DOCUMENT SAC-14-10

ENHANCED MONITORING PROGRAM FOR BIGEYE TUNA CATCHES: PRELIMINARY RESULTS OF PILOT STUDY AND WORKPLAN FOR 2023

Cleridy E. Lennert-Cody, Cristina De La Cadena, Luis Chompoy, Ernesto Altamirano Nieto, Nickolas W. Vogel, Brad A. Wiley, Mark N. Maunder, Alexandre Aires-da-Silva

CONTENTS

SUMMARY 1
BACKGROUND 4
A. EMP PILOT STUDY 4
A.1 Phase 1: Simulations to test sampling designs 4
A.2 Phase 2: field test of the sampling design selected in Phase 1 14
A.3 Comparison of EMP and RDL estimates 15
A.4 Best Scientific Estimate (BSE) methodology for BET catch per trip 17
B. WORKPLAN FOR 2023 17
ACKNOWLEDGEMENTS 19
REFERENCES 19

SUMMARY

Resolution C-21-04, adopted by the Commission in late 2021, contained new measures intended to address conservation concerns related to bigeye tuna (BET), including the establishment of thresholds on individual purse-seine vessel annual BET catches that will trigger additional closure days for a vessel. To help monitor vessel catches relative to these thresholds, the resolution created an Enhanced Monitoring Program (EMP), which included a pilot study to develop and test port-sampling protocols for the estimation of trip-level BET catch. This document presents details of the implementation and results of that pilot study, which was conducted in the ports of Manta and Posorja, Ecuador, during September 2022 – February 2023. The main results of the EMP pilot study can be summarized as follows:

- Intensive sampling of catch composition during the unloading of wells with catches from floating-object sets found large-scale systematic variation in the proportion of BET over the course of unloading of individual wells. This large-scale variation, which needs to be addressed by the EMP within-well sampling protocol, was related, in part, to the number of sets from which catch was loaded into the well.

- A simulation study using the sample data determined that a systematic sampling protocol with 3.33% coverage of units of fish[1] unloaded from a well should be a reasonable compromise between low error and practicality. In actual implementation, this within-well sampling protocol is as follows: sample one out of every 30 units of fish unloaded from a well, from the beginning to the end of the unloading of the well, starting at a randomly selected unit in the first 30 units unloaded.
- Among-well variance was estimated to be roughly an order of magnitude greater than within-well variance, providing support for a decision to not allocate additional resources to estimate within-well variance at this time.
- Taking into consideration results from a second simulation study, which was used to determine the number of wells to sample per trip, the following two-stage sampling protocol will be used by the EMP: 1) at least 6 wells will be sampled per trip, selected at random for the primary catch stratum (or strata) of interest; and, 2) one systematic sample will be collected per well, using the protocol described above, where for each unit of fish sampled, the species identification and length or weight will be obtained for every tropical tuna in the unit.
- This preliminary EMP protocol, which was tested during the latter part of the pilot study, produced reasonably reliable estimates of trip-level BET catch for the primary catch strata of interest, with coefficients of variation largely between 0.22 and 0.39.
- A comparison of BET estimates from observer data to estimates based on the EMP protocol, for the same trips, suggested that the EMP estimates should be more reliable than observer estimates, and that there may be a tendency of some observers to systematically underestimate BET catch.
- Finally, as regards the use of EMP data in computing the Best Scientific Estimates (BSEs) of trip-level BET catch, a comparison of observer BET estimates to cannery BET estimates, using historical data, showed that the cannery estimates were typically smaller than the observer estimates for the same trips. Following on these results, it is inferred that: 1) EMP estimates of BET catch will be more reliable than cannery estimates; and, 2) observer estimates would be preferred over cannery estimates, when available. (A direct comparison of EMP and cannery estimates for trips sampled during the latter part of the pilot study was not possible because cannery data were not yet available for every trip.)

Given the results of the EMP pilot study, it is considered that those data, as well as data collected under the EMP, will be useful for other scientific investigations in support of tuna management, including studies on possible improvements to the sampling protocol of the Commission's regular port-sampling program for estimation of fleet-level catch composition.

The EMP workplan for the remainder of 2023 is outlined at the end of this document. This workplan includes proposals for sensitivity analyses that may lead to refinement of the current EMP sampling protocol. Details of the logistical aspects of the pilot study can be found in Document SAC-14 INF-I.

BACKGROUND

The staff of the IATTC has been collecting data in port on commercial tuna purse-seine vessel catches from the eastern Pacific Ocean (EPO) since 1953 (Tomlinson *et al*. 1992; Suter 2010). Prior to 2000, the data

[1] A unit of fish is defined as either a physical container full of fish or a fixed number of fish unloaded individually from the well, where the fixed number of fish can be thought of as a virtual unit.

collection focused on port-sampling for length composition, by species. In 2000, the data collection was expanded to include port-sampling for species composition (Tomlinson 2002; 2004). Under this regular port-sampling program, the IATTC staff obtains information on the catches of vessels at the end of their trips, including the operational characteristics associated with the wells where the catches are stored. Trips and wells are selected for sampling based on several operational characteristics, such as the set type and area of fishing associated with the catches in each well. Typically, one or two wells per trip are sampled. The current sampling protocol (see the Appendix of Suter 2010) generally involves alternating between counts of a fixed number of fish for species composition and measurement of the fork lengths of a separate, fixed number of fish for length composition, until predetermined totals of fish counts and fish lengths have been reached. These port-sampling data are used in the estimation of the fleet-level catch composition, and in the stock assessments conducted for EPO tuna populations.

Resolution C-21-04, adopted by the Commission at the 98th Meeting of the IATTC in late October 2021, established an Enhanced Monitoring Program (EMP) in order to strengthen the monitoring of bigeye tuna catches (BET; *Thunnus obesus*) of purse-seine vessels, and as support to vessel owners and captains in monitoring their catches, for better compliance with the objectives of the resolution. The conservation measures for tropical tunas during the period 2022-2024 established in the resolution include an extended closure, with potentially additional days based on vessel-specific BET catch amounts[2]. However, the process for vessel-specific catch attribution of BET under the resolution was different in 2022 than what will take place in 2023 and 2024. For 2022, flag CPCs were responsible for the per-vessel attribution of BET catch using the data sources available to them, including observer data, logbook data, and cannery unloading records. For 2023 and 2024, following the completion of the EMP pilot study, the resolution instructs IATTC staff to provide to flag authorities, Best Scientific Estimates (BSEs) of the catch of BET by each vessel following the completion of each fishing trip, using these same data sources, plus the results of EMP sampling, where available. These BSEs are then shared with the flag authorities for confirmation or adjustment before per-trip attribution is finalized. This difference in process between 2022 *versus* 2023 and 2024 was necessary in order to allow the IATTC staff time to conduct the EMP pilot study for the purpose of developing and field-testing port-sampling protocols that are tailored to the estimation of trip-level catch composition.

The sampling protocol used in the EMP must necessarily be different from that used by the regular IATTC port-sampling program because the new conservation measures outlined in Resolution C-21-04 apply at the trip level for individual vessels, whereas the regular port-sampling program takes neither vessels nor trips expressly into consideration in its sampling protocol. The additional conservation measures outlined in Resolution C-21-04 are intended to promote vessel-specific and trip-specific modifications to the normal fishing practices of individual vessels by assigning additional days of closure to vessels that exceed certain thresholds of BET catch. The nature and effect of any modifications to fishing behavior may vary through time for an individual vessel as it approaches a particular BET catch threshold. Therefore, the sampling protocol used by the EMP must focus on collection of data appropriate for estimating trip-level catch composition.

---

[2] Per paragraph 5 of Resolution C-21-04, the BET catch thresholds are as follows: *"… For the years 2023 and 2024, CPCs shall ensure that vessels that exceeded during the previous year the annual catch limit of 1,200 metric tons of bigeye tuna shall increase during the following year by 10 additional days the closure period established in paragraph 3 of this resolution. If during this same period a vessel exceeds the annual catch limit of 1,500 metric tons of bigeye tuna, they shall increase the closure by 13 days; if it exceeds the annual catch limit of 1,800 tons of bigeye tuna, it shall increase its closure by 16 days; if it exceeds the annual catch limit of 2,100 metric tons, it shall increase its closure by 19 days; and if it exceeds the annual catch limit of 2,400 metric tons, it shall increase its closure by 22 days, in addition to the closure stipulated in paragraph 3 of this resolution. …"*

The EMP pilot study occurred in two phases. In Phase 1, data were collected for a simulation study to test sampling designs. The sampling conducted during Phase 1 focused on intensive sampling of catch from specific wells, which was necessary to identify and characterize any trends in the species and size composition of the catch as it is unloaded from a well. Previous IATTC staff research (Wild 1994; SAC-13 INF-E) indicated that trends in catch composition during unloading might occur. Such trends, if present, could lead to bias in the estimates of catch composition at both the well level and the trip level, if not taken into consideration when developing the sampling design. In Phase 2, the sampling design developed in Phase 1 was field-tested to identify and resolve any logistical problems prior to initiation of the EMP in 2023. The field-testing of Phase 2 involved sampling of multiple wells of the same trip, with the same protocol as will be implemented by the EMP in 2023 and 2024. The data collected during Phase 2 were used to estimate the BET catch of those sampled trips. Those estimates were compared to observer estimates for the same trips to assist the IATTC staff in its determination of the preferred data source for the BSE of catch composition for trips not sampled by the EMP.

This document summarizes preliminary results from the EMP pilot study (Section A) and presents a workplan for the remainder of 2023 (Section B). In Section A.1, results of data analyses and simulation studies to test sampling designs, both for fish in a well and wells of a trip, are presented. Section A.2 presents results from field testing of the sampling design selected using the simulation studies of Section A.1. In Section A.3, BET catch estimates are compared to observer estimates for the same wells and trips. Section A.4 describes the planned use of the EMP data for computing trip-level BET BSEs. Given that the EMP began in early March 2023, Section B discusses work that will be undertaken to evaluate possible refinements to the current EMP sampling protocol, including sensitivity analyses for the simulation studies presented in Sections A.1 and A.2.

A. EMP PILOT STUDY

The EMP pilot study was conducted over a 6-month period from September 2022 to February 2023. It was initially scheduled to occur in July – December 2022 (SAC-13 INF-E), but was delayed by two months. As a result, the intensive sampling and simulation work of Phase 1 took place during September – December 2022, and Phase 2, the field testing of the sampling design selected in Phase 1, occurred during January – February 2023. The EMP pilot study took place in Manta and Posorja, Ecuador, due to the large proportion of fleet BET catches unloaded in these ports (SAC-13 INF-L), and the time frame, complexity and budget constraints of the EMP pilot study, which precluded travel to other countries. The selection of trips and wells to be sampled was based on the information contained in the Set Summary data submitted by observers prior to the arrival of vessels in port. All sampling conducted for both phases of the pilot study had a goal of minimizing any negative impacts on the unloading process and catch quality. Additional details regarding data collection protocols, including the sampling instructions and videos of data collection, can be found in SAC-14 INF-I. All statistical analyses and simulations were implemented with the statistical freeware R (R Core Team 2021).

A.1 Phase 1: Simulations to test sampling designs

A.1.1 Phase 1 data collection

Intensive sampling of fish in a well

Trips and wells were selected opportunistically during Phase 1 of the pilot study, with priority placed on sampling trips with BET catch reported by observers and vessels with historically high annual BET catches, in order to arrive at an effective sampling protocol for estimation of BET catch. A decision was made early in Phase 1 to prioritize sampling wells with catch from floating-object (OBJ) sets, particularly those with catch made west of 110°W, over wells with catch from other set types; wells with catch from multiple set types were a low priority for sampling. This decision was motivated by the fact that since 2015, on average, 68% of the annual BET catches reported by observers or in vessel logbooks, for all set types and vessel size classes, came from OBJ sets made west of 110°W (Table 1).

Early on in Phase 1, it was observed that there were two main methods for catch unloading in Manta and Posorja (SAC-14 INF-I). For 89% of vessels surveyed to date (76 out of 85 vessels), the catch in many or all of their wells[3] is unloaded entirely through the well head, either with various forms of small containers in a dry unloading (where chilled brine is pumped out of the well) or as individual fish using the floating properties of the fish in wells with chilled brine. These types of unloadings will, hereafter, be referred to collectively as 'standard' unloadings. With standard unloadings, samplers have access to the catch at the well head, in a safe working environment. The other unloading method, hereafter referred to as a 'cargo net' unloading, involves the use of large cargo nets filled within the well, and transported either directly to a container on the dock or to the main deck of the vessel. An important distinction between standard and cargo net methods is that, without special assistance from the unloading company, sampling of cargo net unloadings was only possible on the main deck of the vessel, which is a challenging working environment with greater safety concerns. In addition, with cargo net unloadings, sampling could not be as intensive due to the quantity of fish in the cargo net; only a limited number of fish could be sampled, regardless of the cargo net's capacity. As a result, only a small number of cargo net unloadings were sampled during Phase 1, and no simulation studies were conducted using data from cargo net unloadings.

The Phase 1 sampling protocol for standard unloadings involved sampling every 10th 'unit' of fish unloaded from a well ("1-out-of-10" protocol), from the start to the end of the unloading (Figure 1). A unit was defined as either a physical container full of fish or a fixed number of fish unloaded individually from the well, where the fixed number of fish can be thought of as a "virtual unit". Virtual units were only used when fish were being unloaded individually from the well, as occurs during flotation or at the beginning of a dry unloading when the unloaders are creating space inside the well. It was necessary to create virtual units in order to try to maintain a consistent level of sampling coverage over the entire unloading. The number of fish used to define a virtual unit was fixed for the entire unloading, and was determined by the samplers at the start of the unloading, by filling the type of container to be used during unloading with fish. This protocol resulted in samples with 10% coverage of units. The sampling started from a randomly selected unit in the first 10 units unloaded from the well, and then every 10th unit thereafter was sampled until the unloading of the well was complete. Every fish in each sampled unit was identified to one of four species groups: BET; yellowfin tuna (YFT; *Thunnus albacares*); skipjack tuna (SKJ; *Katsuwonus pelamis*); or, other fishes. In addition, the fork length (to the nearest millimeter) was measured for all BET, YFT and SKJ in every sampled unit. Intensive sampling at a higher frequency was attempted, sampling one out of every 8 units (12.5% coverage of units; "1-out-of-8" protocol), but was found to be too challenging to implement with a team of 4 samplers when small fish dominated the catch (SAC-14 INF-I).

Intensive sampling data for a total of 71 fully-sampled wells were collected for standard unloadings (Table 2), of which 8 wells were sampled using the 1-out-of-8 protocol and the other 63 wells were sampled using the 1-out-of-10 protocol. The intensive sampling required that the 4-person sampling team was present for the entire unloading, which typically lasted 1-2 days per well, most commonly taking place in 12-hour shifts, from early morning to evening (Figure 2). Per well, typically 30 units or more were sampled and more than 1000 tunas were measured (Figure 2). The 71 wells sampled represent 42 trips of 37 vessels.

[3] Of the 76 vessels that use the standard method of unloading, 64 vessels use it for all of their wells and 12 use it for many of their wells.

Various methods for collecting data from cargo net unloadings were explored during Phase 1 (SAC-14 INF-I). Of the 8 wells for which it was possible to sample the entire unloading (Table 2), the sampling protocol used was either sampling 10% of the fish from each cargo net (3 wells) or a fixed number of fish (e.g., 30 or 40 fish) from every other cargo net (4 wells) or every 10th cargo net (1 well).

A.1.1 Phase 1 data collection

Catch composition for every well of a trip

Although attempts were made to sample multiple wells per trip during Phase 1, it soon became clear that it would not be possible to intensively sample more than a few wells per trip and still allow for sampling of a diverse collection of vessels. However, data from only a few wells per trip was deemed insufficient for conducting simulation studies, given that Class-6 vessels typically have more than 10 wells (SAC-13 INF-E). Therefore, the emphasis of the intensive sampling in Phase 1 was placed on sampling a few wells from as many vessels as possible to try to obtain a data set representative of vessels most likely to be sampled under the EMP, and observer data were used for the simulation study on the number of wells to sample per trip.

Catch estimates for every well of a trip, along with the set types and fishing locations of the corresponding sets from which catch was taken to fill the well, were required for the simulation study. A data set with this information was constructed from data on the permanent observer data base and the data base that contains information recorded by observers on Tuna Tracking forms (TTF). The TTF database is the only database where edited catch data for each set that went into a well are recorded. Therefore, to obtain the necessary data set for the simulations, a computer program was written to match the TTF database to the permanent observer database[4], using the trip numbers and dates (set numbers are not available in the TTF database). For the period 2000 to 2021, only 2.1% of wells in the TTF could not be linked to specific sets in the permanent observer database. The data for years 2017 – 2019 and OBJ sets made west of 110°W were used for the simulation analysis because they represent the most recent pre-pandemic years and the area and set type that has produced much of the annual purse-seine BET catch (Table 1). This data set will include data from trips of vessels smaller than IATTC Class-6, if they carried an observer. Hereafter, this data set will be referred to as "reconstructed RDL observer data".

It is noted that all the information in the constructed data set described in the previous paragraph is available from a preliminary observer data source referred to as the Set Summary or, in Spanish, *Resumen de Lances* (RDL). RDL data, which are available before the vessel unloads in port, are considered preliminary information because they are not subject to the full IATTC observer program data quality control process. After incorporation of edited observer data into the IATTC databases, the information on the RDL can be obtained from edited sources; however, the time frame for this varies. For IATTC-observed trips, the incorporation of edited data is typically complete within a few months of the vessel's arrival to port. For trips covered by observers of national programs, the incorporation of edited data may require more time. Thus, although the RDL data are not subject to the full IATTC observer program data quality control process, at the completion of a fishing trip, they are the only vessel-independent data immediately available on well-specific catch composition, set type composition, and fishing locations, for every well of a trip, and thus, for some purposes, RDL data must be used. In particular, the RDL data will be used for estimating the trip-level BET BSEs (see Section A.4) because of the time frame within which it is anticipated those estimates must be produced following completion of a trip (per Resolution C-21-04). Comparisons

[4] The permanent observer data base is the data base that contains all data collected by observers that have passed through the full IATTC observer program data quality control process.

of RDL data and data from the permanent observer data base have found little or no difference in tuna catch.

A.1.2 Phase 1 exploratory data analysis

Variability among sampled units of the same well

Exploratory analyses were conducted to determine if patterns in catch composition occur over the course of unloading a well, with a focus on variability in the proportion of BET. For each sampled unit, two summary measures of species composition were computed: 1) the proportion of each of the three tropical tuna species in the unit, computed from numbers of fish; and, 2) the proportion of each of the three tropical tuna species in the unit, computed from weight. For the latter, the fork length (in millimeters) of each fish was converted to weight (in kilograms) using the same length-weight relationships as those used to obtain the fleet-level estimates of catch composition. There were 11 fish for which the recorded length data appeared to be in error, out of 147,354 fish (0.0075%) measured. These fish were assigned the average length for their species in the unit where they occurred.

For standard unloadings, preliminary results indicate that there can be considerable variability in the proportion of BET per unit over the course of unloading a well (Figures 3 – 4). For example, different wells with a similar average proportion of BET had different patterns in the proportion of BET per unit, even for wells of the same trip (e.g., in Figure 3, compare the wells of Trips 2 and 3). It was found that the proportion of BET per unit could be high at the start, middle, or end of the unloading (Figure 3). These patterns were evident in both the proportion of BET from numbers and that from weight, which indicates that this variability is not an artifact of converting from numbers to weight. This type of variability was commonly seen in EMP pilot study data (Figure 4). Because the trips and wells sampled in Phase 1 were not selected at random, inference cannot be made about the prevalence of these patterns in the fleet more generally. However, any within-well sampling design must be robust to this type of variability to avoid biasing the estimate of the proportion of BET in a well and for the trip.

Although the level of variability in the proportion of BET per unit is likely due to multiple factors, the number of sets associated with the catch in the well appears to be of importance (Figure 5). For those wells with any catch of BET, when the catch in a sampled well came from only one set, the variability in the proportion of BET among sampled units of the well was significantly smaller than when the catch in the sampled well came from multiple sets [5]. A qualitative comparison of set-specific catch composition to variability in the proportion of BET per sampled unit suggests that for those wells with catch from multiple sets, variability in catch composition among sets may contribute to variability in catch composition during unloading of the well (Figure 5). Given that not all sets have similar amounts of BET catch, even for trips with larger amounts of BET (Section A.1.2), future analyses will quantitatively evaluate the relationship between the variability in the set-specific catch amounts of BET and variability in the proportion of BET per unit during unloading (see Section B).

The sample data available for wells unloaded with cargo nets, although limited, indicate that the well-level sampling protocol to be used for unloadings of this type may need to be similar to that used for standard unloadings. For the 8 sampled wells from 7 different trips (7 different vessels; Figure 6), similar patterns in the proportion of BET over the unloading were seen (Figure 7) as for wells of standard

[5] A permutation test was applied to the data, using the R package *lmPerm* (Wheeler and Torchiano 2016). A one-way analysis of variance model was fitted to test whether the interquartile range (IQR) was the same when the well contained catch from only one set *versus* catch from multiple sets. The *p*-value for this test was 0.008. A permutation test based on fitting a simple linear model to the same data (i.e., IQR = a + b*number of sets) yielded a p-value for the slope of 0.063.

unloadings (Figure 3). In particular, for the four wells that involved sampling 30 or 40 fish from every other cargo net, two had large-scale pattern in the proportion of BET across the unloading.

A.1.2 Phase 1 exploratory data analysis

Variability among wells of the same trip

Based on analysis of the reconstructed RDL observer data for 2017 – 2019, considerable variability was found in the proportion of BET among wells of a trip[6], for many trips, including those that caught larger amounts of BET (Figure 8). For the large-BET catch trips, the interquartile range[7] of the proportion of BET per well often fell between 0.2 and 0.6, even though those same trips had wells with a proportion of BET as small as 0.05 and as large as around 0.8 or greater. Similar patterns were seen in all three years. This level of variability in the proportion of BET among wells of the same trip indicates that the number of wells sampled per trip will be an important consideration in sampling design development. Variability in BET catch per set within a trip (Figure 9), likely contributes to the variability in the proportion of BET among wells of the same trip, given that catch from multiple sets can be loaded into a single well.

Main findings – exploratory data analyses

- The proportion of BET encountered during unloading of a well can be high at the start, middle, or end of the unloading. The within-well sampling design must be robust to this type of variability to avoid biasing the estimate of the proportion of BET in a well and for the trip.
- There can be considerable variability in the proportion of BET among wells of the same trip, which indicates that the number of wells sampled per trip will be an important consideration in sampling design development.

A.1.3 Phase 1 simulation study

Simulation studies were conducted with the well-level and trip-level data described in Section A.1.1 to evaluate the performance of different sampling design options for the two stages of the sampling: 1) sampling wells of a trip; and, 2) sampling units from a well. It is noted that because standard unloadings often involve unloading much of the catch in the well with containers, a sampling protocol based on sampling individual fish would require additional unloader assistance. Therefore, such an approach was not considered for the EMP (see Section B).

In order to evaluate sampling design performance, it was necessary to first determine an estimator of the proportion of BET, and the BET catch amount, for a trip, and their variances. As noted in previous studies (Wild 1994), simple random sampling of fish from a well is not considered practical for logistical reasons, and thus it is assumed that the overall sampling protocol should be a two-stage sampling process, where the first stage 1 (selection of wells from a trip) is stratified random sampling, and the second stage (selection of units from a well) is systematic sampling, starting from a randomly selected first unit. This section is structured as follows: the estimators and their variances are presented in Section A.1.3.1; the simulations testing within-well protocols in Section A.1.3.2; and, the simulations testing protocols for the number of wells per trip in Section A.1.3.3.

---

[6] In terms of amounts of BET per well, across all trips, for OBJ sets made west of 110°W, values ranged from 0 t to maximum of 85 t in 2017 (IQR: 4 t, 23 t), 92 t in 2018 (IQR: 6 t, 26 t), and 124 t in 2019 (IQR: 7 t, 26 t). (To note is that the maximum amount of catch per well is limited by well capacity.) IQR: interquartile range.

[7] The interquartile range is the difference between the two values that correspond to the 75th and 25th percentiles of the data. That is, the interquartile range is defined as the middle 50% of the values of the data, when sorted from smallest to largest.

A.1.3.1 *Catch estimation methodology*

The trip-level estimator of catch for a species that will be used is a stratified estimator, where the strata of a trip are defined by the set types and areas of fishing associated with the catch of the trip. The estimator can be written as

$$\widehat{W}_s = \sum_{t=1}^{T} W_t \, \hat{p}_{t\_s} \qquad (1)$$

where $\widehat{W}_s$ is the estimated catch of species *s* for the trip (in weight), W_t is the weight of tropical tunas (YFT + SKJ + BET) for stratum *t* of the trip (assumed known), *T* is the total number of strata for the trip, and $\hat{p}_{t\_s}$ is the estimated proportion of the catch in stratum *t* of species *s* (from weight). In the case of estimating the species proportion and catch for an individual well, then $T = t = 1$.The data sources used to obtain W_t will depend on availability at the time a catch estimate is required. For trips that carried an observer, W_t will be either the tropical tuna catch reported by the observer for the stratum, or it will be computed from the total tropical tuna catch reported by the cannery, prorated to strata using the observer's data (cannery data provided to IATTC do not contain detailed information on trip-level operational characteristics). For trips that did not carry an observer, W_t will be either the tropical tuna catch reported in the vessel's logbook for the stratum, or it will be computed from the total tropical tuna catch reported by the cannery, prorated to strata using the vessel's logbook data. Set type and area of fishing were used to define *t*. For the area of fishing, the EPO was divided into three regions: west of 110°W; 110°W – 95°W; and, east of 95°W.

The estimate of the species proportion of the catch in stratum *t* is given by:

$$\hat{p}_{t\_s} = \frac{\sum_{i=1}^{n} \left(\frac{M_i}{m_i}\right) \sum_{j}^{m_i} y_{tsij}}{\sum_{i=1}^{n} \left(\frac{M_i}{m_i}\right) \sum_{j}^{m_i} x_{tij}} \qquad (2)$$

where y_{tsij} is the sum of the weight of fish of species *s* in the j^{th} systematic (cluster) sample of the i^{th} well, x_{tij} is the weight of tropical tunas in the j^{th} cluster sample of the i^{th} well (i.e., equal to the sum over *s* of y_{tsij}), m_i is the number of cluster samples taken from the i^{th} well, M_i is the number of possible cluster samples for the i^{th} well, and *n* is the number of sampled wells of the trip for stratum *t*. The quantity $\frac{M_i}{m_i}$ is the inverse of the fraction of possible cluster samples for the well that were sampled, and thus raises the sample data to a total for the well. This estimator is adapted from Cochran (1977, pages 292 – 326) (Lennert-Cody et al. 2022).

Because W_t is in weight, yet the Phase 1 data are in numbers of fish, fish length must be used to convert from numbers to weight. The weight of each fish in the sample will be estimated from its measured length (in mm) and the species-specific length-weight relationships used to estimate the fleet-level catch composition. At present, the weight of each fish will be assumed known when estimating the variance of the species catch; however, future work on methods development could include incorporation of error from the length-weight conversion into the variance on the species catch.

It is noted that in the EMP program, fish weight may be obtained, instead of length, for most fish of a sample. During the last part of Phase 2 of the EMP pilot study, the use of portable scales was tested, because collection of weight data will not only mitigate problems associated with accurately measuring damaged and/or curved fish, but may also be more efficient, requiring fewer samplers per well. Assuming scales will be used in the EMP, conversion to weight will not be necessary, except for the largest fish whose weight exceeds the capacity of the scale (and must therefore be measured), or in the case where use of the scale is not possible for logistical reasons (e.g., inadequate level space near the well head). The use of scales for the second stage sampling is discussed in detail in SAC-14 INF-I. It is noted here that when scales are used, the size composition data collected by the EMP will be different from that currently collected by the IATTC regular port-sampling program, which collects length data. Historically, length data have been collected because the tuna stock assessments require estimates of the length composition of the catch. A topic of future work will be to determine whether collection of weight data instead of length data, more generally, may be beneficial, especially if up-to-date length-weight relationships become available. (See SAC-14 INF-J for the need and opportunities for updating morphometric relationships for priority species in the EPO.)

The estimator of the variance of the catch of species *s* for a trip, assuming strata are independent (and the W_t are known), is given by:

$$Var(\widehat{W}_s) = \sum_{t=1}^{T} {W_t}^2 \, Var(\hat{p}_{t\_s}) \qquad (3)$$

To estimate the variance of $\hat{p}_{t\_s}$ we consider methods for the ratio of two random variables, where cluster samples are selected with equal probability but are of unequal size (see Cochran 1977, pages 303 – 306 and 311 – 313). With simple random sampling of wells within a stratum, the variance of the species proportion in stratum *t* is given by equation 11.51 of Cochran (1977), adjusted to obtain the variance of a proportion, not a total, by dividing by ${W_t}^2$:

$$Var(\hat{p}_{t\_s}) = \frac{1}{{W_t}^2}\left(\frac{N^2}{n}(1-f_1)\frac{\sum_i^n\left(\hat{Y}_{tsi}-\hat{p}_{t\_s}X_{ti}\right)^2}{n-1} + \frac{N}{n}\sum_i^n \frac{M_i^2(1-f_{2i})\hat{S}_{d2i}^2}{m_i}\right) \qquad (4)$$

where

$$\hat{S}_{d2i}^2 = \left(\frac{1}{(m_i-1)}\sum_{j=1}^{m_i}\left[\left(y_{tsij}-\hat{p}_{t\_s}x_{tij}\right)-\left(\bar{\hat{Y}}_{tsi}-\hat{p}_{t\_s}\bar{\bar{X}}_{ti}\right)\right]^2\right) \qquad (5)$$

and *n* and *N* are the number of wells sampled and in the trip, respectively, for stratum *t*, $\hat{Y}_{tsi}$ is the estimated total weight of species *s* in the *i*th well (equal to the product of X_{ti} and the estimated proportion of BET for the *i*th well, which is eq. 2 applied only to the *i*th well), X_{ti} is the total weight of tropical tunas in the *i*th well, f_1 and f_{2i} are the sampling fractions for wells and cluster samples, respectively; $f_1 = \frac{n}{N}$ and $f_{2i} = \frac{m_i}{M_i}$. The means $\bar{\hat{Y}}_{tsi}$ and $\bar{\bar{X}}_{ti}$ are evaluated over the m_i clusters in the *i*th well, and represent the average weight of species *s* in a cluster sample and the average weight of tropical tunas in a cluster sample, respectively. For now, mixed stratum wells are not counted when computing *N*, but this may be revised in the future as it can be considered to underestimate the weighting *N*/*n*.

To be able to estimate $\hat{S}_{d2i}^2$ directly from the sampling design, at least two cluster samples need to be obtained per well (i.e., $m_i \geq 2$); if $m_i = 1$, then $\hat{S}_{d2i}^2$ cannot be evaluated and the second term of the equation for the variance of $\hat{p}_{t\_s}$ cannot be computed. In practice, for practical reasons, a simplifying approximation for evaluating eq. 4 is that $\hat{S}_{d2i}^2$ can be effectively ignored when n/N is small and selection at the first stage (wells of a trip) is at random (Cochran 1977). Although it is not anticipated that n/N will necessarily be small in our situation, resource limitations are expected to make obtaining $m_i > 1$ problematic. Thus, except under special circumstances, the second stage variance (eq. 5) will be ignored when evaluating eq. 4, until such time as it becomes clear that additional samples per well are possible to collect with the resources available; to compensate, the term (1- f_1) in eq. 5 will also be ignored. This may result in an overestimate of the variance (Cochran 1977). To explore the possible impact of the decision to ignore the second stage variance, in Section A.1.3.2 the Phase 1 data for those trips for which more than one well was sampled are used to compare the relative magnitude of within-well and among-well variances.

A practical aspect of defining the strata associated with the sample from an individual well is noted here. Because wells can contain catch from multiple sets, the catch of a well may come from multiple set types or multiple spatial strata. To obtain the tropical tuna catch of a well, this does not pose a problem for trips that carried an observer because the catch of each set in the RDL observer data is broken down by the wells that receive the fish. However, the sample data cannot be similarly dissected. In general, wells with mixed set types will only be sampled if the accuracy of the set types assigned to the catch in the well is in question. As regards spatial strata, the approach taken is to apply a 2° longitude buffer to the stratum definitions mentioned above. If a minor proportion of the catch in the well is within 2° of the area associated with the dominant proportion of the catch in the well, the well is considered to represent one spatial stratum. If the minor proportion of catch in the well is more than 2° outside the area boundaries associated with the dominant catch proportion, the well is considered to represent two spatial strata, and would only be sampled if the accuracy of the areas assigned to the catch in the well is in question.

A.1.3.2 Phase 1 simulation study

Methods: within-well sampling

The simulations conducted to evaluate various options for within-well sampling protocols for standard unloadings, which were based on Phase 1 data, were done by stratum. This is because estimation of the trip-level catch will be done by stratum, in anticipation of insufficient resources to sample every well of a trip. Estimating catch by stratum only requires that an adequate number of wells per stratum be sampled, which is presumably less than every well of the trip. These preliminary simulations focused on wells with catch from the set type and area of the EPO that, in recent years, have generated much of the BET (Table 1): OBJ sets made west of 110°W. A boundary of 110°W was selected to be consistent with the main spatial boundary used in the current BET stock assessment fishery definitions (Xu et al. 2020). Phase 1 data from 39 fully-sampled wells (from 21 trips), for the 1-out-of-10 protocol, were available for the simulations for this stratum. Future simulation work (Section B) will also include the Phase 1 data sampled with the 1-out-of-8 protocol.

Given the impracticality of random selection of fish from a well (Wild 1994), and the large-scale patterns in the proportion BET that can occur during unloading (Figure 3), the simulations for within-well sampling protocols focused on testing various systematic sampling designs. These simulated protocols were intended to approximate actual protocols that would be less intensive than that used in Phase 1 of the EMP pilot study (Figure 1). The sampling protocols tested (Figure 10) involved selecting one sampled unit out of every *k* sampled units. Values for *k* from 2 to 5 were tested. In all cases, the starting unit was chosen at random from among the possible candidates in the first group of sampled units (see Figure 10). For example, for a protocol selecting one out of every 3 sampled units, the first unit selected was one of the first 3 sampled units, and there were three possible systematic (cluster) samples for a well under this simulated protocol: 1) the cluster sample that contains sampled units 1, 4, 7, etc.; 2) the cluster sample that contains sampled units 2, 5, 8, etc.; and, 3) the cluster sample that contains sampled units 3, 6, 9, etc. In preliminary analyses, a protocol based on selection of contiguous blocks of sampled units was also tested, but was not found to provide any improvement over selecting single sampled units, and so was not considered further because of its increased complexity to implement.

The simulated protocols were intended to approximate actual sampling coverage of units, from 2% to 5% (Table 3). For example, the protocol selecting one out of every 2 sampled units ('1-of-2') equates to half the sampling coverage of the original data, which was 10% (Figure 1), and thus is assumed to approximate

5% coverage. Similarly, the protocol selecting one out of every 3 sampled units ('1-of-3') was intended to approximate an actual sampling coverage of 3.33%. Implicit in equating the simulation's coverage to actual coverage is the assumption that the catch composition was fairly constant within the original groups of 10 units represented by each sampled unit (Figure 1). That is, for the original data, each sampled unit is effectively assumed to have nearly identical catch composition to the other 9 units of its group that were not sampled. Future simulations (Section B) will improve upon this by constructing plausible sequences of units within each of the original 10-unit blocks represented by each sampled unit, so that more extensive simulations of within-well protocols can be conducted.

The performance of the within-well simulations was evaluated using two metrics: error and relative error. For each protocol considered, the error and relative error were computed for each cluster sample generated. Error was defined as the estimated proportion of BET in the well from the cluster sample data minus the 'true' proportion of BET in the well. The cluster sample estimate was computed from eq. 2 above (n = 1 and m_i = 1), and the assumed true proportion is that estimated from all of the original data for the well (the full 1-out-of-10 data) using eq. 2 with n = 1 and $M = m$. Relative error was defined as the error divided by the 'true' proportion of BET in the well. To summarize the error and relative error for a protocol across all cluster samples generated for a well, the absolute value was taken and then averaged over the clusters; hereafter, these will be referred to as average absolute error and average absolute relative error. The overall performance of each simulated protocol was summarized across wells by computing the percentage of wells with an average absolute error greater than 5%. For relative error, the overall summary was the percentage of wells with an average absolute relative error above 0.20. It is noted that future simulations (Section B) will evaluate sampling protocol performance using synthetic data sets to avoid the possibility of overestimating performance.

A.1.3.2 Phase 1 simulation study

Results: within-well sampling

Overall, the simulation results (Figure 11, Table 3) indicate that the lower the coverage, the higher the error and relative error, although the performance in terms of average absolute error for the 1-of-3 and 1-of-4 protocols was very similar (differing by the equivalent in percentage points of one well). The range of the largest negative and positive error values also increased with decreasing coverage (Table 3). The average absolute error increased with the true proportion of BET in the well, as might be expected, while the average absolute relative error was largest at lower true BET proportions (Figure 11). Large relative error associated with small estimated proportions is not considered a problem because the contribution to total error is small. There was no clear tendency to under or overestimate (i.e., a tendency for errors to be mostly positive or negative) (Figure 11).

Given these preliminary results, a sampling coverage of 3.33% was considered a reasonable compromise between low error and practicality (Table 3). Sampling to achieve 5% coverage would have been preferred because of the generally lower error and relative error obtained in the simulations. However, it was determined likely to be too intensive to sustain as part of the EMP. The simulated 2.5% coverage had slightly poorer performance, relative to that of the simulated 3.33%, in terms of the error range and average absolute relative error. Thus, in Phase 2 of the EMP pilot study, within-well sampling using a systematic sampling protocol with a coverage of 3.33% of units was tested. The actual implementation of this protocol was to sample one out of every 30 units, from the beginning to the end of the unloading of a well, starting at a randomly selected unit in the first 30 units unloaded from the well (see Section A.2).

For 13 of the 21 trips with wells containing catch from OBJ sets made west of 110°W, 2 – 3 wells were sampled per trip, making possible a comparison between the among-well and within-well variances for those 13 trips under the simulated 3.33% coverage protocol. For this comparison, the among-well variance component of eq. 4

$$\frac{N\ \sum_i^n \left(\hat{Y}_{tsi} - \hat{p}_{t\_s}\hat{X}_{ti}\right)^2}{n-1}$$

was compared to the with-well variance component of eq. 4

$$M\sum_i^n \hat{S}_{d2i}^2$$

where the finite population correction factors, (1- f_1) and (1-f_{2i}), were ignored, and it was assumed $N = n$ (= 2 or 3) and $M = m = 3$. The 13 ratios of among-well variance divided by within-well variance ranged from 0.6 to 459, with a median value of 10.6. This suggests that among-well variance could typically be roughly an order of magnitude greater than within-well variance, providing support for a decision to not allocate additional resources to estimate within-well variance at this time. However, it is noted that the sample size (13 trips) is small, and wells were not selected at random.

Main findings: within-well simulation study

- A systematic sampling protocol with 3.33% coverage of units of fish unloaded from a well is considered a reasonable compromise between low error and practicality. In the field, to collect one systematic sample per well, this protocol would be: sample one out of every 30 units, from the beginning to the end of the unloading of a well, starting at a randomly selected unit in the first 30 units unloaded from the well.
- Among-well variance may be roughly an order of magnitude greater than within-well variance, providing support for a decision to not allocate additional resources to estimate within-well variance at this time.

A.1.3.3 Phase 1 simulation study

Methods: number of wells to sample per trip

These simulations, which used the reconstructed RDL observer data for 2017 – 2019, focused on determining the number of wells to sample from the most important catch stratum of a trip, assuming random selection of wells. To be consistent with the well-level sample data collected in Phase 1, the data used in these simulations were limited to trips of vessels unloading in Manta or Posorja, Ecuador. For each trip, wells filled with catch from OBJ sets made west of 110°W were randomly selected, and the proportion of BET for the stratum (eq. 2[8]) and its coefficient of variation (CV) were estimated. The CV was estimated as the among-well variance, excluding the finite population correction factor (i.e., the first component of eq. 4, without (1-f_1)), divided by the estimate of the proportion of BET for the stratum. The finite population correction factor was not used in this simulation to be conservative in the estimation of variance since the source of data for this simulation (observer data) is not the same source as that which will be used in the EMP (sample data).

The number of wells simulated to be sampled ranged from 2 to 8 (2 is the minimum number of wells necessary to estimate the among-well variance component of eq. 4). The simulations were run for every trip that had at least one more well with catch from the stratum than the simulated sample size

[8] Equation 2 simplifies in this exercise to the sum of the $y_{tsi.}$ over *n*, divided by the sum of the $x_{ti.}$ over *n*.

(otherwise, there is only one sample to draw). For each simulated number of wells to sample, all possible combinations of that number of wells in the stratum for the trip were obtained, up to a limit of 100,000, to avoid computer memory management problems. For each combination of wells sampled for a trip, the proportion of BET for the stratum and its CV were estimated. Those CV values (one for each possible random sample of a fixed number of wells from the trip) were summarized by a box-and-whisker plot of the CV values for the trip.

A.1.3.3 Phase 1 simulation study

Results: number of wells to sample per trip

There were between 140 (2018) to 164 (2017) trips per year with wells that contained catch from OBJ sets made west of 110°W. The number of trips affected by the 100,000-combination limit was relatively small. For example, for 2019, at 6 wells sampled, there were 2 trips with more than 100,000 possible 6-well samples, and at 8 wells sampled, there were 13 trips with more than 100000 possible 8-wells samples. Nonetheless, future simulation work (Section B) will optimize the simulation programs to cover all possible combinations. The number of trips available for each simulation varied with the number of wells to be sampled, from 115 (2018) to 132 (2017) with 3 or more wells in the stratum (for the simulation sampling 2 wells per trip) to 70 (2018) to 88 (2017) trips with 9 or more wells in the stratum (for the simulation sampling 8 wells per trip). As an example of the number of wells per trip with catch from the stratum, see the frequency distribution for 2019 shown in Figure 12.

Preliminary simulation results indicate that for a trip with a larger number of wells with catch from OBJ sets made west of 110°W, the upper quartile of possible CV values may not fall substantially below 0.5 unless 6 or more wells are sampled (Figure 13). Sampling only two wells for a trip results in many trips with a median CV value between 0.5 and 1.0 (Figure 13). If 6 wells were sampled, the mode of the distribution of median trip-specific CV values falls below about 0.35 for each of the three years (Figure 14). Absent a specific level of precision specified in Resolution C-21-04, these results suggest that sampling 6 wells per stratum is a reasonable practical compromise between improving precision and having resources available to sample more trips. Thus, in Phase 2 of the EMP pilot study, the planned sampling was for 6 wells per stratum of a trip, with priority given to sampling wells with catch from the stratum OBJ sets made west of 110°W, and secondarily, OBJ sets made between 95°W and 110°W. If there were fewer than 6 wells for the primary stratum, all wells of the stratum were to be sampled. Simulations to evaluate the number of wells to sample per trip for catches from other strata will be evaluated in future simulation studies (Section B).

Main finding: number of wells to sample per trip

- Sampling 6 wells per stratum for a trip, for the stratum expected to produce the most BET catch, is considered a reasonable practical compromise between improving precision and having resources available to sample more trips. However, sampling more wells per trip, if resources permit, would be desirable.

A.2 Phase 2: field test of the sampling design selected in Phase 1

Following the preliminary simulation results obtained from Phase 1 (Section A.1.3), the sampling protocol selected was tested in Phase 2. That protocol was a two-stage sampling process, where the first stage (selection of wells from a catch stratum for the trip) is stratified random sampling, and the second stage (sampling within a well) is systematic sampling of units from the well, where all tuna of each selected unit are identified to species and measured (or weighed). Specifically, the protocol tested was: 1) sample 6 wells per trip, selected at random from the stratum (or strata) considered likely to generate the most BET catch; and, 2) for each of those 6 wells, collect one systematic sample, sampling one out of every 30 units per well, from a random starting unit in the first 30 units unloaded, until the end of the unloading. A practical modification to (1) can sometimes be necessary if the timing of the unloading of individual wells and/or their location precludes sampling. In such cases, an alternate well (or wells) are randomly selected. The implementation of this protocol during January – February, 2023, and preliminary results are presented in this section. This sampling protocol is currently being used in the EMP, which began in March 2023.

Phase 2: Data collection

The following criteria were used to implement the sampling protocol described in the previous paragraph. Trips (vessels) were selected based on either the trip's BET catches or the vessel's historical BET catches. The stratum generally considered the most important to sample was OBJ sets made west of 110°W, and secondarily, OBJ sets made between 95°W – 110°W. Given the time typically required to unload an individual well (Figure 2), 6 teams of 3 samplers each were assigned to each trip selected for sampling. With teams of 3 samplers, and 20 samplers in total in the EMP pilot study, it was anticipated that typically only one stratum could be sampled per trip, or two strata per trip if fewer than 6 wells of the trip contained catch from the primary stratum of interest.

Phase 2: Exploratory data analysis

A total of 10 trips were sampled during Phase 2 with the preliminary sampling protocol. It was possible to sample about 2 trips per week using this protocol. The teams of samplers sampled anywhere from 6 to 21 units per well and between 150 to 1,700 tunas per well (Figure 15). Variability in the proportion of BET among sampled units of the same well seen in the Phase 2 data (Figures 16 – 17) was similar to that seen in the intensive sampling data collected during Phase 1 (Figures 3 – 4). The proportion of BET per sampled unit could be high at any point over the course of the unloading, and quite variable among sampled units of some wells. This suggests that the lower within-well sampling coverage was not so low as to smooth over large-scale changes in catch composition that may occur during unloading.

Phase 2: Stratum-level BET estimates for sampled trips

As an illustration of the catch estimation methodology (eqs. 2, 4) with the Phase 2 data, an estimate of BET catch for the dominant stratum of each trip was made (Table 4). Estimates, standard errors, and approximate 95% confidence intervals were computed for the following strata: OBJ sets made west of 110°W (three trips), OBJ sets made between 95°W – 110°W (5 trips), and OBJ sets made east of 95°W (two trips). The CVs for the stratum estimates of the amount of BET catch were between 0.22 and 0.39 for 9 of the 10 estimates (Table 4; Figure 18), roughly consistent with the expected precision based on the among-well simulations using the reconstructed RDL observer data (Figures 13 – 14). One trip with little BET catch for the dominant stratum (estimated proportion of BET was 0.0009) had a CV of 0.74; however, this is not considered of concern because the estimated proportion BET for the stratum was very small. For four of the 10 trips, the stratum of primary interest for the trip represented over 90% of the trip catch of tropical tunas (Table 4); i.e., for these four trips, estimating the stratum catch of BET was nearly equivalent to estimating the trip catch of BET. For only two of the 10 trips was the tropical tuna catch of the primary stratum less than 60% of the trip catch. In other words, for only two the 10 trips sampled in Phase 2 would it have been necessary to rely substantially on another data source to generate the BET BSE for the trip. In the case of one trip (Trip A of Table 4), the entire catch of the trip came from OBJ sets made west of 110°W, and hence the EMP estimate for the stratum was also the EMP estimate for the trip, at 538 t of BET.

Main finding: field test of preliminary sampling protocol

- The sampling protocol selected based on Phase 1 simulations was found to produce reasonably reliable estimates of trip-level BET catch for the primary catch strata of interest, with coefficients of variation largely between 0.22 and 0.39.

A.3 Comparison of EMP and RDL estimates

Several comparisons were made between estimates of the proportion of BET from the EMP pilot study data to estimates from the RDL observer data. First, estimates of the proportion of BET per well from the intensive sampling data of Phase 1 were compared to RDL estimates for the same wells. Assuming the estimates based on the Phase 1 data represent the true proportion of BET in the well, the error associated with the RDL estimates, computed as the RDL estimate minus the EMP estimate, was compared to the error range from the Phase 1 simulation study for the 3.33% coverage protocol (Table 3). In addition, a comparison of the well-level estimates of the proportion of BET from the Phase 2 data to the RDL estimates for the same wells was also done. Finally, the stratum-level BET estimates, by trip, computed from the Phase 2 data were compared to RDL estimates for the same trips and strata.

Comparison of the proportion of BET per well from Phase 1 data to the RDL estimates for the same wells (Figure 19) indicates that there is an overall positive correlation between the estimates from the two data sources. Limiting the data to trips with more than one well sampled, and fitting a linear regression model with the square root of the RDL proportion of BET as the dependent variable, the EMP proportion of BET as the independent variable, and the trip identifier included in the model as a factor, yielded a significant positive relationship (slope estimate = 0.57, p-value < 0.01). The factor for trip was not significant at the 5% level (p-value = 0.07); however, differences in slopes among trips was not tested because only 2 – 3 wells were sampled per trip during Phase 1 (see below). The error on the RDL estimates (RDL minus EMP) for wells with catch from OBJ sets made west of 110°W (Figure 20) was skewed towards negative values, with a larger range in error than that obtained from the simulated 3.33% coverage protocol (Table 3). In other words, the range of well-level error for the RDL estimates is larger than that observed for the simulated sampling protocol. When grouped by trip, it can be seen that there are multiple trips with large negative RDL error values (Figure 20).

When more wells were sampled per trip in Phase 2, it became clear that the relationship between RDL and EMP estimates of the proportion of BET per well can vary considerably by trip (Figure 21), which may be due to a difference in abilities among observers and/or other factors. The differences in the RDL and EMP estimates for some wells of several trips, and most wells of Trips A and I, fall considerably outside of the range of the error found for the simulated 3.33% coverage protocol (Table 3), highlighting the importance of collecting EMP data. Refinements to the within-well simulation study with Phase 1 data will be implemented (Section B) to verify this result when the assumption of nearly constant catch composition within the original groups of 10 units is relaxed.

Comparison of EMP and RDL stratum-level estimates illustrates how differences in the proportion of BET at the well-level between the two data sources (Figure 21) translate into stratum-level (trip-level) differences in the estimated BET catch (Table 4). For example, for most wells of Trips E, F and H, the well-level estimates of the proportion of BET were relatively similar to the RDL well-level estimates (Figure 21), with the RDL estimates falling within the approximate 95% confidence intervals for the EMP estimates (Table 4). By contrast, for Trips A and I, the EMP well-level estimates of the proportion of BET were considerably larger than the RDL well-level estimates for most or all of the wells sampled (Figure 21), with the RDL BET estimates for the stratum well below, or near, the lower limit of the approximate 95% confidence intervals on the EMP estimates (Table 4). Given that most of the differences in the well-level estimates for these two trips fall outside the error range from the simulations (Table 3), greater credibility is given to the EMP estimates than the RDL estimates.

Main findings: comparison of EMP and RDL estimates

- The error on the RDL estimates (RDL minus EMP) for wells with catch from OBJ sets made west of 110°W was skewed towards negative values, with a larger range in error than that obtained for the simulated 3.33% coverage protocol.
- Although the EMP estimates can be larger or smaller than the RDL estimates, the most extreme differences tended to be negative (i.e., EMP estimate was considerably greater than RDL estimate), suggesting that for some trips, the RDL estimate may be negatively biased.

A.4 Best Scientific Estimate (BSE) methodology for BET catch per trip

The current plan for estimating the BET BSE by trip is outlined in this section. This plan may be revised should further analysis in 2023 (Section B) indicate that improvements can be made.

Trips sampled by the EMP

For a trip sampled by the EMP, the BET BSE will be based on the sum of the EMP estimate, or estimates, if sufficient wells from more than one catch stratum of the trip can be sampled, plus estimates from other catch sources for any catch strata of the trip not sampled by the EMP. For those trips for which the entire catch of the trip comes from one or two strata, an EMP estimate for the entire trip catch, or the majority of the trip catch, is considered possible, as was illustrated by Phase 2 of the pilot study[9]. If the trip carried an observer, the RDL data will be used to estimate the catch of unsampled strata for the trip. If the trip did not carry an observer, cannery and logbook data will be used; the logbook data are necessary to prorate the cannery total trip catch to strata.

Trips not sampled by the EMP

For trips not sampled by the EMP, RDL data will be used to estimate the catch of BET, if an observer was onboard, or cannery and logbook data, if an observer was not onboard. A direct comparison of EMP and cannery estimates has not yet been done (see Section B). Therefore, as of the drafting of this document, the emphasis on the use of the RDL data for observed trips not sampled by the EMP is motivated by the following observations. First, comparisons of observer and logbook ("CAE") data for Class-6 vessels[10] to cannery estimates, by trip, for 2015 – 2019 (trip departure year), indicated that while the total catch of tropical tunas was statistically very similar among the two data sources (cannery, CAE), there was a statistically significant tendency for the BET catch reported in the cannery data to be less than that reported in the CAE data in all 5 years (Figures 22 – 23, Table 5). Similar patterns were noted previously for data from the early 2000s (IATTC unpublished). Second, the EMP estimates by well and stratum (Table 4, Figures 19 – 21), show that, although the EMP estimates can be larger or smaller than the RDL estimates, the most extreme differences tended to be negative (i.e., EMP estimate was considerably greater than RDL estimate), suggesting that for some trips, the RDL estimate may be negatively biased. Thus, pending further analysis (Section B), it is considered that the cannery estimates may represent an underestimation of BET catch, and therefore the observer RDL estimates will be used, when available.

[9] Improvements to sampling equipment that will be fully implemented in the EMP will help to further streamline the data collection process, possibly allowing for sampling teams of two instead of three samplers, except for standard unloadings that use flotation (SAC-14 INF-I). This will increase sampling capability, both in terms of wells and trips.
[10] From the IATTC Catch-and-Effort (CAE) database, which will be primarily observer data for Class-6 vessels.

B. WORKPLAN FOR 2023

EMP-related work for the remainder of 2023 is outlined in this section. This work is intended to improve upon the analyses and simulations presented in this document, in addition to streamlining the BSE calculations that use EMP data. It is anticipated that task B.1 will be completed before the IATTC Annual Meeting in August 2023, and that it will be possible to generate BET BSEs in a timely manner by that time, as well. Tasks B.2 – B.4 will be started once task B.1 has been completed.

B.1 Computer program development to automate BSE calculations based on EMP and RDL data

Additional computer programs need to be written in order to automate, to the extent possible, the BSE calculations for trips sampled by the EMP. In addition, a database, and data entry and editing routines, need to be developed for the RDL observer data from EMP-sampled trips.

B.2 Identification of factors associated with patterns in catch composition during unloading

Using the Phase 1 data and the reconstructed RDL observer data, further analysis will be conducted to evaluate the relationship between the proportion of BET in sampled units, the by-set catch amounts of BET loaded into the well, and the order in which those sets were loaded into the well. Results from this may make possible an assessment of the prevalence of large-scale patterns in species composition during unloading across all Class-6 vessels, which might in turn lead to refinements of the within-well sampling design.

B.3 Sensitivity analyses for within-well sampling simulations

Sensitivity analyses will be conducted to investigate the effect of assumptions made in the within-well sampling simulations, which may lead to refinements of the current within-well sampling protocol, as well as improved quantification of performance. For these sensitivity analyses, the Phase 1 data will be used to create a synthetic data set of catch unloadings, with catch composition information estimated for every unit, sampled and unsampled. The sensitivity analyses will attempt to:

1) Quantify sampling protocol performance under plausible levels of variability in species composition within the original groups of 10 (8) units.
2) Evaluate tradeoffs between sampling physical units (which often contain different numbers of fish, depending on the species, as well as other factors) and sampling groups of a fixed number of fish (virtual units), as well as the optimal number of fish to sample in virtual units. This sensitivity analysis will allow for an evaluation of any potential for bias, which might arise if, for example, the containers used to unload the well catch could only hold a few fish. (It is noted, however, that sampling a fixed number of fish for unloadings that use containers would require additional cooperation from unloaders, and would also require refining the current sampling protocol to minimize sampler selection bias.)
3) Evaluate the sensitivity of the simulation results to irregularity in the number of unsampled units between sampled units over the course of unloading a well.
4) Estimate the minimum catch of BET in a well that can be reliably detected by the current sampling protocol.
5) Evaluate any effect of set distribution among wells (e.g., mixed-stratum wells; an individual large-catch sets that fill multiple wells) on variance weighting factors that involve N.
6) Evaluate whether sampling mixed stratum wells could lead to selection bias.
7) Compare the error under the protocol tested in Phase 2 of the EMP to that under a simple random sampling protocol for both stages.

8) Evaluate options for more statistically efficient estimators.

B.4 Conduct simulation studies to evaluate the number of wells to sample from important catch strata other than OBJ sets made west of 110°W, such as OBJ sets made between 110°W and 95°W. And additionally, evaluate how best to allocate available resources among wells of different catch strata of a single trip, to obtain the best estimate of trip-level BET catch.

B.5 Complete statistical comparison of cannery estimates to other sources (EMP, observer)

B.6 Improve efficiency of sampling teams for all types of unloadings (see SAC-14 INF-I)

The use of voice recorders for data collection during sampling and the use of precision scales for weighing individual fish was initiated in the EMP in March 2023. The use of these tools will be evaluated and any necessary adjustments to the sampling protocols will be made in a process of continuous improvement. It is anticipated that these sampling tools will allow the EMP to increase its sampling capabilities.

The samplers will be distributed in groups of two for dry unloading and in groups of three for flotation, ensuring the sampling of at least six wells per fishing trip with the available human and financial resources. Sampling of cargo net unloadings will also be implemented and improved, resolving logistical challenges specific to this type of unloading.

B.7 Address logistical challenges related to sampling of vessels (trips) that do not carry an observer

ACKNOWLEDGEMENTS

The authors extend a special thank you to the samplers Jennifer Aguilar, Grace Álvarez, Carlos Bravo, Pablo Delgado, Jonathan Gaibor, Juan Galarza, José Guillén, Javier Mejía, James Méndez, Diego Montehermoso, Marcos Muñoz, Darío Quimi, Alex Santana, Diego Ureta, Wellington Vásquez, Víctor Vinces, Ledin Vizueta, Robinson Zambrano, Tommy Zamora, and Alisson Zúñiga; the Field Office staff of Manta, Erick D. Largacha Delgado, Daniel E. Cevallos-Alarcón, Carlos de la A Florencia, Glenthon Macías Pita, Nilo Pérez, and Alex Urdiales; the Field Office staff of Playas, William E. Paladines Proaño and Felix F. Cruz Vargas; the national observer programs of Ecuador and Panama, and, to Dan W. Fuller, for their effort to help ensure the success of the EMP pilot study, to Marti McCracken for helpful discussions on survey sampling and comments on this document, and to the fishing industry and Ecuadorian authorities for their collaboration on all aspects of this project.

REFERENCES

Cochran, W.G., 1977. Sampling Techniques, third ed. John Wiley and Sons, New York. 428 pp.

Lennert-Cody, C.E., McCracken, M., Siu, S., Oliveros-Ramos, R., Maunder, M.N., Aires-da-Silva, A., Miguel, Carvajal Rodrigues, J. M., Opsomer, J. 2022. Single-cluster sampling designs for shark catch size composition in a Central American longline fishery. Fisheries Research 251 (2022) 106320. https://doi.org/10.1016/j.fishres.2022.106320

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Bob Wheeler and Marco Torchiano (2016). lmPerm: Permutation Tests for Linear Models. R package version 2.1.0. https://CRAN.R-project.org/package=lmPerm

TABLE 1. Percentage of all EPO BET catch of Class 1 – 6 vessels (observer and logbook data; IATTC Catch and Effort (CAE) database), that was produced annually by OBJ sets made west of 110°W, 2015-2021.
TABLA 1. Porcentaje de toda la captura de BET del OPO de buques de clases 1 a 6 (datos de observadores y de bitácora; base de datos de Captura y Esfuerzo (CAE) de la CIAT), que fue producida anualmente por lances OBJ realizados al oeste de 110°O, en el periodo 2015–2021.

2015	2016	2017	2018	2019	2020	2021
52%	74%	65%	58%	74%	74%	76%

TABLE 2. Number of fully-sampled wells (sampled at both 1-out-of-10 and 1-out-of-8), by set type (OBJ; unassociated, NOA) and area (west/east of 110°W) of the catch in the well, for Phase 1 of the EMP pilot study. *: One well with OBJ catch from both east and west of 110°W, unloaded with standard methods, was also sampled, for a total of 71 wells of standard unloadings.
TABLA 2. Número de bodegas muestreadas en su totalidad (tanto con los protocolos 1 de 10 como 1 de 8), por tipo de lance (OBJ; no asociado, NOA) y área (oeste/este de 110°O) de la captura en la bodega, para la Fase 1 del estudio piloto del PRM. *: También se muestreó una bodega con captura OBJ tanto al este como al oeste de 110°O, descargada con métodos estándar, para un total de 71 bodegas de descargas estándar.

Method of unloading	West of 110°W	East of 110°W	Total
Standard unloadings			
OBJ	44	22	66
NOA	1	1	2
OBJ-NOA mix	0	2	2
Total	45	25	70*
Cargo net unloadings			
OBJ	4	3	7
NOA	0	1	1
Total	4	4	8

TABLE 3. Results from the within-well simulations for the 39 sampled OBJ-set wells (sampled at 1-out-of-10) with catch west of 110°W. The error range of each protocol, across all cluster samples of all wells in the simulation, is shown in parentheses in the second column. Note that a difference in percent of 2.5641 (= 100*(1/39)) corresponds to a difference of 1 well.

TABLA 3. Resultados de las simulaciones dentro de la bodega para las 39 bodegas muestreadas de lances OBJ (muestreadas con el protocolo 1 de 10) con capturas al oeste de 110°O. El rango de error de cada protocolo, en todas las muestras de conglomerados de todas las bodegas de la simulación, se muestra entre paréntesis en la segunda columna. Obsérvese que una diferencia porcentual de 2.5641 (= 100*(1/39)) corresponde a una diferencia de 1 bodega.

Simulated protocol	% Wells with mean absolute error > 0.05 (error range, all cluster samples and wells)	% Wells with mean absolute relative error > 0.20
5% (1-of-2)	10.2564 (-7.5% to 6.8%)	7.6923
3.33% (1-of-3)	17.9487 (-11.0% to 11.8%)	7.6923
2.5% (1-of-4)	15.3846 (-13.7% to 13.5%)	12.8205
2% (1-of-5)	30.7692 (-15.1% to 16.5%)	25.6410

TABLE 4. Estimates of BET for the dominant catch stratum of each trip sampled in Phase 2 of EMP pilot study. Approximate 95% confidence intervals (CI) are based on the assumption that the estimated proportion of BET, which is a ratio estimate, is approximately normally distributed. "s.e": standard error. Catch amounts are in metric tons. Trip labels (first column) match those shown in bottom panel of Figure 21.
TABLA 4. Estimaciones de BET para el estrato de captura dominante de cada viaje muestreado en la Fase 2 del estudio piloto del PRM. Los intervalos de confianza (IC) aproximados del 95% se basan en el supuesto de que la proporción estimada de BET, que es una estimación de razón, tiene una distribución aproximadamente normal. "s.e": error estándar (por sus siglas en inglés). Las cantidades de las capturas se expresan en toneladas métricas. Las identificaciones de los viajes (primera columna) coinciden con las que se muestran en el panel inferior de la Figura 21.

Trip	Stratum (% trip tropical tuna catch in stratum)	Number wells sampled (out of those sampleable)	EMP estimate of proportion of BET	EMP estimate of BET (rounded to whole tons)	Approximate 95% CI for BET (Est +/- 1.96*s.e.)	RDL BET
A	OBJ sets west of 110W (100%)	6 out of 18	0.397 (s.e. = 0.0875)	538 t (s.e. = 118.693; CV = 0.22)	(306 t, 771 t)	149 t
B	OBJ sets west of 110W (29%)	4 out of 4	0.055 (s.e. = 0.0195)	18 t (s.e. = 6.386; CV = 0.35)	(5 t, 31 t)	32 t
C	OBJ sets west of 110W (98%)	6 out of 20	0.232 (s.e. = 0.0702)	292 t (s.e. = 88.168; CV = 0.30)	(119 t, 465 t)	325 t
D	OBJ sets 95W-110W (97%)	6 out of 15	0.0009 (s.e. = 0.00691)	1 t (s.e. = 0.629; CV = 0.74)	(0 t, 2t)	10 t
E	OBJ sets 95W-110W (83%)	6 out of 11	0.131 (s.e. = 0.0429)	71 t (s.e. = 23.367; CV = 0.33)	(25 t , 117 t)	90 t
F	OBJ sets 95W-110W (40%)	4 out of 8	0.392 (s.e. = 0.0981)	236 t (s.e. = 59.125 ; CV = 0.25)	(121 t, 352 t)	170 t
G	OBJ sets 95W-110W (63%)	5 out of 5	0.192 (s.e. = 0.0628)	81 t (s.e. = 26.644 ; CV = 0.39)	(19 t, 123 t)	14 t
H	OBJ sets 95W-110W (77%)	5 out of 7	0.289 (s.e. = 0.0962)	146 t (s.e. = 48.504 ; CV = 0.33)	(51 t, 241 t)	98 t
I	OBJ sets east of 95W (84%)	4 out of 10	0.311 (s.e. = 0.1128)	353 t (s.e. = 128.395; CV = 0.36)	(102 t, 605 t)	109 t
J	OBJ sets east of 95W (91%)	6 out of 14	0.147 (s.e. = 0.0560)	204 t (s.e. = 77.650; CV = 0.38)	(51 t, 356 t)	169 t

TABLE 5. Estimated slopes from generalized linear models (GLMs) fitted between trip-level catch estimates from cannery data and catch estimates from the IATTC CAE data base (observer and logbook data) (Class-6 vessels only). The analysis, which was limited to trips with any BET catch in the cannery data, resulted in a loss of BET catch from the CAE of 1% – 2.3%, depending on the year; to avoid this restriction, mixture models may be used in the future for a more refined comparison (see Section B). GLM specifics: the model fitted was cannery catch = $\beta_0 + \beta_1$ * CAE catch, with gamma error and identity link function. "Tropical tuna": YFT + SKJ + BET; "% Dev": percent deviance explained by the model; "s.e.": standard error; "p": p-value. Catches per trip for 2019 are shown in Figures 22 – 23. The p-values, for a test of Ho: the slope from the model for BET is 1.0, were all less than 0.01.

TABLA 5. Pendientes estimadas a partir de modelos lineales generalizados (MLG) ajustados entre las estimaciones de captura a nivel de viaje de los datos de enlatadoras y las estimaciones de captura de la base de datos CAE de la CIAT (datos de observadores y de bitácora) (solo buques de clase 6). El análisis, que se limitó a los viajes con alguna captura de BET en los datos de enlatadoras, dio como resultado una pérdida en la captura de BET de CAE del 1%- 2.3%, dependiendo del año; para evitar esta restricción, es posible que en el futuro se utilicen modelos mixtos para una mejor comparación (ver la Sección B). Especificaciones del MLG: el modelo ajustado fue de captura de enlatadoras =β_0 +β_1 * captura de CAE, con error gamma y función de enlace de identidad. "Atunes tropicales": YFT + SKJ + BET; "% Dev": desviación porcentual explicada por el modelo; "s.e.": error estándar; "p": valor p. Las capturas por viaje para 2019 se muestran en las Figuras 22 y 23. Los valores p, para una prueba de Ho: la pendiente del modelo para BET es 1.0, todos fueron inferiores a 0.01.

Year (number of trips)	BET GLM		Tropical tuna GLM	
	Estimated slope	% Dev	Estimated slope	% Dev
2015 (232)	0.64 (s.e. = 0.104; p < 0.01)	54%	0.96 (s.e. = 0.005; p < 0.01)	99%
2016 (221)	0.64 (s.e. = 0.049; p < 0.01)	62%	0.96 (s.e. = 0.007; p < 0.01)	98%
2017 (250)	0.79 (s.e. = 0.083; p < 0.01)	58%	0.97 (s.e. = 0.004; p < 0.01)	99%
2018 (230)	0.74 (s.e. = 0.053; p < 0.01)	69%	0.99 (s.e. = 0.008; p < 0.01)	99%
2019 (201)	0.75 (s.e. = 0.061; p < 0.01)	69%	0.96 (s.e. = 0.005; p < 0.01)	98%

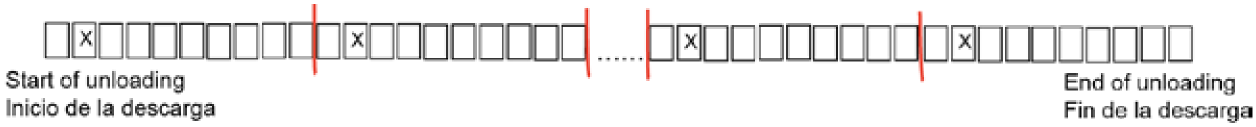


FIGURE 1. Schematic illustrating the EMP pilot study intensive sampling protocol for standard unloadings. Each rectangle represents a unit of fish. A unit was defined as either a physical container full of fish or a fixed number of fish unloaded individually from the well. The red vertical lines delineate consecutive blocks of 10 units. Each "X" indicates a unit that would be sampled, starting from the second unit (selected randomly out of the first 10 units), and then sampling every subsequent 10th unit from that random starting point (i.e., sampling units 2 (selected at random), unit 12, unit 22, unit 32, etc.).

FIGURA 1. Diagrama que ilustra el protocolo de muestreo intensivo del estudio piloto del PRM para descargas estándar. Cada rectángulo representa una unidad de peces. Una unidad se define como un contenedor físico lleno de peces o un número fijo de peces que se descargan de la bodega de manera individual. Las líneas verticales rojas delimitan bloques consecutivos de 10 unidades. Cada "X" indica una unidad que se muestrearía, empezando por la segunda unidad (seleccionada aleatoriamente de entre las 10 primeras unidades) y, posteriormente, muestreando cada 10 unidades subsecuentes a partir de ese punto de partida aleatorio (es decir, muestreando las unidades 2 (seleccionada aleatoriamente), la unidad 12, la unidad 22, la unidad 32, etc.).

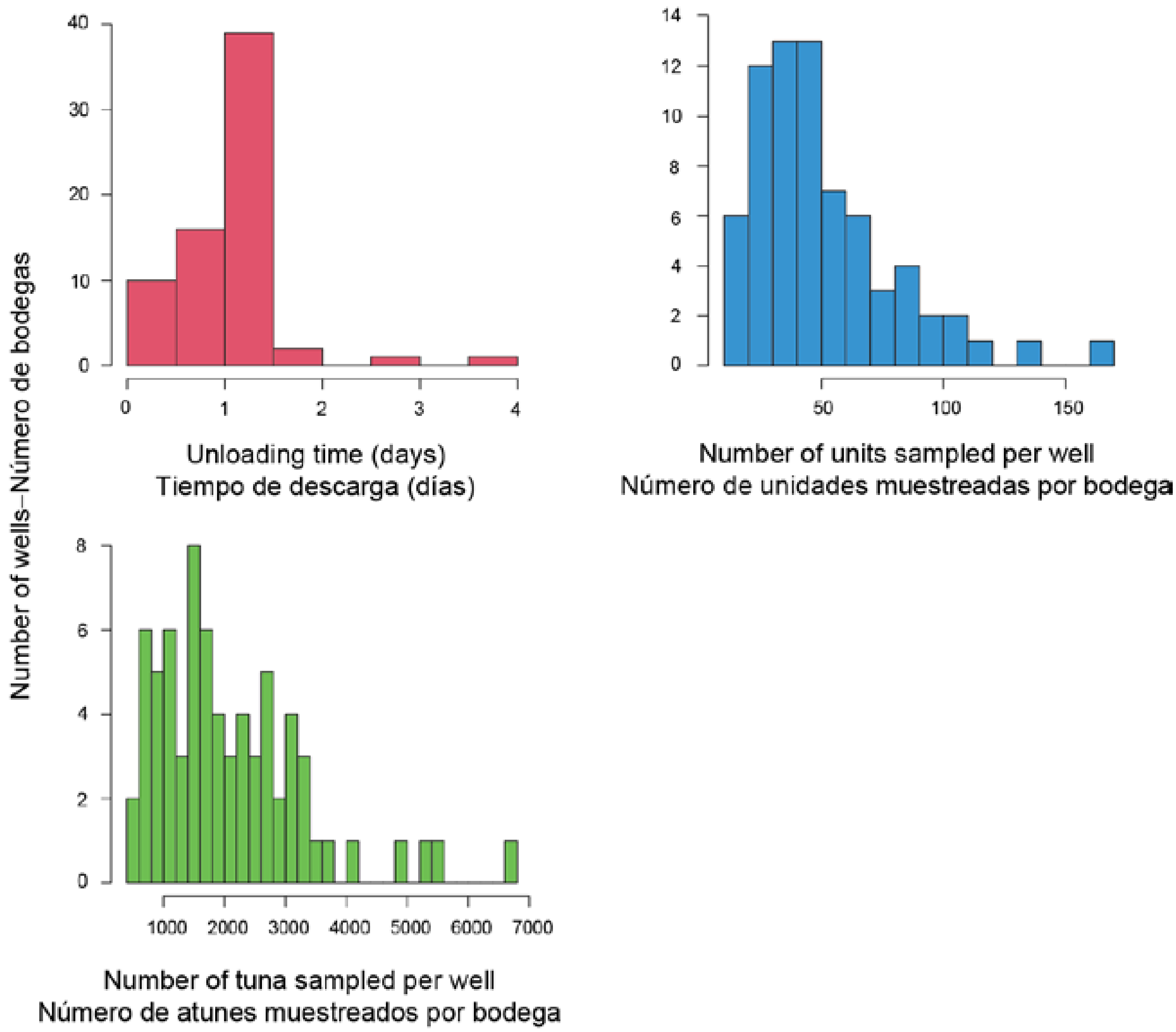


FIGURE 2. Histograms showing the unloading times (top left), numbers of units sampled (top right) and numbers of tuna sampled (bottom left), for the 71 intensively sampled standard-unloading wells of Phase 1.
FIGURA 2. Histogramas que muestran los tiempos de descarga (arriba a la izquierda), el número de unidades muestreadas (arriba a la derecha) y el número de atunes muestreados (abajo a la izquierda), para las 71 bodegas de descarga estándar que se muestrearon intensivamente en la Fase 1.

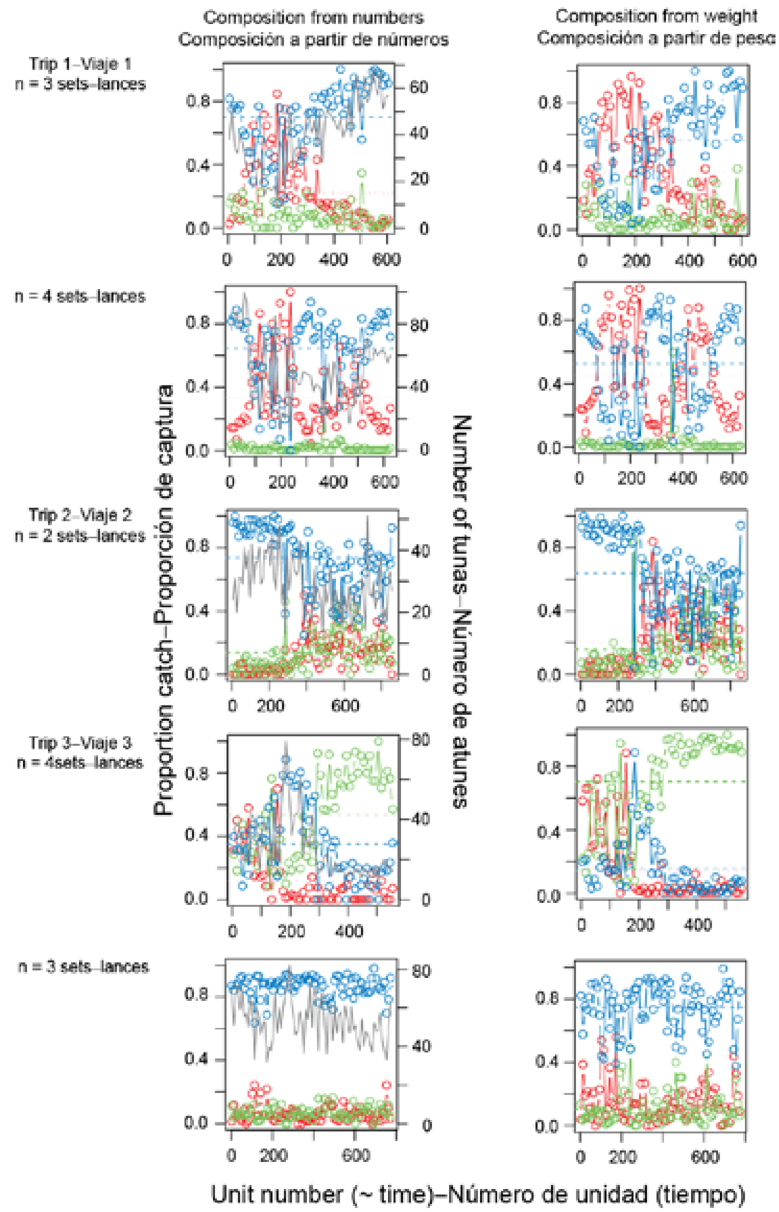


FIGURE 3. Tropical tuna species composition, by sampled unit, computed from numbers of fish (left-hand side) and from weight (right-hand side), for 6 standard-unloading wells of three trips (each row is a different well). There are two wells shown for Trip 1, one well for Trip 2, and 2 wells for Trip 3. Red: BET; green: YFT; blue: SKJ; gray: number of the three species combined. Dashed lines indicate the average species proportion (computed over sampled units of the well). The x-axis shows the unit number, from the start (left) to the end (right) of the unloading. 'n': number of purse-seine sets associated with the catch in the well.

FIGURA 3. Composición por especie de atunes tropicales, por unidad muestreada, calculada a partir del número de peces (lado izquierdo) y del peso (lado derecho), para seis bodegas de descarga estándar de tres viajes (cada fila es una bodega diferente). Se muestran dos bodegas para el Viaje 1, una para el Viaje 2 y dos para el Viaje 3. Rojo: BET; verde: YFT; azul: SKJ; gris: número de las tres especies combinadas. Las líneas punteadas indican la proporción promedio de las especies (calculada a partir de las unidades muestreadas de la bodega). El eje *x* muestra el número de la unidad, desde el inicio (izquierda) hasta el final (derecha) de la descarga. 'n': número de lances de cerco asociados a la captura en la bodega.

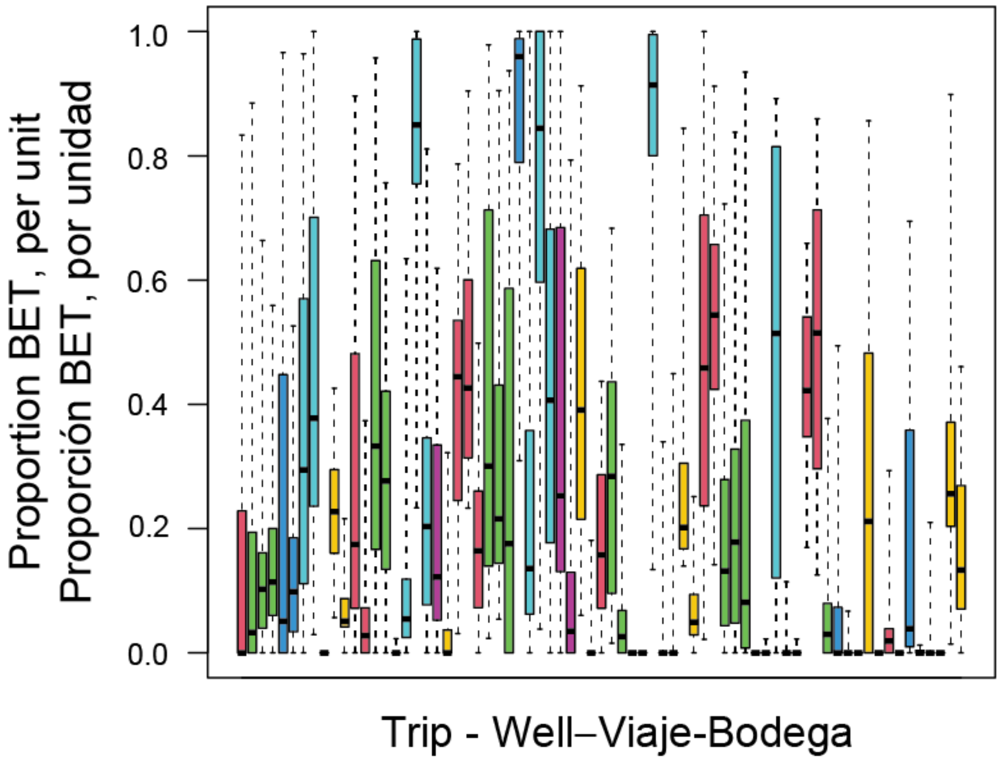


FIGURE 4. Box-and-whisker plots of the proportion of BET, per sampled unit, for each of the 71 standard-unloading wells. Each box-and-whisker plot represents a separate well. Wells sampled from the same trip are arranged consecutively and have the same color. Horizontal black bars indicate the median proportion, the colored box indicates the interquartile range, and the whiskers extend to the extremes of the unit proportions for the well.

FIGURA 4. Gráfica de cajas y bigotes de la proporción de BET, por unidad muestreada, para cada una de las 71 bodegas de descarga estándar. Cada gráfica representa una bodega distinta. Las bodegas muestreadas del mismo viaje se ordenan consecutivamente y tienen el mismo color. Las barras negras horizontales indican la proporción mediana, la caja de color indica el rango intercuartil y los bigotes se extienden hasta los extremos de las proporciones de la unidad de la bodega.

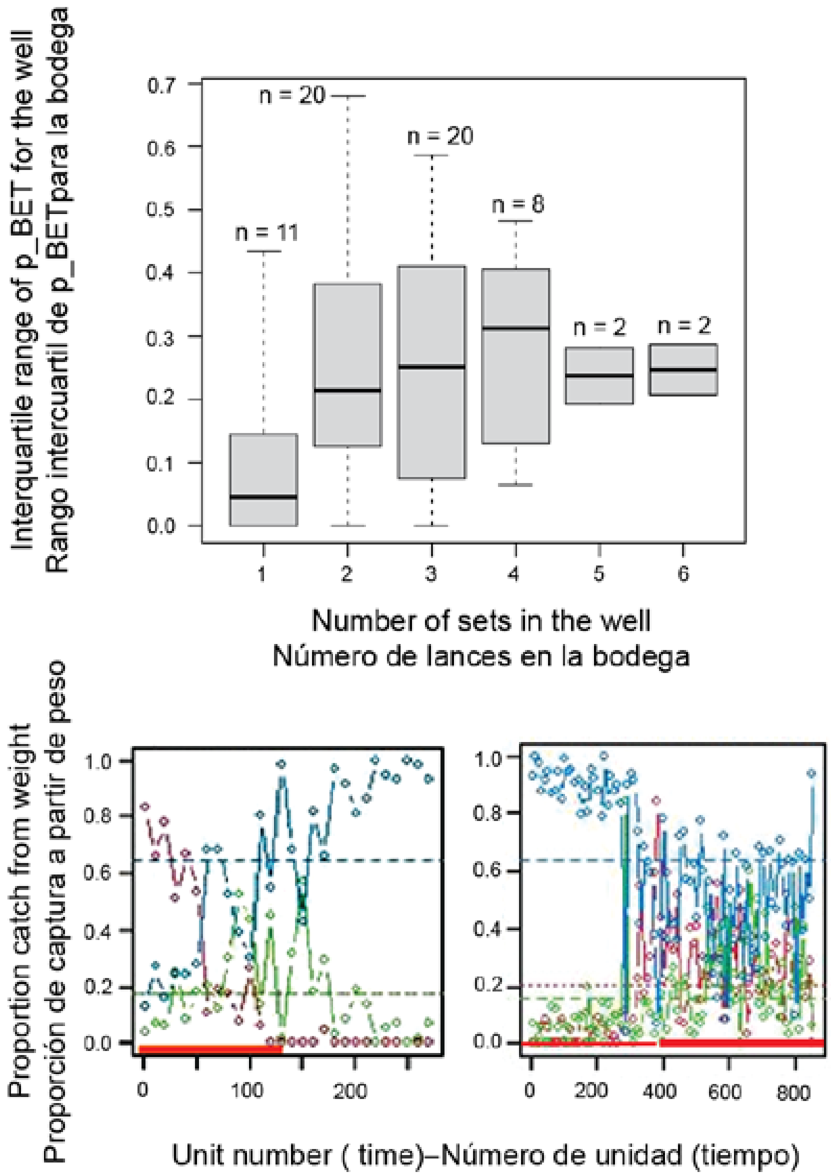


FIGURE 5. Box-and-whisker plots (top) of the interquartile ranges of the proportion of BET by unit ('p-BET'), for standard-unloading wells that had any amount of BET, grouped according to the number of purse-seine sets from which catch was loaded into the well ('n': number of wells). The interquartile range is the difference between the values of the proportion of BET that correspond to the upper and lower edges of the colored boxes shown in Figure 4. (bottom) Two examples of species composition per sampled unit (from weight), along with a qualitative indicator of the distribution of BET loaded into the well (orange bar; the thicker the bar, the more BET per set); catch in these two wells was from OBJ sets.

FIGURA 5. Gráficas de cajas y bigotes (arriba) de los rangos intercuartiles de la proporción de BET por unidad ('p-BET'), para las bodegas de descarga estándar que tenían cualquier cantidad de BET, agrupadas según el número de lances de cerco de los que se cargaron las capturas en la bodega ('n': número de bodegas). El rango intercuartil es la diferencia entre los valores de la proporción de BET que corresponden a los bordes superior e inferior de las cajas de color que se muestran en la Figura 4. (Abajo) Dos ejemplos de composición por especie por unidad muestreada (a partir del peso), junto con un indicador cualitativo de la distribución de BET cargado en la bodega (barra anaranjada; cuanto más gruesa es la barra, mayor es la cantidad de BET por lance); la captura en estas dos bodegas procedía de lances OBJ.

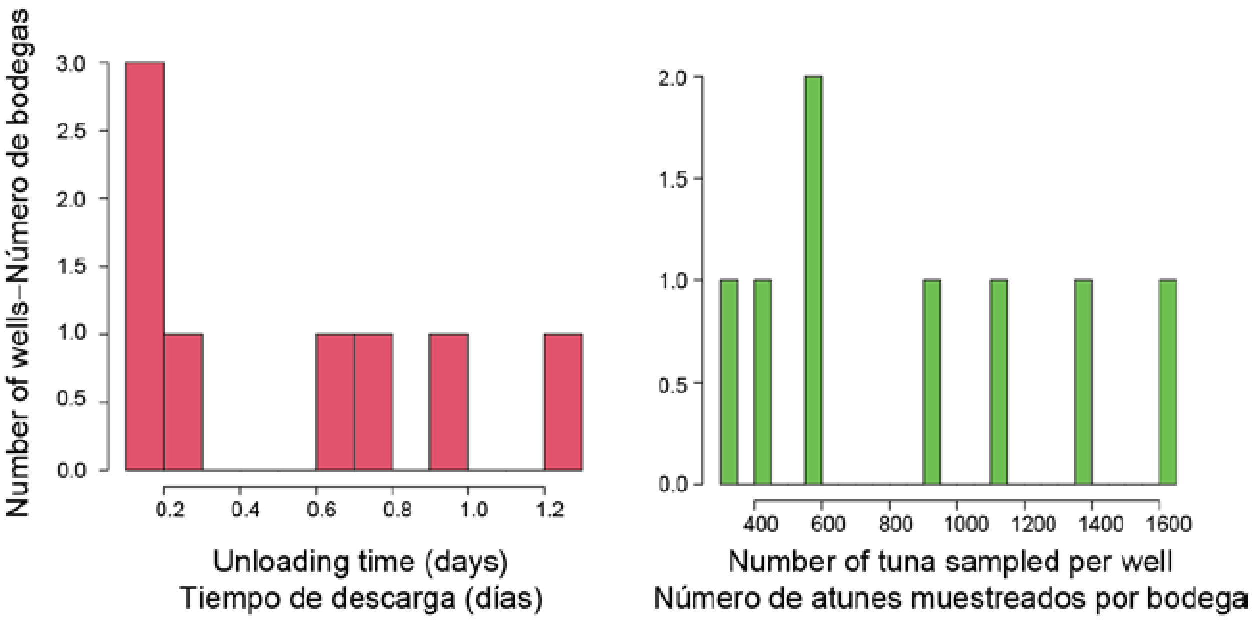


FIGURE 6. Histograms showing the unloading times (left) and numbers of tuna sampled (right), for the 8 fully-sampled wells of Phase 1 that were unloaded using cargo nets.

FIGURA 6. Histogramas que muestran los tiempos de descarga (izquierda) y el número de atunes muestreados (derecha), para las ocho bodegas muestreadas en su totalidad en la Fase 1 que fueron descargadas utilizando redes de carga.

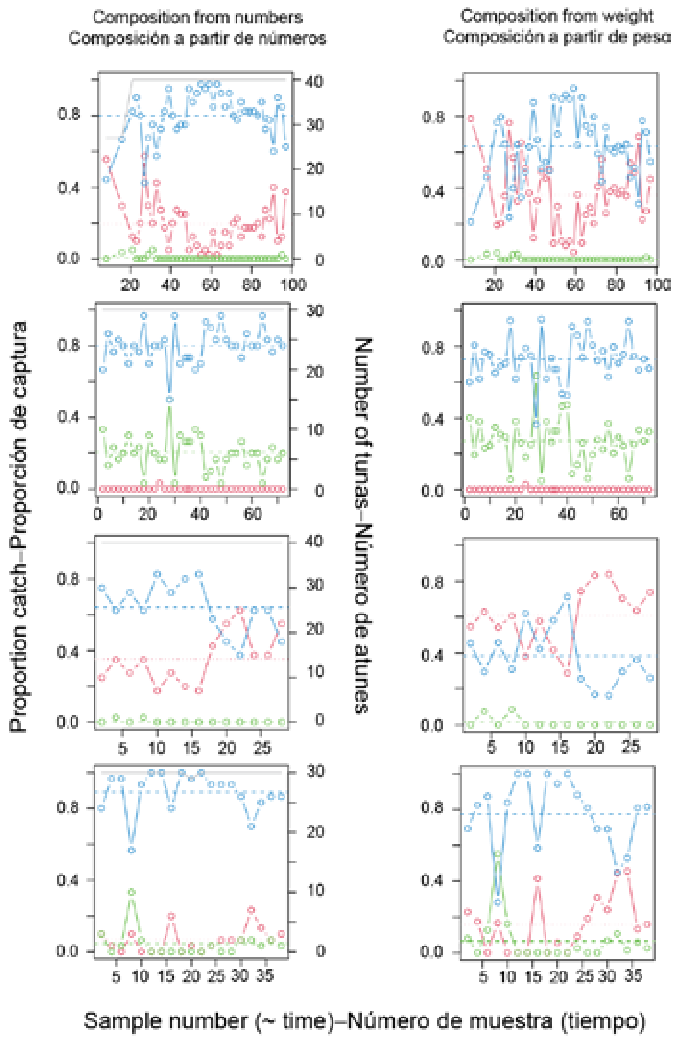


FIGURE 7. Tropical tuna species composition for four wells of 3 trips unloaded using cargo nets, by sampled group of fish ('sample number'), computed from numbers of fish (left-hand side) and from weight (right-hand side). Each row represents a well; the last two rows correspond to two wells from the same trip. With exception of the second well (second row), which had catch from one set, the other three wells were each loaded with catch from two sets. The first three wells had catch from OBJ sets made west of 110°W, and the last from OBJ sets made east of 110°W. Red: BET; green: YFT; blue: SKJ; gray: number of the three species combined. Dashed lines indicate the average species proportion (computed over sampled groups of fish from the well). The x-axis shows the sample number, from the start (left) to the end (right) of the unloading.

FIGURA 7. Composición por especie de atunes tropicales para cuatro bodegas de tres viajes descargados con redes de carga, por grupo de peces muestreados ("número de muestra"), calculada a partir del número de peces (lado izquierdo) y del peso (lado derecho). Cada fila representa una bodega; las dos últimas filas corresponden a dos bodegas del mismo viaje. A excepción de la segunda bodega (segunda fila), que tenía captura de un lance, las otras tres bodegas estaban cargadas con capturas de dos lances cada una. Las tres primeras bodegas tenían captura de lances OBJ realizados al oeste de 110°O, y la última de lances OBJ realizados al este de 110°O. Rojo: BET; verde: YFT; azul: SKJ; gris: número de las tres especies combinadas. Las líneas punteadas indican la proporción promedio de las especies (calculada a partir de los grupos de peces muestreados de la bodega). El eje *x* muestra el número de la muestra, desde el inicio (izquierda) hasta el final (derecha) de la descarga.

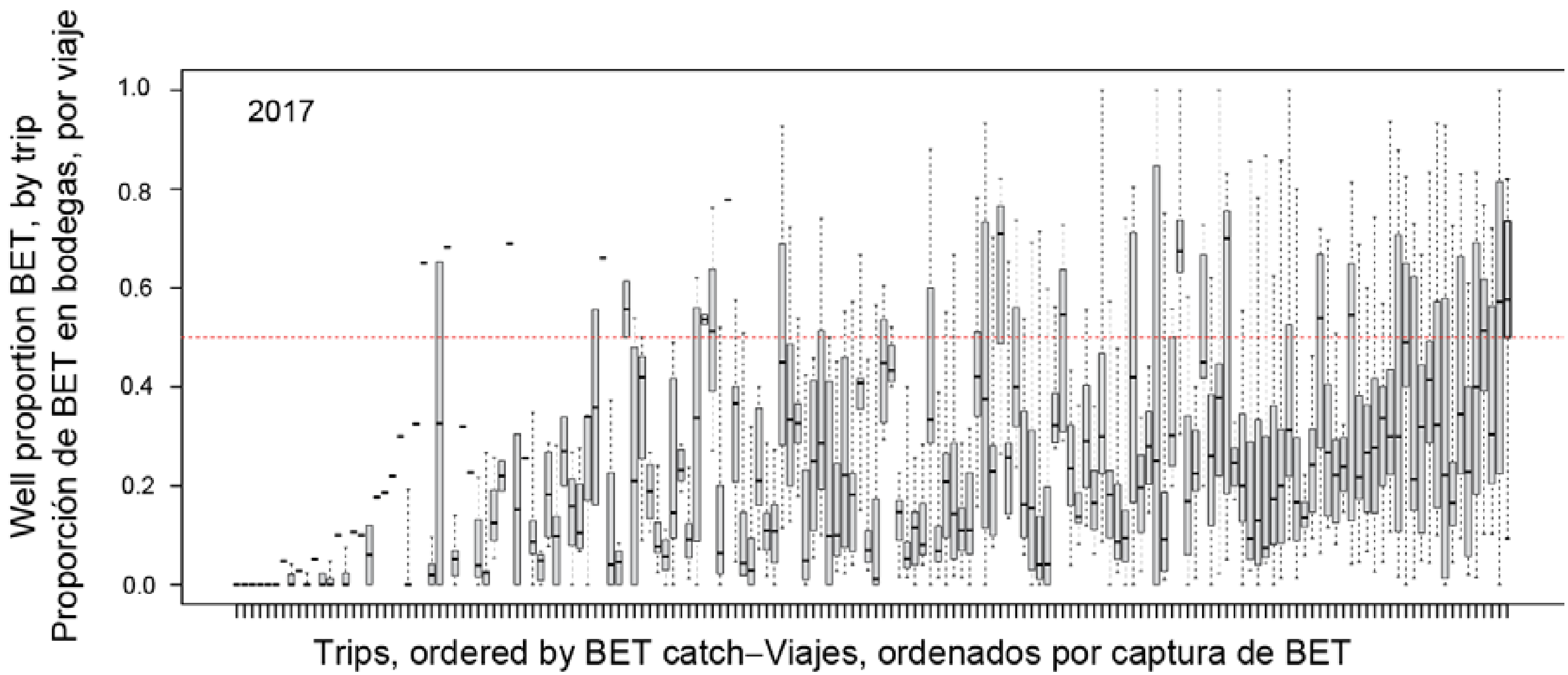
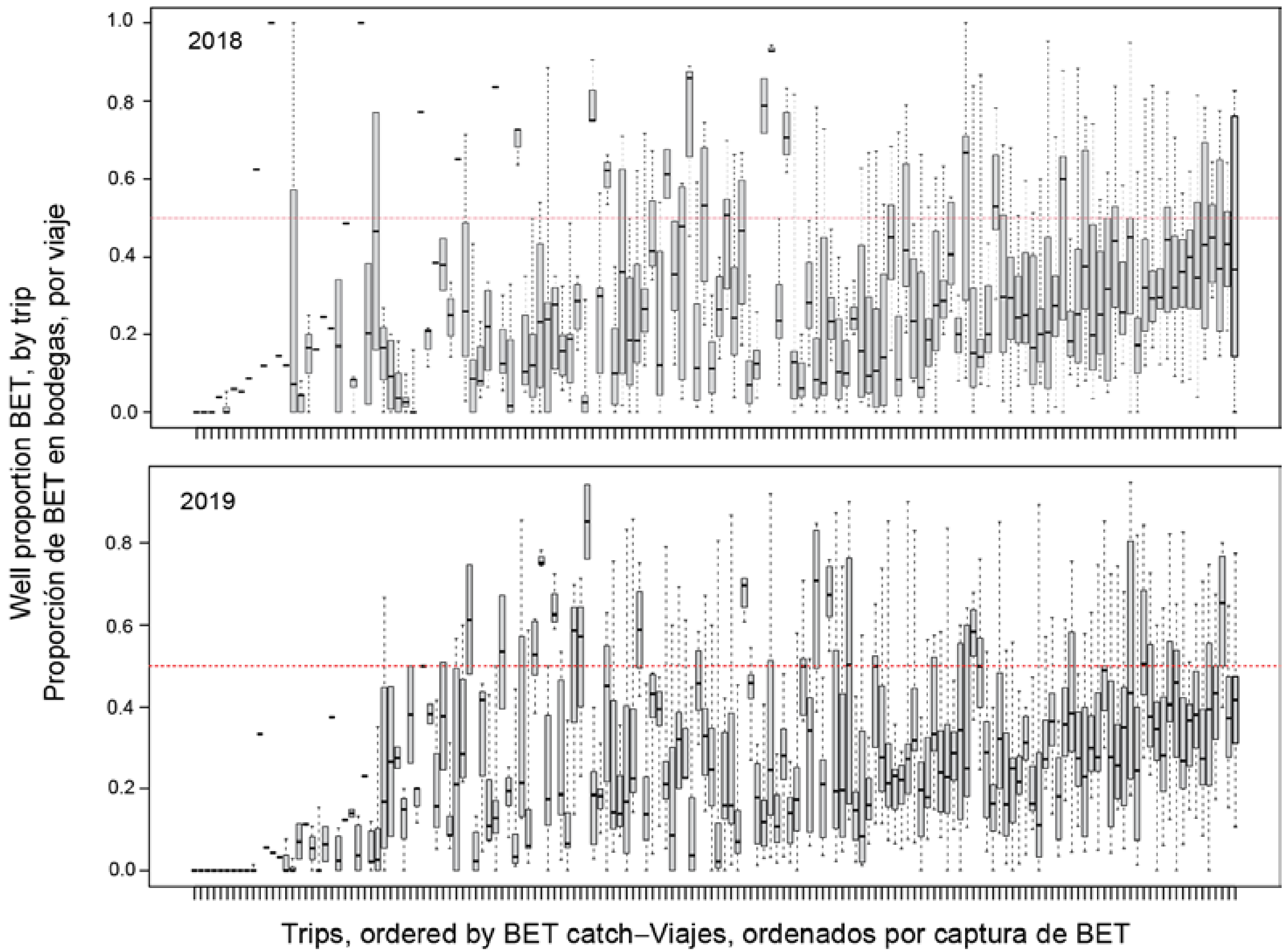


FIGURE 8. Box-and-whisker plots of the proportion of BET per well, by trip, from reconstructed RDL observer data, for wells with catch from OBJ sets made west of 110°W, 2017 – 2019. For each year, trips are ordered by BET catch amounts, from smallest (left) to largest (right). Each box-and-whisker plot represents a single trip. The black horizontal bar within each box indicates the median proportion of BET per well for the trip, the box represents the interquartile range (lower 25th percentile to upper 75th percentile), and the whiskers extend to the largest and smallest values.
FIGURA 8. Gráficas de cajas y bigotes de la proporción de BET por bodega, por viaje, a partir de los datos de observadores del RDL reconstruidos, para bodegas con captura de lances OBJ realizados al oeste de 110°O, en el periodo 2017-2019. Para cada año, los viajes se ordenan por cantidades de captura de BET, de menor (izquierda) a mayor (derecha). Cada gráfica representa una bodega distinta. La barra horizontal negra dentro de cada caja indica la proporción mediana de BET por bodega para el viaje, la caja representa el rango intercuartil (del percentil 25 inferior al percentil 75 superior) y los bigotes se extienden a los valores más grandes y más pequeños.

FIGURE 8. continued.



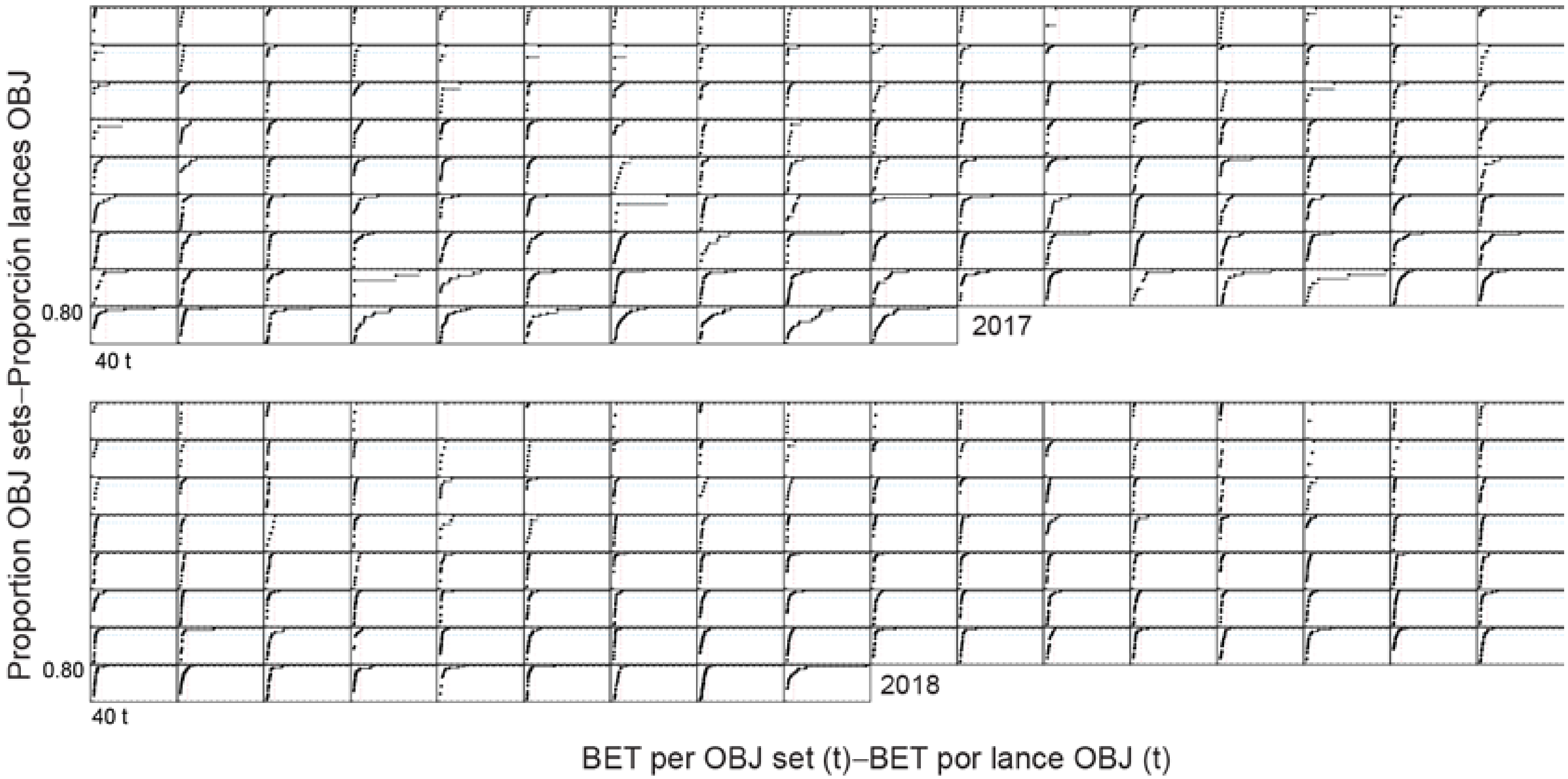
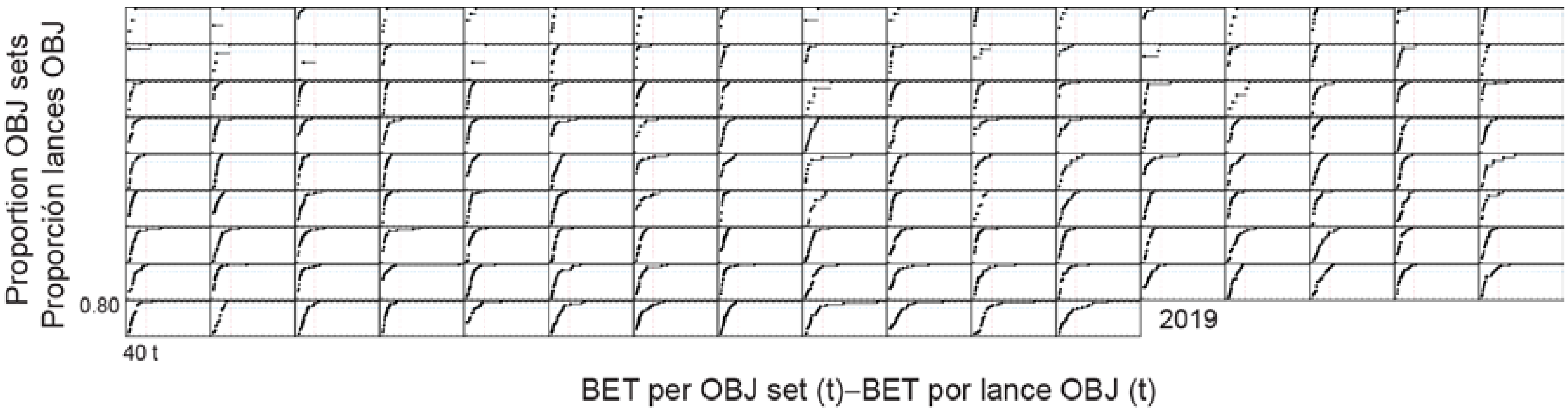


FIGURE 9. Empirical cumulative distribution function (ECDF) plots of BET catch per OBJ set, by trip, from reconstructed RDL observer data, for sets made west of 110°W, 2017-2019. Each panel is a trip. Dashed red vertical lines indicate the location on the x-axis for that panel of the value 40 t. Dashed blue horizontal lines indicate the location on the y-axis for that panel of the value 0.80 for the proportion of OBJ sets. Only trips with more than 20 t of BET catch in OBJ sets west of 110°W are shown. Trips are ordered by catch amounts of BET, from lowest (upper left) to greatest (lower right).

FIGURA 9. Gráficas de función de distribución acumulativa empírica (ECDF, por sus siglas en inglés) de la captura de BET por lance OBJ, por viaje, a partir de los datos de observadores del RDL reconstruidos, para lances realizados al oeste de 110°O, en el periodo 2017-2019. Cada panel es un viaje. Las líneas rojas punteadas verticales indican la ubicación en el eje *x* para ese panel del valor 40 t. Las líneas azules punteadas horizontales indican la ubicación en el eje *y* para ese panel del valor 0.80 para la proporción de lances OBJ. Únicamente se muestran los viajes con más de 20 t de captura de BET en lances OBJ al oeste de 110°O. Los viajes se ordenan por cantidad de captura de BET, de menor (arriba a la izquierda) a mayor (abajo a la derecha).

FIGURE 9. continued.



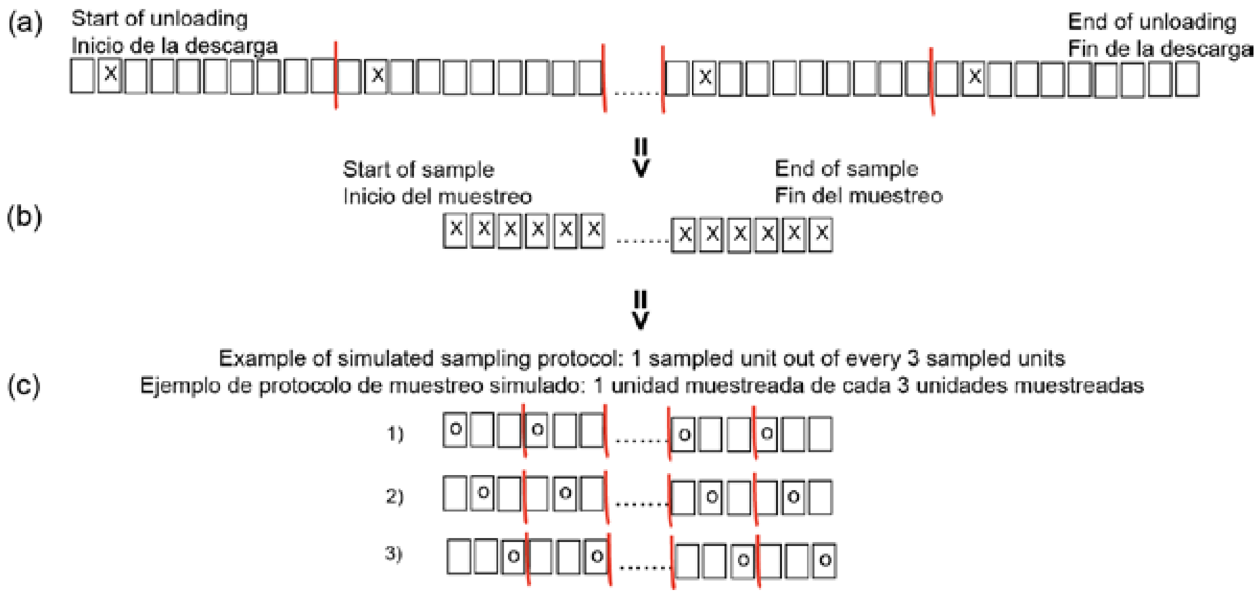


FIGURE 10. Schematic illustrating the within-well sampling protocols tested with the Phase 1 data. Each rectangle represents a unit of fish (either virtual or physical), and each "X" or "o" indicates a unit that would be sampled. (a) (from Figure 1) Schematic illustrating the EMP pilot study intensive sampling protocol for standard unloadings, sampling every 10th unit from a random starting unit (unit #2, in this case); (b) the resulting data set of sampled units from (a); and, (c) the three cluster samples that would be produced from (b) sampling every third sampled unit (i.e., the 1-of-3 protocol).

FIGURA 10. Diagrama que ilustra los protocolos de muestreo dentro de la bodega que se probaron con los datos de la Fase 1. Cada rectángulo representa una unidad de peces (virtual o física), y cada "X" u "o" indica una unidad que se muestrearía. (a) (de la Figura 1) Diagrama que ilustra el protocolo de muestreo intensivo del estudio piloto del PRM para descargas estándar, muestreando cada 10 unidades a partir de una unidad inicial aleatoria (unidad núm. 2, en este caso); (b) el conjunto de datos resultante de las unidades muestreadas a partir de (a); y (c) las tres muestras de conglomerados que se producirían a partir de (b) muestreando cada tercera unidad muestreada (es decir, el protocolo 1 de 3).

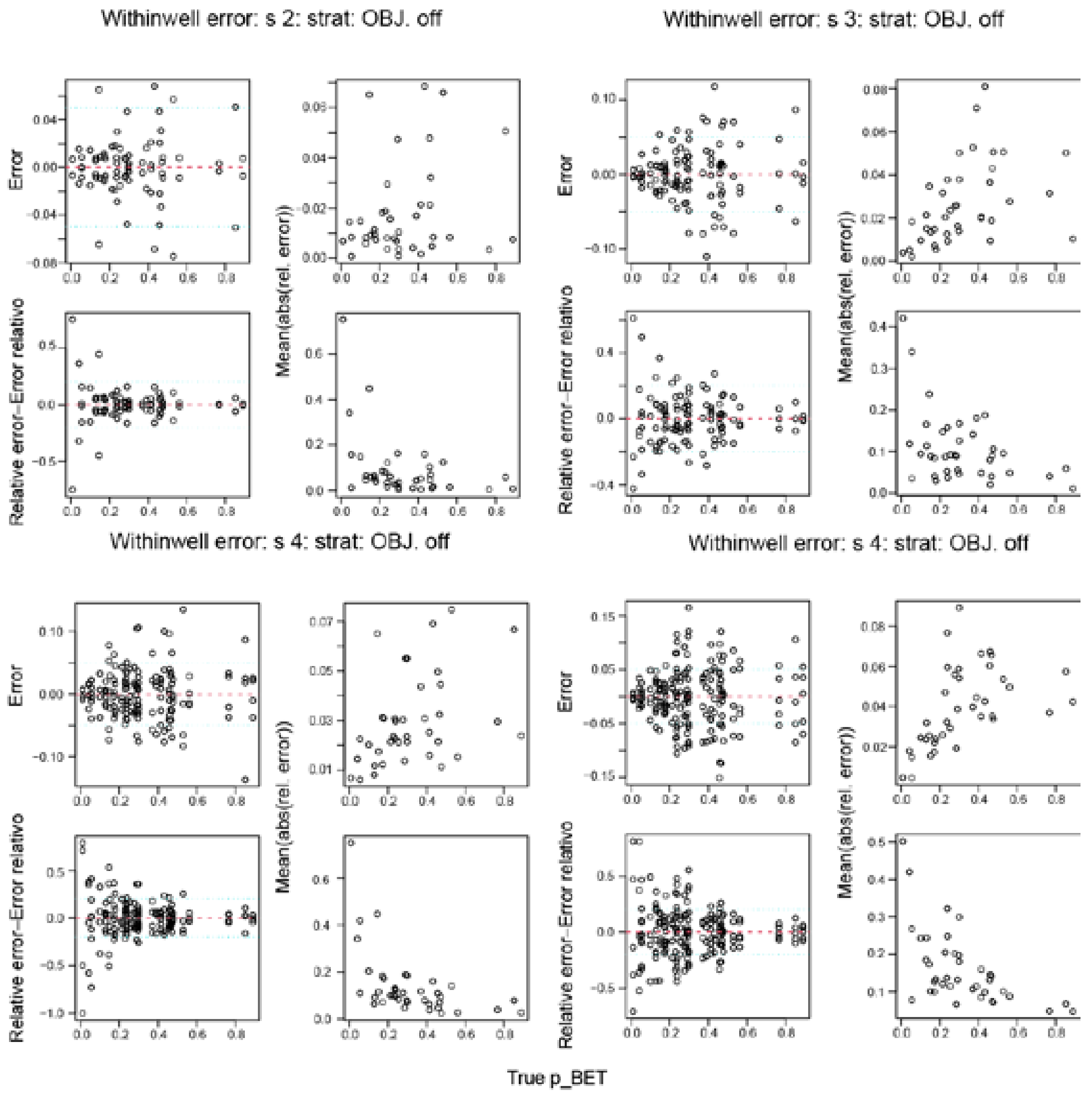


FIGURE 11. Within-well simulation results based on Phase 1 data for 5% coverage (1-of-2 units; upper left 2x2 panel), 3.33% coverage (1-of-3 units, upper right 2x2 panel), 2.5% coverage (1-of-4 units; lower left 2x2 panel) and 2% coverage (1-of-5 units; lower right 2x2 panel). In each 2x2 panel of graphs is shown: error (upper left), relative error (lower left), average absolute error (upper right) and average absolute relative error (lower right). 'True p_BET': the estimate of the proportion of BET in the well from the full intensive sampling data, assumed to be the true proportion of BET in the well. Blue dashed lines are at 0.05 for error and 0.20 for relative error. The red dashed line is at 0.

FIGURA 11. Resultados de la simulación dentro de la bodega basados en los datos de la Fase 1 para una cobertura del 5% (1 de 2 unidades; panel superior izquierdo de 2x2), una cobertura del 3.33% (1 de 3 unidades; panel superior derecho de 2x2), una cobertura del 2.5% (1 de 4 unidades; panel inferior izquierdo de 2x2) y una cobertura del 2% (1 de 5 unidades; panel inferior derecho de 2x2). En cada panel de gráficas de 2x2 se muestra: error (arriba a la izquierda), error relativo (abajo a la izquierda), error absoluto promedio (arriba a la derecha) y error relativo absoluto promedio (abajo a la derecha). 'True p_BET': la estimación de la proporción de BET en la bodega a partir de los datos de muestreo intensivo completo, que se supone que es la verdadera proporción de BET en la bodega. Las líneas punteadas azules corresponden a 0.05 para el error y a 0.20 para el error relativo. La línea punteada roja está en 0.

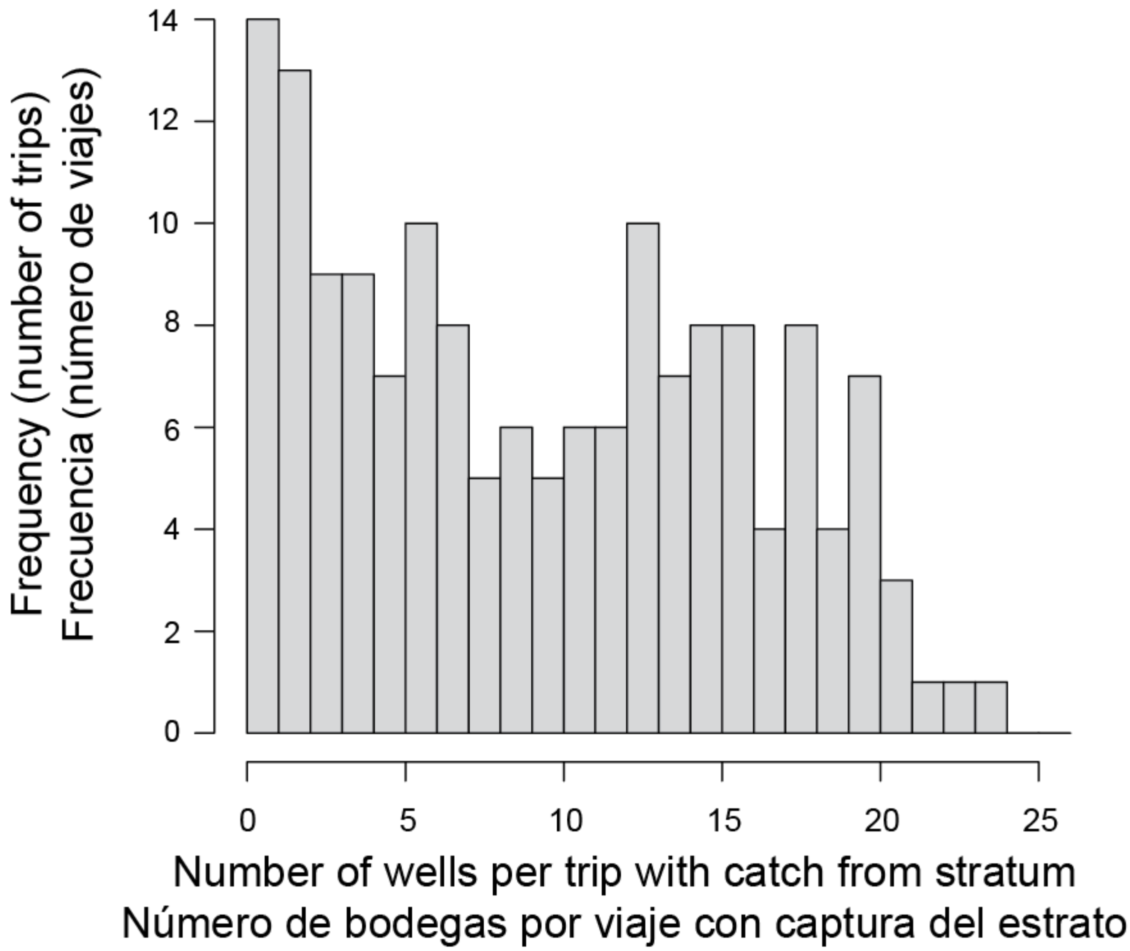


FIGURE 12. Number of wells per trip (from the reconstructed RDL observer data) with catch from the stratum OBJ sets made west of 110°W, 2019.
FIGURA 12. Número de bodegas por viaje (a partir de los datos de observadores del RDL reconstruidos) con captura de lances OBJ del estrato realizados al oeste de 110°O, 2019.

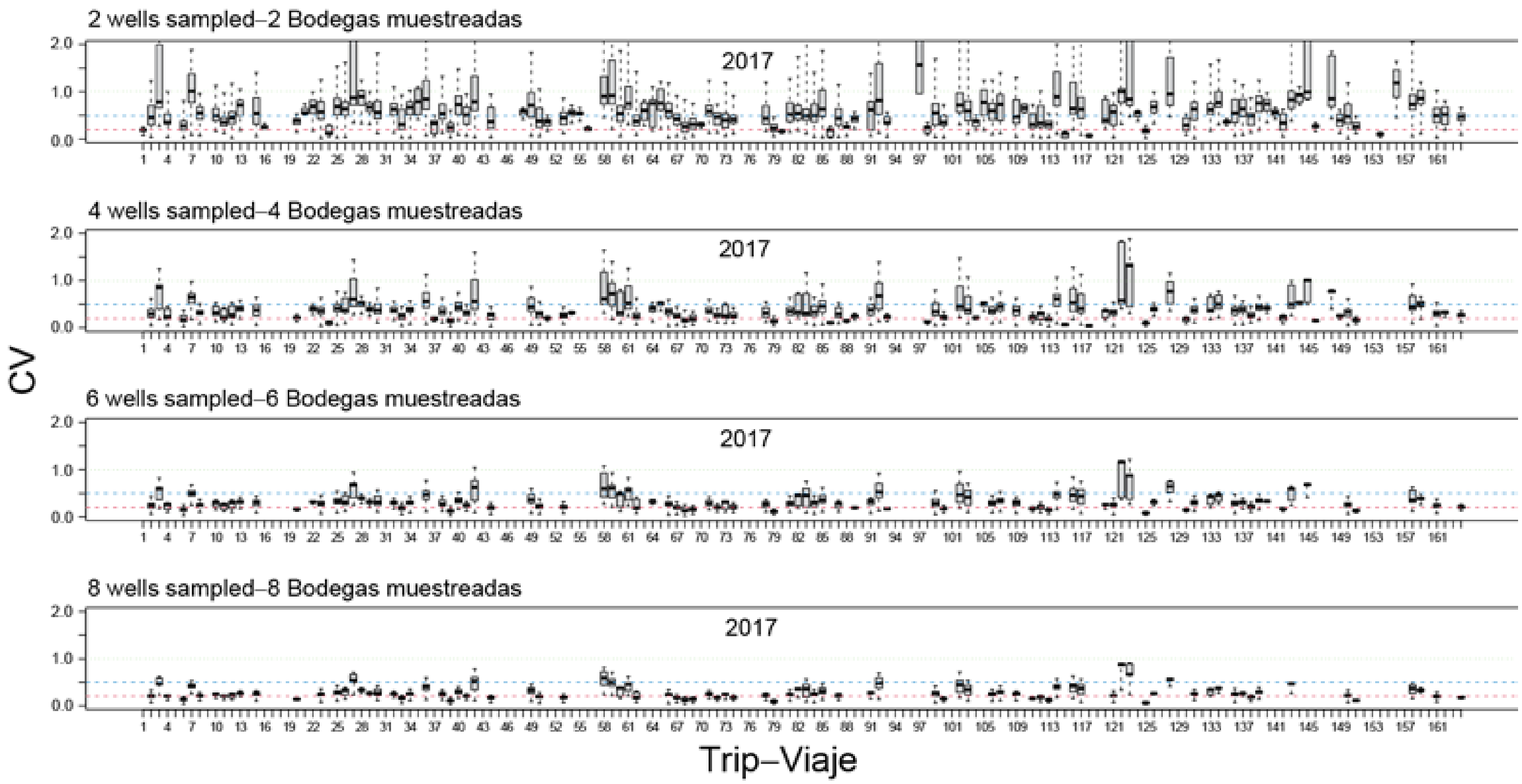


FIGURE 13. Box-and-whisker plots of the coefficient of variation (CV) values for each random sample of 2 wells per trip (top graph), 4 wells per trip (second graph), 6 wells per trip (third graph) and 8 wells per trip (bottom graph), from the simulations based on reconstructed RDL observer data for 2017 – 2019. Red dashed line: CV of 0.2; blue dashed line: CV of 0.5; green dashed line: CV of 1.0. Not all trips had at least 8 wells with catch from the stratum; as the number of wells to sample increases, some trips will not appear in subsequent graphs because the number of wells in the stratum was fewer than the number of wells to sample.

FIGURA 13. Gráficas de cajas y bigotes de los valores de coeficiente de variación (CV) para cada muestra aleatoria de dos bodegas por viaje (gráfica superior), cuatro bodegas por viaje (segunda gráfica), seis bodegas por viaje (tercera gráfica) y ocho bodegas por viaje (gráfica inferior), a partir de las simulaciones basadas en los datos de observadores del RDL reconstruidos para el periodo 2017-2019. Línea punteada roja: CV de 0.2; línea punteada azul: CV de 0.5; línea punteada verde: CV de 1.0. No todos los viajes tenían al menos ocho bodegas con captura del estrato; a medida que aumenta el número de bodegas para muestrear, algunos viajes no aparecerán en las siguientes gráficas porque el número de bodegas del estrato era inferior al número de bodegas para muestrear.

FIGURE 13. continued.

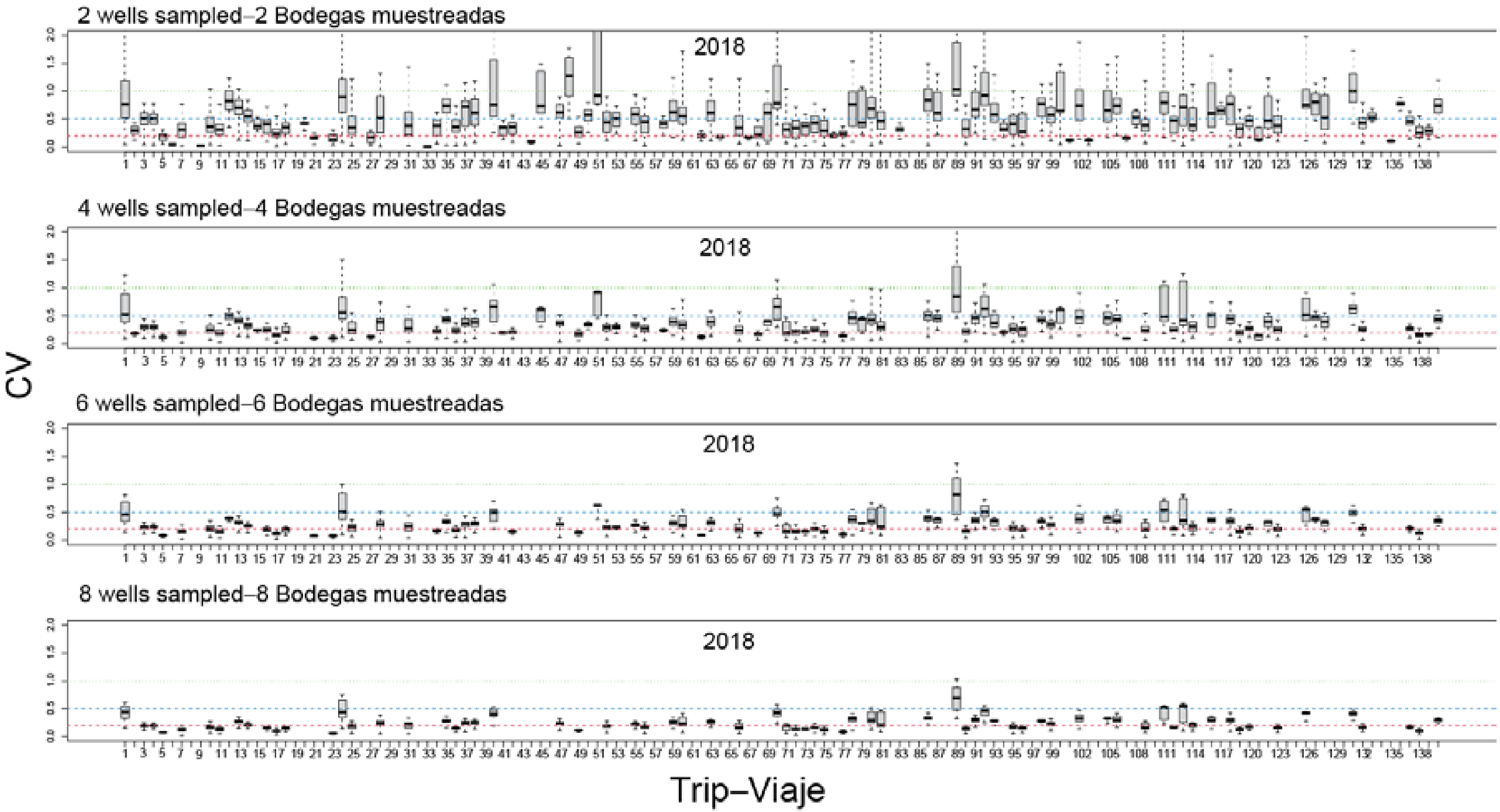
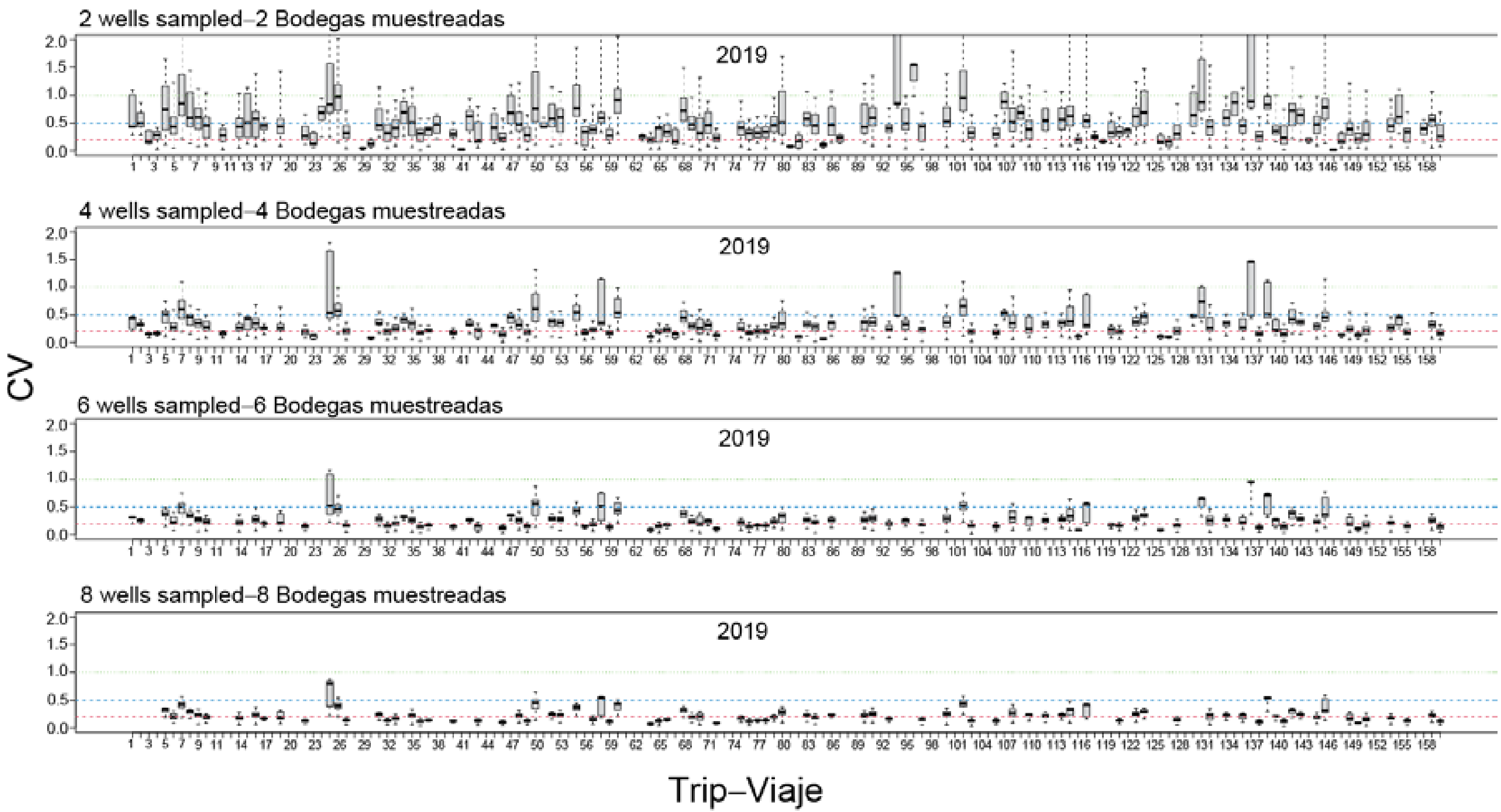


FIGURE 13. continued.



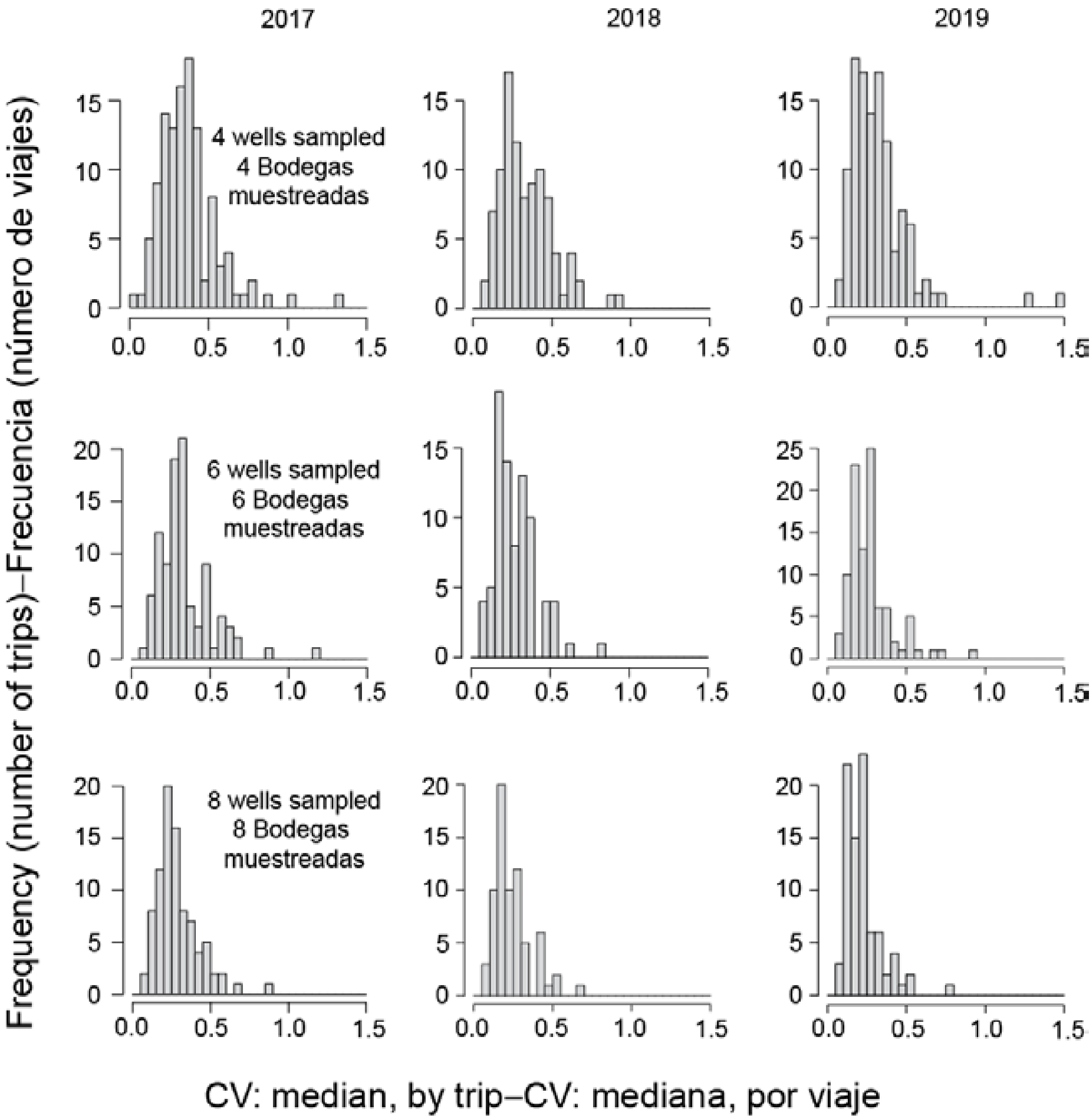


FIGURE 14. Histograms of the median CV, by trip, across possible samples of 4 wells (top), 6 wells (middle) and 8 wells (bottom), for wells with catch from OBJ sets made west of 110°W, for 2017 – 2019. These histograms show the distribution of median CV values indicated in Figure 13 by the horizontal black bar within each gray box.

FIGURA 14. Histogramas de la mediana del CV, por viaje, a través de posibles muestras de cuatro bodegas (arriba), seis bodegas (centro) y ocho bodegas (abajo), para bodegas con captura de lances OBJ realizados al oeste de 110°O, para el periodo 2017-2019. Estos histogramas muestran la distribución de los valores medianos del CV que se indican en la Figura 13 mediante la barra negra horizontal dentro de cada caja gris.

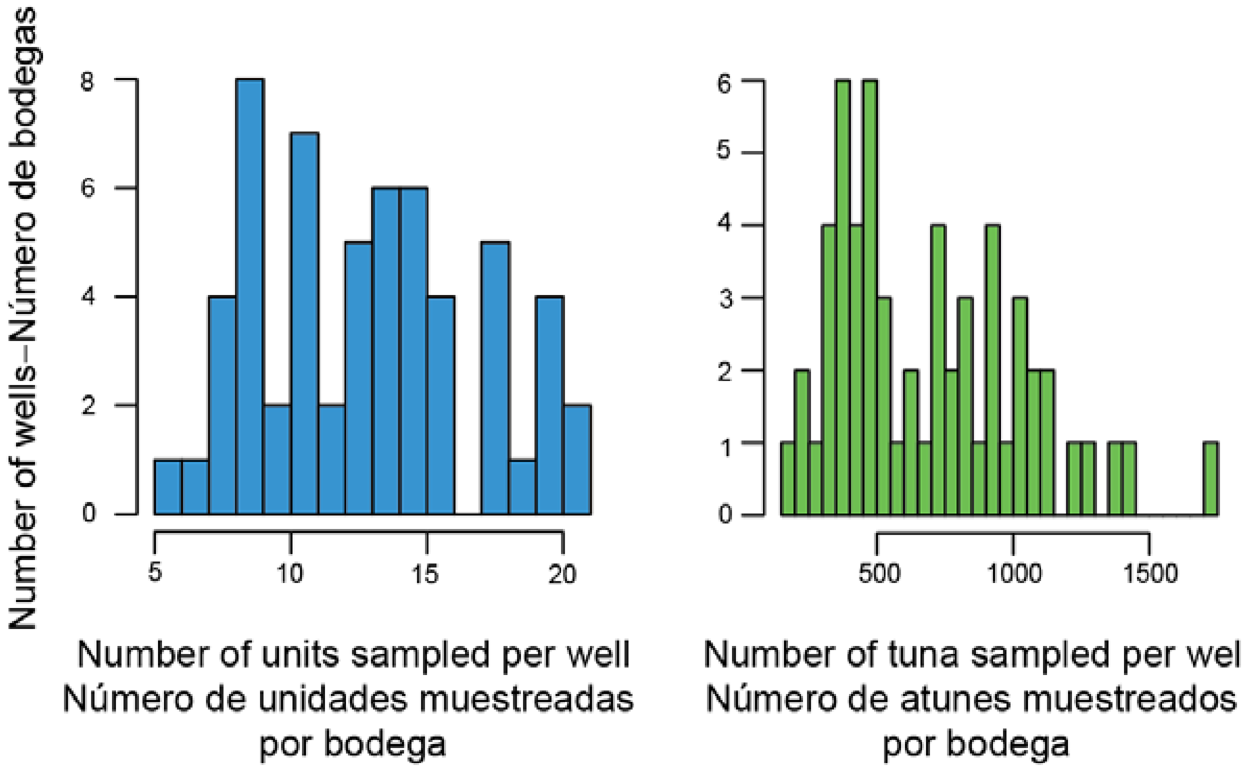


FIGURE 15. Number of units sampled per well (left) and number of tuna sampled per well (right) under the preliminary EMP protocol tested in Phase 2.

FIGURA 15. Número de unidades muestreadas por bodega (izquierda) y número de atunes muestreados por bodega (derecha) según el protocolo preliminar del PRM que se probó en la Fase 2.

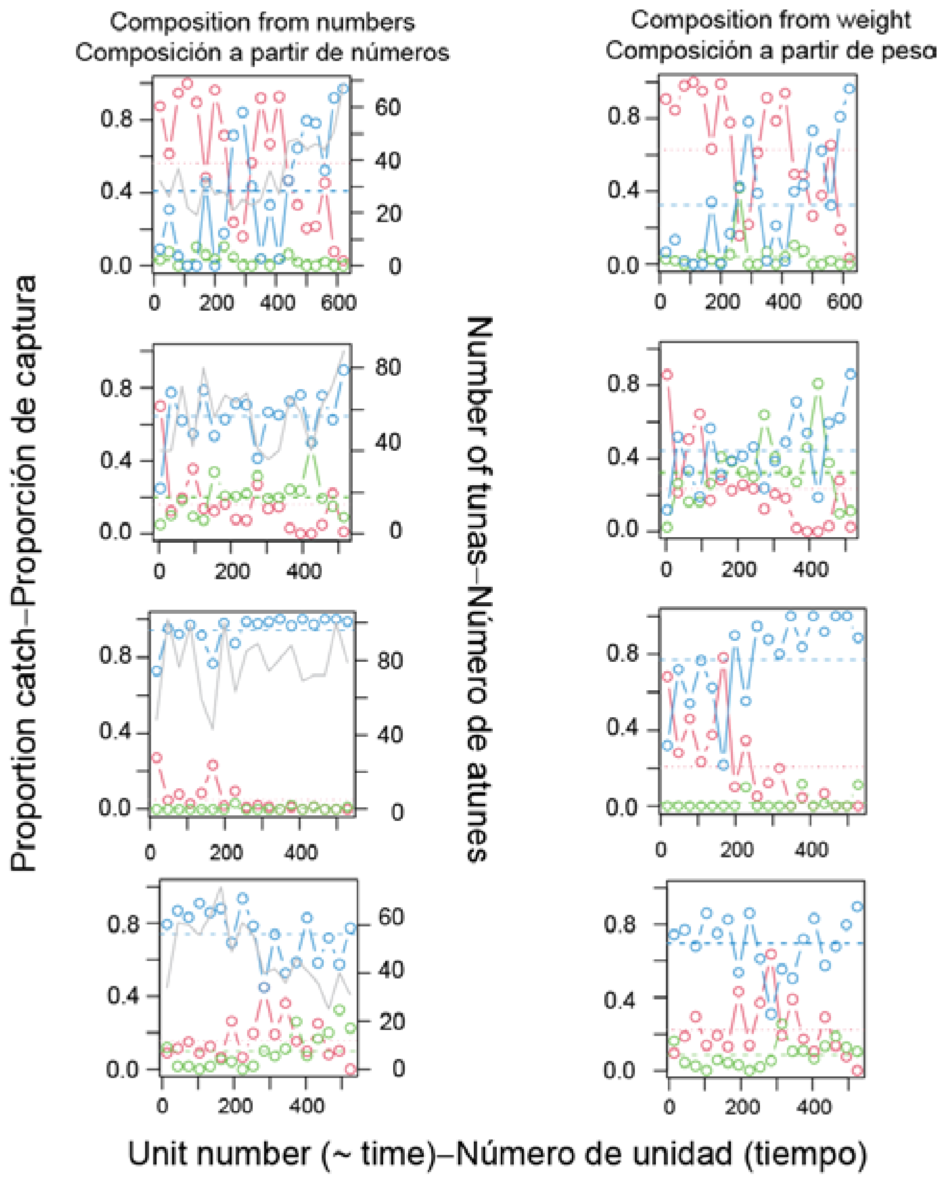


FIGURE 16. Tropical tuna species composition, by sampled unit, computed from numbers of fish (left-hand side) and from weight (right-hand side), for 4 wells of the same trip, sampled under the preliminary EMP sampling protocol tested in Phase 2. All four wells contained catch from floating-object sets made west of 110W (4 sets loaded into the well shown in the top panel, and 3 sets into each of the other three wells). Red: BET; green: YFT; blue: SKJ; gray: number of tropical tunas. Dashed lines indicate the average species proportion (computed over units). The x-axis shows the unit number, from the start (left) to the end (right) of the unloading.

FIGURA 16. Composición por especie de atunes tropicales, por unidad muestreada, calculada a partir del número de peces (lado izquierdo) y del peso (lado derecho), para cuatro bodegas del mismo viaje, muestreadas según el protocolo preliminar de muestreo del PRM que se probó en la Fase 2. Las cuatro bodegas contenían captura de lances sobre objetos flotantes realizados al oeste de 110°O (cuatro lances cargados en la bodega se muestran en el panel superior y tres lances en cada una de las otras tres bodegas). Rojo: BET; verde: YFT; azul: SKJ; gris: número de atunes tropicales. Las líneas punteadas indican la proporción promedio de las especies (calculada a partir de las unidades). El eje *x* muestra el número de la unidad, desde el inicio (izquierda) hasta el final (derecha) de la descarga.

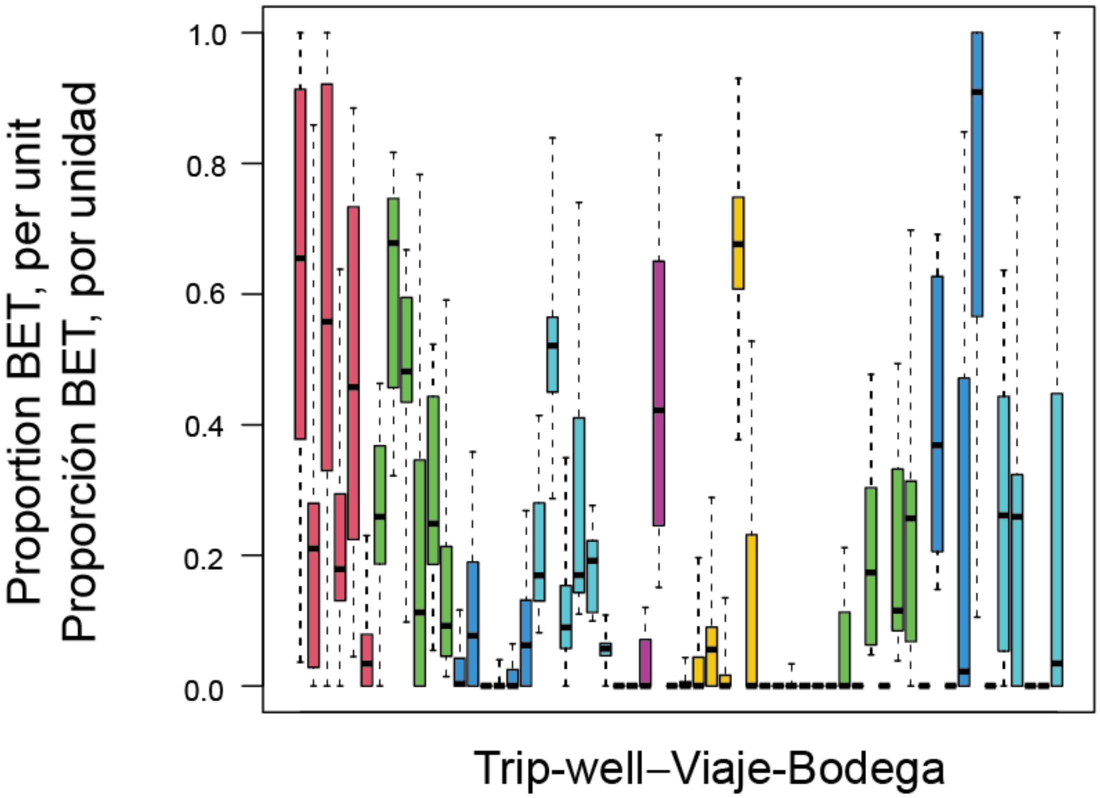


FIGURE 17. Box-and-whisker plots of the proportion of BET, per unit, of each of the 58 wells sampled under the preliminary EMP sampling protocol. Wells sampled from the same trip are arranged together and have the same color. Horizontal black bars indicate the median proportion, the colored box indicates the interquartile range, and the whiskers extend to the extremes of the unit proportions for the well.
FIGURA 17. Gráficas de cajas y bigotes de la proporción de BET, por unidad, para cada una de las 58 bodegas muestreadas según el protocolo preliminar de muestreo del PRM. Las bodegas muestreadas del mismo viaje se ordenan juntas y tienen el mismo color. Las barras negras horizontales indican la proporción mediana, la caja de color indica el rango intercuartil y los bigotes se extienden hasta los extremos de las proporciones de la unidad de la bodega.

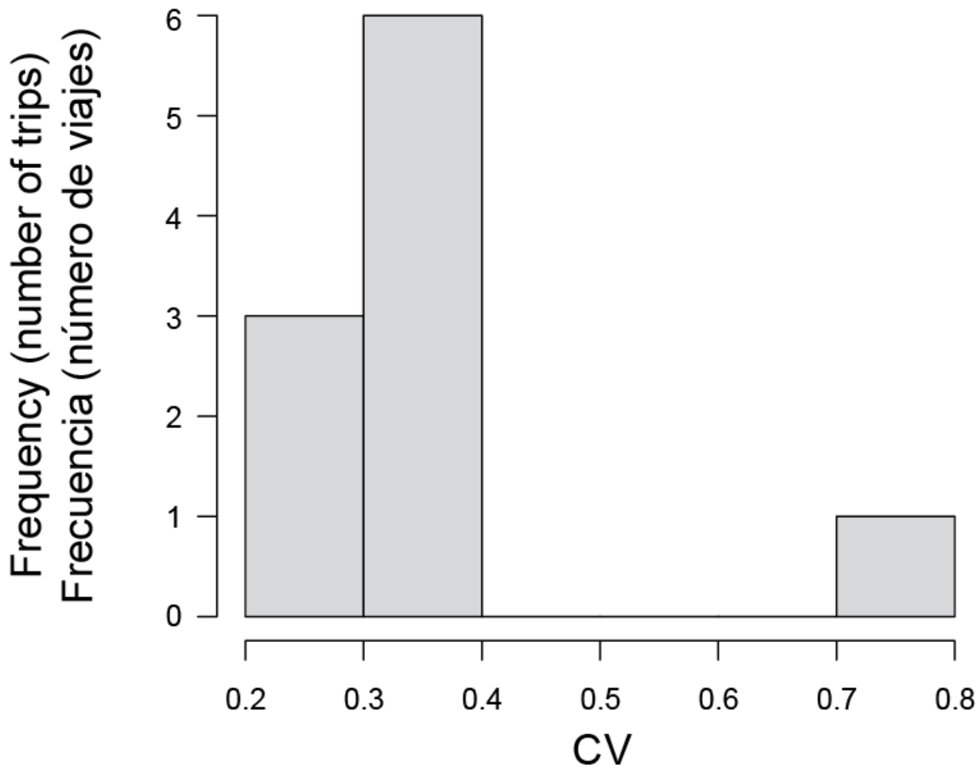


FIGURE 18. Histogram of the coefficient of variation (CV) values, on the proportion BET for a stratum, from Table 4.

FIGURA 18. Histograma de los valores de coeficiente de variación (CV), sobre la proporción de BET para un estrato, de la Tabla 4.

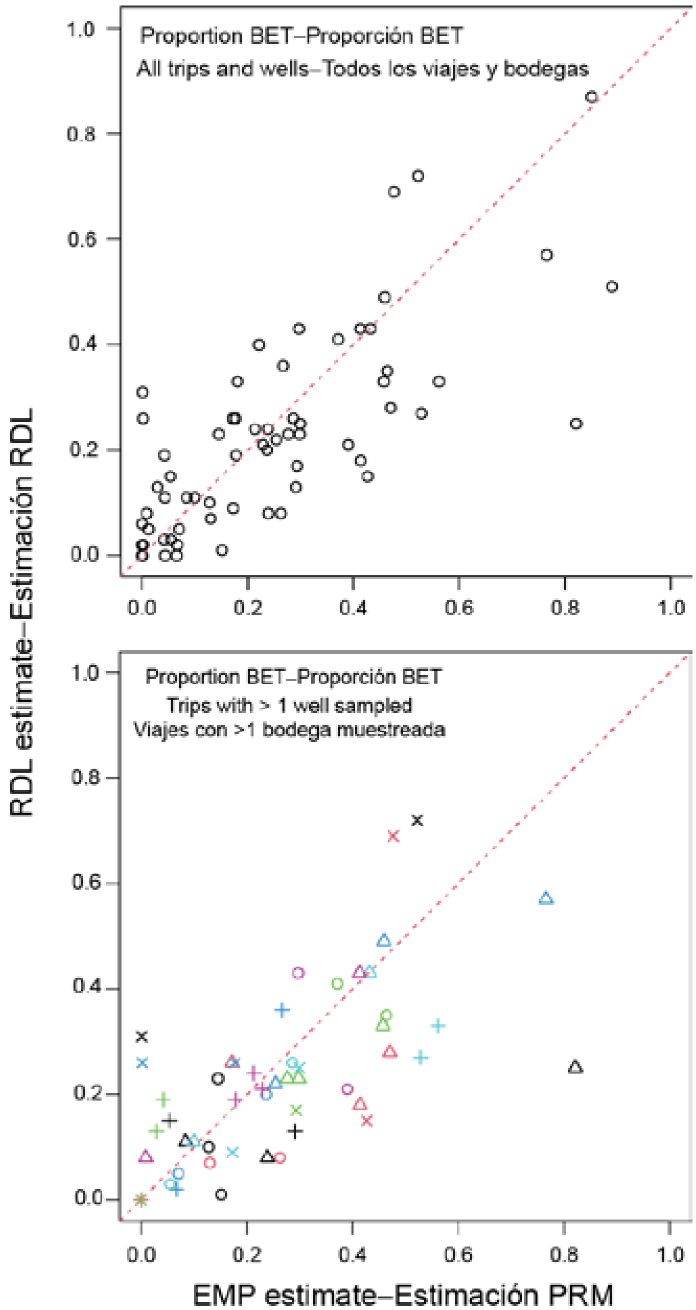


FIGURE 19. Estimated proportion of BET, by well, for the EMP Phase 1 intensive sampling data plotted against the estimates from RDL observer data for the same wells. Top: all wells sampled (71 wells from 42 trips); bottom: only wells of trips with more than one well sampled (2-3 wells per trip for 52 wells from 23 trips; 17 trips with 2 wells sampled and 6 trips with 3 wells sampled); wells of the same trip have the same plot symbol and color.

FIGURA 19. Proporción estimada de BET, por bodega, para los datos de muestreo intensivo de la Fase 1 del PRM comparada con las estimaciones de los datos de observadores del RDL para las mismas bodegas. Arriba: todas las bodegas muestreadas (71 bodegas de 42 viajes); abajo: solo bodegas de viajes con más de una bodega muestreada (de dos a tres bodegas por viaje para 52 bodegas de 23 viajes; 17 viajes con dos bodegas muestreadas y seis viajes con tres bodegas muestreadas); las bodegas del mismo viaje tienen el mismo símbolo y color en la gráfica.

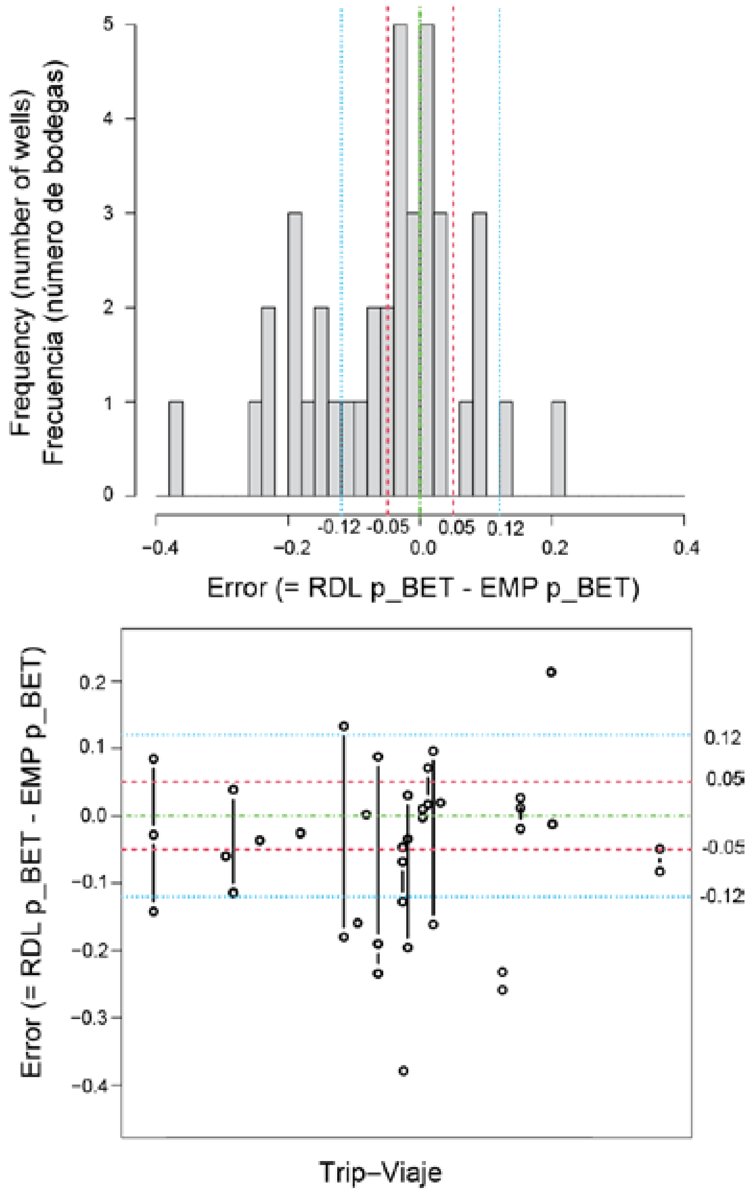


FIGURE 20. Error (RDL estimate of proportion of BET – EMP Phase 1 estimate of proportion of BET) for wells with catch from OBJ sets made west of 110°W. Top: histogram of by-well error values; bottom: by-well error values grouped by trip (when more than one well was sampled per trip, the values are joined by a vertical line). The red dashed lines correspond to an error of +/-5%, for comparison to performance of the simulated 1-of-3 protocol (i.e., 3.33% coverage of units) (Table 3, Figure 11). The blue dashed lines indicate the maximum error found for any cluster sample under the simulated 1-of-3 (Table 3). The green dashed line is at the value 0.

FIGURA 20. Error (estimación del RDL de la proporción de BET – estimación del PRM de la Fase 1 de la proporción de BET) para bodegas con captura de lances OBJ realizados al oeste de 110°O. Arriba: histograma de valores de error por bodega; abajo: valores de error por bodega agrupados por viaje (cuando se muestreó más de una bodega por viaje, los valores aparecen unidos por una línea vertical). Las líneas punteadas rojas corresponden a un error de +/-5%, para comparar con el rendimiento del protocolo simulado 1 de 3 (es decir, cobertura del 3.33% de las unidades) (Tabla 3, Figura 11). Las líneas punteadas azules indican el error máximo encontrado para cualquier muestra de conglomerados según la simulación 1 de 3 (Tabla 3). La línea punteada verde está en el valor 0.

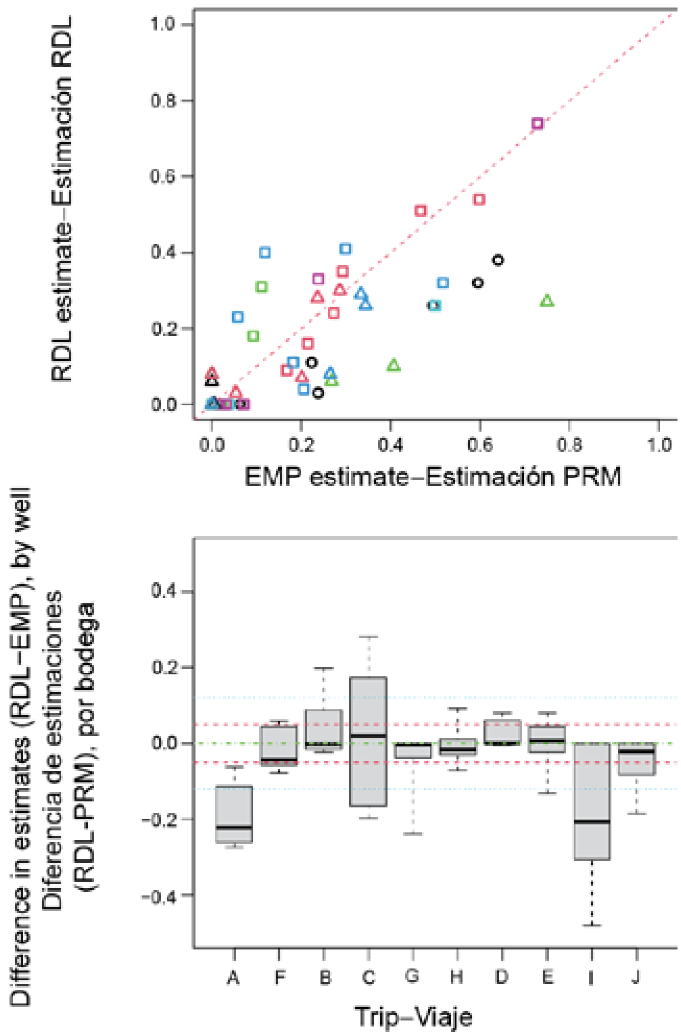


FIGURE 21. Comparison of EMP estimates of the proportion of BET from the preliminary EMP protocol tested in Phase 2 *versus* the RDL well-level estimates of the proportion of BET, for the same wells. Top: wells of the same trip are indicated by the same plot symbol and color (the red dashed line is the 1-to-1 line); bottom: box-and-whisker plot of the differences between the two estimates (RDL – EMP), by trip. The blue dashed lines in the bottom graph indicate the maximum error from the simulated 1-of-3 protocol (i.e., 3.33% coverage of units) (Table 3) and the red dashed lines correspond to an error of +/-5%, for comparison to performance of the simulated 1-of-3 protocol (Table 3, Figure 11). The green dashed line is at the value 0. Trip labels in bottom panel match those shown in Table 4.

FIGURA 21. Comparación de las estimaciones del PRM de la proporción de BET a partir del protocolo preliminar del PRM que se probó en la Fase 2 frente a las estimaciones del RDL a nivel de bodega de la proporción de BET, para las mismas bodegas. Arriba: las bodegas del mismo viaje se indican con el mismo símbolo y color en la gráfica (la línea punteada roja es la línea 1 a 1); abajo: gráfica de cajas y bigotes de las diferencias entre las dos estimaciones (RDL – PRM), por viaje. Las líneas punteadas azules de la gráfica inferior indican el error máximo del protocolo simulado 1 de 3 (es decir, cobertura del 3.33% de las unidades) (Tabla 3) y las líneas punteadas rojas corresponden a un error de +/-5%, para comparar con el rendimiento del protocolo simulado 1 de 3 (Tabla 3, Figura 11). La línea punteada verde está en el valor 0. Las identificaciones de los viajes en el panel inferior coinciden con las de la Tabla 4.

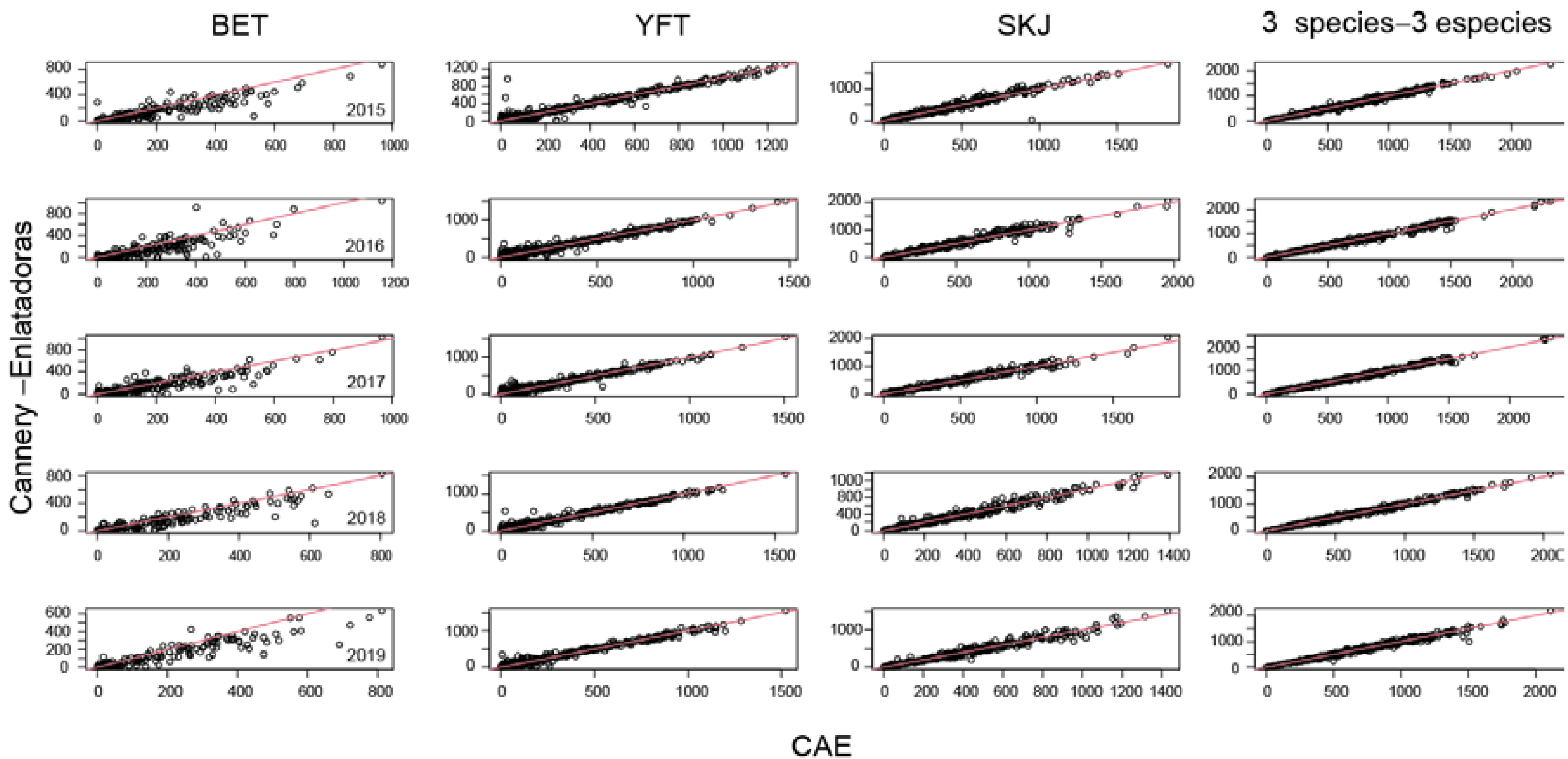


FIGURE 22. Comparison of trip-level estimates of BET from observer and logbook (CAE) data *versus* cannery data, for Class-6 vessels by year, 2015-2019. The dashed red lines indicate the 1-to-1 line.

FIGURA 22. Comparación de las estimaciones de BET a nivel de viaje a partir de los datos de observadores y de bitácora (CAE) frente a los datos de enlatadoras, para los buques de clase 6 por año, para el periodo 2015-2019. Las líneas punteadas rojas indican la línea 1 a 1.

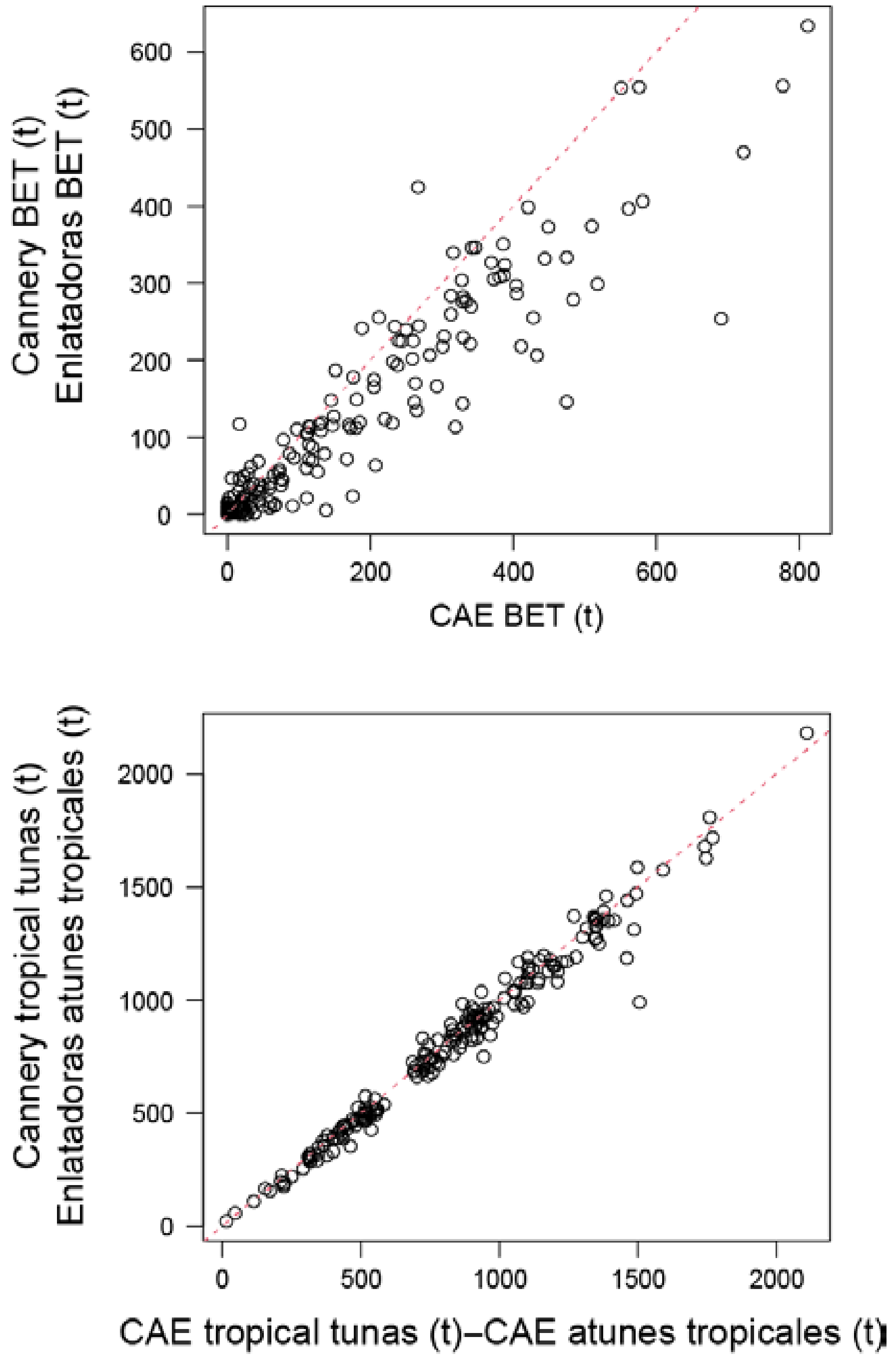


FIGURE 23. Comparison of trip-level BET and tropical tunas estimates from observer and logbook (CAE) data to cannery estimates, for those trips for which the cannery BET estimate was greater than zero, Class-6 vessels, 2019. Tropical tuna catch is the sum of catches of YFT, SKJ and BET. The dashed red lines indicate the 1-to-1 line. The estimated slopes of these two relationships are shown in Table 5.

FIGURA 23. Comparación de las estimaciones de BET y atunes tropicales a nivel de viaje a partir de los datos de observadores y de bitácora (CAE) con las estimaciones de las enlatadoras, para aquellos viajes en los que la estimación de BET de las enlatadoras fue superior a cero, buques de clase 6, 2019. La captura de atunes tropicales es la suma de las capturas de YFT, SKJ y BET. Las líneas punteadas rojas indican la línea 1 a 1. Las pendientes estimadas de estas dos relaciones se muestran en la Tabla 5.