples according to the sampling design described by Lima-Green and Moreira (2012) and the results of these simulations were compared with the true catch in São Paulo. All information used was obtained from the Fishing Activity Monitoring Program (“PMAP”), coordinated by researchers from Fisheries Institute of the Department of Agriculture and Food Supply of the state of São Paulo.

For comparison purposes, the sample in Miranda et al. (2016) that resulted in catch estimates with the smallest mean squared error was replicated to apply the inferential Bayesian models developed in this study. This sample consisted of around 54000 fishing landings and 77 fishing ports, of which 36 are located in census strata and 41 are divided among the 19 sampled strata.

The criteria to define the population of interest (133 fishing ports, 14 municipalities, 83137 fishing landings), the adopted sampling design (stratification) and the sampling methodology (Lima-Green and Moreira 2012) as well as the location of all municipalities and fishing ports along the entire coast of São Paulo are described in detail in Miranda et al. (2016).

Bayesian Inference for Finite Populations

Using a Bayesian approach, five inferential models were adjusted to the simulated sample of the fishing data. **Model 1** was designed primarily to estimate the number of landings and total landed catch per municipality in the state of São Paulo. The number of landings and catches per fish category were set by the proportion of each fish category in the total number of landings of the sampled strata. **Model 2** was initially designed to estimate the total number of landings and catches per fish category for the state of São Paulo, and later, through summation, obtain the estimated totals by municipalities of São Paulo.

Initially, two versions of model 2 were elaborated: **Model 2A**, based on a presence-absence model of the fish categories in the fishing landings, combined with the total landed catch model of these same categories; and **Model 2B**, which follows the same logic of model 2A, but considers the number of fishing landings as a known covariate in all fishing ports, sampled or not. **Model 2Am** and **Model 2Bm** were formulated from models 2A and 2B, respectively, with analogue structures, but with

input data and inference broken down by month to better apprehend the seasonality of some fish categories.

Data analysis

Posterior parameter estimates were obtained using Monte Carlo methods by stochastic process of Markov chain Monte Carlo (MCMC) (Gelman et al. 2014). Three MCMC chains were simulated with 20000 iterations each. The retention interval in the sample was one iteration for every five (thinning) and the 5000 initial iterations were ignored (burn-in), resulting in a posterior sample with 12000 values for each parameter of the models. Estimates of the number of fishing landings and total landed catch were expressed as the summaries of posterior predictive distributions. The convergence of chains was evaluated using the Gelman-Rubin diagnostic (Brooks and Gelman 1998), where $\hat{R}$ values ≤ 1.1 suggest convergence. All analyses were performed using software R (R Core Team 2017) and JAGS (Plummer 2016), connected by the R2jags package (Su and Yajima 2015). The scripts developed for models 1, 2A, and 2B were included in the supplementary material of this article.

The notation and principal equations used to formulate the inferential and predictive models for monitoring fishing activity are presented in Table 1. The hierarchical sequence of relationships between the observed data, parameters estimated by the models, and estimates of the catch and number of landings was described using Directed Acyclic Graphs (DAG), where the known constants and observed data are represented by rectangles. The ellipses represent random variables or latent parameters, both with an associated probability distribution. The dotted arrows represent stochastic dependency and the solid arrows represent known deterministic relationships.

General structure of the inferential models

In JAGS, the Normal and log-Normal distributions are parameterized by the mean μ and precision τ , defined as the inverse of the variance ($\tau = \sigma^{-2}$). In the rest of the text, parameterized distributions with τ will always be marked with an asterisk, according to the criterion adopted in Kinas and Andrade (2010).

To obtain the estimates by municipality, within a sampled stratum h , the total number of landings in a fishing port (D_{hp}) was described by the Poisson distribution, where the parameter η_h represents the mean of the number of landings from the ports within h . Indeed, D_{hp} was described by its generalization in the form of Negative Binomial distribution (Poisson-Gamma), where the variance always exceeds the mean (Kinas and Andrade 2010). This was a measure adopted to deal with the high variability between ports and months, even within the same stratum.

Total landed catch of a fishing port Y_{hp} conditioned to the corresponding D_{hp} was obtained from the sum of catches per landing C_{hpd} . If the catches have a mean (m) and variance (v), then we assume log-Normal distribution $C_{hpd} \sim LN^*(\mu, \tau)$, with parameters obtained from m and v , according to equations 1 and 2 (Table 1). If $Y_{hp} = \sum_{d=1}^{D_{hp}} C_{hpd}$ expresses the total catch for D_{hp} , independent and identically distributed landings, then we approximate its distribution by $Y_{hp} \sim LN^*(\mu_D, \tau_D)$, where μ_D and τ_D are defined according to equations 3 and 4 (Table 1). $P(Y_{hp} = 0) = 1$ was defined whenever $D_{hp} = 0$.

For estimates by fish category (j) in the state of São Paulo, within stratum h , the number of landings in a fishing port p in (h, j) (D_{hpj}), given the total number of landings of a fishing port (D_{hp}) was estimated with a presence-absence model. Thus, D_{hp} became one of the parameters of the Binomial distribution that describes D_{hpj} , and

the proportion of each fish category in relation to total landed catches in stratum $(\rho_{(D)_{hj}})$ is the second parameter of this distribution, that is, $D_{hpj} \sim \text{Bin}(D_{hp}, \rho_{(D)_{hj}})$. In $\rho_{(D)_{hj}}$, the inclusion of index D is needed to differentiate it from the proportion of each fish category in relation to total landed catch within the stratum $(\rho_{(Y)_{hj}})$, which will be used later in this text.

The total landed catch conditioned to the total landings of a fishing port, per fish category (Y_{hpj}) , was described with a log-Normal distribution, with parameters μ_{Dj} and τ_{Dj} calculated in analogy to those described in equations 3 and 4 (Table 1), but for each analyzed fish category. Here, $P(Y_{hpj} = 0) = 1$ was also defined whenever $D_{hpj} = 0$.

Input data

The information obtained from sampling and available for the adjustment of model 1 were: vectors with number of landings $(D_{ihp}^{(obs)})$ and with total landed catch in tons $(Y_{ihp}^{(obs)})$ in the monitored ports in (i, h) for 2011. The constant $K = (0.01 * (D_{ihp}^{(obs)} + 1))$ was added for each $Y_{ihp}^{(obs)}$ to avoid zeros in the input data. This constant was subsequently subtracted in the prediction of the catches from non-observed ports. The number of sampled ports m_h was also included as necessary data to fit the model. Fishing information of 2010 was used to determine prior distributions of the parameters, as detailed below.

The information provided as data for models 2A and 2Am included a vector with the number of landings in 2011 at the observed ports $(D_{hp}^{(obs)})$ and vectors with the number of landings and total landed catch by fish category, port, and sampled stratum for 2011 $(D_{hpj}^{(obs)})$ e $Y_{hpj}^{(obs)}$, respectively). Here, $Y_{hpj}^{(obs)}$ was similarly added to the

constant K . The number of sampled ports m_h and total number of fish categories ($n.sp$) were also included as necessary data to fit the model.

For models 2B and 2Bm, the input data were the same as for models 2A and 2Am. However, D_{hp} became a vector with known values for all M_h ports. The values of the observed ports (1 to m_h) were used in the inferential model and the $(M_h - m_h)$ non-observed ports were used in the predictive model.

Prior probability distributions

In the inferential model 1, the set of parameters (θ_{ih}) for which we obtained a posterior sample was composed by $(\eta_{ih}, m_{ih} \text{ e } v_{ih})$. Informative prior distributions (Hall 2011) was used and based on fishing information of São Paulo in 2010, as shown in Table 2. The active fishing ports in 2010 were stratified under the same criteria used for 2011. Only the data of the observed ports in the simulated sample was used to obtain the parameters of the prior distributions.

The values of α_{ih}^{10} and β_{ih}^{10} were set to describe a Gamma distribution with mode equal to the average number of landings in (i, h) for 2010 and assuming a mean two times greater than the mode. Values of m_{min}^{10} and m_{max}^{10} were obtained as $\pm 30\%$ of the average individual catch (m_{ih}^{10}) in (i, h) for 2010, guaranteeing the minimum interval (0.01, 0.03) when m_{ih}^{10} was equal or too close to zero and around the m_{ih}^{10} itself in other cases. And the values of v_{min}^{10} and v_{max}^{10} were designed to guarantee the lower limit of the variance interval between individual catches (v_{ih}^{10}) in (i, h) of 2010 equal to zero and v_{max}^{10} around v_{ih}^{10} and greater than zero.

In the inferential model 2 (versions 2A and 2Am), θ_h was composed of $(\eta_h, m_{hj}, v_{hj} \text{ e } \rho_{(D)_{hj}})$. Informative prior distributions were also used, based on fishing data of São Paulo in 2010, as shown in Table 2. The values of m_{min}^{10} and m_{max}^{10} were calculated using the same criteria of model 1, but in relation to individual average

catch (m_{hj}^{10}) in (h, j) of 2010, as well as v_{min}^{10} and v_{max}^{10} . The prior distribution of $\rho_{(D)_{hj}}$ was defined as a Uniform distribution in $[0,1]$. In the versions of models 2B and 2Bm, η_h was no longer a part of θ_h since D_{hp} became a known information.

Definition of the inferential models

The different versions of the inferential models followed the general structure previously described and formally defined as shown in Table 3.

The DAG in Figure 1 defines the structure of inferential model 1 and the hierarchical sequences of relationships between the observed data and the model parameters in (i, h, p) . The DAG in Figure 2 does the same for the inferential models 2A, 2Am, 2B and 2Bm, but with the parameters of the model in (h, d, j) . Inferential models 2B and 2Bm were practically the same as models 2A and 2Am, but without the need to model η_h and D_{hp} (Fig. 2, disregarding the part in blue).

Predictive models

The multiple imputation of data from the non-observed ports was performed according to the model, based on the estimated parameters in the form of posterior distributions.

The structure of the predictive models and the relationships between the estimated parameters and estimates of total landed catch and number of landings by municipality and fish category were described by the DAGs in Figure 3 (model 1) and Figure 4 (models 2A, 2Am, 2B, and 2Bm). Each simulated value (component of the posterior distribution of the parameters) became a constant that provides, within the hierarchical sequence, parameters of the (posterior) predictive probability distributions. The sequence of calculations is described below.

In model 1, considering (i, h) and for each posterior sample value of the estimated parameters, $M_h - m_h$ values of $\hat{Y}_{ihp}^{(miss)}$ and $\hat{D}_{ihp}^{(miss)}$ were simulated. It

should be noted that $\hat{Y}_{ihp}^{(miss)}$ was simulated whenever $\hat{D}_{ihp}^{(miss)} > 0$; otherwise it was set equal to zero. Once it was done, total landed catch of the observed ports in stratum h ($Y_{ih}^{(obs)}$) and the following predictive distributions were obtained: landed catch of the non-observed ports within stratum h ($\hat{Y}_{ih}^{(miss)}$) and total landed catch of stratum h ($\hat{Y}_{ih}$) (Equation 5), total landed catch for municipality i $\hat{Y}_i$ (Equation 6), and total landed catch for the state of São Paulo ($\hat{Y}_{SP}$) (Equation 7). It was also obtained the number of landings in the observed ports of stratum h ($D_{ih}^{(obs)}$), the predictive distributions of the number of landings of non-observed ports of stratum h ($\hat{D}_{ih}^{(miss)}$), and total landings of stratum h ($\hat{D}_{ih}$) (Equation 9). Estimates of the number of landings for municipality i ($\hat{D}_i$) and for the state of São Paulo ($\hat{D}_{SP}$) were obtained in a similar manner to the process described in equations 6 and 7, but using the values of $\hat{D}$. Total landed catch and total number of landings by fish category in the state of São Paulo ($\hat{Y}_j$ and $\hat{D}_j$) (equations 8 and 10) were obtained considering the proportions of each fish category in relation to total landed catch ($\rho_{(Y)hj}$) and total number of landings ($\rho_{(D)hj}$) of the observed ports in each sampled stratum h .

In models 2A and 2Am, the prediction of total landed catch and number of landings by fish category in the state of São Paulo was similar to that of model 1, with the obvious and necessary adaptations. The predictive models 2B and 2Bm were practically the same of 2A and 2Am, but without the need to estimate $\hat{D}_{hp}^{(miss)}$ since D_{hp} turned into a covariate with all the known values.

Finally, it was possible to obtain the total landed catch by fish category of stratum h ($\hat{Y}_{hj}$) (Equation 11) and the total catch by fish category for the state of São Paulo ($\hat{Y}_j$) (Equation 12). The same estimates were made for the total number of

landings in the ports: $\widehat{D}_{hj}$ (Equation 14) and $\widehat{D}_j$, obtained in the same way as the method described in Equation 12. In these models, $\widehat{Y}_i$ and $\widehat{D}_i$ were obtained according to Equation 13 and analogous to the description of Equation 6, respectively. The estimate of $\widehat{D}_i$ cannot be obtained from the sum of the number of landings of all the fish categories since the same landing can have more than one category, resulting in $\sum_{j=1}^{n.sp} \widehat{D}_{hj} > \widehat{D}_h$. The estimates of number of landings ($\widehat{D}_{SP}$) and total landed catch for the state of São Paulo ($\widehat{Y}_{SP}$) were obtained using the same equations defined in predictive model 1.

RESULTS

The estimated total landed catch by municipalities and the summary of their posterior predictive distributions are shown in Table 4. Only municipalities with fishing ports allocated in the sampled strata were considered. Seven municipalities had very good $\widehat{Y}_i$ estimates, with fairly low (<6.0%) bias and coefficient of variation (CV). However, in Iguape, Ilha Comprida, and Ilhabela, where the lowest CV occurred, the credibility interval of the estimated catch did not include the true total landed catch of 2011. This result was also observed for Cananéia and Caraguatatuba, which are municipalities with more biased catch estimates (134.5% and 22.0% higher than the true value, respectively). The highest CV values were observed in the municipalities of Santos/Guarujá (12.7%) and Ubatuba (12.2%). Estimates of the total number of landings were quite close to the true values in most municipalities, but they were overestimated in Cananéia and very underestimated in Santos/Guarujá. The variability of $\widehat{D}_i$ was very small in all municipalities and the coverage of the credibility interval for 2011 was low.

After obtaining the estimates for the municipalities in São Paulo, the next step was to formulate an inferential model to obtain the catch and number of landings by fish category ($\hat{Y}_j$ and $\hat{D}_j$), model 2A, whose results are shown in tables 5 and 6. The 20 main fish categories landed on the coast of São Paulo in 2011 (identified at lowest taxonomic rank possible) were considered in the analysis. Landed catches of species with low market value or of small individual of commercial species from all fisheries (with or without target species) were grouped in the category "Landed mixture of fishes" and all other categories have been lumped together in "Others".

Total landed catch estimates were considered accurate for various categories, such as *Micropogonias furnieri*, *Xiphopenaeus kroyeri*, Ariidae, among others. In contrast, the catch was over- or underestimated for some species, especially *Crassostrea brasiliiana*, *Mugil liza*, *Selene* spp, and *Doryteuthis* spp. The variability was greater in the estimated catch of *M. liza*, *Selene* spp, and *Caranx crysos*. In relation to the estimated number of landings, CV was quite low in all cases and the bias was greater in the estimates of *C. brasiliiana*, *Macrodon atricauda*, and *X. kroyeri*. The coverage of the credibility interval was lower in the estimate of $\hat{D}_j$ than $\hat{Y}_j$, especially when the bias were a little higher.

The results of these two models revealed that: the values of $\hat{Y}_j$ were worse for those fish categories that, knowingly, have a variation in the landed catches throughout the year on the coast of São Paulo; and $\hat{D}_i$ and $\hat{D}_j$ estimates were more biased compared to the catch estimates. In view of these results, new versions of the inferential models were designed, namely the 2Am model with monthly information of the sampled strata to assess the influence of fishing seasonality in all their estimates; model 2B sought to cancel the influence of the bias of $\hat{D}_{hp}$ in the $\hat{D}_j$ and $\hat{D}_i$ estimates; while the 2Bm model joint these two characteristics: seasonality and known values of D_{hp} . The estimates of

$\hat{Y}_i$ of the models 2A, 2Am, 2B, and 2Bm are presented in Table 7, while the estimates of $\hat{Y}_j$ e $\hat{D}_j$ of the models 1A, 2Am, 2B, and 2Bm are in tables 8 and 9. The results presented from this point on refer to the municipalities and fish categories with striking differences between the estimates of the models. The full version of all the tables was included in the supplementary material of this article.

In general, model 2Am had the best results for the estimates of $\hat{Y}_i$. Compared to model 1, the estimates of total landed catch by municipality improved in all the models for Santos/Guarujá and Ubatuba, and improved considerably for Cananéia (Table 7). In contrast, the $\hat{Y}_i$ estimates for Iguape and São Sebastião were worse in the other model versions compared to the estimates of model 1.

Compared to model 2A, the estimates of $\hat{Y}_j$ obtained from models 2Am and 2Bm were notably better for *M. liza*, *Selene* spp, and *C. brasiliiana*, and better, but very similar, for *S. brasiliensis*, *Doryteuthis* spp (Table 8). Models 1 and 2A had the best $\hat{Y}_j$ results for *Anchoviella lepidentostole*; however, for most of these fish categories, the best $\hat{Y}_j$ estimates were observed in model 2B. With respect to the $\hat{D}_j$ estimates, model 2A had the best results for most fish categories (5/9), while model 1 was the best adjusted for the estimates of *S. brasiliensis*, *M. liza*, and *A. lepidentostole* (Table 9). All model versions had low credibility interval coverage in the estimates of $\hat{Y}_j$, and especially of $\hat{D}_j$.

DISCUSSION

The inferential model proposed in this study had to deal with all complexity of the fishing activity. It was prepared to obtain estimates of landed catch and number of landings for 11 municipalities and 20 fish categories with completely different fishing characteristics. The results revealed that the model was very well-fitted to the data of

various municipalities and fish categories on the coast of São Paulo, and provided good estimates for catch and number of landings, but not for all cases.

The difficulties identified in the process of creating the model included the inability of the model to deal with the high variability of observed data, the presence of zeros in the input data, and the likely influence of the seasonality of species in the catch estimates. Proposals to circumvent these model-structuring problems are addressed below.

Catch estimates of some fish categories of model 2A were not precise and accurate, as in the case of *M. liza* and *Selene* spp. These categories have in common the fact that they are school-forming species and most of them have pelagic habits, at least at some point in their lifecycle (Menezes et al. 2003). They also had strong seasonality and interannual variation (Mendonça and Miranda 2008), being more common in the landings of some specific fishing ports on the coast of São Paulo. Models 2Am and 2Bm, which incorporate monthly information in the input data, were designed to reduce the variability of the estimates and to control the influence of harvest and protection periods (seasonal fishing ban). The best results recorded for models 2Am and 2Bm compared to model 2A for some of these fish categories proved that this solution was successful in these cases. Estimates by municipality where these species are commonly landed (e.g. São Sebastião) also improved.

In contrast, dealing with monthly data by fish category and fishing port revealed a problem previously detected in model 2A and common in fish data: the presence of several zeros. In this case, the zeros do not represent non-observed data in the sampling process, but rather the fact that not all the species occur in all the months and observed fishing ports. Consequently, in models 2A and 2Am, the $\hat{D}_{hpj}$ estimate incorporated, through a Binomial distribution, the presence probability of the fish categories in the

fishing landings $(\rho_{(D)_{hj}})$. Moreover, in the inference it was defined that $P(Y = 0) = 1$ whenever $D = 0$ and in the prediction, $\hat{Y}$ was estimated only when $\hat{D} > 0$. These measures enhanced the precision and accuracy of the estimates in all the model versions. However, for those fish categories that did not have seasonality and/or are less frequent in the fishing landings, the best results were observed in models 2A and 2B, instead of in models 2Am and 2Bm, where zeros are more frequent. This result may be evidence that the occurrence of zeros in the input data of the models is not fully resolved.

A key component of the Bayesian approach is the inclusion, in the form of prior distributions, of all past relevant and available information on the model parameters. Bayesian inference obtains acceptable results since the model is a realistic description of the observed data, especially in poor data situations, when the design-based approach proves problematic. Prior distributions are important to compensate this insufficiency in data, directing the inferences through the posterior and predictive distributions (Kinas and Andrade 2007, Hall 2011, Gelman et al. 2014). In this work, the use of informative prior distributions based on fishing data of 2010 of the coast of São Paulo was yet another measure adopted in the formulation of the inferential models.

While the use of these prior distributions directed the models and proved critical to deal with the high variability and the presence of zeros in the observed data, it was also a problem for the strata where prior information was very different from present information. This problem was observed in the total catch estimates obtained by model 1 for the municipalities of Cananéia, where the intervals specified in the prior distributions of the mean and variance did not cover the values obtained using data from 2011. This can also be one of the reasons for the bad estimated catch results for some categories of fish obtained by the other model versions. On the other hand, a large

variability in the number and specific composition of landings and in the total landed catch between two consecutive years is quite common situation in fishing activity. This variability can be circumvented by considering data from a set of previous years.

The excessive influence of the prior distributions and the lack of convergence of the MCMC chains observed in some strata and months in different model versions are issues that still need to be better understood and properly addressed. However, only when versions 2B and 2Bm were designed, it was possible to perceive that one of the possible motives for the lack of model adjustment was in the way in which the fishing ports were initially stratified. It was though expected that the adjustments would improve with the inclusion of number of landings as a known covariate, which surprisingly did not occur. It was conjectured that the strata created per municipality to apply the methodology proposed by IBGE (Lima-Green and Moreira 2012, Miranda et al. 2016) and adopted here to compare the results of both approaches were not at all suited for model-based inference.

To obtain good total catch estimates by municipality using a design-based sampling methodology, the variable municipality must be considered in the sampling plan (Lumley 2010). Consequently, the adopted stratification was suitable for the methodology proposed by Lima-Green and Moreira (2012). In this stratification, the sampled strata were divided per municipality and composed of relatively homogeneous fishing ports in relation to total landed catch, but not necessarily in relation to the number of landings. Given the inclusion of the municipalities in the definition of strata, 41 fishing ports were divided among the 19 sampled strata, which resulted in only two ports per stratum, on average. In Miranda et al. (2016) and in this work, the adopted sample design was the main reason why estimates of the number of landings were less accurate.

Moreover, the relationship between Y and D was inexistent in many of these strata. To illustrate this situation, a stratum of Santos/Guarujá with two sampled ports that landed approximately 50 t each in 2011, what is fairly homogeneous in terms of catch. In one of these ports, where larger fishing vessels landed their catches, only 15 landings were recorded, while 1040 landings were recorded in the other port used by smaller fishing vessels. The lack of this relationship in the original stratification adopted in this work, both in the strata and in the sample is shown in Figure 5A.

The properties and advantages of the model-based Bayesian approach are valid when the model is well specified for the observed data (Zangeneh and Little 2015). But it is also a fact that the sampling design can and should be planned to facilitate adjustment of the observed data to the model being proposed. Although it is not a simple task, the sampling mechanism can be incorporated into the model.

Based on everything that has been exposed here, it is believed that in order to use model-based inference in fishing activity, a stratification that is totally different from that one adopted in this work should be considered. The basic idea of this new proposal (model 3) is to use the linear relationship between $\ln(Y)$ and $\ln(D)$ of all the fishing ports to design a new stratification of the fishing ports, resulting in a paradigm change on how the census stratum was defined. It no longer contains the most important fishing ports in terms of total landed catch, but those which not behave in accordance to the proposed inferential model 3. Therefore, to establish the new sampling design, the outliers of the established linear relationship were identified and allocated in census stratum (18 ports), while all other ports (115) were allocated in sampled strata, as presented in Figure 5B. To ensure that data are observed throughout the entire range on landing counts, three sampled strata were defined as follows: (1) ports with less than 200 fishing landings ($M_h = 42$); (2) ports with 201 to 500 fishing landings ($M_h = 35$);

and (3) ports with more than 500 fishing landings ($M_h = 38$). The sample of the fishing ports was simulated using a simple random sampling, proportional to the size of the strata (20% of ports in each h). Thus, the selected sample consisted of 42 fishing ports, of which 18 were allocated in the census stratum and 24 in the sampled strata (Fig. 5B).

To identify the outliers and divide the remaining ports into three large sampled strata, we had to assume that there is general information about total landed catch and number of landings in all the considered fishing ports. This may be difficult to arrange in practice. The goal here, however, is to assess its suitability for sampling with a model-based approach. The fact is that a sampling design (conventional or model-based) can always be improved when relevant information is available at the time of its elaboration.

Some important features of this new sampling design proposed for model 3 should be highlighted, namely: (1) a striking reduction in the number of sampled strata, now larger and more homogeneous with regard to total landed catch and also to total number of landings by fishing ports; (2) reduction of 45% in the sample size, which now includes 31.6% of the total fishing ports; and (3) in the census stratum, the number of ports dropped by half the number of the original stratification. The original 36 fishing ports with 73.9% of the total landed catches in São Paulo dropped to only 18 ports that, however, landed 83.1% of the total catch of the state.

The input data of model 3 was the information of all the observed ports about the total landed catch (Y_{hp}), the number of landings (D_{hp}), and the stratum h to which these ports belong. The set of parameters (θ) for which we obtained a posterior sample consisted of (η_h , τ , β_0 and β_1) where η_h is the mean of the number of landings in the ports of stratum h , τ is the precision, and β_0 and β_1 are the parameters of the linear regression between $\ln(D_{hp})$ and $\ln(Y_{hp})$.

The proposed inferential model was formally defined as described below:

For each sampled port p , within stratum h :

$$D_{hp} \sim Poi(\eta_h)$$

$$Y_{hp} \sim LN^*(\mu_{hp}, \tau)$$

$$\mu_{hp} = \beta_0 + \beta_1 * \ln(D_{hp} + 1)$$

The vague prior distributions for each sampled stratum h were defined as:

$$\beta_0 \sim N^*(0, 1E - 5)$$

$$\beta_1 \sim N^*(0, 1E - 5)$$

$$\tau \sim Gamma(0.01, 0.01)$$

$$\eta_h \sim Gamma(0.01, 0.01)$$

The estimates of $\hat{Y}_i$ obtained using the alternative sampling plan and model 3 were far more accurate in the municipality of Cananéia in comparison to model 1 (Tables 4 and 10). Good results for model 3 were also observed in Santos/Guarujá, Iguape, Ubatuba, and Ilhabela (Table 10). These five municipalities were the most important in relation to the total landed catch from the São Paulo coast in 2011 and jointly represented 93% of the total catch of that year. In the other municipalities, CV and bias were quite high. These municipalities have few fishing ports and low total landed catch. Clearly, this first version of model 3 did not deal well with this situation, but, on the other hand, it can be assumed that 93% of the problem is already well solved for this data set. The estimates of $\hat{D}_i$ were not as accurate as expected once that the number of total landings was more homogeneous in the new stratification proposal. The adoption of GLMs with the inclusion of covariates, such as the number of fishing vessels and fishermen or the types of fishing fleet that operate at all each port or even the seasons of the year, can be a way to consider the specificity of each fishing port and to improve the accuracy of their estimates. This should be the next step in the formulation of this model.

This version of the model 3 is a simplified first idea that will be further developed in future work. This development includes a version to obtain estimate by fish category and other attributes. Nevertheless, good estimates for all five major fishing municipalities of São Paulo were obtained with an expressive increase in sampling efficiency and with a likely higher cost reduction for fisheries monitoring than that obtained with the original sampling design.

All the attempts to fit the different model versions to the data confirmed the huge potential of the model-based approach for the inferential fisheries monitoring. The possibility of including previous information and different covariates is a notable advantage of Bayesian inference, as well as being essential for dealing with small samples and highly variable information.

We propose Bayesian inference as a good alternative to sampling surveys for fisheries monitoring. Future work should include: an exploratory analysis of the information aiming at the elaboration of a sampling plan more harmonized with the model; and the exploration of covariates when designing a sampling strategy and formulating a model in order to have a better calibrate model at the time of predicting outcomes for non-sampled fishing ports.

ACKNOWLEDGEMENTS

We express our thanks to the fisheries scientists and technical team of the Fishing Activity Monitoring Program (PMAP) and to the Fisheries Institute of the Department of Agriculture and Food Supply of São Paulo State.

REFERENCES

- Aragão J.A.N., Martins S. 2006. Censo Estrutural da Pesca – Coleta de Dados e Estimação de Desembarques de Pescado – IBAMA, Brasília/DF. 180 pp.
- <http://www.propesq.pesca.sp.gov.br/propesq/web/app.php/publico/10/conteudo>
- Ávila-da-Silva A.O., Carneiro M.H., Fagundes L. 1999. Sistema gerenciador de banco de dados de controle estatístico de produção pesqueira marinha – ProPesq. In: Anais do XI Congresso Brasileiro de Engenharia de Pesca e I Congresso Latinoamericano de Engenharia de Pesca. 2:824-832. Recife/PE (17-21/01/1999).
- Ávila-da-Silva A.O., Carneiro M.H., Mendonça J.T., Bastos G.C.C., Miranda L.M., Ribeiro W.R., Santos S. 2017. Produção Pesqueira Marinha e Estuarina do Estado de São Paulo – Dezembro de 2016. Inf. Pesq. de São Paulo. 80: 1-4.
- Bolfarine H., Bussab, W.O. 2005. Elementos de Amostragem, 1ª Ed. Editora Blucher, São Paulo/SP. 274 pp.
- Brooks S.P., Gelman A. 1998. Alternative methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7:434-455.
- Bussab W.O., Morettin P.A. 2012. Estatística Básica. 7ª Ed. Editora Saraiva, São Paulo/SP. 540 pp.
- Caddy J.F., Bazigos J.P. 1985. Practical guidelines for statistical monitoring of fisheries in manpower limited situations. FAO Fish Tech Pap. nº 257. Roma. 86 pp.
- Cadima E.L. 2003. Fish stock assessment manual. FAO Fish Tech Pap. nº 393. Roma. 161 pp.
- Cadima E.L., Caramelo A.M., Afonso-Dias M., Conte de Barros P., Tandstad M.O., de Leiva-Loreno J.I. 2005. Sampling methods applied to fisheries science: a manual. FAO Fish Tech Pap. nº 434. Rome. 88 pp.
- Cochran W.G. 1977. Sampling Techniques. Wiley: New York. 428 pp.

- Dias-Neto J. 2010. Pesca no Brasil e seus aspectos institucionais – um registro para o futuro. *Revista CEPSUL - Biodiversidade e Conservação Marinha*, Itajaí – SC. 1 (1): 66-80.
- Dias-Neto J. 2011. Números e Baionetas – A Nova Estatística da Produção Pesqueira do Brasil. Erro Estatístico ou Equívoco Político?. *Pesca & Mar - Informativo SAPERJ* (março/abril). Rio de Janeiro/RJ. 132: 31-34.
- FAO 1999. Guidelines for the routine collection of capture fishery data. *FAO Fish Tech Pap.* nº 382. Roma. 113 pp.
- Freire K.M.F., Aragão J.A.N., Araújo A.R.R., Velasco-Canziani G., Carneiro M.H., Gonçalves F.D.S., Keunecke K.A., Mendonca J.T., Moro P.S., Motta F.S., Olavo G., pezzuto P.R., Sant’Ana R.F., Santos R.A., Trindade-Santos I., Vasconcelos J.A., Vianna M., Divovich E. 2015. Reconstruction of catch statistics for Brazilian marine waters (1950-2010). *Fisheries Centre Research Reports*, 23: 3-30.
- Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., Rubin D.B. 2014. *Bayesian Data Analysis*. 3ª Ed. Texts in Statistical Science Series. CHAPMAN & HALL/CRC PRESS. Boca Raton, Florida, USA. 639 pp.
- Hall B. 2011. Bayesian inference.
<https://cran.r-project.org/web/packages/LaplacesDemon/vignettes/BayesianInference.pdf>
- Hilborn, R., Walters, C.J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Chapman and Hall, New York. 570 pp.
- Issac V.J., Ruffino M.L., Mello P. 2000. Considerações sobre o Método de Amostragem para a Coleta de Dados sobre Captura e Esforço Pesqueiro no Médio Amazonas. IBAMA – Coleção Meio Ambiente. Série Estudos Pesca. 22: 175-199.
- Issac V.J., Espírito Santo R.V., Nunes J.L.G. 2008. A estatística pesqueira no litoral do Pará: resultados divergentes. *Panamjas*, 3(3): 205-213.

- Kéry M. 2010. Introduction to WinBUGS for Ecologists: a Bayesian approach to regression, ANOVA, mixed models and related analyses. 1ª Ed. Academic Press. Burlington/USA. 302 pp.
- Kinas P.G., Andrade H.A. 2007. Bayesian statistics for fishery stock assessment and management: a synthesis. *Panamjas*, 2(2): 103-112.
- Kinas P.G., Andrade H.A. 2010. Introdução à Análise Bayesiana (com R). Editora maisQnada. Porto Alegre/RS. 240 pp.
- Kruschke J.K. 2015. Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. 2ª Ed. Elsevier. London/England. 759 pp.
- Lima-Green A.P., Moreira G.G. 2012. Metodologia Estatística da Pesca – Pesca Embarcada. Textos para Discussão – Diretoria de Pesquisas (IBGE), Rio de Janeiro/RJ. 40: 1-52.
- Little R.J. 2004. To model or not to model? Competing modes of inference for finite population sampling. *J. Amer. Statist. Assoc.* 99(466): 546-556.
- Lumley T. 2010. Complex surveys: a guide to analysis using R. Hoboken, N.J.: John Wiley. 276 pp.
- McCarthy M.A. 2007. Bayesian methods for ecology. Cambridge University Press. Cambridge/England. 296 pp.
- Maunder M.N., Punt A.E. 2004. Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*. 70: 141–159.
- Mendonça J.T., Miranda L.V. 2008. Estatística pesqueira do litoral sul do estado de São Paulo: subsídios para gestão compartilhada. *Panamjas*, 3(3): 152-173.
- Menezes N. A., Buckup P. A., Figueiredo J. L., Moura R. L. 2003. Catálogo de Peixes Marinhos do Brasil. Museu de Zoologia da Universidade de São Paulo, São Paulo.

- Miranda L.V., Kinas P.G., Moreira G.G., Namora R.C., Carneiro M. H. 2016. Survey sampling for fisheries monitoring in Brazil: implementation and analysis. *Braz. J. Oceanogr.* 64(4): 401-414.
- Pauly D. 2013. Does catch reflect abundance? Yes, it is a crucial signal. *Nature*. 494: 303-306.
- Pauly D., Zeller, D. 2016. Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nature communications*. 7:10244.
- Plummer M. 2016. JAGS Version 4.2.0 user manual.
<https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/>
- R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
<http://www.R-project.org>
- Rubin D.B. 1976. Inference and missing data. *Biometrika*. 581-592.
- Su Y.S, Yajima M. 2015. R2jags: Using R to run 'JAGS'. R package version 0.5-7.
<https://cran.r-project.org/web/packages/R2jags/index.html>
- Zangeneh S.Z., Little R.J. 2015. Bayesian inference for the finite population total from a heteroscedastic probability proportional to size sample. *J. Surv. Stat. Methodol.* 3(2): 162-192.

TABLES

Table 1. – Notation and main equations used in the Bayesian Inference for Finite Population for fisheries monitoring in Brazil.

Symbology	
i	– Municipality
p	– Port
h	– Stratum
H	– Total number of sampled strata ($H = 19$)
M_h	– Total number of ports in sampled stratum h .
m_h	– Total number of sampled ports in sampled stratum h .
$(M_h - m_h)$	– Total number of non-observed ports in sampled stratum h .
m	– Months
j	– Fish Categories
$n.sp$	– The 20 most important fish categories (in relation to landed catch) of São Paulo State in 2011, plus “Others” and “Byproduct”. ($n.sp = 22$)
x	– Posterior sample obtained by MCMC. ($x = 12000$)
Equations	
Inferential models	
Eq. 1	$\mu = \ln(m) - 0.5 * \left(\ln \left(1 + \frac{v}{m^2} \right) \right)$
	Location parameter of lognormal distribution for a unit landing's catch (C) within h , where m is the mean of C and v is its variance.
Eq. 2	$\tau = \left(\ln \left(1 + \frac{v}{m^2} \right) \right)^{-1}$
	Precision parameter of lognormal distribution for a unit landing's catch (C) within h , where m is the mean of C and v is its variance.
Eq. 3	$\mu_D = \mu + \left(\frac{1}{2} * (\tau^{-1} - \tau_D^{-1}) \right) + \ln(D)$
	Location parameter of lognormal distribution for total catch (Y), where D is the total number of landings, both within (h, p) .
Eq. 4	$\tau_D = \left[\ln \left(\frac{\exp(\tau^{-1}) - 1}{D} + 1 \right) \right]^{-1}$
	Precision parameter of lognormal distribution for total catch (Y), where D is the total number of landings, both within (h, p) .
Predictive models	
<i>Model 1 - Catches</i>	
Eq. 5	$\hat{Y}_{ih} = \sum_{p=1}^{m_{hi}} Y_{ihp}^{(obs)} + \sum_{p=m_{hi}+1}^{M_{hi}} \left(\hat{Y}_{ihp}^{(miss)} - K \right) =$ $= Y_{ih}^{(obs)} + \hat{Y}_{ih}^{(miss)}$
	Estimated total landed catch in sampled stratum h and municipality i , where the sum is over the observed ports ($Y_{ih}^{(obs)}$) and not sampled ports (missing information) ($\hat{Y}_{ih}^{(miss)}$), within (i, h) . $K = (0.01 * (d_{ihp}^{(obs)} + 1))$
Eq. 6	$\hat{Y}_i = \left(\sum_{h=1}^{H_i} \hat{Y}_{ih} \right) + Y_i^{eg}$
	Estimated total landed catch for municipality i , where Y_i^{eg} is the total landed catch of the census strata from same i .

Eq. 7
$$\hat{Y}_{SP} = \sum_{i=1}^I \hat{Y}_i$$

Estimated total landed catch for São Paulo State as a whole.

Eq. 8
$$\hat{Y}_j = \left(\sum_{h=1}^{H_i} \left(\rho_{(Y)ihj} * \hat{Y}_{ih} \right) \right) + Y_j^{eg}$$

Estimated total landed catch for fish category j , where $\rho_{(Y)ihj}$ is the proportion of each fish category in relation to the total landed catch of the observed ports within (i, h) and Y_j^{eg} is the total landed catch of the census strata by fish category.

Model 1 - Landings

Eq. 9
$$\begin{aligned} \hat{D}_{ih} &= \sum_{p=1}^{m_{hi}} D_{ihp}^{(obs)} + \sum_{p=m_{hi}+1}^{M_{hi}} \hat{D}_{ihp}^{(miss)} = \\ &= D_{ih}^{(obs)} + \hat{D}_{ih}^{(miss)} \end{aligned}$$

Estimated total number of landings in sampled stratum h and municipality i , where the sum is over the observed ports ($D_{ih}^{(obs)}$) and not sampled ports (missing information) ($\hat{D}_{ih}^{(miss)}$), within (i, h) .

Eq. 10
$$\hat{D}_j = \left(\sum_{h=1}^{H_i} \left(\rho_{(D)ihj} * \hat{D}_{ih} \right) \right) + D_j^{eg}$$

Estimated total number of landings for fish category j , where $\rho_{(D)ihj}$ is the proportion of each fish category in relation to the total number of landings of the observed ports within (i, h) and D_j^{eg} is the total number of landings of the census strata by fish category.

Model 2 and versions - Catches

Eq. 11
$$\begin{aligned} \hat{Y}_{hj} &= \sum_{p=1}^{m_{hi}} Y_{hjp}^{(obs)} + \sum_{p=m_{hi}+1}^{M_{hi}} \left(\hat{Y}_{hjp}^{(miss)} - K \right) \\ &= Y_{hj}^{(obs)} + \hat{Y}_{hj}^{(miss)} \end{aligned}$$

Estimated total landed catch in sampled stratum h and by fish category j , where the sum is over the observed ports ($Y_{hj}^{(obs)}$) and not sampled ports (missing information) ($\hat{Y}_{hj}^{(miss)}$), within (h, j) . $K = (0.01 * (d_{hjp}^{(obs)} + 1))$

Eq. 12
$$\hat{Y}_j = \left(\sum_{h=1}^H \hat{Y}_{hj} \right) + Y_j^{eg}$$

Estimated total number of landings by fish category j , where Y_j^{eg} is the total number of landings of the census strata by j .

Eq. 13
$$\hat{Y}_i = \left(\sum_{h=1}^H \sum_{j=1}^{n.sp} \hat{Y}_{hj} \right) + Y_i^{eg}$$

Estimated total number of landings for municipality i , where Y_i^{eg} is the total number of landings of the census strata from same i .

Model 2 and versions - Landings

Eq. 14
$$\begin{aligned} \hat{D}_{hj} &= \sum_{p=1}^{m_{hi}} D_{hjp}^{(obs)} + \sum_{p=m_{hi}+1}^{M_{hi}} \hat{D}_{hjp}^{(miss)} = \\ &= D_{hj}^{(obs)} + \hat{D}_{hj}^{(miss)} \end{aligned}$$

Estimated total number of landings in sampled stratum h and by fish category j , where the sum is over the observed ports ($D_{hj}^{(obs)}$) and not sampled ports (missing information) ($\hat{D}_{hj}^{(miss)}$), within (h, j) .

Table 2. – Summary of prior distributions for all model parameters in the Bayesian inferential models.

Model	Parameter	Description	Prior
Model 1	η_{ih}	Mean of the numbers of landings within (i, h)	<i>Gamma</i> ($\alpha_{ih}^{10}, \beta_{ih}^{10}$)
	m_{ih}	Mean of a unit landing's catch within (i, h)	<i>Unif</i> ($m_{min}^{10}, m_{max}^{10}$)
	Where:		
	$m_{min}^{10} = \max(0.01, m_{ih}^{10} - (m_{ih}^{10} * 0.3))$		
	$m_{max}^{10} = \max(0.03, m_{ih}^{10} + (m_{ih}^{10} * 0.3))$		
Model 2 and versions	v_{ih}	Variance of a unit landing's catch within (i, h)	<i>Unif</i> ($v_{min}^{10}, v_{max}^{10}$)
	Where:		
	$v_{min}^{10} = 0$		
	$v_{max}^{10} = v_{ih}^{10} + (v_{ih}^{10} * 0.3) + 0.0001$		
Model 2 and versions	η_h	Mean of the numbers of landings within (h)	<i>Gamma</i> ($\alpha_h^{10}, \beta_h^{10}$)
	m_{hj}	Mean of a unit landing's catch within (h, j)	<i>Unif</i> ($m_{min}^{10}, m_{max}^{10}$)
	Where:		
	$m_{min}^{10} = \max(0.01, m_{hj}^{10} - (m_{hj}^{10} * 0.3))$		
	$m_{max}^{10} = \max(0.03, m_{hj}^{10} + (m_{hj}^{10} * 0.3))$		
Model 2 and versions	v_{hj}	Variance of a unit landing's catch within (h, j)	<i>Unif</i> ($v_{min}^{10}, v_{max}^{10}$)
	Where:		
	$v_{min}^{10} = 0$		
	$v_{max}^{10} = v_{hj}^{10} + (v_{hj}^{10} * 0.3) + 0.0001$		
	$\rho_{(D)_{hj}}$	Proportion of occurrence in relation to the total number of landings of the observed ports within (h, j)	<i>Unif</i> (0, 1)

Table 3. – Formal definition of the different Bayesian inferential models.

Model 1	Models 2A and 2Am*
Within municipality i and sampled stratum h	Within sampled stratum h (* and month)
For each observed port ($p = 1$ to m_h):	For each observed port ($p = 1$ to m_h):
$D_{ihp} \sim Poi(\eta_{ih})$	$D_{hjp} \sim Poi(\eta_h)$
If ($D_{ihp} \geq 1$)	For each fish category ($j = 1$ to $n.sp$):
$Y_{ihp} D_{ihp} \sim LN^*(\mu_D, \tau_D)$	$D_{hjp} \sim Bin(D_{hjp}, \rho_{(D)hj})$
If ($D_{ihp} = 0$)	If ($D_{hjp} \geq 1$)
$P(Y_{ihp} = 0) = 1$	$Y_{hjp} D_{hjp} \sim LN^*(\mu_{Dj}, \tau_{Dj})$
	If ($D_{hjp} = 0$)
	$P(Y_{hjp} = 0) = 1$
Models 2B and 2Bm*	
Within sampled stratum h (* and month)	
For each observed port ($p = 1$ to m_h):	
For each fish category ($j = 1$ to $n.sp$):	
$D_{hjp} \sim Bin(D_{hjp}^{(obs)}, \rho_{(D)hj})$	
If ($D_{hjp} \geq 1$)	
$Y_{hjp} D_{hjp} \sim LN^*(\mu_{Dj}, \tau_{Dj})$	
If ($D_{hjp} = 0$)	
$P(Y_{hjp} = 0) = 1$	

Table 4. – Results of the model 1 – total landed catches (A) and total number of landings (B) by municipality for São Paulo coast over 2011. Catches are expressed in tons; $q_{2.5}$ and $q_{97.5}$ – limits of the credibility interval; CV - coefficient of variation; $\%B$ – Bias expressed as percentage of catch 2011.

(A)	Municipality	Y_i	$\hat{Y}_i$	$q_{2.5}$	Median	$q_{97.5}$	ICr_{cover}^*	$CV_{\hat{Y}_i}$	$\%B_{\hat{Y}_i}$
	Cananéia	3289.96	7715.58	7267.86	7710.18	8185.67	out	3.02	134.52
	Caraguatatuba	139.58	170.31	162.59	170.02	179.53	out	2.55	22.02
	Iguape	1078.18	1137.27	1110.67	1136.59	1167.35	out	1.27	5.48
	Ilha Comprida	56.97	60.34	59.51	60.32	61.28	out	0.75	5.92
	Ilhabela	702.15	696.39	693.34	696.25	700.27	out	0.25	-0.82
	Mongaguá	53.94	55.13	53.65	55.09	56.84	in	1.47	2.21
	Peruíbe	147.66	151.50	146.68	151.27	157.89	in	1.88	2.60
	Praia Grande	81.00	81.84	79.04	81.79	84.94	in	1.83	1.04
	Santos/Guarujá	11423.33	11249.49	10246.93	10875.93	14405.84	in	12.68	-1.52
(B)	São Sebastião	556.93	552.68	538.34	552.44	568.58	in	1.39	-0.76
	Ubatuba	2064.56	2240.24	2052.12	2061.03	2855.08	in	12.25	8.51
	SP	19961.79	24478.31	23155.67	24174.52	27663.17	out	6.01	22.63
	Municipality	D_i	$\hat{D}_i$	$q_{2.5}$	Median	$q_{97.5}$	ICr_{cover}^*	$CV_{\hat{D}_i}$	$\%B_{\hat{D}_i}$
	Cananéia	23896	26592	26125	26591	27062	out	0.91	11.28
	Caraguatatuba	2627	2486	2440	2486	2534	out	0.96	-5.36
	Iguape	21032	20006	19766	20006	20251	out	0.62	-4.88
	Ilha Comprida	2534	2680	2646	2679	2716	out	0.66	5.75
	Ilhabela	1970	1906	1865	1905	1950	out	1.14	-3.26
	Mongaguá	1323	1318	1285	1318	1354	in	1.33	-0.36
	Peruíbe	2540	2514	2464	2514	2565	in	1.02	-1.03
	Praia Grande	2385	2357	2304	2356	2412	in	1.19	-1.18
	Santos/Guarujá	9800	4652	4494	4652	4810	out	1.73	-52.53
	São Sebastião	5784	5991	5835	5990	6154	out	1.36	3.57
	Ubatuba	6082	6554	6454	6554	6655	out	0.77	7.76
	SP	83137	80219	79623	80216	80817	out	0.38	-3.51

Table 5. – Results of the model 2A – total landed catches (Y) by fish category for São Paulo coast over 2011. Catches are expressed in tons; $q_{2.5}$ and $q_{97.5}$ – limits of the credibility interval; CV - coefficient of variation; $\%B$ – Bias expressed as percentage of catch 2011.

Minor taxonomic rank	Y_j	$\hat{Y}_j$	$q_{2.5}$	Median	$q_{97.5}$	ICr_{cover}^*	$CV_{\hat{Y}_j}$	$\%B_{\hat{Y}_j}$
<i>Sardinella brasilienses</i>	4105.09	4220.83	4067.02	4193.34	4532.39	in	2.92	2.82
<i>Xiphopenaeus kroyeri</i>	2759.11	2660.69	2603.51	2660.94	2718.57	out	1.12	-3.57
<i>Micropogonias furnieri</i>	2102.17	2226.11	2125.05	2220.58	2359.46	out	2.71	5.90
<i>Cynoscion jamaicensis</i>	1036.34	1078.72	992.74	1075.12	1182.89	in	4.54	4.09
<i>Macrodon atricauda</i>	705.18	640.93	630.24	638.99	662.09	out	1.69	-9.11
<i>Menticirrhus</i> spp	646.99	764.35	697.61	759.38	856.55	out	5.40	18.14
<i>Anchoviella lepidentostole</i>	612.01	689.47	672.59	689.12	708.45	out	1.33	12.66
<i>Selene</i> spp	480.40	617.62	492.32	585.57	930.08	out	19.70	28.56
Ariidae	454.35	436.25	421.90	435.52	454.95	in	1.93	-3.98
<i>Octopus vulgaris</i>	355.26	335.57	316.02	334.20	362.17	in	3.57	-5.54
<i>Prionotus</i> spp	352.43	366.92	352.36	366.29	385.15	in	2.29	4.11
<i>Opisthonema oglinum</i>	343.43	342.94	342.23	342.86	344.09	in	0.14	-0.14
<i>Mugil liza</i>	319.68	497.32	396.66	489.88	639.40	out	12.55	55.57
<i>Oligoplites</i> spp	290.77	290.86	280.95	289.68	307.94	in	2.39	0.03
<i>Doryteuthis</i> spp	256.73	197.45	195.08	197.37	200.38	out	0.69	-23.09
<i>Trichiurus lepturus</i>	255.64	283.47	267.31	282.68	304.79	out	3.35	10.89
<i>Farfantepenaeus</i> spp	249.70	210.00	201.84	209.28	222.31	out	2.55	-15.90
<i>Urophycis</i> spp	193.85	229.74	203.34	226.86	272.67	out	7.77	18.51
<i>Crassostrea brasiliiana</i>	173.82	270.82	262.53	270.84	279.21	out	1.57	55.80
<i>Caranx crysos</i>	170.84	162.63	146.41	157.13	209.66	in	10.21	-4.81
Landed mixture of fishes *	882.45	893.61	762.08	846.83	1313.88	in	18.32	1.26
Others	3215.56	3681.97	3189.28	3664.15	4223.95	in	7.03	14.50
SP	19961.80	21098.25	20394.90	21060.12	22016.05	out	1.99	5.69

* Fishes with low market value or of small individual of commercial species from all fisheries, however, landed and marketed.

Table 6. – Results of the model 2A – total number of landings (D) by fish category for São Paulo coast over 2011. $q_{2.5}$ and $q_{97.5}$ – limits of the credibility interval; CV - coefficient of variation; $\%B$ – Bias expressed as percentage of catch 2011.

Minor taxonomic rank	D_j	$\hat{D}_j$	$q_{2.5}$	Median	$q_{97.5}$	ICr_{cover}^*	$CV_{\hat{D}_j}$	$\%B_{\hat{D}_j}$
<i>Sardinella brasilienses</i>	205	234	216	233	258	out	4.58	14.15
<i>Xiphopenaeus kroyeri</i>	17618	14628	14440	14626	14826	out	0.67	-16.97
<i>Micropogonias furnieri</i>	11311	10202	10046	10200	10358	out	0.78	-9.80
<i>Cynoscion jamaicensis</i>	1003	1039	1012	1038	1070	out	1.41	3.59
<i>Macrodon atricauda</i>	8277	6157	6048	6156	6271	out	0.93	-25.61
<i>Menticirrhus</i> spp	3357	3365	3302	3364	3430	in	0.96	0.24
<i>Anchoviella lepidentostole</i>	11630	11730	11524	11729	11940	in	0.90	0.86
<i>Selene</i> spp	849	783	740	782	831	out	2.97	-7.77
Ariidae	10835	10104	9888	10103	10327	out	1.11	-6.75
<i>Octopus vulgaris</i>	722	692	670	691	720	out	1.82	-4.16
<i>Prionotus</i> spp	600	667	642	666	697	out	2.14	11.17
<i>Opisthonema oglinum</i>	888	777	744	776	813	out	2.24	-12.50
<i>Mugil liza</i>	7053	6454	6295	6453	6619	out	1.28	-8.49
<i>Oligoplites</i> spp	3110	2825	2744	2824	2910	out	1.50	-9.16
<i>Doryteuthis</i> spp	2546	2445	2407	2445	2487	out	0.83	-3.97
<i>Trichiurus lepturus</i>	3865	4166	4071	4166	4265	out	1.18	7.79
<i>Farfantepenaeus</i> spp	872	871	849	871	899	in	1.48	-0.11
<i>Urophycis</i> spp	259	291	272	290	315	out	3.81	12.36
<i>Crassostrea brasiliiana</i>	5354	10614	10291	10613	10946	out	1.58	98.24
<i>Caranx crysos</i>	744	809	754	808	870	out	3.68	8.74
Landed mixture of fishes *	10661	8754	8659	8753	8850	out	0.56	-17.89
Others	39721	37475	37121	37474	37837	out	0.49	-5.65
SP	83137	80228	79644	80224	80829	out	0.38	-3.50

* Fishes with low market value or of small individual of commercial species from all fisheries, however, landed and marketed.

Table 7. – Results of the models 2A, 2Am, 2B e 2Bm – total landed catches (Y) by municipality for São Paulo coast over 2011. Catches are expressed in tons; $q_{2.5}$ and $q_{97.5}$ – limits of the credibility interval; CV - coefficient of variation; $\%B$ – Bias expressed as percentage of catch 2011.

	<i>Municipality</i>	Y_i	$\hat{Y}_i$	$q_{2.5}$	<i>Median</i>	$q_{97.5}$	<i>ICr</i> <i>cover*</i>	$CV_{\hat{Y}_i}$	$\%B_{\hat{Y}_i}$
Model 2A	<i>Cananéia</i>	3289.96	3519.68	3453.81	3519.42	3586.79	out	0.97	6.98
	<i>Iguape</i>	1078.18	1534.39	1113.48	1539.46	1901.97	out	12.25	42.31
	<i>Santos/Guarujá</i>	11423.33	11567.44	11009.09	11520.23	12404.66	in	3.19	1.26
	<i>São Sebastião</i>	556.93	805.26	699.78	797.57	954.75	out	8.05	44.59
	<i>Ubatuba</i>	2064.56	2062.87	2049.07	2057.19	2111.91	in	1.17	-0.08
	<i>SP</i>	19961.79	21098.25	20394.90	21060.12	22016.05	out	1.99	5.69
Model 2Am	<i>Cananéia</i>	3289.96	2949.59	2928.83	2948.34	2977.59	out	0.43	-10.35
	<i>Iguape</i>	1078.18	1233.53	1112.09	1224.34	1403.14	out	6.05	14.41
	<i>Santos/Guarujá</i>	11423.33	11425.02	11075.32	11388.61	12002.08	in	2.14	0.01
	<i>São Sebastião</i>	556.93	481.04	460.40	476.15	531.73	out	4.09	-13.63
	<i>Ubatuba</i>	2064.56	2118.22	2071.52	2110.51	2210.12	out	1.72	2.60
	<i>SP</i>	19961.79	19799.46	19411.76	19767.76	20392.28	in	1.31	-0.81
Model 2B	<i>Cananéia</i>	3289.96	3376.19	3323.54	3375.81	3429.41	out	0.80	2.62
	<i>Iguape</i>	1078.18	1785.92	1193.09	1792.31	2216.62	out	12.76	65.64
	<i>Santos/Guarujá</i>	11423.33	11420.01	11105.01	11378.26	11990.33	in	2.07	-0.03
	<i>São Sebastião</i>	556.93	715.03	637.02	708.46	830.42	out	6.98	28.39
	<i>Ubatuba</i>	2064.56	2045.66	2035.70	2040.49	2087.51	in	1.10	-0.92
	<i>SP</i>	19961.79	20975.90	20320.99	20959.10	21685.3	out	1.60	5.08
Model 2Bm	<i>Cananéia</i>	3289.96	3351.10	3302.17	3349.90	3407.41	out	0.80	1.86
	<i>Iguape</i>	1078.18	1845.59	1644.87	1837.99	2092.97	out	6.21	71.18
	<i>Santos/Guarujá</i>	11423.33	11382.47	11091.40	11347.73	11870.14	in	1.95	-0.36
	<i>São Sebastião</i>	556.93	633.03	592.01	629.39	695.57	out	4.28	13.66
	<i>Ubatuba</i>	2064.56	2076.23	2047.58	2069.48	2142.52	in	1.25	0.57
	<i>SP</i>	19961.79	20912.39	20529.92	20884.16	21442.98	out	1.22	4.76

Table 8. – Results of the models 1, 2Am, 2B e 2Bm – total landed catches (Y) by fish category for São Paulo coast over 2011. Catches are expressed in tons; $q_{2.5}$ and $q_{97.5}$ – limits of the credibility interval; CV - coefficient of variation; $\%B$ – Bias expressed as percentage of catch 2011.

	Minor taxonomic rank	Y_j	$\hat{Y}_j$	$q_{2.5}$	Median	$q_{97.5}$	ICr cover*	$CV_{\hat{Y}_j}$	$\%B_{\hat{Y}_j}$
Model 1	<i>Sardinella brasilienses</i>	4105.09	4161.83	4071.50	4128.32	4446.18	in	3.08	1.38
	<i>Xiphopenaeus kroyeri</i>	2759.11	3783.31	3615.92	3775.56	3989.81	out	2.54	37.12
	<i>Micropogonias furnieri</i>	2102.17	2515.18	2257.91	2443.06	3213.76	out	12.70	19.65
	<i>Macrodon atricauda</i>	705.18	734.93	719.54	732.18	767.07	out	2.04	4.22
	<i>Anchoviella lepidentostole</i>	612.01	681.00	665.84	680.64	698.05	out	1.20	11.27
	<i>Selene</i> spp	480.40	497.29	477.24	489.85	560.42	in	5.74	3.52
	<i>Mugil liza</i>	319.68	371.75	353.65	370.25	395.94	out	2.97	16.29
	<i>Doryteuthis</i> spp	256.73	319.04	295.59	313.43	361.85	out	6.38	24.27
	<i>Crassostrea brasiliana</i>	173.82	1462.18	1349.48	1461.01	1581.27	out	4.01	741.20
	SP	19961.80	24514.73	23188.62	24212.47	27704.13	out	6.00	22.81
Model 2Am	<i>Sardinella brasilienses</i>	4105.09	4131.77	4066.63	4117.65	4285.55	in	1.53	0.65
	<i>Xiphopenaeus kroyeri</i>	2759.11	2301.90	2292.65	2301.64	2312.77	out	0.22	-16.57
	<i>Micropogonias furnieri</i>	2102.17	2309.49	2161.28	2286.60	2615.99	out	5.21	9.86
	<i>Macrodon atricauda</i>	705.18	620.68	617.65	620.00	627.35	out	0.70	-11.98
	<i>Anchoviella lepidentostole</i>	612.01	872.70	753.12	863.83	1041.16	out	8.46	42.60
	<i>Selene</i> spp	480.40	493.32	470.30	483.82	572.99	in	7.04	2.69
	<i>Mugil liza</i>	319.68	211.74	193.80	206.62	261.06	out	8.90	-33.77
	<i>Doryteuthis</i> spp	256.73	194.17	193.04	194.10	195.66	out	0.35	-24.37
	<i>Crassostrea brasiliana</i>	173.82	166.52	164.88	166.48	168.32	out	0.53	-4.20
	SP	19961.80	19799.47	19411.76	19767.76	20392.28	in	1.31	-0.81
Model 2B	<i>Sardinella brasilienses</i>	4105.09	4170.50	4058.08	4146.35	4423.83	in	2.38	1.59
	<i>Xiphopenaeus kroyeri</i>	2759.11	2827.73	2773.38	2827.41	2882.96	out	1.00	2.49
	<i>Micropogonias furnieri</i>	2102.17	2162.86	2094.06	2159.00	2254.18	in	1.89	2.89
	<i>Macrodon atricauda</i>	705.18	645.21	634.22	643.62	664.21	out	2.09	-8.50
	<i>Anchoviella lepidentostole</i>	612.01	733.10	715.78	732.60	753.31	out	1.31	19.79
	<i>Selene</i> spp	480.40	572.61	481.19	544.95	831.80	out	17.37	19.19
	<i>Mugil liza</i>	319.68	388.53	317.23	381.88	497.39	in	12.00	21.54
	<i>Doryteuthis</i> spp	256.73	196.64	194.55	196.55	199.29	out	0.61	-23.41
	<i>Crassostrea brasiliana</i>	173.82	246.69	240.78	246.67	252.79	out	1.25	41.92
	SP	19961.80	20975.90	20321.00	20959.10	21685.31	out	1.60	5.08
Model 2Bm	<i>Sardinella brasilienses</i>	4105.09	4098.96	4061.36	4101.86	4187.92	in	0.99	-0.15
	<i>Xiphopenaeus kroyeri</i>	2759.11	2608.18	2585.40	2607.93	2632.16	out	0.46	-5.47
	<i>Micropogonias furnieri</i>	2102.17	2280.84	2139.26	2253.69	2586.43	out	5.42	8.50
	<i>Macrodon atricauda</i>	705.18	647.58	639.76	646.96	658.32	out	0.87	-8.17
	<i>Anchoviella lepidentostole</i>	612.01	1268.36	1108.78	1261.13	1476.76	out	7.46	107.24
	<i>Selene</i> spp	480.40	478.01	472.14	476.38	495.15	in	1.49	-0.50
	<i>Mugil liza</i>	319.68	269.47	246.22	266.54	311.46	out	6.54	-15.71
	<i>Doryteuthis</i> spp	256.73	198.01	196.12	197.94	200.34	out	0.54	-22.87
	<i>Crassostrea brasiliana</i>	173.82	237.81	231.98	237.78	243.89	out	1.29	36.81
	SP	19961.80	20912.39	20529.92	20884.16	21442.98	out	1.22	4.76

Table 9. – Results of the models 1, 2Am, 2B e 2Bm – total number of landings (D) by fish category for São Paulo coast over 2011. $q_{2.5}$ and $q_{97.5}$ – limits of the credibility interval; CV - coefficient of variation; $\%B$ – Bias expressed as percentage of catch 2011.

	Minor taxonomic rank	D_j	$\hat{D}_j$	$q_{2.5}$	Median	$q_{97.5}$	ICr cover*	$CV_{\hat{D}_j}$	$\%B_{\hat{D}_j}$
Model 1	<i>Sardinella brasilienses</i>	205	205	204	205	207	in	0.39	0.00
	<i>Xiphopenaeus kroyeri</i>	17618	14591	14436	14591	14750	out	0.55	-17.18
	<i>Micropogonias furnieri</i>	11311	10118	10027	10117	10212	out	0.46	-10.55
	<i>Macrodon atricauda</i>	8277	6127	6054	6127	6201	out	0.62	-25.98
	<i>Anchoviella lepidentostole</i>	11630	11699	11523	11698	11878	in	0.77	0.59
	<i>Selene</i> spp	849	751	741	751	762	out	0.71	-11.54
	<i>Mugil liza</i>	7053	6533	6470	6533	6597	out	0.50	-7.37
	<i>Doryteuthis</i> spp	2546	2441	2433	2441	2450	out	0.19	-4.12
	<i>Crassostrea brasiliana</i>	5354	10624	10392	10623	10856	out	1.13	98.43
	SP	83137	80219	79623	80216	80817	out	0.38	-3.51
Model 2Am	<i>Sardinella brasilienses</i>	205	300	277	299	325	out	4.10	46.34
	<i>Xiphopenaeus kroyeri</i>	17618	12451	12401	12451	12503	out	0.21	-29.33
	<i>Micropogonias furnieri</i>	11311	8887	8824	8887	8951	out	0.36	-21.43
	<i>Macrodon atricauda</i>	8277	5612	5571	5611	5654	out	0.38	-32.20
	<i>Anchoviella lepidentostole</i>	11630	9132	9091	9132	9171	out	0.22	-21.48
	<i>Selene</i> spp	849	748	719	748	779	out	2.02	-11.90
	<i>Mugil liza</i>	7053	5145	5096	5144	5196	out	0.50	-27.05
	<i>Doryteuthis</i> spp	2546	2445	2418	2445	2475	out	0.61	-3.97
	<i>Crassostrea brasiliana</i>	5354	5288	5231	5288	5345	out	0.55	-1.23
	SP	83137	53746	53744	53744	53744	out	0.46	-35.35
Model 2B	<i>Sardinella brasilienses</i>	205	235	214	234	264	out	5.39	14.63
	<i>Xiphopenaeus kroyeri</i>	17618	19423	19293	19424	19546	out	0.33	10.25
	<i>Micropogonias furnieri</i>	11311	10025	9890	10024	10166	out	0.70	-11.37
	<i>Macrodon atricauda</i>	8277	6567	6437	6565	6708	out	1.06	-20.66
	<i>Anchoviella lepidentostole</i>	11630	12426	12292	12427	12559	out	0.55	6.84
	<i>Selene</i> spp	849	780	735	779	830	out	3.12	-8.13
	<i>Mugil liza</i>	7053	6049	5921	6049	6179	out	1.08	-14.24
	<i>Doryteuthis</i> spp	2546	2445	2401	2443	2498	out	1.02	-3.97
	<i>Crassostrea brasiliana</i>	5354	9360	9166	9360	9553	out	1.06	74.82
	SP	83137	83137	-	-	-	-	-	-
Model 2Bm	<i>Sardinella brasilienses</i>	205	652	546	650	767	out	8.75	218.05
	<i>Xiphopenaeus kroyeri</i>	17618	19541	19380	19541	19704	out	0.42	10.91
	<i>Micropogonias furnieri</i>	11311	10355	10185	10353	10534	out	0.87	-8.45
	<i>Macrodon atricauda</i>	8277	6897	6735	6895	7066	out	1.22	-16.67
	<i>Anchoviella lepidentostole</i>	11630	12306	12158	12306	12460	out	0.62	5.81
	<i>Selene</i> spp	849	1196	1084	1194	1321	out	5.07	40.87
	<i>Mugil liza</i>	7053	6371	6211	6368	6544	out	1.34	-9.67
	<i>Doryteuthis</i> spp	2546	2875	2761	2874	2999	out	2.12	12.92
	<i>Crassostrea brasiliana</i>	5354	9374	9150	9372	9604	out	1.24	75.08
	SP	83137	83137	-	-	-	-	-	-

Table 10. – Results of the model 3 – total landed catches (A) and total number of landings (B) by municipality for São Paulo coast over 2011. Catches are expressed in tons; $q_{2.5}$ and $q_{97.5}$ – limits of the credibility interval; CV - coefficient of variation; $\%B$ – Bias expressed as percentage of catch 2011.

	<i>Municipality</i>	Y_i	$\hat{Y}_i$	$q_{2.5}$	<i>Median</i>	$q_{97.5}$	<i>ICr cover</i> *	$CV_{\hat{Y}_i}$	$\%B_{\hat{Y}_i}$
(A)	<i>Bertioga</i>	216.13	71.93	13.71	53.40	238.31	<i>in</i>	102.82	-66.72
	<i>Cananéia</i>	3289.96	3415.27	3005.79	3351.45	4210.69	<i>in</i>	9.67	3.81
	<i>Caraguatatuba</i>	139.58	169.00	56.40	144.13	427.2	<i>in</i>	68.20	21.08
	<i>Iguape</i>	1078.18	1063.34	817.06	1019.00	1582.25	<i>in</i>	19.33	-1.38
	<i>Ilha Comprida</i>	56.97	125.36	46.36	107.06	310.07	<i>in</i>	62.36	120.05
	<i>Ilhabela</i>	702.15	697.01	685.98	694.82	721.29	<i>in</i>	1.38	-0.73
	<i>Itanhaém</i>	45.47	36.80	11.71	31.07	97.77	<i>in</i>	63.71	-19.06
	<i>Mongaguá</i>	53.94	55.55	26.38	50.03	117.48	<i>in</i>	44.07	2.99
	<i>Peruíbe</i>	147.66	184.15	75.12	159.08	434.14	<i>in</i>	58.43	24.71
	<i>Praia Grande</i>	81.00	140.70	57.07	123.58	327.91	<i>in</i>	56.40	73.71
	<i>Santos/Guarujá</i>	11423.33	11469.95	11412.33	11451.46	11638.38	<i>in</i>	0.57	0.41
	<i>São Sebastião</i>	556.93	496.04	331.12	463.25	859.53	<i>in</i>	29.29	-10.93
	<i>São Vicente</i>	105.93	88.07	22.86	69.71	258.63	<i>in</i>	84.14	-16.86
	<i>Ubatuba</i>	2064.56	1852.90	1774.22	1835.59	2034.21	<i>out</i>	4.03	-10.25
	<i>SP</i>	19961.79	19866.09	18824.21	19723.44	21759.57	<i>in</i>	3.89	-0.48

	<i>Municipality</i>	D_i	$\hat{D}_i$	$q_{2.5}$	<i>Median</i>	$q_{97.5}$	<i>ICr cover</i> *	$CV_{\hat{D}_i}$	$\%B_{\hat{D}_i}$
(B)	<i>Bertioga</i>	1300	1656	1573	1655	1739	<i>out</i>	2.58	27.37
	<i>Cananéia</i>	23896	26612	26201	26609	27022	<i>out</i>	0.78	11.37
	<i>Caraguatatuba</i>	2627	3867	3730	3868	4006	<i>out</i>	1.82	47.21
	<i>Iguape</i>	21032	21943	21688	21944	22199	<i>out</i>	0.60	4.33
	<i>Ilha Comprida</i>	2534	2866	2749	2866	2985	<i>out</i>	2.09	13.11
	<i>Ilhabela</i>	1970	2015	1970	2014	2062	<i>in</i>	1.17	2.28
	<i>Itanhaém</i>	805	838	775	838	903	<i>in</i>	3.89	4.08
	<i>Mongaguá</i>	1323	1335	1259	1335	1411	<i>in</i>	2.90	0.90
	<i>Peruíbe</i>	2540	4017	3883	4017	4152	<i>out</i>	1.72	58.15
	<i>Praia Grande</i>	2385	3238	3112	3237	3370	<i>out</i>	2.03	35.75
	<i>Santos/Guarujá</i>	9800	10638	10556	10637	10721	<i>out</i>	0.39	8.55
	<i>São Sebastião</i>	5784	8387	8190	8386	8585	<i>out</i>	1.20	45.00
	<i>São Vicente</i>	1059	2029	1937	2029	2120	<i>out</i>	2.35	91.59
	<i>Ubatuba</i>	6082	4291	4166	4291	4414	<i>out</i>	1.46	-29.45
	<i>SP</i>	83137	93731	92689	93728	94791	<i>out</i>	0.57	12.74

FIGURES

Fig. 1. – Directed acyclic graph (DAG) of the Bayesian inferential model 1 to obtain the posterior distributions of all parameters given the observed data. Where $p = port = (1, 2, \dots, m_{ih})$.

Fig. 2. – Directed acyclic graph (DAG) of the Bayesian inferential model 2 (models 2A and 2Am in blue and black and models 2B e 2Bm in orange and black) to obtain the posterior distributions of all parameters given the observed data. Where $p = port = (1, 2, \dots, m_{ih})$ and $j =$ fish categories.

Fig. 3. – Directed acyclic graph (DAG) of the Bayesian predictive model 1 to obtain the predictive posterior distributions of the number of landings and the total landed catch by municipality, given the parameters estimated by the inferential model. Where $p = port = (m_{ih} + 1, \dots, M_{ih})$, $x =$ size of the posterior sample obtained by MCMC and $j =$ fish categories.

Fig. 4. – Directed acyclic graph (DAG) of the Bayesian predictive model 2 (models 2A and 2Am in blue and black and models 2B e 2Bm in orange and black) to obtain the predictive posterior distributions of the number of landings and the total landed catch by fish categories, given the parameters estimated by the inferential model. Where $p = port = (m_h + 1, \dots, M_h)$, $x =$ size of the posterior sample obtained by MCMC and $j =$ fish categories.

Fig. 5. – Linear relationship between the logarithms of the number of landings and the total landed catch of the fishing ports of the São Paulo State in 2011. (A) Original adopted sample design; and (B) New proposal of sample design.

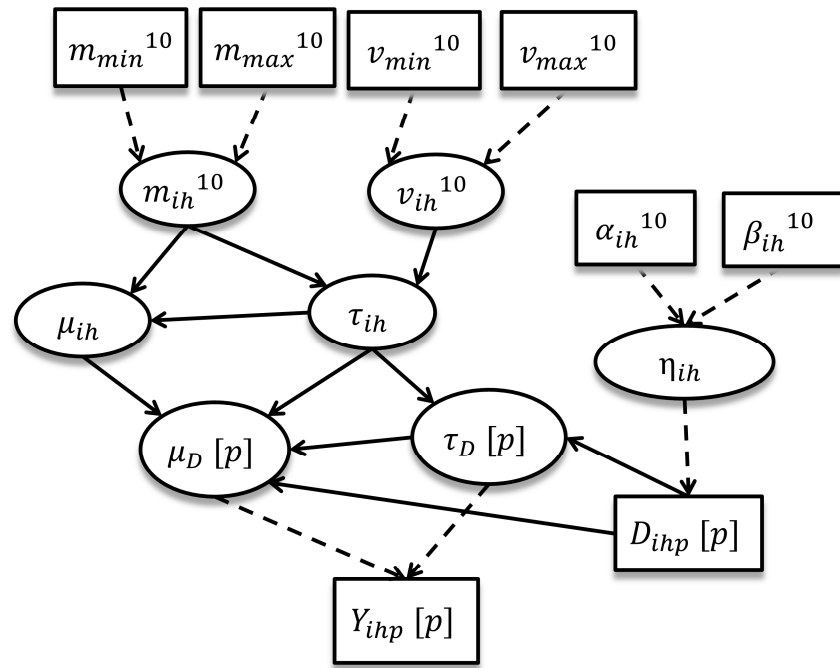


Figure 1

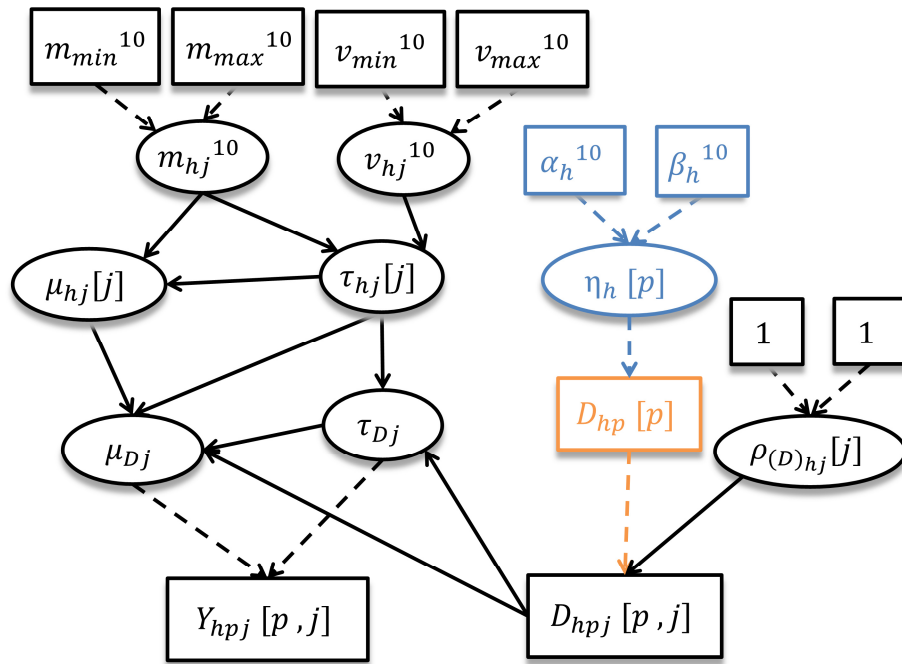


Figure 2

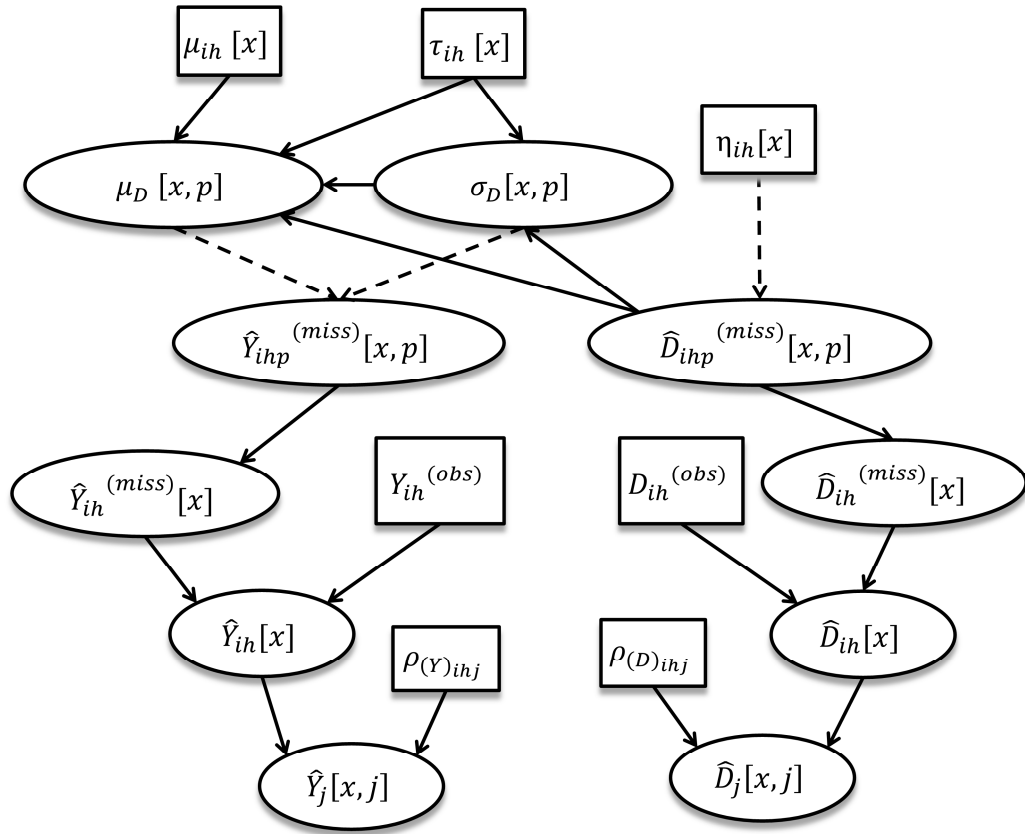


Figure 3

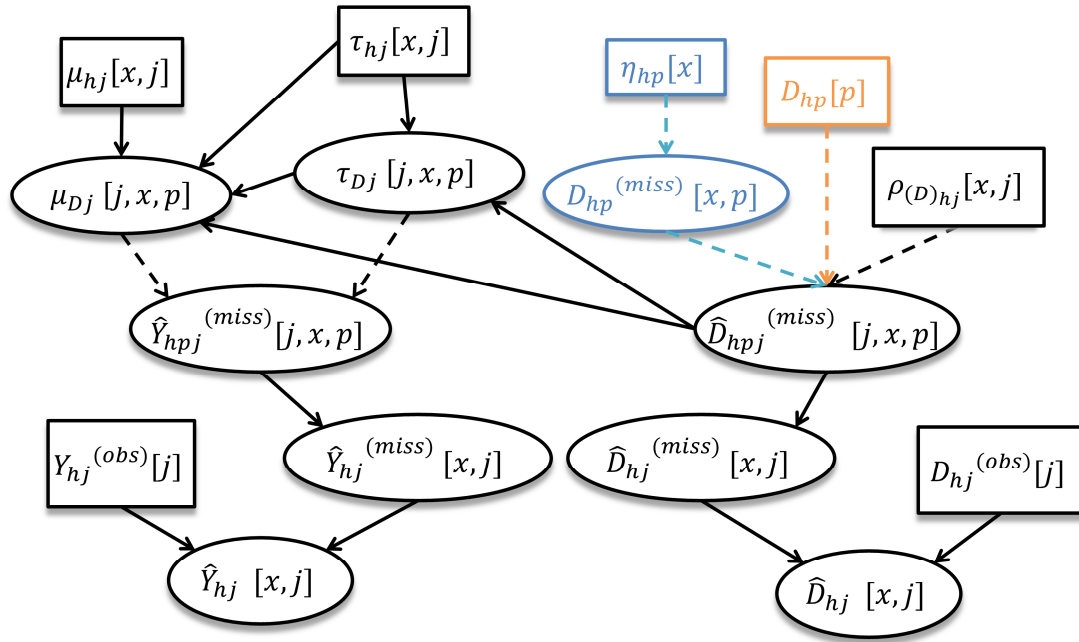


Figure 4

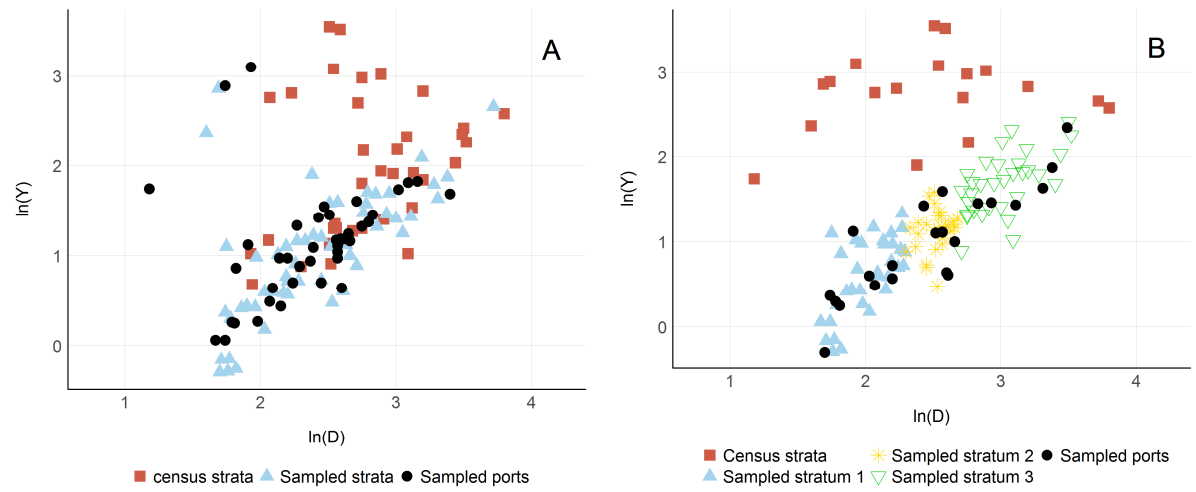


Figure 5