Pacific Islands Fisheries Science Center

PREDICTION OF FUTURE BYCATCH OF SEA TURTLES AND CERTAIN CETACEANS IN THE HAWAII DEEP-SET LONGLINE FISHERY [1]

Marti L. McCracken

Pacific Islands Fisheries Science Center
National Marine Fisheries Service
NOAA Inouye Regional Center
1845 Wasp Blvd., Building 176
Honolulu, Hawaii 96818

[1] PIFSC Internal Report IR-13-029
Issued 1 May 2014

1. INTRODUCTION

In this report, prepared at the request of the NOAA Fisheries Pacific Islands Regional Office (PIRO), predicted future annual levels of bycatch in the Hawaii deep-set longline fishery are provided for the main Hawaiian Islands insular stock of false killer whales; leatherback, loggerhead, green, and olive ridley sea turtles; and humpback and sperm whales. Additionally, predictions of the future annual levels of bycatch resulting in a classification of serious injury or mortality are provided for the three cetacean populations. The data, methods, and assumptions used to derive these predictions are described within this report.

The Hawaii-based longline fishery is a limited entry fishery with a maximum of 164 permits and is split into two management components: a shallow-set fishery (targeting primarily swordfish) and a deep-set fishery (targeting primarily tunas, most commonly bigeye tuna). In this report, a Hawaii longline fishing trip is defined as any commercial fishing trip by a vessel with a Hawaii longline permit that departs or returns at a Hawaii port. An exception to this rule is a trip that departs from a Hawaii port and lands in American Samoa. In this case, the trip is assigned to the American Samoa longline fishery. The shallow-set fishery consists of trips that were declared to the NOAA National Marine Fisheries Service (NMFS) by the captain, prior to departure, as shallow-set trips. The deep-set longline fishery consists of all other trips and must comply with the regulations for this fishery. A permitted vessel may participate in both fisheries. All the predictions provided in this report refer to the deep-set fishery.

Bycatch rates used in this report are derived from data on incidental takes of protected species collected by NMFS observers deployed on sampled deep-set Hawaii longline trips. The Marine Mammal Protection Act (MMPA) and the Endangered Species Act (ESA) define "take" in slightly different ways, but basically, "take" means to catch, kill, or harm a marine mammal or protected species in any way. An "incidental take" means a take that is incidental to, and not the purpose of, the carrying out of an otherwise lawful activity. Herein, "bycatch" refers to the total number of take events in which an animal is hooked or entangled in the longline gear. Under this definition, bycatch is a component of the total incidental take in the Hawaii longline deep-set fishery. For marine mammals, the definition of bycatch is extended to include observed calves of an adult marine mammal taken as bycatch. As there have been no such observed calves for the species being analyzed, I do not go into details about incorporating these events into the bycatch estimates.

Following the guidelines outlined by NMFS (2012*a*), an observed marine mammal incidental take is classified as a death or into a relevant injury category (i.e., serious or non-serious). Herein, the category "death or serious injury" is abbreviated as DSI. For cetaceans, the predictions computed here are for both the total bycatch and the number of bycatch events resulting in a DSI.

There are a couple of practical constraints on the definition of bycatch used herein. First, observers are instructed to record all observed hooked or entangled animals during haulback of the longline gear (Pacific Islands Regional Office Observer Progam, 2013). Animals observed hooked or entangled that are freed before being landed on deck are included in this definition. However, hooked or entangled animals that are removed (e.g. by predators) or freed (e.g. by escape or drop-off) from the longline prior to the longline becoming visible on the haulback would not be observable and therefore could not be recorded, unless warranted by convincing circumstantial evidence of their capture. These "missed" animals are not included in the bycatch estimates or predictions as there is no practical way to quantify them. Nor do the predictions include animals that are not hooked or entangled but are in some other unobserved way caught, killed, or harmed by the activity of deep-set fishing. Such events are not included because it is not feasible to monitor all aspects of a trip. If an observer witnesses an interaction with a marine mammal or sea turtle that does not fall into the category of hooked or entangled, the interaction is reviewed by NMFS to determine if that interaction should be included in the bycatch estimates. There has only been one observed take between 2002 and 2012 that was not categorized as hooked or entangled. This incident involved a marine mammal that was briefly restrained by the gear and non-seriously injured. Upon review, it was decided to include this interaction in the bycatch estimates and record this interaction in the catch table (table of all species that were hooked or entangled during an observed haul).

Second, the predictions of bycatch refer to the total number of bycatch events, which may exceed the number of individual animals that are caught. It is possible for an animal to be observed caught, then freed or released, and then caught again during the same year. For example, a loggerhead sea turtle was observed to be caught twice during a shallow-set trip in 2012. These two events are considered two bycatch events. Similarly, the DSI predictions refer to the total number of bycatch events resulting in a DSI. This number may exceed the number of individual animals that are DSI.

The predictions in this report were derived using a Bayesian inferential approach. In the next section some of the underlying assumptions and caveats behind the predictions are discussed. In Section 3 the data collection process for the data used to develop the prior predictive density is described. The prior predictive density is defined and the different Bayesian models used to compute these densities are introduced in Section 4. The requested predictions for each species are given in Section 5. The sections covering the predictions for each species are written so they are self-contained with regards to the results.

2. UNDERLYING ASSUMPTIONS AND CAVEATS

2.1. **Effort Scenario.** Except for predictions concerning false killer whales, the predictions presented in this report assume that the only alteration in the fishery or ecosystem that may affect bycatch is an increase in effort. The scenario for future effort provided by PIRO in its Biological Evaluation (BE) (NMFS, 2013) is that future fishing effort may be the same as that in 2012, but with the addition of small increases in the number of hooks and sets deployed annually. Specifically, including these increments, the effort scenario hypothesized 1,305 trips[2], 18,592 sets, and 46,117,532 hooks. The scenario implies that, on average, trips will be composed of more sets and that more hooks will be deployed on a set. Although effort is expected to fluctuate between years, predictions are made assuming that future annual effort is constant and equal to PIRO's projected scenario.

When using a statistical model to make predictions for new observations, the predictions for new cases are referred to as interpolations or extrapolations. Interpolation means the prediction is for new cases with predictor variables not too different from the sampled values of the predictor variables used to construct the fitted model; whereas, extrapolation refers to predictions based on predictor variables outside the range observed in the sample used to construct the model. For an interpolation to be useful, only a few assumptions must

[2]There were 1,305 deep-set Hawaii longline fishing trips landing in 2012; this is the number of trips used in the predictions. The BE indicated there were 1,523 trips in 2012; this figure was based on a different method of aggregation and was not used.

be satisfied, principally that the new cases behave like those in the dataset used to fit the model. For valid extrapolation we also require that the estimated prediction function is relevant to cases with predictors outside of the interpolatory range. For the extrapolations in this report to be relevant, not only does the predictor model have to be relevant for values of effort outside the range of observed effort, it also needs to be relevant to the future composition of trips (more hooks per set and more sets per trip).

2.2. **Bycatch as a Function of Effort.** To predict future annual bycatch levels, we are assuming that bycatch levels are proportional to a function of effort. This is a simple model that makes the assumption that bycatch can be controlled by changing the amount of effort. Although it may seem logical that there is a causal relationship between bycatch levels and effort, it is unlikely that the causal relationship between bycatch levels of endangered species and effort is so simplistic. During a fishing operation, the maximal number of a species that can be caught is the number of that species exposed to the operation. For endangered species, this is likely to be very few animals, if any. As a fishing operation typically involves deploying thousands of hooks, it seems that the number caught would have a stronger relationship to the number of animals exposed to the operation than to the number of hooks set during the operation. For example, in a randomized experiment where three animals are exposed to a single longline set with 1,000 hooks and, alternatively, a longline set with 1,500 hooks, where the conditions are such that each animal is exposed to numerous hooks, would we really expect there to be a difference in the bycatch levels between 1,000 and 1,500 hooks? Now consider effort summarized over several trips and dozens of sets, then effort gives a measure over a wider area, longer time period, and wider range of fishing practices. At this scale, effort may have a stronger association with bycatch levels and such a simplistic equation could be useful for prediction. However, such a model relies on the stringent assumption that the casual conditions that are captured by effort in the data used to construct the model will be similar in the future. Nevertheless, as predictions of future bycatch based on future projected effort have been requested by management, a simple proportional relationship between fishing effort and bycatch has been imposed.

2.3. **Prohibited Area in Main Hawaiian Islands.** For the main Hawaiian Islands insular false killer whale (IFKW) stock, the predictions assume the main Hawaiian Islands longline fishing prohibited area is that which went into effect on December 31, 2012 (NMFS, 2012*b*) and permanently closed a portion of this area which had previously been seasonally open to longline fishing. The imposed geographic range of the IFKW stock is the waters within 140 km of the main Hawaiian Islands, meaning that any observed false killer whale (FKW) takes outside the 140 km boundary are assumed not to be from the insular stock. Under this assumption, the only zones that remain open to longline fishing where a FKW from the IFKW stock may interact with a longline trip are the zones within the 140 km boundary and outside the longline fishing prohibited area (Figure 1). Herein, the zones that remain open to longline fishing and are within the 140 km boundary are referred to collectively as the “open area”. All predictions of future levels of bycatch and DSI given in this report for the FKW are therefore only for the open area.

2.4. **Effort in the Open Area.** A scenario specific for effort in the open area was not prescribed by PIRO for this analysis. The effort trends within the 140 km boundary and the open area for 2008-2012 given in Table 1 do not match those for the entire deep-set fishery (NMFS, 2013). So for this analysis it is assumed that the fleets effort within the

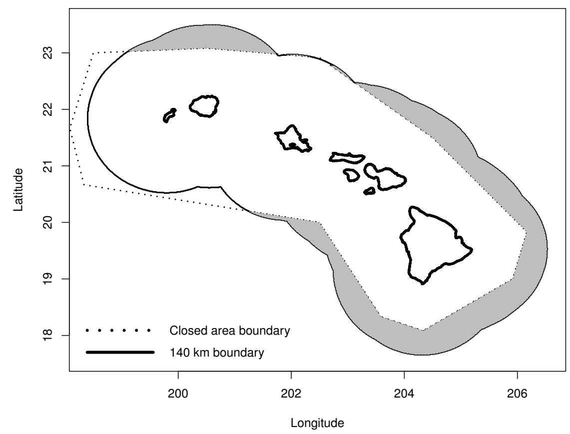

Figure 1. The boundaries for the main Hawaiian Islands insular false killer whale stock (140 km boundary) and the longline fishing prohibited area (closed area). The predictions for future annual bycatch levels for the insular false killer whale stock are for the area that is within the 140 km boundary and outside the closed area boundary; i.e., the open area (shaded in grey).

open area will be similar to that of past years, although a range of effort (from lower to higher) is also explored. As the zones within the open area are small compared to the length of a deployed longline, prior to the regulations that established the longline fishing prohibited area it was not uncommon for a longline set to span a zone and cross the boundaries of the open area. To approximate past effort within the open area, the four points representing the recorded begin and end locations (longitude and latitude) of the gear setting and haulback were used. A set was assumed to have been in an area (e.g., within 140 km boundary or within the open area) if at least one of the four points was within the area, and a trip was assumed to have been in the area if at least one set was considered to be within the area. The number of hooks within a set in an area was approximated as the number of hooks multiplied by the proportion of the four points in the area. For example, if a set had 2,000 hooks and three or the four points (0.75) were within the open area, the approximate effort in the open area was 1,500 hooks. As a set or trip can be assigned to more than one region and hooks cannot, using hooks seems like a more logical measure of effort in this situation. Although annual effort within the open area is expected to fluctuate, predictions within this report assumed a constant effort of 600,000 hooks. The effort of 600,000 hooks is slightly less than the average (635,620 hooks) and greater than the median (499,440 hooks) for the effort given in Table 1. Except for assuming the new (for 2013 and onwards) prohibited area and the future effort in the open area, it is assumed that all fishing and ecological factors that affect bycatch remain at historical levels.

TABLE 1. Hawaii deep-set longline fishing effort in the range of the IFKW stock computed using the Hawaii longline logbook database (https://inport.nmfs.noaa.gov/inport/item/2721). Effort for a trip was assigned to the year the trip arrived back at port. The headings "140km" and "open" refer to the area within the 140 km boundary and the open area, respectively.

year	trips†		sets		hooks	
	140km	open	140km	open	140km	open
2008	259	235	1,085	855	1,893,507	1,014,775
2009	218	193	871	614	1,595,487	750,906
2010	134	118	534	339	1,057,813	433,036
2011	164	138	542	371	1,064,569	499,440
2012	170	138	501	357	928,818	479,944

†Number of trips returning during the year that exerted some effort (sets, hooks) within the open area or 140 km area.

2.5. **Probability of a DSI.** When predicting future levels of bycatch events resulting in a DSI, it is assumed that levels of past classification proportions are representative of future classification proportions. In other words, it is assumed that there will be no changes in the fishery or ecosystem that will affect the probability that a bycatch event will result in a DSI.

3. SAMPLE SELECTION AND DATA STRUCTURE

3.1. **The Two-Stage Probability Sampling Design.** A sample of deep-set longline trips on which scientific observers were deployed was used to develop the models for computing the predictive densities. When developing a Bayesian model, the structure of the data used to construct the model needs to be taken into account. As the data collection method determines the data structure, I now explain how trips were selected for observer deployment.

From mid-year 2002 through mid-year 2012, a probability sampling design was used to randomly select the trips. A probability sample implies that all trips have a probability of being sampled and the sampling probabilities are known or approximated accurately. To reach a balance between obtaining a probability sample and being cost effective, the sampling design used to sample the fleet had a two-stage sampling protocol. This two-stage design accommodated fluctuating coverage levels while utilizing observers efficiently. This flexibility was advantageous as coverage levels typically varied throughout the year because of fluctuations in the fleet's activity level, demands of 100% coverage in the Hawaii longline shallow-set fishery, and an influx of new observers after their completion of NMFS observer training. Additionally, the time between an observer's deployments needed to be minimized as they were not paid during this period. The alternative of paying them while they were waiting to be deployed would have increased the cost of the observer program.

3.1.1. *The First Stage.* The first stage of the sampling protocol was a systematic sample. The sampling frame used to draw the systematic sample was the list of notifications of intended departures of deep-set trips. Before departing on a fishing trip, longline vessels were required to notify the PIRO observer program contractor at least 72 hours prior to their intended departure date. Consequently, the list of notifications is a record of all deep-set longline trips that were undertaken. To enable sample selection, the PIRO contractor numbered notifications sequentially in the order in which they were received. Herein, this assigned number is referred to as the notification number. Prior to the beginning of a quarter, a new systematic sample of notification numbers was drawn by the NOAA Fisheries Pacific Islands Fisheries Science Center (PIFSC) and supplied to the contractor. The trips associated with these selected notification numbers were designated to be sampled.

The systematic sample required having an observer available to be deployed whenever a selected trip was ready to depart. Achieving this requirement under the full targeted coverage, typically 20%, required having enough observers on contract to accommodate higher levels of fleet activity and paying them between deployments. These requirements could not be met under the level of funding; therefore, the systematic sample was drawn at 5% smaller coverage than the targeted level.

3.1.2. *The Second Stage.* The additional trips needed to reach the full targeted level were selected using the secondary sampling protocol. This second stage was used when all trips selected by the systematic sample were already covered and an observer was ready to be deployed. In this instance, a trip was randomly selected with equal probability from the notifications still eligible for observer placement. If more than one observer needed to be assigned, the appropriate number of trips was sampled with equal probability from this pool of notifications. The coverage obtained by this second stage was flexible and dependent on the need to deploy observers. This secondary sampling protocol departs from traditional probability sampling because the days when additional samples were drawn were not randomly selected but determined by the need to deploy observers; consequently, the sampling probabilities are unknown. To approximate the sampling probabilities the contractor's notification records were used. Examination of these records revealed periods of time when coverage appeared to have been greater or less than the full targeted coverage.

3.1.3. *Systematic-Plus Sample.* Herein, this two-stage sampling design is called a systematic-plus design. The samples selected by the systematic sample are referred to as the systematic sample and the samples drawn additional to the systematic sample are referred to as the plus sample. These two samples compose the systematic-plus sample. The systematic-plus sample is not a traditional two-stage sample design as the term "two-stage design" typically refers to a design where primary units are selected during the first stage, and during the second stage, elements within the selected primary units are selected. With the systematic-plus sample, the second stage involves selecting notifications that are not sampled during the first stage.

3.2. **Implementation of the Systematic-Plus Sample.** Now consider a few details concerning the implementation of this design in the deep-set fishery. To allow periodic adjustments to the level of coverage by the systematic sample, a new systematic sample was drawn quarterly. The quarterly samples were drawn at 15% coverage, except for the samples covering the first quarters of years 2005-2009 that were drawn at 10% coverage. The lower coverage ensured observers were available to cover the deep-set systematic sample and all the trips by the shallow-set fleet (typically most active during the first quarter). The lower coverage in the first quarter was offset using the plus sampling protocol throughout the year so that the required annual coverage of at least 20% deep-set trips was achieved. Because a new systematic sample was being drawn each quarter, a year's sample was a stratified sample with a systematic-plus sample within each stratum (quarter).

On 27 August 2012 there was a change in the regulations of the Hawaii longline deep set fishery that imposed new limits on swordfish landed (NMFS, 2012*c*). The new limits are as follows: (1) With a NMFS observer onboard, there is no limit on the number of swordfish landed or possessed on a trip, regardless of the type of hook used; (2) With no NMFS observer onboard, the limit is 25 swordfish landed or possessed on a trip, if the vessel uses only circle hooks; (3) With no NMFS observer onboard, and if the vessel uses any hooks other than circle hooks, the limit is 10 swordfish landed or possessed on a trip. In essence, this regulation created three components of the deep-set fleet defined by the number of swordfish a trip can keep. Regardless what design is used to select notifications, the first component (no limit on swordfish kept) will have 100% coverage and the other two components will have no observer coverage. Prior to the new regulations all deep-set trips had a trip limit of 10 swordfish landed. The regulation limiting number of swordfish landed was put into place to discourage trips from targeting swordfish, which typically implies setting the gear shallow. The shallow setting of gear has historically resulted in different observed catch rates for the protected species. The exclusion of part of the fleet from the sample gives rise to the potential of exclusion bias and places limits on how much information our sample can provide about the population (total effort of the deep-set fleet). Extrapolating from our sample to the population requires making assumptions about the population that cannot be confirmed from the sample.

Although the systematic-plus sample is still used to sample the notifications, the new regulations changed what the sample represents. Prior to the new regulations, the sample was a random sample of all deep-set trips fishing under a uniform set of rules and requirements. Under the new regulations, what is being randomly drawn is a selection of trips that will have an observer onboard and not have a trip limit on swordfish landings.

3.3. **Properties of the Systematic-Plus Sample.** For now ignore the consequences of these new regulations and consider some of the properties of the systematic-plus sample. First, consider just the systematic sample as this sample by itself constitutes a probability sample of the fleet. The systematic sample is a one-stage cluster sample where all elements in the selected clusters are sampled. For example, suppose there are a total of 100 trips and a systematic sample at 20% coverage with 5 starting points is to be drawn from these 100 trips. Using sets of notification numbers, the 25 clusters that define the population are $\{1, 26, 51, 76\}, \{2, 27, 52, 77\}, \ldots, \{25, 50, 75, 100\}$. To draw a sample of 5 clusters, only five starting points (the first notification number in a cluster) between 1 and 25 need to be drawn to define the selected clusters. Hereafter, the clusters defined by a systematic sample will be referred to as the systematic clusters. The systematic samples used to sample the deep-set fleet used simple random sampling without replacement to select 5 systematic clusters. Therefore, the primary sample units are the systematic clusters and the notifications (trips) are the primary elements of the systematic clusters.

A desirable property of the data collection process when using Bayesian inference is ignorabililty. Informally speaking, ignorability means that a model can explain the inclusion process (process that determines if a unit is observed or missed) and that the inclusion process can be ignored after the model accounts for it; that is, conditional on measured covariates, the missing data are not related to the variable of interest. However, the inclusion process does need to be modeled for the purpose of simulating replicated datasets and model checking (Gelman et al., 2004). The simple random sample is a design that is ignorable and known with no covariates. Designs that are ignorable and known given covariates include the stratified design with simple random sampling within strata and the two-stage cluster design with simple random sampling at both stages. With regards to the systematic samples of the deep-set fleet, a year's sample is a stratified sample where within strata a one-stage cluster sample is drawn using simple random sampling. Thus, the design is ignorable given known covariates. For a detailed discussion of ignorability see Gelman et al. (2004) and Little and Rubin (2002).

So far, I've assumed that all selected samples are observed. However, there have been instances when a trip selected by the systematic sample was not observed. The typical reason for not deploying an observer was that no observers were available. When the number of missed samples is small, one can often perform a good analysis assuming that the missing data are ignorable (conditional on fully observed covariates). As the fraction of missing information increases, the ignorability assumption becomes more critical (Gelman et al., 2004). One component of the definition of ignorability states that the data are missing at random. An informal definition of missing at random is that missingness is not related to the data values. In the earlier years of the systematic-plus sample there would be periods where observer coverage fell below the targeted coverage of the systematic sample and several trips selected by the systematic sample were missed. In recent years a small percentage, under 7%, of trips selected by the systematic sample have been missed.

Now consider the plus sample. Standing alone, the plus sample does not necessarily produce a probability sample as there are typically periods when no plus samples are selected so some trips will not have a chance of being selected. In terms of the complete systematic-plus sample, trips are typically selected with unequal probability as a result of the plus sample. For example, trips that have a potential of being selected by the plus sample will have a greater probability of being selected than trips that are never subjected to being drawn by the plus sample. Random sampling with unequal sampling probabilities is ignorable conditional on the covariates that determine the sampling probability, as long as the covariates are fully observed in the general population (Gelman et al., 2004). One possible strategy is to consider the periods that appear to have a constant level of observer coverage as post-stratification. Two obstacles with this strategy are that most post-strata would have no observed bycatch and it is unclear how to model the inclusion process for the purpose of simulating replicated datasets and model checking as the systematic sample would run through the different post-strata. A key idea for Bayesian inference with complicated designs is to include in the model of the sampling distribution (or data distribution) enough explanatory variables such that the design is ignorable. The problem with this strategy is that the information recorded in a vessel's logbook is very limited and bycatch events are so rare that models quickly become over parameterized.

3.4. **Datasets used for Analysis of Bycatch.** Because of the uncertainty on how to proceed using the systematic-plus sample, the plus sample was excluded. The majority of observed bycatches of the species of concern were on trips selected by the systematic sample. As the samples from 2002 through 2007 had long periods where coverage fell below the targeted coverage of the systematic sample, the data from these years were not used. For years 2008 through 2010 there were a few short periods where coverage fell below the targeted coverage of the systematic sample. As there were trips selected by the plus sample near these periods, these trips were substituted for the missed trips in the systematic sample.

In 2008 the coverage during most of the last month of the fourth quarter was below the targeted coverage of the systematic sample so the data were truncated near the beginning of December. During all years, there were trips selected by the systematic sample that were missed even when coverage was above the targeted level of the systematic sample. In these cases, trips selected by the plus sample that were near these missed samples were used as substitutes. The percentage of notifications selected by the systematic sample that were missed was 3% for year 2008 (excluding notifications in December), 6% for year 2009, and only 1% to 0% for the years 2010 through 2012. As explained above, mid-year 2012 a change in regulations for the fishery resulted in part of the deep-set fishery not being subject to selection. Nevertheless, these data were included and treated as if representative of the two components of the fleet not sampled.

As there were so few observed bycatch events for the species of concern, stringent distributional assumptions were made when modeling the data. With these stringent assumptions it would be advantageous to have a sample that would be robust to model misspecification. A self-weighting sample would be preferred as it can provide robustness to some forms of model misspecification (Lohr, 2010). A self-weighting sample is a sample in which the probability of inclusion (probability the unit is included in the sample) are equivalent. If each quarter had a systematic sample at 15% coverage the sample would be self-weighting as each systematic cluster, and thus each trip, would have a 15% chance of being included in the sample. Therefore, to create a self-weighting sample, we can exclude the three quarters with only 10% coverage or use the plus samples to create a psuedo systematic sample at 15% coverage. As each of the three quarters had sufficient plus samples to create a psuedo systematic sample at 15% coverage, three pseudo systematic samples were created. This was achieved by adding trips selected by the plus sample within the same stratum to each sampled systematic cluster so that it contained the number of trips that would have been included had the systematic sample been drawn at 15% coverage. As the trips in the plus sample were generally spread out throughout the first quarter, the adjusted systematic clusters should approximate systematic clusters at 15% coverage. With regards to predicting bycatch, this adjusted systematic sample should provide some level of robustness to temporal and similar fluctuations of bycatch rates. Herein, this dataset will be referred to as the adjusted systematic sample and the dataset that includes the unadjusted systematic samples will be referred to as the systematic sample. The effort observed for both of these samples is provided in Table 2.

3.5. **Additional Datasets used for Analysis of DSI.** For inference concerning the future annual predicted levels of DSI, the injuries of observed bycaught marine mammals during 2007-2011 were classified under the current policy directive for classification (NMFS, 2012*a*). The observed bycatches of sperm and humpback whales prior to 2007 were also classified under the current policy directive. The sperm, humpback, and Bryde's whale observations in the deep-set and shallow-set fisheries were pooled for analysis related to DSI. As these three species are not known to depredate catch or bait and their observed interactions have been entanglements, pooling their injury conditions likely introduces little, if any, bias; whereas, without pooling, the predictions of DSI would likely be less accurate as a result of each species having only a couple of observed interactions. For inference concerning FKW, injury classifications of blackfish were also considered. A blackfish is a cetacean that is either a false killer whale or a short-finned pilot whale but which of these two is unknown. FKW and blackfish observed prior to 2007 and classified under the previous policy directive for

Table 2. Observed effort in the Hawaii deep-set longline fishery computed using the Pacific Islands Region longline observer database (http://ias.pifsc.noaa.gov/lds/lods.html). The effort is given for the systematic sample and the adjusted systematic sample (adj.systematic). Trips are assigned to the year when their notification was sampled.

year	systematic			adj.systematic		
	trips	sets	hooks	trips	sets	hooks
2008	166	2,295	5,140,728	184	2,504	5,593,824
2009	179	2,518	5,735,715	194	2,733	6,185,006
2010	152	2,109	4,738,414	167	2,290	5,123,772
2011	194	2,698	6,359,670	194	2,698	6,359,670
2012	201	2,720	6,540,941	201	2,720	6,540,941

Table 3. Observed counts in categories of injury classifications, where large cetaceans include the sperm whale, humpback whale, and Bryde's whale. The categories represented are DSI (dead or serious injury), NSI (non-serious injury), and PR (prorated).

2007-2011			pre-2007		
DSI	NSI	PR	DSI	NSI	PR
large cetaceans					
0	1	2	0	2	3
false killer whales					
20	2	-	20	1	-
blackfish					
5	1	-	5	0	-

classification were used as prior information. Table 3 provides the counts in each category for observed bycatches prior to 2007 and after 2007. The 2012 marine mammal observed interactions were not classified prior to this report. In 2012 there were no observed bycatches of humpback or sperm whales but there were observed bycatches of FKW and unidentified marine mammals.

4. BAYESIAN INFERENCE AND PREDICTION MODELS

The first part of this section provides a brief introduction to Bayesian inference. Following this introduction, the different models used to compute the predictive probability densities are defined.

4.1. **Bayesian Inference.** A Bayesian approach to inference starts with the formulation of a model that we hope is adequate to describe the situation of interest. Specifically, the model provides a joint probability density of the variable of interest y and the unknown parameters θ of the data density. This joint density can be written as a product of two densities that are commonly referred to as the prior density $g(\theta)$ and the data density $f(y \mid \theta)$. The intent of the prior density $g(\theta)$ is to capture our knowledge or beliefs about these parameters before observing the data. In the absence of prior knowledge about a parameter, or if we are interested in letting the data speak for themselves without the influence of our prior knowledge, a noninformative prior is chosen.

The prior and data densities are all the ingredients needed for Bayesian analysis as Bayes' rule combines $g(\theta)$ and $f(y \mid \theta)$ to produce the posterior density of θ:

$$g(\theta \mid y) = \frac{f(y \mid \theta)g(\theta)}{f(y)}, \tag{1}$$

where $f(y)$ is the marginal density of the data. As $f(y)$ is the normalizing constant, the posterior density of θ can be characterized without computing $f(y)$. That is, Equation 1 can be expressed as

$$g(\theta \mid y) \propto f(y \mid \theta)g(\theta).$$

This simple expression of the posterior density is at the core of Bayesian inference.

To compute the posterior density, the likelihood of the data distribution $L(Y \mid \theta)$ is computed after observing some data Y, where Y represents a vector of observed values $(y_1, y_2, \ldots, y_n)$. Given the likelihood and the prior, the posterior density of θ is completely determined by the unique distribution function proportional to the product of these two: $g(\theta \mid Y) \propto L(Y \mid \theta)g(\theta)$. Once this expression is formulated, the next step is to perform the necessary computations to summarize the posterior density in appropriate ways.

While we do not need the normalizing constant $f(Y) = \int g(\theta)L(Y \mid \theta)d\theta$ to define the posterior density, we do need to know it if we want to calculate details of the posterior density, such as the mean, variance, and percentiles. For the posterior densities evaluated in this report either evaluating the normalizing constant was not needed or it was straightforward.

All Bayesian inference is based on posterior densities, including predictions of hypothetical data, such as future annual bycatch levels. Specifically, the predictive density is the posterior density of hypothetical replicates of the data, $\tilde{y}$, generated under the same conditions as produced y. Although the exact values of the parameters are unknown, we have their posterior densities that express what we believe the parameters to be given the data we've already seen. The predictive density averages all of the various predictive probabilities over the different possible parameter values, weighting them according to how strongly we believe in them. Formally, the predictive density of $\tilde{y}$ is given by

$$f(\tilde{y}) = \int f(\tilde{y} \mid \theta)g(\theta)d\theta.$$

If $g(\theta)$ is a prior density, $f(\tilde{y})$ is referred to as the prior predictive density, and if $g(\theta)$ is a posterior density, $f(\tilde{y})$ is referred to as the posterior predictive density.

If the intent of this report were to quantify (predict) the bycatch of the unobserved portion of the fleet for past years, the Bayesian approach to finite population modeling would be to treat the unobserved portion of the fleet as missing data. As observed data from past years are available to compute the posterior density, the posterior predictive density would be used to predict the bycatch of the unobserved portion of the fleet. The intent of this report is to quantify the future annual levels of bycatch. Under the Bayesian framework, a convenient source of prior information is the observed samples from the past years that are assumed to be generated under the same conditions as expected in the future. As we have yet to sample the fleet in future years, it is the prior predictive density that is computed.

One can check the adequacy of the proposed model by inspecting the posterior predictive density. This can be done by locating the observed data values in the predictive density. If very few of the observed values are located in the tails of the predictive density, then the model seems reasonable. On the other hand, an observed value in the extreme tail portion of the predictive density casts doubt on the adequacy of the model.

4.2. **Models for Predicting Bycatch.** To estimate the future bycatch based on the assumed future effort, the unknown parameters of interest are the bycatch rates (quantity of bycatch per unit effort). The first model makes the assumption that the bycatch events for the systematic clusters are independent Poisson variates whose mean is the cluster's effort multiplied by a common bycatch rate, the unknown parameter. The second model goes a step further and makes the stringent assumption that the bycatch for the trips are independent identically distributed (iid) Poisson variates, implying a common bycatch rate per trip. These two models are simple models that make assumptions concerning a common bycatch rate throughout the past data and into the future. Based on the assumption of a common bycatch rate, the covariates defining the strata can be dropped for the purpose of calculating the posterior density. To add some robustness to the violation of a common bycatch rate, the adjusted systematic dataset was used. The third model allows the bycatch rate to vary between strata. All computations required to predict bycatch were done in the statistical software program R (R Core Team, 2013). Several functions available in the LearnBayes R library (Albert, 2009) were used for model construction and diagnostics.

4.2.1. *Notation.* Prior to defining the models, some notation needs to be defined. Hereafter, let y_{ij} represent the observed number of bycatch events in the j^{th} systematic cluster of the i^{th} stratum and let e_{ij} represent the total effort for this systematic cluster. Additionally, let $y_{i+} = \sum_{j=1}^{5} y_{ij}$ (the total observed bycatch for a stratum) and $e_{i+} = \sum_{j=1}^{5} e_{ij}$. Furthermore, let y_{yr} and e_{yr} represent the observed bycatch and effort, respectively, within year yr for $yr = 2008, \ldots, 2012$; and let $\tilde{y}$ represent an unknown hypothetical value of bycatch from a unit with known effort $\tilde{e}$. As the objective is to predict the future annual level of bycatch, denote the annual level of bycatch as τ and its predicted value as $\tilde{\tau}$. Denote the projected future annual effort as $\tilde{e}_{\tau}$. At times $\tilde{\tau}$ will have a subscript identifying the species of the prediction.

4.2.2. *Poisson/Gamma Model.* The first model assumes that the y_{ij}s are independent Poisson$(e_{ij}\lambda)$ variates for $i = 1, \ldots, 20$ and $j = 1, 2, 3, 4, 5$. The common bycatch rate per unit of effort λ is assigned a standard noninformative prior of the form

$$g(\lambda) \propto \frac{1}{\lambda}.$$

The posterior density of λ is given by

$$g(\lambda \mid y) \propto \lambda^{\sum_{i=1}^{20} \sum_{j=1}^{5} (y_{ij} - 1)} \exp\left(-\lambda \sum_{i=1}^{20} \sum_{j=1}^{5} e_{ij} \right),$$

which is the gamma distribution with parameters $\alpha = \sum_{i=1}^{20} \sum_{j=1}^{5} y_{ij}$ and $\beta = \sum_{i=1}^{20} \sum_{j=1}^{5} e_{ij}$ (Albert, 2009). The predictive density is given by

$$f(\tilde{y} \mid \tilde{e}) = \int f_P(\tilde{y} \mid \tilde{e}\lambda) g(\lambda) d\lambda,$$

where $f_P(\tilde{y} \mid \tilde{e}\lambda)$ is the Poisson data distribution with mean $\tilde{e}\lambda$.

To simulate the posterior predictive density of $\tilde{y}$, we first simulate a large number of draws of λ from its gamma(α, β) posterior density and then, conditional on λ, simulate draws of $\tilde{y}$ from a Poisson distribution with mean $\tilde{e}\lambda$. When computing future predictions, the gamma(α, β) posterior density becomes the updated prior density. Using this updated prior gamma density, the prior predictive density is generated in a similar manner as the posterior density. Of interest is the prior predictive density of annual bycatch in future years. Under the independence and common bycatch rate assumption, $\tilde{\tau} \sim$ Poisson$(\lambda\tilde{e}_\tau)$; therefore, the prior predictive density of annual bycatch can be generated by substituting $\tilde{e}_\tau$ into the algorithm above. As the predictive density of bycatch is discrete, the predictive probabilities can be computed by computing the frequency of each predictive value of bycatch and dividing these by the number of predictive values generated. Hereafter, this model will be referred to as the Poisson/gamma model.

4.2.3. *Conditional Binomial Model.* The second model was developed to handle the situation where no bycatch events were observed during 2008-2012, i.e., $\sum_{yr=2008}^{2012} y_{yr} = 0$. Let y_t represent a trip's bycatch. This model assumes that the y_ts for all trips completed by the fleet over the period of interest are iid Poisson(λ) variates. Now, let τ_{obs} represent the total observed bycatch, τ_{unobs} the total unobserved bycatch, and τ the true total bycatch, where $\tau = \tau_{obs} + \tau_{unobs}$. Furthermore, let n_t represent the number of trips observed and N_t the number of trips completed during the sampling period. Under the iid Poisson assumption, $\tau_{obs} \sim$ Poisson$(n_t\lambda)$ and $\tau \sim$ Poisson$(N_t\lambda)$. It follows that $(\tau_{obs} \mid \tau) \sim$ binomial$(n = \tau, p = n_t/N_t)$. With this reformulation, τ is now the unknown parameter in the data distribution. Assign τ a discrete uniform distribution $[0, \tau_{max}]$ prior, where τ_{max} is what we believe is the maximal value of τ. The posterior density of $g(\tau \mid \tau_{obs})$ is

$$g(\tau \mid \tau_{obs}) = \frac{g(\tau_{obs} \mid \tau)}{\sum_{\tau=0}^{\tau_{max}} g(\tau_{obs} \mid \tau)},$$

where

$$g(\tau_{obs} \mid \tau) = \binom{\tau}{\tau_{obs}} \left(\frac{n_t}{N_t}\right)^{\tau_{obs}} \left(1 - \frac{n_t}{N_t}\right)^{\tau_{unobs}}.$$

For predicting future levels of bycatch, let $\tilde{N}_t$ represent the projected effort expressed as the number of trips. Since $\tilde{N}_t < N_t$ the predictive density of $\tilde{\tau}$ is given by

$$f(\tilde{\tau} \mid \tilde{N}_t) = \sum f_B(\tilde{\tau} \mid \tau, \tilde{N}_t, N_t) g(\tau \mid \tau_{obs}),$$

where $f_B(\tilde{\tau} \mid \tau, \tilde{N}_t, N_t)$ is the binomial density

$$\binom{\tau}{\tilde{\tau}} \left(\frac{\tilde{N}_t}{N_t}\right)^{\tilde{\tau}} \left(1 - \frac{\tilde{N}_t}{N_t}\right)^{\tau - \tilde{\tau}}. \tag{2}$$

Random draws from $f(\tilde{\tau} \mid \tilde{N}_t)$ are generated by simulating draws of τ from $g(\tau \mid \tau_{obs})$ and then simulating draws of $\tilde{\tau}$ from a binomial distribution conditional on parameter τ (Equation 2). Hereafter, this model will be referred to as the conditional binomial model.

4.2.4. *Poisson/Gamma Exchangeability Model.* Unlike the first two models that assume a common bycatch rate per unit of effort, the third model assumes that there is random fluctuation in the true bycatch rate. This model simultaneously estimates the true bycatch rate λ_i for each of the 20 strata and is developed in Albert (2009) and recreated here in terms of our situation. Under the assumption that bycatch rates are equal across strata such that $y_{i+} \sim \text{Poisson}(e_{i+}\lambda)$ for $i = 1, \ldots, 20$, the maximum likelihood estimate of λ is

$$\frac{\sum_{i=1}^{20} y_{i+}}{\sum_{i=1}^{20} e_{i+}}.$$

This pooled estimate is based on the stringent assumption that the true bycatch rate is the same across the strata and into the future. An alternative is to simply estimate the true rates within each stratum by using the individual observed bycatch rates,

$$\frac{y_{1+}}{e_{1+}}, \ldots, \frac{y_{20+}}{e_{20+}}.$$

Unfortunately, as bycatch is so rare there are typically very few, if any, observed bycatches within a stratum, consequently, the observed rates can be poor estimates. This is especially apparent when there are no observed bycatch events and the estimated bycatch rate would be 0, likely underestimating the true bycatch rate within the stratum. A third possibility is the compromise estimate

$$(1-\beta)\frac{y_{i+}}{e_{i+}} + \beta\frac{\sum_{i=1}^{20} y_{i+}}{\sum_{i=1}^{20} e_{i+}}. \tag{3}$$

This estimate shrinks or moves the individual estimates y_{i+}/e_{i+} toward the pooled estimate $\sum y_{i+}/\sum e_{i+}$, where the parameter $0 < \beta < 1$ determines the size of the shrinkage. This shrinkage estimate is a natural by-product of the application of an exchangeable prior model to the true bycatch rates.

As the name implies, this model assumes that the true bycatch rates $\lambda_1, \ldots, \lambda_{20}$ are exchangeable. Informally, exchangeable means that the stratum labels are interchangeable. An example where this assumption would not be appropriate is when there is a seasonal pattern such that the true bycatch rate in the first quarter $(\lambda_1, \lambda_5, \lambda_9, \lambda_{13})$ is different than the bycatch rate from the other quarters. In this circumstance, the stratum labels from the first quarter could not be interchanged with the other quarters. If it is not appropriate to assume exchangeability across quarters, in theory, the exchangeable model could be used within quarters. For the two species where this model is used, the exchangeablility assumption appears to be reasonable.

Proceeding with the assumption that a stratum's observed bycatch follows a Poisson distribution with mean $e_{i+}\lambda_i$. The prior density is constructed in a hierarchical fashion. At the first stage of the prior, the true bycatch rates $\lambda_1, \ldots, \lambda_{20}$ are assumed to be a random sample from a gamma$(\alpha, \alpha/\mu)$ distribution of the form

$$g(\lambda \mid \alpha, \mu) = \frac{(\alpha/\mu)^{\alpha}\lambda^{\alpha-1}\exp(-\alpha\lambda/\mu)}{\Gamma(\alpha)}, \lambda > 0.$$

The prior mean and variance of λ are given by μ and μ^2/α, respectively. At the second stage of the prior, the hyperparameters μ and α are assumed independent. The parameter μ is assigned a vague prior of the form

$$g(\mu) \propto \frac{1}{\mu}, \mu > 0,$$

and the parameter α is assigned a relatively flat prior of the form

$$g(\alpha) = \frac{z_0}{(\alpha + z_0)^2}, \alpha > 0.$$

The parameter z_0 is the median of α and needs to be specified.

The posterior density of λ_i is gamma$(y_{i+} + \alpha, e_{i+} + \alpha/\mu)$. The posterior mean of λ_i, conditional on α and μ, is

$$E(\lambda_i \mid y_{i+}, \alpha, \mu) = \frac{y_{i+} + \alpha}{e_{i+} + \alpha/\mu} = (1 - B_i)\frac{y_{i+}}{e_{i+}} + B_i\mu, \tag{4}$$

where

$$B_i = \frac{\alpha}{\alpha + e_{i+}\mu}.$$

As Equations 4 and 3 have the same form, we can interpret the posterior mean of the true rate λ_i as a shrinkage estimator, where B_i is the shrinkage fraction of the posterior mean away from the usual estimate y_{i+}/e_{i+}, toward the prior mean μ.

The marginal posterior density of (α, μ) is

$$p(\alpha, \mu \mid \text{data}) = K \frac{1}{\Gamma^{20}(\alpha)} \prod_{i=1}^{20} \left[\frac{(\alpha/\mu)^\alpha \Gamma(\alpha + y_{i+})}{(\alpha/\mu + e_{i+})^{(\alpha + y_{i+})}} \right] \frac{z_0}{(\alpha + z_0)^2} \frac{1}{\mu},$$

where K is a proportionality constant.

Random draws from the predictive density are simulated as follows:

- Simulate draws of (μ, α) from the marginal density of the hyperparameters μ and α.
- Simulate draws of λ_i from its gamma distribution conditional on the values of μ and α.
- Simulate draws of a quarter's (stratum's) bycatch from its Poisson distribution conditional on the value of λ_i.

Random draws from the predictive density for a year's bycatch require simulating draws from each quarter and then adding these across quarters. When generating the predictive density of future annual bycatch levels, a quarter's effort was a fourth of the year's effort.

To simulate draws from the marginal density of the hyperparameters, I used the R functions *poissgamexch* and *gibbs* in the LearnBayes package (Albert, 2009). The *gibbs* function uses Gibbs sampling to sample from the distribution of interest. An explanation of Gibbs sampling is beyond the scope of this report; see Albert (2009) and Gelman et al. (2004) for details on Gibbs sampling. For a more detailed discussion of this model see Albert (2009), and for more information concerning Bayesian hierarchical models see Gelman et al. (2004). Hereafter, this model will be referred to as the Poisson/gamma exchangeability model.

4.3. **Models for Predicting the Number of Bycatch that Result in a DSI.** This section applies only to predictions concerning marine mammals. Let $\tilde{d}$ represent an unknown hypothetical value of DSI. As the objective is to predict the future annual levels of bycatch events resulting in DSI, denote the annual level of DSI as δ and its predicted value as $\tilde{\delta}$.

4.3.1. *Predicting δ for Humpback and Sperm Whales.* First consider predicting the number of humpback and sperm whale bycatch events that result in a DSI. The three injury classifications of interest for predicting future levels of DSI are (1) death or serious injury (DSI), (2) non-serious injury (NSI), and (3) a prorated category where an injury has a probability ψ of being a DSI (PR). The assumption is made that bycatch events (past and future) have common probabilities θ_1,θ_2, and θ_3 of being classified in these three categories, respectively. Let m_1, m_2, and m_3 represent the observed count in each category and $M = (m_1, m_2, m_3)$ the vector of these three counts. The model used to predict future levels of δ begins with assuming that the data distribution of M is the multinomial distribution with sample size $m_+ = \sum_{i=1}^{3} m_i$ and respective probabilities θ_1,θ_2, and θ_3, where $\sum_{i=1}^{3} \theta_i = 1$. Based on these assumptions

$$f(M \mid \theta) \propto \prod_{i=1}^{3} \theta_i^{m_i}.$$

The vector $\theta = (\theta_1, \theta_2, \theta_3)$ is assigned the uniform prior distribution (assigns equal probability to any vector θ satisfying $\sum_{i=1}^{3} \theta_i = 1$). The resulting posterior density of θ is proportional to

$$g(\theta) = \theta_1^{m_1} \theta_2^{m_2} \theta_3^{m_3}$$

(Gelman et al., 2004), which is known as the Dirichlet distribution with parameters $(m_1 + 1, m_2+1, m_3+1)$. To simulate draws from this Dirichlet distribution the R function *rdirichlet* in the package LearnBayes (Albert, 2009) was used. For a more extensive discussion on the multinomial model and the Dirichlet distribution see Gelman et al. (2004).

For the next step let ψ be the probability that a bycatch event that is classified as PR is from the DSI category. Assuming that θ and ψ are independent, the probability of interest is $\rho = \theta_2 + \theta_3\psi$. To simulate draws from the posterior density of ρ, we need to simulate draws from the posterior density of ψ to combine with the simulated draws from θ. In the policy document that provided guidelines for classifying injuries (NMFS, 2012*a*), the value of ψ was determined to be .75 based on a sample of 114 documented entanglement events of which 85 were determined to be a serious injury. It appears that this probability was derived assuming these events were iid binomially distributed. As these data were collected over different fisheries, this assumption presumes that ψ does not fluctuate spatially, temporally, or between fisheries and species. I proceed using this imposed assumption. To formulate the posterior density of ψ, assume

$$f(x \mid \psi) = \binom{n}{x} \psi^x (1-\psi)^{n-x},$$

where n is the number of sampled entanglements and x is the number of these determined to be DSI, and assign ψ a relatively flat beta(2.07,2.07) prior. The resulting posterior density of ψ is the beta distribution with parameters $2.07 + x$ and $2.07 + (n - x)$.

To simulate draws from the posterior distribution of ρ, simulated draws from the Dirichlet posterior density of θ are combined with simulated draws from the beta posterior density of ψ using the equation $\theta_2 + \theta_3\psi$.

At this junction, the methods for constructing $g(\rho)$ and $f(\tilde{y})$ (Section 4.2) have been described. The next step is to formulate the predictive density of $f(\tilde{d})$. Assume ρ and $\tilde{y}$ are independent, the predictive density of $\tilde{d}$ is given by

$$f(\tilde{d}) = \int f_B(\tilde{d} \mid \tilde{y}, \rho) f(\tilde{y}) g(\rho) d\tilde{y} d\rho,$$

where

$$f_B(\tilde{d} \mid \tilde{y}, \rho) = \binom{\tilde{y}}{\tilde{d}} \rho^{\tilde{d}} (1-\rho)^{\tilde{y}-\tilde{d}}.$$

For $\tilde{\delta}$, the prior density of ρ is updated with the posterior density $g(\rho)$ and $f(\tilde{y})$ is the prior predictive density $f(\tilde{\tau})$. Random draws from the predictive density of $\tilde{\delta}$ are simulated as follows:

- Simulate draws of $\tilde{\tau}$ from its predictive density.
- Simulate draws of ρ from its updated prior density.
- Simulate draws from a binomial distribution with parameters conditional on $\tilde{\tau}$ and ρ.

4.3.2. *Predicting δ for FKW.* Now consider formulating the predictive density of $\tilde{\delta}$ for FKW. For FKW the third injury category does not apply, thus the injury classifications can be collapsed into the 2 categories (1) DSI and (2) NSI. The predictive density is formulated in a similar manner as for the humpback and sperm whales but without the third category. The assumed data distribution of M is the binomial distribution with sample size $m_+ = \sum_{i=1}^{2} m_i$ and probability of a DSI θ: $f(M \mid \theta) \propto \theta^{m_1}(1-\theta)^{m_2}$.

Although there were changes in the guidelines for injury determinations, they seem to have had little effect on the probability that a FKW bycatch event results in a DSI. Therefore, the classifications of observed FKW bycatches prior to 2007 are used to develop a prior density. The prior density of θ is assigned the form beta$(s+1, f+1)$, where s is the number of DSI and f is the number of NSI prior to 2007. Based on this formulation of the prior, the posterior density of θ is beta$(s+m_1+1, f+m_2+1)$.

Under the assumption that $\tilde{y}$ and θ are independent, $f(\tilde{d}) = \int f_B(\tilde{d} \mid \tilde{y}, \theta) f(\tilde{y}) g(\theta) d\tilde{y} d\theta$, where

$$f_B(\tilde{d} \mid \tilde{y}, \theta) = \binom{\tilde{y}}{\tilde{d}} \theta^{\tilde{d}} (1-\theta)^{\tilde{y}-\tilde{d}}.$$

For $\tilde{\delta}_{FKW}$, the prior density of θ is updated with the posterior density of $g(\theta)$ and $f(\tilde{y})$ is the prior predictive density $f(\tilde{\tau})$ (Section 4.2). Random draws from the predictive density of $\tilde{\delta}_{FKW}$ are simulated as follows:

- Simulate draws of $\tilde{\tau}_{FKW}$ from its predictive density.
- Simulate draws of θ from its updated prior density.
- Simulate draws from a binomial distribution conditional on parameters $\tilde{\tau}_{FKW}$ and θ.

4.4. **Model for Determining if a FKW Bycatch is from the IFKW Population.** At this juncture, the methods for constructing the predictive densities of $\tilde{\tau}_{FKW}$ and $\tilde{\delta}_{FKW}$ have been described, and to achieve the objective of predicting $\tilde{\tau}$ and $\tilde{\delta}$ for the IFKW, the density of the probability that a FKW caught in the open area is from the IFKW stock needs to be constructed. Herein, let ι denote this probability. A database of sufficient size where FKW bycaught inside the 140 km boundary are identified as from the IFKW stock, or another stock, does not exist. Instead, the mathematical model that NOAA Fisheries has decided upon (Carretta et al., 2011) and information on the spatial distribution of effort within the open area are used to create a prior density of ι. The mathematical model used to determine ι is a function of the minimal distance to shore from the location where the animal is first observed bycaught and the abundance estimates of the IFKW and Hawaii FKW pelagic stocks. McCracken (2010) describes this mathematical model and its assumptions and compares it to other models that were considered. For the present analysis, this mathematical model was updated with the most current abundance estimates. The old FKW stock abundance estimates were 123 individuals in the IFKW stock and 484 individuals in the Hawaii pelagic stock within the U.S. EEZ surrounding Hawaii (Carretta et al., 2010). The updated abundance estimates are 151 individuals in the IFKW stock and 1,503 individuals in the Hawaii pelagic stock within the U.S. EEZ surrounding Hawaii (Bradford et al., 2012). The mathematical model is described by the logistic curves shown in Figure 2. As the mathematical model is dependent on the capture locations, a distribution of these capture locations is required. Because there have been no confirmed captures inside the open area, it was assumed that the distribution of capture locations resembles the distribution of effort within the open area. Unfortunately, constructing the distribution of effort is not straightforward as a longline will typically span over a mile and can traverse the boundaries of the open area. To approximate the distribution of effort, the set and haul begin and end locations of fishing operations that occurred in this open area were used. Using the set and haul locations from the 2008-2012 observer data within the open area, the minimal distance from each point to shore, and subsequently ι, were computed. The collection of these probabilities, rounded to the third decimal, was used to create an empirical probability density of ι. Figure 3 shows the set and haul begin and end locations that occurred in the open area. Figure 4 is a graphical depiction of the empirical density of ι.

To predict future levels of IFKW bycatch, τ_{IFKW}, this empirical density is used as the prior density of ι. Given that a FKW has been bycaught, let $z = 1$ if it is from the IFKW stock and $z = 0$ if it is from the pelagic stock. Assuming that $\tilde{\tau}_{FKW}$ (predicted annual bycatch of FKW in the open area) and ι are independent, the predictive density of $\tilde{\tau}_{IFKW}$ is given by

$$f(\tilde{\tau}_{IFKW}) = \int f(\tilde{\tau}_{IFKW} \mid \tilde{\tau}_{FKW}, \iota) f(\tilde{\tau}_{FKW}) g(\iota) d\tilde{\tau}_{FKW} d\iota,$$

where

$$f(\tilde{\tau}_{IFKW} \mid \tilde{\tau}_{FKW}, \iota) = \sum_{i=1}^{\tilde{\tau}_{FKW}} I(z_i = 1)$$

and $z_i \sim \text{Bernoulli}(\iota_i)$. Random draws from the predictive density of $\tilde{\tau}_{IFKW}$ are simulated as follows:

- Simulate draws of $\tilde{\tau}_{FKW}$ from its predictive density.
- Simulate $\tilde{\tau}_{FKW}$ draws from a Bernoulli distribution conditional on ι, where a new random draw of ι from its prior density is made for each Bernoulli draw.
- The sum of the Bernoulli variates is $\tilde{\tau}_{IFKW}$.

4.5. **Simulating Random Draws from the Predictive Density of $\tilde{\delta}_{IFKW}$.** At this juncture, all the densities necessary to create the predictive densities of $\tilde{\delta}_{IFKW}$ have been established. The predictive density for $\tilde{\delta}_{IFKW}$ follows the same formulation as the predictive density of $\tilde{\delta}_{FKW}$ (Section 4.3.2) with $\tilde{\tau}_{IFKW}$ substituted for $\tilde{\tau}_{FKW}$. Random draws from the predictive density of $\tilde{\delta}_{IFKW}$ are simulated as follows.

- Simulate draws of $\tilde{\tau}_{IFKW}$ from its predictive density (preceding algorithm).
- Simulate draws of θ from its updated prior density (Section 4.3.2).

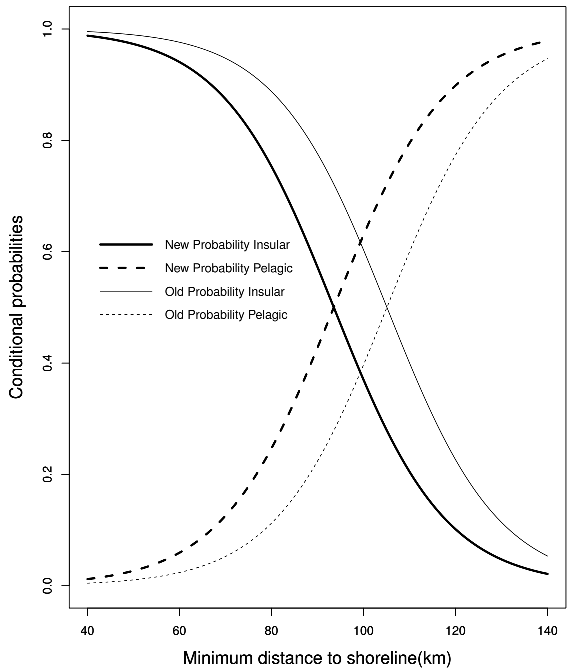

Figure 2. New and old conditional probabilities that a bycaught false killer whale is from the main Hawaiian Islands insular stock or pelagic stock.

Figure 3. Begin and end locations of observed sets and hauls in open area.

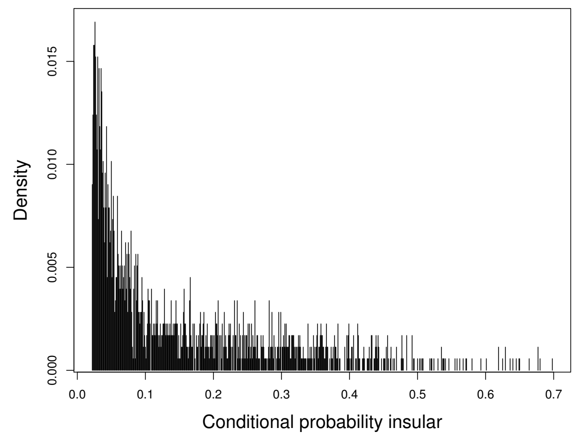

FIGURE 4. The empirical probability density function of the probability that a FKW bycaught in the open area is from the IFKW stock (ι).

- Simulate draws of $\tilde{\delta}_{IFKW}$ from a binomial distribution conditional on parameters $\tilde{\tau}_{IFKW}$ and θ.

5. PREDICTIONS

The predictive density of the annual future level of bycatch is presented for all seven species in this section. Prior to presenting the results by species, details concerning model fitting, model diagnostics, and a summary of the results that are common between species are presented.

5.1. **Statistical Details.**

5.1.1. *Measures of Effort.* The measures of effort considered were number of trips, number of sets, number of hooks deployed, and functions of these variables of effort. One such variable was

$$e_i^\star = \frac{m_i}{M_i} \frac{N.hooks_i}{(\sum_{i=1}^{20} N.hooks_i/20)},$$

where for stratum i, m_i was the number of clusters sampled, M_i was the total number of clusters, and $N.hooks_i$ was the total number of hooks deployed. The first part of this equation represents the sampling fraction, and when using the adjusted systematic sample, m_i/M_i is constant and plays no role. When predicting future levels of τ, $m_i/M_i = 1$ and $N.hooks_i$ is the projected number of hooks. Most predictions were not sensitive to the measure of effort modeled.

To make predictions based on the projected effort of 1,305 trips and ignore that the effort within a trip is projected to increase compared to the effort within a trip in 2012 seems naive. Still, there are times when using the number of trips as the measure of effort may be advantegous. For example, when there has been no observed bycatch between 2008 and 2012 using the conditional binomial model may be advantageous, however, this model uses trips as the measure of effort (Section 4.2.3). To assume unconditionally that the increase in effort within a trip will not increase the bycatch rate per trip may result in the under-prediction of future levels of bycatch. Therefore, adjusting the projected number of trips upward to account for the increase effort seems advisable if under-prediction is a concern. How best to adjust the number of projected trips upward is unclear.

In this report, we are concerned with how the projected future levels of effort and composition of trips compare to the effort data from the 2008-2012 period that provided bycatch data used to construct the predictive models. The recorded annual effort based on year of landing (arrival in port and catch is offloaded) as it is usually defined for annual bycatch estimation for years 2008 through 2012 is given in Table 4[3]. If we compare the projected future effort to the 2008-2012 average annual effort, the increase to 18,592 sets is a 8.1% increase in sets, and to 46,117,532 hooks is a 15.5% increase in hooks. Thus, we might consider adjusting 1,305 by 8.1%, 1,411 trips, or 15.5%, 1,507 trips. Alternatively, we could first adjust the projected number of sets for the increased number of hooks per set and then adjust the number of trips by the adjusted number of sets per trip. Specifically, the adjusted number of sets is the projected number of sets multiplied by the ratio of the projected hooks per set to the five year average hooks per set, 18,592(2,480.5/2,321.3)= 19,867 sets. Then, the adjusted number of trips is the adjusted number of sets divided by the five year average sets per trip, (19,867/13.677)=1,453 trips. The figure of 1,453 trips is very similar to the product of the projected number of trips times the projected hooks per trip divided by the average hooks per trips, 1,305(35,339/31,754)=1,452 trips. Herein, to distinguish from the projected number of 1,305 trips, the adjusted number of projected trips will be referred to as the effective number of projected trips.

Since management requested one value of the average predicted bycatch, when using the conditional binomial model appeared advantageous, 1,453 effective projected trips is used to compute the predictive distribution for future annual levels of bycatch. Furthermore, in this section, 1,453 effective projected trips is used whenever future effort in terms of number of trips is required. Because it is unclear how to derive the effective number of projected trips, a sensitivity analysis that evaluated the mean of the predictive distribution under different levels of projected effort was conducted. This study is presented in Appendix A. As expected, the distribution of the mean of the predictive distribution shifts upward but very gradually as a result of bycatch being so rare (Appendix A).

5.1.2. *General Details of Generating Predictive Distributions.* Whenever simulated draws from a distribution were generated, 10,000 draws were simulated. For the Poisson/gamma exchangeability model, the first 2,000 draws from the marginal density of (μ, α) were discarded as is commonly done when using Gibbs sampling. As all other distributions had a familiar distribution, draws from these distributions could be simulated using the R random variate function for the distribution.

[3]Effort in this table is slightly different from the effort table in the BE as a result of a different rules for assigning effort to the year.

TABLE 4. Hawaii deep-set longline fishery annual effort in recent years with trips and components of effort (number of sets, total number of hooks deployed, and functions of these values) assigned to the year of landing (date the trip arrived back at port and offloaded its catch). Included is the average annual effort over 2008-2012 and the projected future effort.

year	trips	sets	hooks	sets/trips	hooks/sets	hooks/trips
2008	1,333	17,875	40,063,212	13.410	2,241.3	30,055
2009	1,225	17,001	38,177,005	13.878	2,245.6	31,165
2010	1,179	16,077	37,244,654	13.636	2,316.6	31,590
2011	1,246	16,888	40,022,132	13.554	2,369.9	32,120
2012	1,305	18,151	44,160,245	13.909	2,432.9	33,839
average	1,258	17,199	39,933,450	13.677	2,321.3	31,754
future	1,305	18,592	46,117,532	14.247	2,480.5	35,339

5.1.3. *Upper Bound, τ_{max}, of Discrete Uniform Prior for the Conditional Binomial Model.* The conditional binomial model was only used when there was no more than 1 observed bycatch event during 2008-2012. When fitting the conditional binomial model to the five years of data, the value of τ_{max} needs to be specified (Section 4.2.3). In this situation, τ_{max} is the maximal number of bycatches over five years that we believe is possible. To evaluate the sensitivity of the conditional binomial model to the specified value of τ_{max}, the model was constructed using the specified values $\tau_{max} = 25, 50, \text{and } 100$. Comparing the models with $\tau_{max} = 50$ and $\tau_{max} = 100$, there was very little difference between the predictive densities. When comparing the models with $\tau_{max} = 25$ and $\tau_{max} = 50$, there were slight differences between the predictive densities and these differences seemed to be slightly more pronounced when there was 1 observed bycatch. These results suggest that the conditional model is not sensitive to the value of τ_{max} if it is larger than the upper limits suggested by the data distribution.

5.1.4. *Summary Statistics of Predictive Density.* Along with the mean and median of the predictive density being computed, a credible interval was also computed. A $100(1-\alpha)\%$ credible interval is an interval that contains $100(1-\alpha)\%$ of the probability density's mass. For example, a 95% credible interval for $\tilde{\tau}$ that is given as [0,5] means that 95% of the predictive density lies between 0 and 5. With discrete distributions it is not uncommon that an interval with the exact desired coverage, such as 95%, does not exist. In this situation, the interval provided is the interval that has the smallest coverage that is at least 95%. The exact coverage is given for all credible intervals provided in this report.

5.1.5. *Evaluating Different Models.* One method used to compare the performance of the different models was to compute the probability of observing a predicted observation $\tilde{y}_{yr}$ at least as extreme as the observed data value y_{yr}, $\min\left(P(\tilde{y}_{yr} \leq y_{yr}), P(\tilde{y}_{yr} \geq y_{yr})\right)$. If an observed value is in the extreme tail of the predictive density, the predictive probability will be small and the validity of the Bayesian model becomes questionable. On the other hand, if none of the predictive probabilities are small then the observed values are consistent with the predictive density and the model seems adequate. A table of these predictive probabilities of "at least as extreme" bycatch observations is provided for each species.

Table 5. Predictive probabilities of "at least as extreme" observations of loggerhead sea turtle bycatch in the adjusted systematic sample. The conditional binomial and Poisson/gamma models are denoted as BIN and PG, respectively. The variable of effort (number of trips or hooks) for the PG models and the values of τ_{max} (25 or 50) for the BIN models are shown in parentheses.

year	observed	BIN(50)	BIN(25)	PG(trips)	PG(hooks)
2008	0	0.71	0.74	0.83	0.96
2009	0	0.70	0.74	0.83	0.95
2010	1	0.26	0.24	0.15	0.04
2011	0	0.71	0.73	0.84	0.95
2012	0	0.69	0.73	0.81	0.95

If the primary regulatory concern is the control of the risk to the endangered species population and more than one model provides a predictive density that is consistent with the observations between 2008 and 2012, then the predictive density with the largest mass to the right of the other distributions is considered the most conservative predictive density. This presumes that over-predicting future bycatch is preferable to under-predicting.

5.2. **Loggerhead Sea Turtle.** During 2008-2012 there was one observed loggerhead sea turtle bycatch event in the deep-set fishery. Previously, there were four observed bycatch events in 2002 (two of the four were on trips that did not comply with the regulations) and one in 2007. With so few observed takes the Poisson/gamma exchangeability model was not considered. Predictive densities were computed using the Poisson/gamma and the conditional binomial model. Table 5 gives the predictive probabilities of "at least as extreme" loggerhead bycatch observations for years 2008-2012. The Poisson/gamma model with effort expressed as number of hooks has a small probability, 0.04, for the year when the one observed bycatch occurred. This small probability indicates that the model is inadequate for explaining the relationship between effort and bycatch as the value of 1 is in the extreme tail of the predictive density. Comparing the Poisson/gamma model with the conditional binomial model, the conditional binomial model has the larger probabilities when there is one observed bycatch event and the smaller probabilities when no bycatch is observed. There is little difference between the probabilities when comparing the conditional binomial model with τ_{max} equal to 50 and 25. If the bycatch rate during 2002-2007 is similar to that in 2008-2012, the Poisson/gamma model is likely underestimating the bycatch rate. Although both conditional binomial models appear to adequately fit the observer data, a $\tau_{max} = 25$ seems small considering it is the maximal bycatch possible over five years.

Table 6 provides the means, medians, and 95% (at least) credible intervals for the three different models. The most conservative predictive density is the density constructed using the conditional binomial model with $\tau_{max} = 50$. This is the predictive density I would recommend if it is necessary to select one. Figure 5 is a graphical depiction of the predictive density of annual bycatch for the conditional binomial model with $\tau_{max} = 50$. The predictive density has a long tail with most of its mass below 10 bycatch events.

TABLE 6. The means, medians, and credible intervals, including their coverage, are given for the predicted densities of annual loggerhead sea turtle bycatch ($\tilde{\tau}$). The conditional binomial and Poisson/gamma models are denoted as BIN and PG, respectively. For the PG models the variable of effort is number of trips. The value of τ_{max} (25 or 50) for the BIN models is shown in parentheses.

model	mean	median	credible interval	coverage
PG(trips)	1.5	1	[0,5]	95%
BIN(25)	2.5	2	[0,7]	98%
BIN(50)	2.8	2	[0,8]	97%

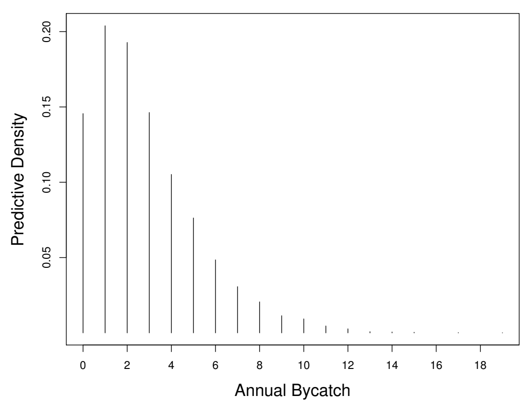

FIGURE 5. Predictive density of the annual loggerhead sea turtles bycatch ($\tilde{\tau}$) constructed using the conditional binomial model with $\tau_{max} = 50$.

5.3. **Green Sea Turtle.** During 2008-2012 there was one observed green sea turtle bycatch event in the deep-set fishery. Previously, there were four observed bycatch events during 2002-2007: 2 in 2006, 1 in 2004, and 1 in 2002. With so few observed takes the Poisson/gamma exchangeability model was not considered. Table 7 gives the predictive probabilities of "at least as extreme" green sea turtle bycatch observations for years 2008-2012. There is very little difference between the Poisson/gamma models and between the conditional binomial models. Comparing the Poisson/gamma model with the conditional binomial model, the conditional binomial model has the larger probabilities when there is one observed bycatch event and the smaller probabilities when no bycatch is observed. If the bycatch rate during 2002-2007 is similar to that in 2008-2012, the Poisson/gamma model is likely underestimating the bycatch rate. Although both conditional binomial models appear to adequately fit the

Table 7. Predictive probabilities of "at least as extreme" observations of green sea turtle bycatch in the adjusted systematic sample. The conditional binomial and Poisson/gamma models are denoted as BIN and PG, respectively. The variable of effort is number of trips or hooks for the PG models. The value of τ_{max} (25 or 50) for the BIN models is given in parentheses.

year	observed	BIN(50)	BIN(25)	PG(trips)	PG(hooks)
2008	0	0.71	0.75	0.83	0.84
2009	0	0.71	0.74	0.83	0.83
2010	0	0.73	0.76	0.85	0.85
2011	1	0.30	0.27	0.18	0.18
2012	0	0.70	0.73	0.82	0.82

Table 8. The means, medians, and credible intervals, including their coverage, are given for the predicted densities of the annual green sea turtle bycatch ($\tilde{\tau}$). The conditional binomial and Poisson/gamma models are denoted as BIN and PG, respectively. For the PG models the variable of effort (number of trips and hooks) and for the BIN models the value of τ_{max} (25 and 50) are given in parentheses.

model	mean	median	credible interval	coverage
PG(hooks)	1.6	1	[0,6]	97%
PG(trips)	1.6	1	[0,6]	97%
BIN(25)	2.4	2	[0,6]	96%
BIN(50)	2.8	2	[0,8]	97%

observer data, a $\tau_{max} = 25$ seems small considering it is the maximal bycatch possible over five years.

Table 8 provides the means, medians, and 95% (at least) credible intervals of $\tilde{\tau}$ for the four different models. The most conservative predictive density is the density constructed using the conditional binomial model with $\tau_{max} = 50$. This is the predictive density I would recommend if it is necessary to select one. Figure 6 is a graphical depiction of the predictive density of annual bycatch for the conditional binomial model with $\tau_{max} = 50$. The predictive density has a long tail with most of its mass below 10 bycatch events.

As loggerhead and green sea turtles have the same number of observed bycatches during 2008-2012, in theory, the Poisson/gamma models using the same measures of effort are equivalent and the conditional binomial models with the same values of τ_{max} are equivalent. Any difference between these theoretically equivalent predictive densities is a result of simulation error.

5.4. **Leatherback Sea Turtle.** During 2008-2012 there were 7 observed leatherback sea turtle bycatch events, all but one of which were within the systematic sample. During 2002-2007 there were 11 observed leatherback sea turtle bycatch events, however, one of the two observations in 2002 was on a vessel not complying with the current regulations.

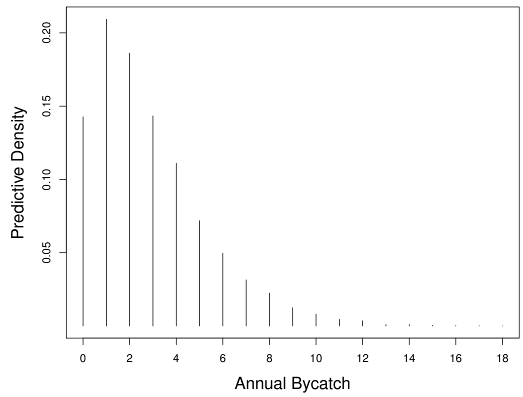

FIGURE 6. Predictive density of the annual green sea turtle bycatch ($\tilde{\tau}$) constructed using the conditional binomial model with $\tau_{max} = 50$.

The Poisson/gamma and the Poisson/gamma exchangeability model were both considered. The adjusted systematic sample was used to construct the Poisson/gamma model. The Poisson/gamma exchangeability model was formulated using the systematic and adjusted systematic sample to examine if there were any anomalous differences in the results between these two data sets. Differences between comparable predictive probabilities are evidence that the assumed relationship between bycatch and effort is inappropriate or the adjusted systematic sample is not representative of systematic samples consistently drawn at 15% coverage. A problem with the Poisson/gamma exchangeability model was that suspiciously very large numbers of bycatch (in the hundreds) were occasionally predicted. Expressing effort as $e^{\star}$ slightly reduced the frequency of these suspiciously large numbers. As there appeared to be a problem with the exchangeability model in the far right tail of the predictive density of annual bycatch, the top 0.5% of the largest predicted bycatches was discarded. This level of truncation still left some questionable large predicted values of bycatch, reducing the risk of underestimating the extreme right tail of the predictive density.

Figure 7 is the contour plot of the posterior density of ($\log(\mu)$,$\log(\alpha)$) constructed using the adjusted systematic sample and $e^{\star}$ as the measure of effort. The contour lines are drawn at 10%, 1%, and 0.1%. A simulated random sample from this density is shown on top of the contour plot. This figure confirms that the samples are from the posterior distribution as the mass of points are within the inner contours, hiding the contour at 10%.

Table 9 gives the predictive probabilities of 'at least as extreme" leatherback bycatch observations for years 2008-2012. As results are similar between the different measures of effort, only the results for number of successful trips, number of hooks, and $e^{\star}$ are shown. Year 2008 has no observed bycaught leatherbacks and 2011 has three observed bycatches; all other years have one observed bycatch. Although the observed bycatch levels are identical

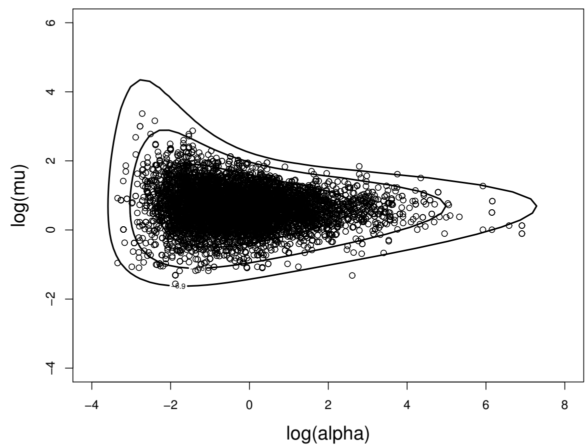

FIGURE 7. Contour plot of the posterior density of (log(μ),log(α)) for leatherback sea turtles constructed using the adjusted systematic sample. A sample of simulated draws from the distribution lies on top.

between the systematic and adjusted systematic datasets, their effort levels are different. The Poisson/gamma exchangeability models have slightly higher predictive probabilities in 2008 and 2011, the years of minimal and maximal observed bycatch, respectively. For the other years, the Poisson/gamma has slightly higher predictive probabilities, except for year 2012 where all models have very similar probabilies. When comparing the exchangeability models fitted to the systematic and adjusted systematic datasets, there appears to be very little difference between the probabilities. This is reassuring as we expect the probabilities to be similar if the adjusted systematic sample resembles systematic samples consistently drawn at 15% coverage and the relationship between bycatch and effort is adequate.

Figure 8 compares the predictive density of $\tilde{\tau}$ under future levels of effort constructed using the Poisson/gamma and Poisson/gamma exchangeability models. The exchangeability models are flatter with a longer tail compared to the Poisson/gamma model. Although the exchangeability model occasionally produced unrealistic values of $\tilde{\tau}$, it appears to do better in predicting the tail of the density.

Table 10 provides the means, medians, and credible intervals for the predictive densities. The means are smaller and the confidence intervals shorter for the Poisson/gamma model. The results are robust to the measure of effort and the dataset (systematic or adjusted systematic) used to construct the Poisson/gamma exchangeability model. If one model is to be recommended, I would lean towards the exchangeability model using the systematic data with effort expressed as $e^{\star}$ for two reasons. First, the Poisson/gamma model likely underestimates the extent of the right tail as a consequence of assuming a constant bycatch

TABLE 9. Predictive probabilities of "at least as extreme" observations of leatherback sea turtle bycatch. The Poisson/gamma and Poisson/gamma exchangeability models are denoted as PG and EXPG, respectively. Shown in parentheses is the dataset used to construct the model: systematic sample (sys) or adjusted systematic sample (adj.sys). The observed values (obs) are from the specified sample: the systematic or adjusted systematic sample.

		PG(adj.sys)			EXPG(sys)			EXPG(adj.sys)		
year	obs	$e^\star$	hooks	trips	$e^\star$	hooks	trips	$e^\star$	hooks	trips
systematic sample										
2008	0	0.36	0.39	0.37	0.42	0.43	0.42	0.43	0.45	0.46
2009	1	0.63	0.65	0.65	0.59	0.60	0.60	0.55	0.58	0.58
2010	1	0.61	0.58	0.60	0.57	0.54	0.56	0.55	0.52	0.55
2011	3	0.13	0.14	0.15	0.19	0.20	0.20	0.17	0.19	0.17
2012	1	0.63	0.64	0.64	0.63	0.63	0.63	0.62	0.63	0.62
adjusted systematic sample										
2008	0	0.33	0.36	0.34	0.39	0.41	0.39	0.40	0.42	0.41
2009	1	0.66	0.65	0.65	0.59	0.63	0.62	0.59	0.62	0.61
2010	1	0.64	0.61	0.63	0.60	0.57	0.58	0.57	0.55	0.57
2011	3	0.13	0.15	0.15	0.18	0.19	0.18	0.17	0.18	0.18
2012	1	0.63	0.63	0.65	0.63	0.63	0.63	0.62	0.63	0.62

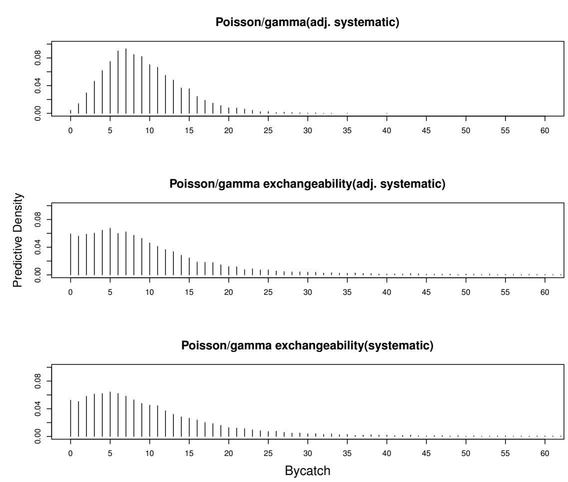

FIGURE 8. Predictive densities of annual bycatch ($\tilde{\tau}$) for leatherback sea turtles for the Poisson/gamma and Poisson/gamma exchangeability models.

Table 10. The means, medians, and credible intervals, including their coverage, are provided for predicted densities of annual leatherback sea turtles bycatch ($\tilde{\tau}$). The Poisson/gamma and Poisson/gamma exchangeability models are denoted as PG and EXPG, respectively. Shown in parentheses is the dataset used to construct the model: systematic sample (sys) or adjusted systematic sample (adj.sys).

model	effort	mean	median	credible interval	coverage
PG(adj.sys)	$e^{\star}$	9.2	9	[1,19]	96%
PG(adj.sys)	hooks	9.2	8	[1,18]	95%
PG(adj.sys)	trips	9.3	9	[1,19]	96%
EXPG(sys)	$e^{\star}$	10.5	8	[0,31]	95%
EXPG(sys)	hooks	10.2	8	[0,31]	95%
EXPG(sys)	trips	10.0	8	[0,29]	95%
EXPG(adj.sys)	$e^{\star}$	10.2	8	[0,29]	95%
EXPG(adj.sys)	hooks	9.7	7	[0,29]	95%
EXPG(adj.sys)	trips	9.7	8	[0,26]	95%

rate. While the exchangeability models may overestimate the extent of the right tail, this seems preferable to underestimation. Without truncating the predicted values generated using the recommended exchangeability model, the mean, median, and 95% credible interval of the predictive density are 12.0, 8, and [0,31], respectively. As the mean is sensitive to outliers, truncating the data has the greatest effect on this value. Second, although simulation error may account for the small difference between the exchangeability model constructed under the systematic and adjusted systematic sample, the adjusted systematic sample has psuedo systematic samples for three quarters,

5.5. **Olive Ridley Sea Turtle.** During 2008-2012 there were 23 observed olive ridley sea turtle bycatch events, 20 of these were observed by the systematic sample. There were 43 observed olive ridley sea turtle bycatch events during 2002-2007. The Poisson/gamma and the Poisson/gamma exchangebility model were both considered. The adjusted systematic sample was used to construct the Poisson/gamma model. The Poisson/gamma exchangeability model was formulated using the systematic and adjusted systematic sample to examine if there were any anomalous differences in the results between these two data sets. Differences between comparable predictive probabilities are evidence that the assumed relationship between bycatch and effort is inappropriate or the adjusted systematic sample is not representative of systematic samples consistently drawn at 15% coverage. A problem with the Poisson/gamma exchangeability model was that suspiciously very large numbers of bycatch (in the hundreds) were occasionally predicted; these unrealistic values were generated notably less frequently than when modeling leatherback sea turtle bycatch. Expressing effort as $e^{\star}$ slightly reduced the frequency of these suspiciously large numbers. As there appeared to be a problem with the exchangeability model in the far right tail of the predictive density of annual bycatch, the top 0.5% of the largest predicted bycatches were discarded. This level of truncation still left some questionable large predicted values of bycatch, reducing the risk of underestimating the extreme right tail of the predictive density.

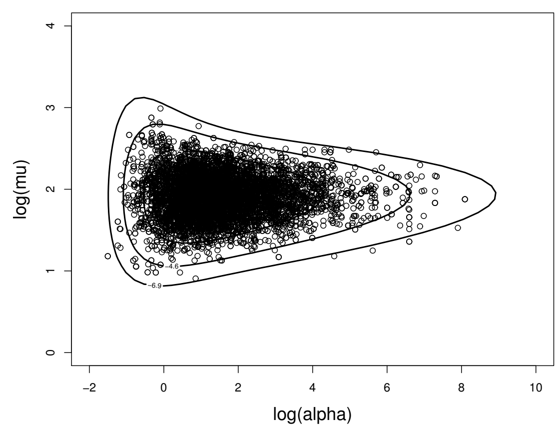

FIGURE 9. Contour plot of the posterior density of (log(μ),log(α)) for the olive ridley sea turtle constructed using the adjusted systematic sample. A sample of simulated draws from the distribution lies on top.

Figure 9 is the contour plot of the posterior density of (log(μ),log(α)) derived using the adjusted systematic sample and $e^{\star}$ as the measure of effort. The contour lines are drawn at 10%, 1%, and 0.1%. A simulated random sample from this density is shown on top of the contour plot. This figure confirms that the samples are from the posterior distribution as the mass of points are within the inner contours, hiding the contour at 10%.

Table 11 gives the predictive probabilities of 'at least as extreme" olive ridley bycatch observations for years 2008-2012. The probabilities are computed using the observed bycatch from the systematic sample and the adjusted systematic sample. In year 2008 the systematic sample has only one observed bycatch; whereas, the adjusted systematic sample has two observed adjusted bycatch events. The Poisson/gamma exchangeability models have a slightly higher probability of predicting the observed bycatch in 2008 and 2011, the years of minimal and maximal observed bycatch, respectively. For the other years, there is very little, if any, difference between the different models. When comparing the exchangeability models fitted to the systematic and adjusted systematic datasets, there appears to be very little difference between the probabilities. This is reassuring as we expect the probabilities to be similar if the adjusted systematic sample resembles systematic samples consistently drawn at 15% coverage and the relationship between bycatch and effort is adequate.

Figure 10 compares the predictive density of ($\tilde{\tau}$) under future levels of effort constructed using the Poisson/gamma and Poisson/gamma exchangeability models. The exchangeability models are flatter with a longer tail compared to the Poisson/gamma model, although these differences are not as extreme as seen for the predictions of leatherback bycatch. Although

Table 11. Predictive probabilities of "at least as extreme" observations of olive ridley sea turtle bycatch. The Poisson/gamma and Poisson/gamma exchangeability models are denoted as PG and EXPG, respectively. Shown in parentheses is the dataset used to construct the model: systematic sample (sys) or adjusted systematic sample (adj.sys). The observed values (obs) are from the specified sample: the systematic or adjusted systematic sample.

		PG(adj.sys)			EXPG(sys)			EXPG(adj.sys)		
year	obs	$e^\star$	hooks	trips	$e^\star$	hooks	trips	$e^\star$	hooks	trips
systematic sample										
2008	1	0.12	0.14	0.14	0.17	0.19	0.17	0.16	0.18	0.17
2009	4	0.50	0.56	0.55	0.48	0.52	0.52	0.48	0.54	0.53
2010	2	0.32	0.37	0.35	0.37	0.41	0.40	0.36	0.41	0.39
2011	7	0.16	0.19	0.17	0.18	0.20	0.19	0.18	0.21	0.20
2012	6	0.33	0.33	0.31	0.33	0.32	0.32	0.34	0.33	0.32
adjusted systematic sample										
2008	2	0.24	0.27	0.25	0.28	0.32	0.29	0.31	0.29	0.30
2009	4	0.51	0.57	0.57	0.51	0.57	0.58	0.58	0.57	0.56
2010	2	0.29	0.32	0.30	0.33	0.36	0.34	0.35	0.33	0.33
2011	7	0.15	0.18	0.17	0.18	0.20	0.19	0.21	0.20	0.21
2012	6	0.33	0.32	0.30	0.33	0.32	0.33	0.33	0.32	0.31

the exchangeability model occasionally produced unrealistic values of $\tilde{\tau}$, it appears to do better in predicting the tail of the density.

Table 12 provides the means, medians, and credible intervals for the predictive densities of $\tilde{\tau}$. The means are smaller and the credible intervals shorter for the Poisson/gamma model. The results are robust to the measure of effort and the dataset (systematic or adjusted systematic) used to construct the Poisson/gamma exchangeability model. If one model is to be recommended, I would lean towards the exchangeability model using the adjusted systematic data with effort expressed as $e^\star$. Although the adjusted systematic data has psuedo systematic samples for three quarters, it has one additional observed bycatch event and is the most conservative predictive density. Without truncating the predicted values generated using the recommended exchangeability model, the mean, median, and 95% credible interval of the predictive density are 32.7, 31, and [7,63], respectively. Truncating the data had little effect on these three values.

5.6. **Humpback Whale.** During 2008-2012 there were no observed bycaught humpback whales in the deep-set fishery. Previously, there were two observed bycatch events during 2002-2007: 1 in 2002 and 1 in 2004. As there were no observed bycatches, the Poisson/gamma and Poisson/gamma exchangeability models were not considered. Table 13 gives the predictive probabilities of "at least as extreme" humpback bycatch observations for years 2008-2012. There is little difference between the probabilities for the conditional binomial model with τ_{max} equal to 25 or 50. Table 14 provides the means, medians, and 95% (at least) credible intervals for the two different models. Based on observations during 2008-2012, the predictive density created using $\tau_{max} = 25$ seems to have a slight advantage as its density has less

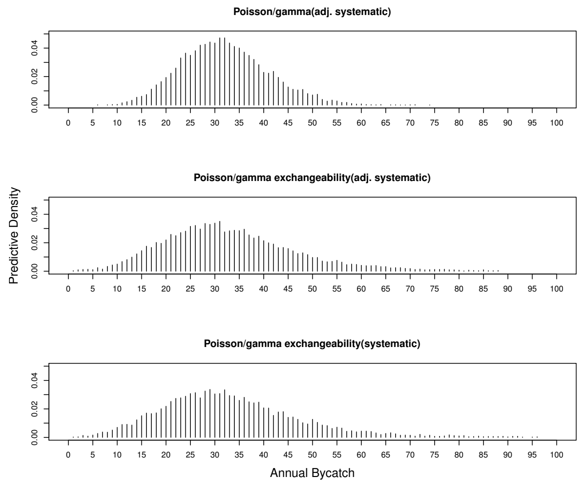

FIGURE 10. Predictive densities of annual olive ridley sea turtle bycatch ($\tilde{\tau}$) for the Poisson/gamma and Poisson/gamma exchangeability models.

TABLE 12. The means, medians, and credible intervals, including their coverage, are provided for predicted densities of olive ridley sea turtle annual bycatch ($\tilde{\tau}$). The Poisson/gamma Poisson/gamma exchangeability models are denoted as PG and EXPG, respectively. Shown in parentheses is the dataset used to construct the model: systematic sample (sys) or adjusted systematic sample (adj.sys).

model	effort	mean	median	credible interval	coverage
PG(adj.sys)	$e^\star$	32.1	31	[16,51]	95%
PG(adj.sys)	hooks	32.4	32	[15,50]	95%
PG(adj.sys)	trips	32.5	32	[16,51]	95%
EXPG(sys)	$e^\star$	31.8	30	[8,59]	95%
EXPG(sys)	hooks	32.3	31	[9,64]	95%
EXPG(sys)	trips	32.7	31	[7,60]	95%
EXPG(adj.sys)	$e^\star$	32.3	31	[8,61]	95%
EXPG(adj.sys)	hooks	32.5	31	[9,61]	95%
EXP(adj.sys)	trips	33.3	31	[9,64]	95%

mass above zero (higher probability of no bycatch). If the bycatch rates during 2002-2007 are similar to bycatch rates after 2007 then the density with $\tau_{max} = 50$ would likely fit the 2002-2007 observations better. If it is necessary to select one model, I recommend using

Table 13. Predictive probabilities of "at least as extreme" observations of humpback whale bycatch in the adjusted systematic sample. The conditional binomial model is denoted as BIN. The value of τ_{max} (25 or 50) for the BIN models is shown in parentheses.

year	observed	BIN(50)	BIN(25)
2008	0	0.86	0.87
2009	0	0.85	0.86
2010	0	0.87	0.88
2011	0	0.85	0.86
2012	0	0.85	0.87

Table 14. The means, medians, and credible intervals, including their coverage, are given for the predicted densities of annual humpback whale bycatch ($\tilde{\tau}$). The conditional binomial model is denoted as BIN. The value of τ_{max} is 25 and 50 for the BIN models.

model	mean	median	credible interval	coverage
BIN(25)	1.2	1	[0,5]	97%
BIN(50)	1.3	1	[0,5]	96%

Table 15. The means, medians, and credible intervals, including their coverage, are given for the predicted densities of annual humpback whale DSI ($\tilde{\delta}$). The conditional binomial model is denoted as BIN. The value of τ_{max} (25 or 50) for the BIN models is given in parentheses.

model	mean	median	credible interval	coverage
BIN(25)	0.6	0	[0,3]	97%
BIN(50)	0.7	0	[0,3]	97%

the conditional binomial model with $\tau_{max} = 50$ as it is more conservative. Figure 11 is a graphical depiction of the predictive density of $\tilde{\tau}$ for the conditional binomial model with $\tau_{max} = 50$.

The methods described in Section 4.3 were used to compute the posterior density of ρ (probability a bycatch results in DSI). Figure 12 provides a graphical display of the posterior densities of each category's probability computed using the Dirichlet model. The posterior, data, and prior density of ψ (probability that a PR results in a DSI) computed using the binomial/beta model are graphically displayed in Figure 13. This figure shows that the prior has very little influence on the posterior density. The predicted posterior density of $\tilde{\delta}$ for humpback whales is provided graphically in Figure 14. Summary statistics of the predicted densities of $\tilde{\delta}$ are provided in Table 15.

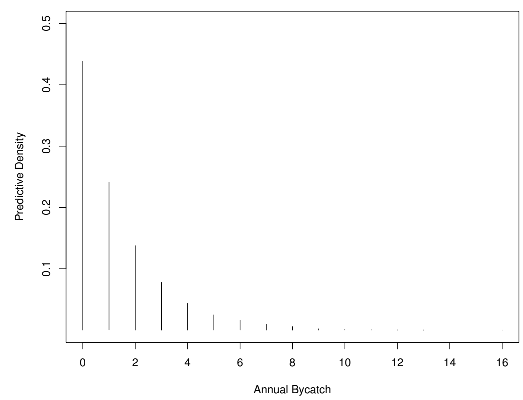

FIGURE 11. Predictive density of annual humpback whale bycatch ($\tilde{\tau}$) computed using the conditional binomial model with $\tau_{max} = 50$.

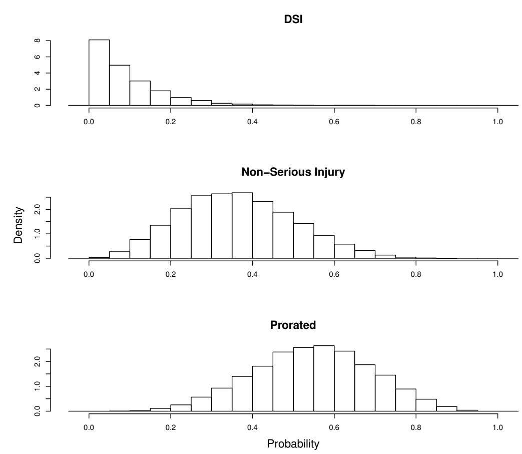

FIGURE 12. Posterior density of probabilities for DSI, NSI, and PR injury categories for humpback and sperm whales.

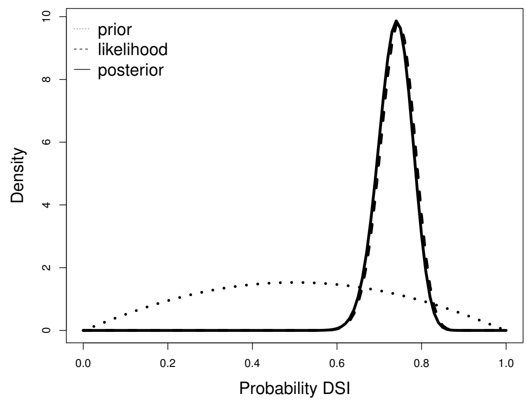

FIGURE 13. The prior, data (likelihood) and posterior densities (densities) for the probability a PR is a DSI. These densities apply to the humpback and sperm whales.

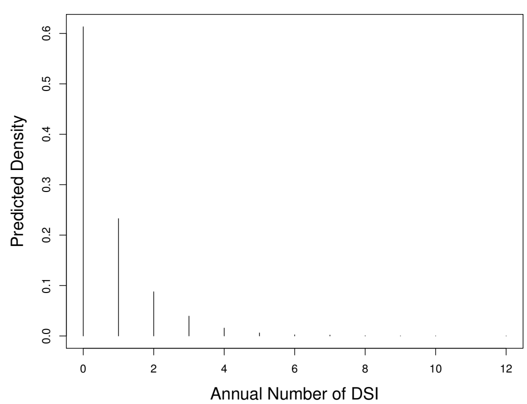

FIGURE 14. Predicted density of annual humpback whale DSI ($\tilde{\delta}$) computed using the conditional binomial model with $\tau_{max} = 50$.

Table 16. Predictive probabilities of "at least as extreme" observations of sperm whale bycatch in the adjusted systematic sample. The conditional binomial model and Poisson/gamma model are denoted as BIN and PG, respectively. The variable of effort (number of trips or hooks) for the PG models and the value of τ_{max} (25 or 50) for the BIN models are shown in parentheses.

year	observed	BIN(50)	BIN(25)	PG(trips)	PG(hooks)
2008	0	0.72	0.74	0.84	0.84
2009	0	0.71	0.74	0.83	0.83
2010	0	0.75	0.77	0.85	0.86
2011	1	0.29	0.26	0.17	0.18
2012	0	0.70	0.73	0.82	0.81

Table 17. The means, medians, and credible intervals, including their coverage, are given for the predicted densities of annual sperm whale bycatch ($\tilde{\tau}$). The conditional binomial model and Poisson/gamma model are denoted as BIN and PG, respectively. The variables of effort (number of trips or hooks) for the PG models and the value of τ_{max} (25 or 50) for the BIN models are given in parentheses.

model	mean	median	credible interval	coverage
PG(hooks)	1.6	1	[0,5]	95%
PG(trips)	1.6	1	[0,6]	95%
BIN(25)	2.5	2	[0,6]	96%
BIN(50)	2.8	2	[0,8]	97%

5.7. **Sperm Whale.** During 2008-2012 there was one observed sperm whale bycatch event in the deep-set fishery, and during 2002-2007 there were no observed bycaught sperm whales. With so few observed takes the Poisson/gamma exchangebility model was not considered. Table 16 gives the predictive probabilities of "at least as extreme" sperm whale bycatch observations for years 2008-2012. When comparing the conditional binomial and Poisson/gamma models, the conditional binomial model has larger probabilities for the year of the one observed bycatch and smaller probabilities for the other years. As year 2011 is also the year of the one observed green sea turtle bycatch event, any difference between the two tables of "at least as extreme" probabilities is a consequence of simulation error.

Table 17 provides the means, medians, and 95% (at least) credible intervals for the four different models. If necessary to select one model, I recommend using the most conservative predictive density of $\tilde{\tau}$ constructed using the conditional binomial model with $\tau_{max} = 50$.

The posterior density of each injury category's probability and the posterior density of ρ (probability a bycatch results in a DSI) used for humpback whales was used for the sperm whales. Summary statistics of the predicted densities are provided in Table 18. The predicted posterior density of $\tilde{\delta}$ for sperm whales constructed using the conditional binomial model with $\tau_{max} = 50$ for $\tilde{\tau}$ is provided graphically in Figure 15.

TABLE 18. The means, medians, and credible intervals, including their coverage, are given for the predicted densities of the annual DSI ($\tilde{\delta}$) of the sperm whales. The conditional binomial model and Poisson/gamma model are denoted as BIN and PG, respectively. The variables of effort (the number of trips and hooks) for the PG models and the value of τ_{max} (25 or 50) for the BIN models are given in parentheses.

model	mean	median	credible interval	coverage
PG(hooks)	0.8	0	[0,3]	96%
PG(trips)	0.8	0	[0,3]	96%
BIN(25)	1.2	1	[0,4]	98%
BIN(50)	1.4	1	[0,4]	96%

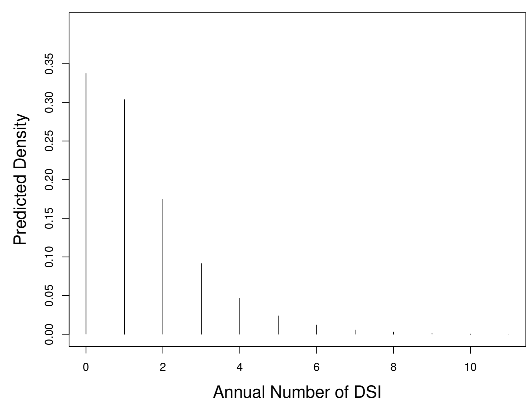

FIGURE 15. Predictive density of the annual sperm whale DSI ($\tilde{\delta}$) constructed using the conditional binomial model with $\tau_{max} = 50$ for $\tilde{\tau}$.

5.8. **Main Hawaiian Islands Insular Stock of False Killer Whale.** During 2008-2012 there were 2 observed FKW and 1 observed blackfish bycatch events within the 140 km boundary and no observed bycatch events within the open area. There were 3 observed FKW and 3 observed blackfish bycatch events within the 140 km boundary during 2002-2007. Two of the 3 observed FKW bycatch events may have occurred in the open area, as there were no capture locations recorded and some of the associated set and/or haul coordinates were located in this area.

As our primary objective is to predict the annual bycatch (τ) and DSI (δ) within the open area and there are no observed FKW takes within this area during 2008-2012, one of the first questions to address is the appropriateness of using data collected outside this

TABLE 19. Predictive probabilities of "at least as extreme" observations of FKW bycatch in the adjusted systematic sample that occurred within the 140 km boundary for different predictive densities. Using the adjusted systematic sample, the predictive densities were constructed using the Poisson/gamma model fitted to all observations (total), observations within the boundaries of Hawaii's EEZ (hawaii), and observations within the 140 km boundaries (140 km).

year	observed	total	hawaii	140 km
2008	0	.904	.819	.618
2009	0	.841	.723	.467
2010	0	.907	.820	.605
2011	1	.084	.158	.350
2012	2	.006	.022	.129

area when predicting τ. The disadvantages of restricting ourselves to data collected within the open area are that the bycatch rate would likely be underestimated as a result of too few observations for such a rare event and a model that could produce a predictive density without any observed bycatch events would need to be employed. As we want to use the number of hooks as the unit of effort for predicting future bycatch in the open area (see Section 2), the conditional binomial model is not appropriate. It is natural to consider using the data within the 140 km boundaries as this is the assumed boundary of the IFKW stock. All three observed bycatch events within the 140 km boundary occurred within the systematic sample. Using the methods described in McCracken (2010), the observed blackfish has an estimated probability of 0.97 of being a FKW. As this blackfish is very likely a FKW, hereafter, it is considered a FKW. Fitting the Poisson/gamma model to observations in the adjusted systematic sample that occurred within the 140 km boundary (approximately 936 thousand hooks), the predictive probability of observing no bycatch during 2008-2012 in the open area is 0.312. As this value is not in the tail area of the predictive density, the observation of no bycatch in the open area is consistent with this model.

Now, consider using data collected outside the assumed boundary of the IFKW. During 2008-2012, within the adjusted systematic sample there were 18 observed bycaught FKW with approximately 29,803 thousand hooks deployed, and within the boundaries of Hawaii's EEZ there were 11 observed bycaught FKW with approximately 9,311 thousand hooks deployed. A Poisson/gamma model was fitted to these two data sets (within Hawaii's EEZ and total fishing grounds) and the predictive probabilities of observing "at least as extreme" observed bycatch events of FKW within the 140 km boundaries computed. These probabilities are given in Table 19. The very small predictive probabilities for year 2012 (.022 and .006) indicate that these two models are inadequate in fitting this observation. Based on the findings here and in the above paragraph, the decision was to use the Poisson/gamma model fitted to observations in the adjusted systematic sample that occurred within the 140 km boundary. Hereafter, all results refer to this model.

Predictive densities of τ_{FKW} in the open area were constructed for future effort levels of 500 thousand, 600 thousand, and 1 million hooks. This range of values approximately covers the median, mean, and maximum effort within the open area during 2008-2012. Figure 16

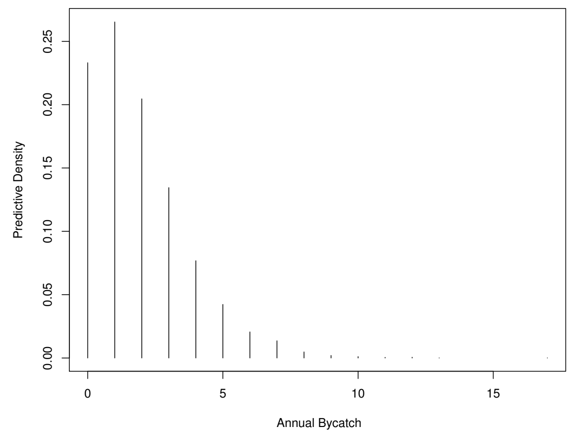

Figure 16. Predictive density of annual FKW bycatch ($\tilde{\tau}_{FKW}$) in the open area based on 600 thousand hooks.

Table 20. Means, medians, and credible intervals, including their coverage, are given for the predicted densities of annual FKW bycatch ($\tilde{\tau}_{FKW}$) in the open area. Future levels of effort considered are 500 thousand, 600 thousand, and 1 million hooks.

effort (1000s hooks)	mean	median	credible interval	coverage
500	1.61	1	[0,5]	96%
600	1.90	2	[0,5]	98%
1000	3.24	3	[0,8]	96%

provides a graphical display of the predictive density based on 600 thousand hooks, and Table 20 provides the predicted means, medians, and 95% (at least) credible intervals for these predictive densities based on the three levels of future effort.

The posterior density of ι (the probability that a FKW bycaught within the 140 km boundary is a IFKW) and θ (the probability that a bycatch resulted in a DSI classification) were constructed as described in Section 4.4 and Section 4.3, respectively. Figure 17 is a graphical display of the prior, data, and posterior densities of θ when the prior and data distribution either included or, alternatively, excluded blackfish classifications (see Section 3). The two posterior densities are very similar.

The methods described in Section 4.5 were used to construct predictive densities of $\tilde{\delta}_{IFKW}$. Table 21 provides the means and 95% (at least) credible intervals for the predictive densities

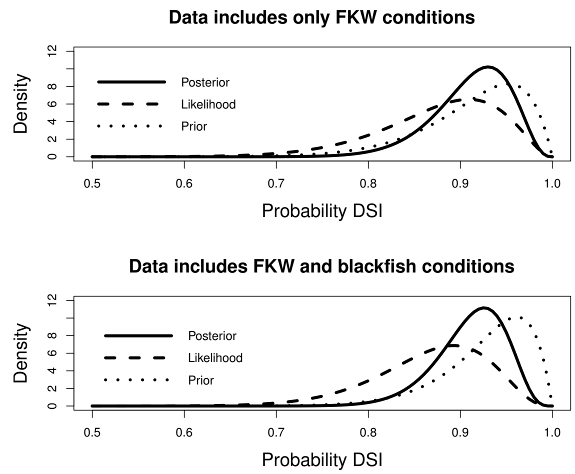

FIGURE 17. The prior, data (likelihood), and posterior density of the probability a false killer whale bycatch results in a DSI classification (θ).

of $\tilde{\delta}_{IFKW}$ created using the three levels of effort and the two injury classification datasets (excluding and including blackfish). As the median is 0 for all these densities, it is not included in the table. Since the two posterior densities of θ in Figure 17 are so similar, it is not surprising that the predictive densities of $\tilde{\delta}_{IFKW}$ appear to be very similar between the two injury classifications datasets within the same level of effort.

For a future effort of 600 thousand hooks (our assumed projected effort for the future) in the open area and the injury classification dataset excluding blackfish (results in the more conservative predictive density), the predictive densities of $\tilde{\tau}_{IFKW}$ and $\tilde{\delta}_{IFKW}$ have means of 0.30 and 0.27, respectively. These are the recommended predictive densities if required to select one of each. Figure 18 is a graphical depiction of the recommended predictive densities of $\tilde{\tau}_{IFKW}$ and $\tilde{\delta}_{IFKW}$. Since the management of marine mammal bycatch is based on a five year average, the same predictive density for $\tilde{\delta}_{IFKW}$ is used to construct a predictive density of the five year average of $\tilde{\delta}_{IFKW}$. This is done by drawing five samples of $\tilde{\delta}_{IFKW}$ from this predictive density, summing over these five samples, and then dividing by 5 and repeating this process 10,000 times. As the quantity of DSI is an integer with a small range, the resulting predictive density is discrete with a small range. Table 22 provides the predictive density for the 5 year average of $\tilde{\delta}_{IFKW}$.

As there is another proposed stock of FKW referred to as the Northwestern Hawaiian Islands false killer whale (NHIFKW) stock, management requested information on what proportion of the predicted IFKW bycatch would be expected to occur within the area that overlaps the proposed range of the NHIFKW stock. This partitioning of the predictions is given in Appendix B.

TABLE 21. Means, medians, and 95% credible intervals (CI), including their coverage, are given for the predicted densities of annual IFKW bycatch ($\tilde{\tau}_{IFKW}$) and annual IFKW DSI ($\tilde{\delta}_{IFKW}$). Future levels of effort considered are 500 thousand, 600 thousand, and 1 million hooks. The predictive densities of $\tilde{\delta}_{IFKW}$ were constructed using (1) only FKW injury classifications (FKW) and (2) FKW and blackfish classifications (FKW.BF).

	$\tilde{\tau}_{IFKW}$			$\tilde{\delta}_{IFKW}$		
dataset	mean	CI	coverage	mean	CI	coverage
future effort of 500,000 hooks						
FKW.BF	0.25	[0,1]	97%	0.22	[0,1]	97%
FKW	0.24	[0,1]	97%	0.22	[0,1]	97%
future effort of 600,000 hooks						
FKW.BF	0.29	[0,1]	96%	0.27	[0,1]	97%
FKW	0.30	[0,1]	95%	0.27	[0,1]	96%
future effort of 1,000,000 hooks						
FKW.BF	0.49	[0,2]	98%	0.44	[0,2]	98%
FKW	0.49	[0,2]	98%	0.45	[0,1]	98%

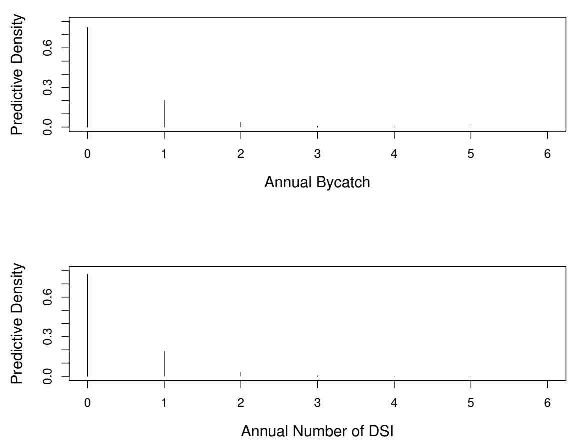

FIGURE 18. Predictive densities of annual IFKW bycatch ($\tilde{\tau}_{IFKW}$) and annual IFKW DSI ($\tilde{\delta}_{IFKW}$).

Table 22. Predictive density of the 5 year average of annual IFKW DSI.

average	predictive probability
0.0	0.2774
0.2	0.3422
0.4	0.2182
0.6	0.1035
0.8	0.0416
1.0	0.0127
1.2	0.0032
1.4	0.0010
1.6	0.0002

6. CONCLUSIONS

In this report, many predictions have been given and many assumptions and caveats related to these predictions have been listed. Several important caveats should be highlighted. First, the models used to create predictive densities assume that levels of bycatch are proportional to functions of effort. To add a measure of robustness to these models, the models were constructed using one of two subsets of the 2008-2012 observer data. One subset of the data was self-weighting across strata (the systematic sample), and the other was self-weighting across and within strata (the adjusted systematic sample). Diagnostics indicated that predictive densities of bycatch that assumed a proportional relationship with effort were adequate for predicting bycatch during 2008-2012. Second, the predictive density based on effort as the predictor variable assumes that the associations between effort and bycatch levels captured in the period ahead will be similar to those in existence during the period upon which the model is based, years 2008 to 2012, and that the model is relevant to the extrapolation of the projected effort scenario. Because the effort within a trip is projected to increase, the projected number of trips was adjusted upward to account for this increased effort. Appendix A examines the sensitivity of the predictions to this adjustment. Third, predictive densities were constructed assuming a constant level of yearly effort, when in fact, effort will vary. Therefore, the predictive densities of $\tilde{\tau}$ and $\tilde{\delta}$ will likely show more variation than that predicted. Fourth, the predictive densities of $\tilde{\delta}$ assume that the past probabilities that a bycatch results in a DSI are representative of the future probabilities.

Appendix A. Sensitivity Analysis to Number of Projected Trips

To evaluate the sensitivity of the mean and median of the prior predictive distribution of future annual bycatch to levels of projected annual effort ranging from 1,300 to 1,525 trips, a study was undertaken where the distributions of the mean and median for different levels of projected annual effort were examined. Specifically, the levels of effort examined were from 1,300 to 1,525 trips increasing by intervals of 25 trips. For each level of annual effort, the prior predictive distribution of annual bycatch was generated 2,000 times. As the loggerhead sea turtle, green sea turtle, and sperm whale each had only one observed bycatch event between 2008-2012, the prior predictive distribution based on the conditional binomial model is theoretically equivalent between these three species. Therefore, the conditional binomial

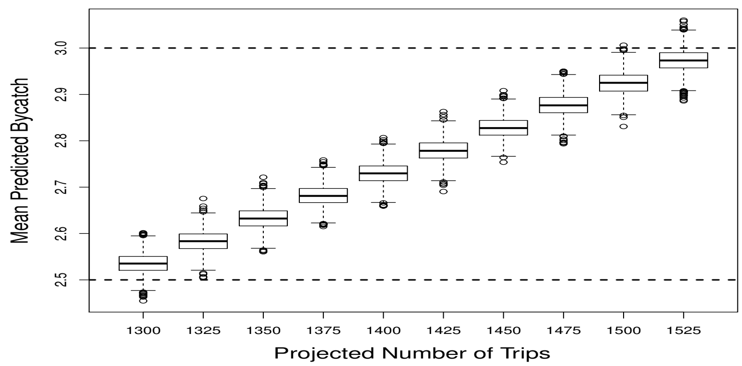

FIGURE 19. Distributions of the mean of the predictive distribution of future annual bycatch for loggerhead and green sea turtles, and sperm whales (1 observed bycatch event between 2008-2012) based on the conditional binomial model. The horizontal dash lines bound the values of 2.5 and 3.0.

model($\tau_{max} = 50$) was fitted to 1 observation and 194 trips (observed effort between 2008-2012 for the adjusted systematic sample) and 10,000 random draws from the corresponding prior predictive distributions of annual bycatch generated as described in Section 4.2.3. The mean and median of the 10,000 draws were recorded as they represent random draws from the distributions of the mean and median of the prior predictive distribution of future annual bycatch. Herein, the distributions of the mean and median of the prior predictive distribution will be referred to as the distribution of the mean and median, respectively. For the humpback whale there were no observed humpback whale bycatch events between 2008-2012, and the conditional binomial model ($\tau_{max} = 50$) was fitted accordingly.

Figures 19 and 20 are box-and-whisker plots of the distribution of the mean over different levels of predicted effort. The bottom and top of the individual boxes represent the first and third quartiles, and the band inside the box is the median. The vertical dashed lines are called the "whiskers" and show 1.5 times the interquartile range (the difference between the first and third quartiles) of the data. Points outside 1.5 times the interquartile range are plotted individually and considered outliers.

First consider Figure 19, the plot corresponding to the conditional binomial model based on 1 observed bycatch event. The horizontal dashed lines are located at the values of 2.5 and 3.0. As expected, the predictive distribution of the mean shifts upward as effort increases; however, this shift is gradual as all the box-and-whisker plots lie between the values of 2.4 and 3.1 despite effort increasing from 1,300 to 1,525 trips. This figure shows that the mean of the predictive distribution of annual bycatch has a very high probability of falling between 2.5 and 3.0 for different levels of projected values of effort between 1,300 and 1,525 trips.

Next consider Figure 20, the plot corresponding to the conditional binomial model based on 0 observed bycatch events. The horizontal dashed lines are located at the values of 1.1 and

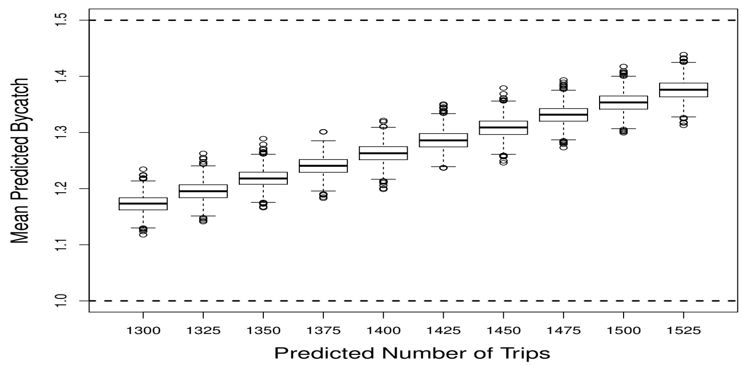

FIGURE 20. Distributions of the mean of the predictive distribution of future annual humpback whale bycatch (0 observed bycatch events between 2008-2012) based on the conditional binomial model. The horizontal dash lines bound the values of 1.0 and 1.5.

1.5. As expected, the distribution of the mean shifts upward as effort increases; however, this shift is gradual as all the box-and-whisker plots lie between the values of 1.0 and 1.5 despite effort increasing from 1,300 to 1,525. This figure shows that the mean of the predictive distribution of annual bycatch has a very high probability of falling between 1.15 and 1.40 for different levels of projected values of effort between 1,300 and 1,525.

Now consider the distribution of the median. For each of the predictive distributions based on the conditional binomial model and 1 observed bycatch, the total mass of the distribution is at 2 for all levels of projected effort. For each the predictive distributions based on the conditional binomial model and 0 observed bycatch, the total mass of the distribution is at 1 for all levels of projected effort. In summary, the values for the median of the prior predicted distribution of annual bycatch for the four species of concern are not sensitive to values of projected effort between 1,300 and 1,525 trips.

A similar study was undertaken to evaluate the sensitivity of the mean and median of the prior predictive distribution of future annual DSI to levels of projected annual effort ranging from 1,300 to 1,525 trips. The same algorithm was used as in the study concerning future annual bycatch, except that after generating the prior predictive distribution of annual bycatch, the prior predictive distribution of annual DSI was generated as described in Section 4.3. The mean and median of the 10,000 draws were recorded as they represent random draws from the distributions of the mean and median of the prior predictive distribution of future annual DSI.

Figures 21 and 22 are box-and-whisker plots of the distribution of the mean of the predictive distribution of future annual DSI. For sperm whales, Figure 21 shows that the mean of the predictive distribution of future annual DSI has a very high probability of falling between 1.2 and 1.5 for different levels of projected values of effort between 1,300 and 1,525 trips.

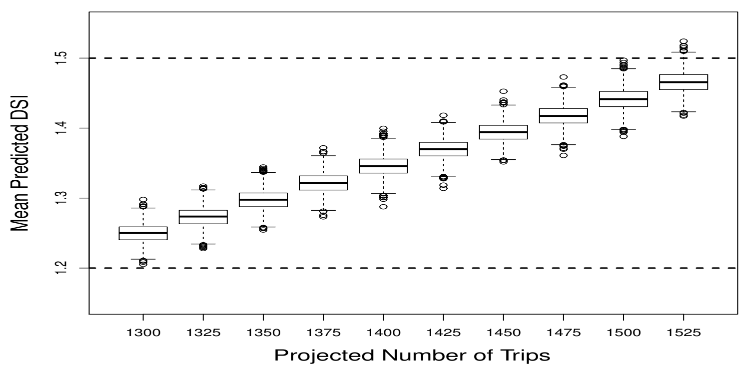

Figure 21. Distributions of the mean of the predicted distribution of future annual sperm whale DSI (1 observed bycatch event between 2008-2012) based on the conditional binomial model. The horizontal dash lines bound the values of 1.2 and 1.5.

Similarly, Figure 22 shows that the mean of the predictive distribution of annual humpback DSI has a very high probability of falling between 0.55 and 0.70. The total mass of the distributions of the median of the predictive distribution of annual DSI of sperm whales are at the value 1 for all levels of projected effort, and for humpback whales, the total mass of the distributions of the median are at 0 for all levels of projected effort.

Appendix B. Partition of predicted IFKW bycatch that overlaps NWHIFKW range

In addition to the predictions for the IFKW, management requested that these predictions be split between two areas: (1) the portion of the open area that falls within the proposed boundaries of the Northwestern Hawaiian Islands false killer whale (NWHIFKW) stock and (2) the portion of the open area that does not overlap the proposed boundaries of the NWHIFKW stock. Figure 23 shows the relevant boundaries. Based on the set and haul begin and end locations within the open area that were part of the adjusted systematic sample, the average yearly proportion of locations that were in the area of the NWHIFKW stock boundaries was 3.3%. This implies that for the assumed 600 thousand hooks within the open area, an estimated 20,000 would be within the NWHIFKW stock boundaries and 580,000 would be outside the NWHIFKW stock boundaries. Using these amounts of effort, the predictive densities of $\tilde{\tau}_{IFKW}$ and $\tilde{\delta}_{IFKW}$ suggest a mean of 0.29 IFKW bycatches and 0.26 IFKW DSI in the area outside the boundaries of the NWHIFKW stock boundaries, and an average of 0.01 bycatches and 0.01 DSI in the area within the NWHIFKW stock boundary. These are imprecise estimates based only on distribution of effort within the referenced boundaries, not on interactions between NWHIFKW and IFKW stocks. Therefore, caution should be exercised when applying these estimates to bycatch projections.

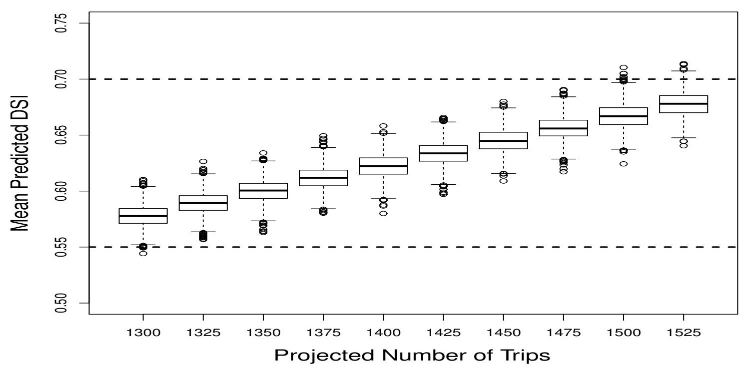

FIGURE 22. Distributions of the mean of the predictive distribution of future annual humpback whale DSI (0 observed bycatch events between 2008-2012) based on the conditional binomial model. The horizontal dash lines bound the values of 0.55 and 0.7.

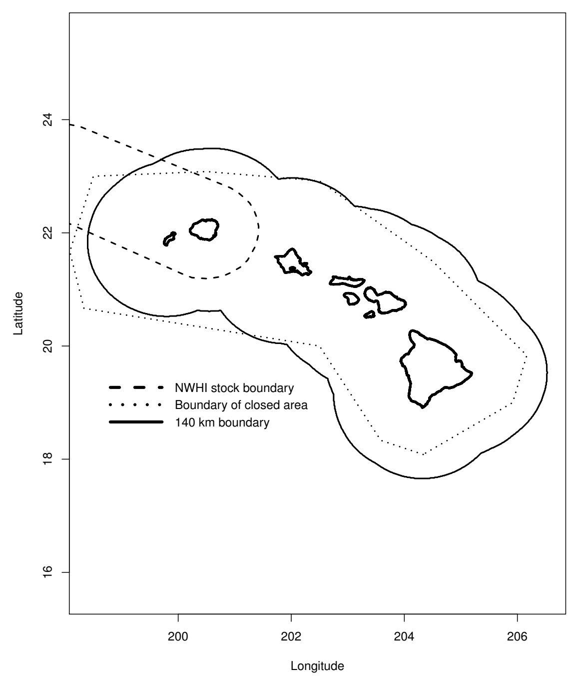

FIGURE 23. Boundaries for the proposed NWHI stock, closed area, and the 140 km. The overlap area between the NWHI stock boundary and the area within the 140 km boundary that remains open is the triangular shape region in the upper left corner.

References

Albert, A. (2009), *Bayesian Computation with R*, 2nd edn, New York: Springer.

Bradford, A. L., Forney, K. A., Oleson, E. M., and Barlow, J. (2012), Line-transect abundance estimates of false killer whales (Pseudorca crassidens) in the pelagic region of the Hawaiian exclusive economic zone and in the insular waters of the Northwestern Hawaiian Islands, Administrative Report H-12-02, Pacific Islands Fisheries Science Center, National Marine Fisheries Service, Honolulu, Hawaii.

Carretta, J. V., Forney, K. A., Lowry, M. S., Barlow, J., Baker, J., Johnson, D., Hanson, B., L, B. R., J, R., K, M. D., K, R., M, M. M., D, L., and L, C. (2010), U.S Pacific Marine Mammal Stock Assessments: 2009, NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-453, U.S. Department of Commerce.

Carretta, J. V., Forney, K. A., Oleson, E., Martien, K., M, M. M., Lowry, M. S., Barlow, J., Baker, J., Johnson, D., Hanson, B., Lynch, D., Carswell, L., L, B. R., J, R., K, M. D., K, R., and Hill, M. C. (2011), U.S. Pacific Marine Mammal Stock Assessments: 2011, NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-488, U.S. Department of Commerce, La Jolla, California.

Gelman, A., Carlin, J. B., Stern, S. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, 2nd edn, New York: Chapman and Hall.

Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, 2nd edn, Haboken: Wiley.

Lohr, S. L. (2010), *Sampling: Design and Analysis*, 2nd edn, Boston: Brooks/Cole.

McCracken, M. L. (2010), Adjustments to False Killer Whale and Short-finned Pilot Whale Bycatch Estimates, Working Paper WP-10-007, Pacific Islands Science Center, National Marine Fisheries Service, Honolulu, Hawaii.

NMFS (2012*a*), Process for Distinguishing Serious from Non-Serious Injury of Marine Mammals, Policy Directive 02-038, U.S. Department of Commerce.

NMFS (2012*b*), Taking of Marine Mammals Incidental to Commercial Fishing Operations; False Killer Whale Take Reduction Place; Final Rule, Federal Register Vol. 77, No. 230, U.S. Department of Commerce.

NMFS (2012*c*), Western Pacific Pelagic Fisheries: Revised Swordfish Trip Limits in the Hawaii Deep-set Longline Fishery; Final Rule, Federal Register Vol. 77, FR 43721, U.S. Department of Commerce.

NMFS (2013), Biological Evaluation of the Hawaii Deep-set Longline Fishery, Technical report, Pacific Islands Regional Office, National Marine Fisheries Service, Honolulu, Hawaii.

Pacific Islands Regional Office Observer Progam (2013), *Hawaii Longline Observer Program Field Manual: LM.13.02*, Pacific Islands Regional Office, National Marine Fisheries Service, Honolulu, Hawaii.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*