NOAA Technical Memorandum NMFS-AFSC-397

Deployment Performance Review of the 2018 North Pacific Observer Program

P. Ganz, S. Barbeaux, J. Cahalan, J. Gasper,
S. Lowe, R. Webster, and C. Faunce

September 2019

U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Alaska Fisheries Science Center

The National Marine Fisheries Service's Alaska Fisheries Science Center uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series reflect sound professional work and may be referenced in the formal scientific and technical literature.

The NMFS-AFSC Technical Memorandum series of the Alaska Fisheries Science Center continues the NMFS-F/NWC series established in 1970 by the Northwest Fisheries Center. The NMFS-NWFSC series is currently used by the Northwest Fisheries Science Center.

This document should be cited as follows:

Ganz, P., S. Barbeaux, J. Cahalan, J. Gasper, S. Lowe, R. Webster, and C. Faunce. 2019. Deployment performance review of the 2018 North Pacific Observer Program. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-397, 64 p.

This document is available online at:

Document available: https://www.afsc.noaa.gov/Publications/AFSC-TM/NOAA-TM-AFSC-397.pdf

Deployment Performance Review of the 2018 North Pacific Observer Program

P. Ganz[1], S. Barbeaux[2], J. Cahalan[3], J. Gasper[1], S. Lowe[2], R. Webster[4], and C. Faunce[5,*]

[1]Sustainable Fisheries Division
Alaska Regional Office
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
709 West 9th Street
Juneau, AK, 99801

[2]Resource Ecology and Fisheries Management Division
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE
Seattle, WA 98115

[3]Pacific States Marine Fisheries Commission
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE
Seattle, WA 98115

[4]International Pacific Halibut Commission
2320 West Commodore Way, Suite 300
Seattle, WA 98199

[5]Fisheries Monitoring and Analysis Division
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE
Seattle, WA 98115

* Corresponding author
E-mail: craig.faunce@noaa.gov

U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Alaska Fisheries Science Center

NOAA Techncial Memorandum NOAA-TM-AFSC-397

September 2019

ABSTRACT

This report contains the analyses and findings of the Alaska Fisheries Science Center's Fisheries Monitoring and Analysis Division's Observer Science Committee (OSC) on the efficiency and effectiveness of observer deployment following the 2018 Annual Deployment Plan. Responses to comments by the North Pacific Fishery Management Council's Science and Statistical Committee from the 2017 version of this report, and recommendations to improve data quality and guide the 2020 Annual Deployment Plan are also included. In 2018, there were 11 strata to evaluate: one full coverage stratum, five partial coverage observer strata defined by gear and tender designation, one partial coverage Electronic Monitoring (EM) stratum for hook-and-line (HAL) vessels, two EM strata in pre-implementation for vessels fishing pot gear with and without tendering, one zero coverage EM research stratum, and one zero coverage stratum. Observers were deployed under trip-selection on 159 full coverage vessels that fished for 3,400 trips and 602 partial coverage vessels that fished 5,318 trips. This was the first year in which data from the *EM HAL* stratum were used in catch estimation and this represents the first time EM data have been used in this way in Alaska. The *EM HAL* stratum was the third largest stratum by trip count in 2018 with 767 total trips fished. Although the resulting data are not used by the National Marine Fisheries Service (NMFS), "pre-implementation" EM systems were successfully deployed onto 18 pot vessels that fished for 164 trips. EM systems were deployed for research onto three vessels that fished for 23 trips. A total of 361 vessels fished 1,725 trips with no chance of being observed or monitored (zero selection stratum).

Coverage rates in full coverage, zero coverage, and the pre-implementation EM strata for pot vessels met their expectations for the year. Coverage rates in partial coverage met their expectations in three out of six gear- and tender-based strata, with coverage rates being lower than expected for the *HAL* and *EM HAL* strata, and higher than expected for the trawl tender stratum. The reasons for low coverage differed between the *HAL* and *EM HAL* strata. Trips were selected for observer coverage at the expected

rate for the *HAL* stratum, but observed at a lower rate than expected, suggesting a disconnect between the trip selection process and actual fishing activity. Trips within the *EM HAL* stratum were selected for monitoring at a rate higher than expected once cancellations, inherited coverage from cancellations, and waivers were taken into account, but actually monitored at a rate much lower than expected due to a shortage of personnel available to review video. As a result, the National Marine Fisheries Service (NMFS) plans to work with the Pacific States Marine Fisheries Commission in devoting more personnel to this task going forward.

The OSC recommends that the NMFS evaluate the Observer Declare and Deploy System used for trip logging and selection in order to address issues with temporal bias such as those seen in the *HAL* stratum. There was less evidence of data gaps and spatial bias in 2018 than in 2017. These results led to the OSC recommendation that the 15% minimum coverage hurdle that was in place in 2018 be continued into the future. Finally, while the OSC recommended a re-evaluation of the strata defined by gear and tendering status (to possibly eliminate these strata due to low fishing effort and sample size), it did not support a further stratification of trawl trips by pelagic and non-pelagic gear types. Both recommendations are due in part to the concern that it is logistically difficult to deploy observers according to these trip characteristics, as these trip characteristics are not always known with certainty before a trip begins. Additional details on the differences between pelagic and non-pelagic trips are provided in the Appendix.

CONTENTS

INTRODUCTION

Background of the North Pacific Groundfish and Halibut Observer Program

Fisheries observers collect independent information that is used to determine the effects of fishing on natural resources. The National Marine Fisheries Service (NMFS) uses its observer program in Alaska to enable the use of tools such as catch quotas to manage against the over- or under-harvest of fishes. The data collection by observers is currently the only reliable and verifiable method for collecting fishery discard information used to estimate total catch, as well as seabird and marine mammal interactions with fisheries. In addition, observers collect biological information such as length, sex, weight, age structures (e.g., otoliths, spines, scales, and vertebrae), and stomachs to support ecosystem studies and stock assessments.

The observer program in the North Pacific has a long history. Observers were first deployed onto fishing vessels in the Bering Sea in 1973 and into the remainder of the North Pacific in 1975 (Nelson et al. 1981, Wall et al. 1981). Fisheries in the North Pacific were initially prosecuted exclusively by foreign and later by "joint venture" operations where a developing domestic fleet of catcher vessels delivered to foreign-owned processing vessels. During the foreign and joint venture operations, foreign vessels carried fisheries observers at their expense while domestic vessels were exempted from this observer coverage. As foreign vessels' rights to fish in the U.S. Exclusive Economic Zone (EEZ) were reduced over time and the domestic fishery grew, it became evident to managers that observer coverage would be necessary for the emerging domestic fleet. At the onset of fully domestic fishery operations in 1990, the North Pacific Groundfish Observer Program was established as an interim observer program with rules governing observer coverage codified in regulations. This interim program would be extended four times over the next 20 years by the North Pacific Fishery Management Council (Council) - the last without a sunset date.

The regulations established in 1990 required vessels 60-125 feet in length (overall) and all vessels fishing pot gear to carry observers at their own cost for 30% of their fishing days in a calendar quarter plus at least one trip in each fishery they participate in (termed the “30% fleet”), and vessels greater than 125 feet in length to carry an observer for 100% of their fishing days at their expense. Some vessels were not required to carry observers. These included vessels less than 60 feet, vessels fishing jig gear or vessels fishing with trawl gear that delivered unsorted codends to processing vessels (termed “catcher processors” or CPs if the vessel also had catching ability and “mothership” or M if the vessel did not) and vessels that fished for Pacific halibut (*Hippoglossus stenolepis*). For shoreside processors, the rules governing observer coverage were based on the estimated tonnage processed in a calendar month: plants that processed less than 500 metric tons (t) a month were exempted from coverage, those that processed between 500 t and 1,000 t a month were required to be observed for 30% of the calendar days, and those that processed more than 1,000 t a month were required to be observed for each day in the month.

Soon after the establishment of the domestic observer program, concerns over the ability and incentive for fishers to manipulate observer coverage in a way that might bias catch estimates and other analytic products prompted efforts by NMFS and the Council to provide a mechanism for NMFS to gain control over where and when observers were deployed (Faunce and Barbeaux 2011). From 1992 to 2008, several attempts to “restructure” the program were made. In 2010, the Council unanimously decided to move forward with the restructured observer program. In 2012, the Final Rule 77 FR 70062 was published to implement Amendment 86 to the Fishery Management Plan for Groundfish of the Bering Sea and Aleutian Islands (BSAI) Management Area and Amendment 76 to the Fishery Management Plan for Groundfish of the Gulf of Alaska (GOA). Amendments 86/76 added a funding and deployment system for observer coverage to the existing North Pacific Groundfish Observer Program and amended existing observer coverage requirements for vessels and processing plants. The “restructured” North Pacific Groundfish and Halibut Observer Program (hereafter termed “Observer Program”) began in 2013 with the randomization of deployments among trips and vessels. Since 2013, the Observer Program has employed a hierarchical sampling design with randomization at all levels (trips > haul > species composition > length, age, sex, maturity and tissues for genetic analysis). In 2018, the use of electronic monitoring (EM) was added as an additional catch monitoring tool, with the understanding that some data elements collected by observers would not be collected using EM systems.

THE ANNUAL DEPLOYMENT PLAN AND REVIEW

Analysis and evaluation of the data collected by observers is an ongoing process. The NMFS considers Council input in making decisions as to the amount of coverage (i.e., selection probabilities that are assigned to each partial-coverage category). These decisions are based on available funding, the cost of observer coverage, and anticipated effort. The restructure of the Observer Program established new annual reporting processes. Each June, the NMFS provides the Council with a comprehensive evaluation of past years' observer deployments, costs, sampling levels, and implementation issues as well as recommended changes for the coming year. The June deployment performance review aims to identify areas where improvements are needed to 1) collect the data necessary to manage the groundfish and halibut fisheries; 2) maintain the scientific goals of unbiased data collection; and 3) accomplish the most effective and efficient use of the funds collected through the observer fee. The annual deployment performance review is an opportunity to inform the Council and the public of how well various aspects of the program are working, and consequently lead to recommendations for improvement as appropriate. The NMFS also prepares the Observer Program Annual Deployment Plan (ADP) each fall. The ADP defines deployment strata and establishes selection rates given available budgets and anticipated fishing effort. A draft ADP is released by September of each year to allow review by the Council's Groundfish Plan Teams, as well as the Council and its Scientific and Statistical Committee (SSC). Based on input from its advisory bodies and the public, the Council may choose to clarify objectives and provide recommendations to NMFS for the ADP. Upon analysis of the Council recommendations, NMFS will make any necessary adjustments to finalize the ADP and release it to the public. The ADP is released to the public prior to the December Council meeting.

Observer Science Committee

Each year the Alaska Fisheries Science Center's (AFSC) Fisheries Monitoring and Analysis (FMA) Division establishes an *ad hoc* Observer Science Committee (OSC) for the North Pacific Observer Program. The OSC provides scientific advice in the areas of regulatory management, natural science, mathematics, and statistics as they relate to observer deployment and sampling in the groundfish and halibut fisheries of the BSAI and the GOA. The OSC members have analytical and scientific expertise relating to observer sampling of groundfish and halibut fisheries of the BSAI and GOA and use of the collected data. If possible, the OSC is represented by at least one member of the AFSC/FMA (Observer Program) Division, one member of the AFSC/Stock Assessment and Multispecies Assessments Program, one member of the Alaska Regional Office (AKRO) Sustainable Fisheries Division, and one member of the International Pacific Halibut Commission (IPHC).

The Sampling Design of the Observer Program

Since 2013, the Observer Program has used a stratified hierarchical sampling design with randomization at all levels. Stratification is used to increase the efficiency of sampling by observers and to address logistical issues associated with deployment. By grouping similar fishing activities into strata and sampling those strata appropriately, the variance of resulting estimates may also be decreased. Sampling strata are defined in the ADP and are designed such that a unit of deployment (trip) is generally unique to a stratum.

Within a stratum, observers are deployed randomly to either vessels for a predetermined period of time (termed vessel-selection), or to individual fishing trips (termed trip-selection). In both cases, this initial deployment to the fishery is the first level of the sampling hierarchy and defines the primary sampling unit (PSU; either vessel-periods or individual trips). The list of all PSUs in a stratum defines the sampling frame and should equate to the population of interest for that sampling stratum (e.g., all trips taken by trawl vessels fishing in the U.S. EEZ off Alaska). In cases where the sampling frame (list of

PSUs) for a stratum does not include all the elements of the stratum (i.e., where some fishing occurs in the stratum but is not captured by the sample frame), the resulting information from sampling may not represent the population of trips. The magnitude and direction of the bias will depend on how different the fishing activities in the sample frame are from actual fishing activity.

For each observed trip, if all hauls cannot be sampled for logistical reasons, hauls are randomly selected to be sampled. This is the next level in the hierarchy; the secondary sampling units are defined as hauls within a trip. Randomization of haul selection is designed to allow observers to record and transmit data, attend to other non-sampling responsibilities, and to allow observers time to sleep and eat. Randomization of haul selection also gives EM video reviewers the ability to optimize the amount of video that can be reviewed from each trip. Haul selection is determined using the random sampling tables and random break tables provided by NMFS. For each haul, fishing location and effort (e.g., number of hooks) are recorded, while marine mammal and seabird interactions are primarily recorded on randomly selected hauls. The ability of EM to capture marine mammal and seabird interactions is less than that of observers due to the fixed location in which EM equipment is placed.

For the randomly selected hauls for each trip, a random sample of the catch is collected (observers) or selected for video review (EM), and data from those samples are used to determine the species composition and amount of discarded catch. These samples of catch within each haul are the third level of the sampling hierarchy. While observers are trained to collect multiple large samples of catch, the number and size of samples taken from each haul will depend on the vessel configuration, fishing operations, and diversity of catch. The size of EM samples is largely determined by the number of video reviewers available relative to the amount of video to be reviewed.

At the fourth level of the sampling hierarchy, a predetermined number of individual fish of predetermined species is randomly selected from the species composition sample and measured. Lastly, at the fifth sampling level, a random selection of fish is used to collect otoliths, reproductive maturity assessments, stomach contents, genetic tissues, and other biological specimens. The number and species of fish selected for measurement and biological specimen collection are specified each year by the

AFSC's stock assessment scientists. Sampling rates for genetic tissue collection by observers (e.g., 1 of 10 Chinook salmon caught as bycatch) are set each year by the Genetics Program of the AFSC's Auke Bay Laboratory. Sampling at the fourth and fifth levels of the sampling hierarchy does not occur with EM.

In summary, the overall sample design used by the Observer Program is a stratified design where, within each stratum, NMFS randomly selects primary units (vessels or trips) to be monitored. Within each selected trip, hauls are randomly selected to be further sampled, and marine mammal and seabird interaction data are collected. From each selected haul, a random sample of the catch is collected to obtain species composition and disposition data. From each species composition sample, individual fish are randomly selected and measured by observers only. Finally, from these measured fish, additional fish are randomly selected for the collection of biological specimens by observers only. More information on the sampling design used by observers and the relationship between the sample design and catch estimation can be found in Cahalan et al. (2014) and the 2018 Observer Sampling Manual (AFSC 2017). The focus of this report is deployment related and the resulting evaluation is at the trip level of the sampling hierarchy.

THE 2018 ANNUAL DEPLOYMENT PLAN

The following briefly summarizes the final 2018 ADP (NMFS 2017b). In general, all vessels that participated in cooperatives or acted as catcher-processors or motherships were fully observed at the trip-level and constitute the full-coverage category of the fleet. In 2016, NMFS published new regulations to allow the owner of a trawl catcher vessel to annually request that NMFS place requesting vessels in the full coverage category for all directed fishing for groundfish using trawl gear in the Bering Sea and Aleutian Islands management area (BSAI) in the following calendar year. This regulated process has replaced an interim policy. For the 2017 calendar year, NMFS received and approved requests and has placed 34 catcher vessels in the full coverage category for all directed fishing for groundfish using trawl gear in the BSAI management area (NMFS 2017b). The NMFS used only the trip-selection method (i.e., no vessel-selection) to assign observers and EM to vessels in the partial-coverage category for 2018. The partial-coverage category includes vessels greater than or equal to 40 feet (ft) length overall (LOA) and not in the full coverage category. There were six sampling strata in the partial coverage category in 2018:

1. Hook-and-line vessels (*HAL* stratum).
2. Hook-and-line vessels with EM (*EM HAL* stratum).
3. Pot vessels not delivering to tenders (*POT - No Tender* stratum).
4. Pot vessels delivering to tenders (*POT - Tender* stratum).
5. Trawl vessels not delivering to tenders (*TRW - No Tender* stratum).
6. Trawl vessels delivering to tenders (*TRW - Tender* stratum).

The year 2018 was the first year that data collected through EM was used for catch accounting. Data from *EM HAL* vessels comprised the entirety of the implemented EM program used for catch accounting in 2018. Data from EM vessels fishing pot gear was still under pre-implementation in 2018. Vessels had to volunteer to participate in the EM Program by 1 November 2017. The NMFS then selected vessels to be included in the EM Program following an evaluation of available funding (NMFS 2017b). NMFS also sought vessels to participate in EM research and development activities. Vessels that volunteered for the EM Program or EM research activities and were selected by the NMFS were not required to carry observers but were required to continue to log their fishing trips into the Observer Declare and Deploy System (ODDS). In this report, the deployment of EM onto vessels fishing pot gear was not evaluated to the same degree as the deployment of EM onto *EM HAL* vessels that were part of the implemented partial coverage program.

PERFORMANCE REVIEW OBJECTIVES

The following sections contain the OSC review of the deployment of observers in 2018 relative to the intended sampling plan and goals of the 2018 ADP (NMFS 2017b). This report identifies where potential mechanisms for biases exist and provides recommendations for further evaluation, including potential improvements to the observer deployment process that should be considered during the development of the 2020 ADP.

The goal of the Observer Program is to monitor commercial fishing activities in the Alaska EEZ for the purpose of producing bycatch estimates that are used for in-season management and for collecting biological information that is used in stock assessments and ecosystem modeling efforts. It accomplishes this goal through the random deployment of observers and EM systems into fisheries to collect representative data. This analysis focuses on the randomization of observer and EM deployments and how departures from a random sample affect data quality. This analysis focuses on the random deployment of observers and EM onto fishing trips, which is the primary sampling unit in the observer program sampling methodology. This analysis does not evaluate further sampling units, such as biological samples, which are collected by observers but not EM.

The following items from the 2018 ADP are identified as objectives for evaluation in this report:

1. Deploy for the planned number of sea-days. This objective will be considered to be met if the actual number of sea-days expended falls within the range of values from simulated sampling provided in the 2018 ADP. The Observer Program's budget was expected to cover 4,470 days in 2018.
2. Deploy at the coverage rates specified in the 2018 ADP. Following the 2018 ADP, ODDS was programmed to randomly select logged trips at a rate of 20.18% in the *TRW - No Tender* stratum, 17.26% in the *HAL* stratum, 16.21% in the *POT - No Tender* stratum, 16.67% in the *TRW - Tender* stratum, 17.39% in the *POT - Tender* stratum, and 30% in the EM strata. Under a

randomized deployment scheme, actual partial coverage rates are expected to fall 95% of the time within a 95% confidence interval computed from the realized coverage rates (under the assumption of a binomial distribution for observed trips).

3. Collect tissue samples from Chinook and chum salmon as specified in the 2018 Observer Sampling Manual to support the goal of collecting genetic samples from salmon caught as bycatch in groundfish fisheries to identify stock of origin. The sampling protocol established in the 2014 ADP (NMFS 2013) was used in 2018. Under this protocol, observers on vessels delivering to shoreside processors in the GOA trawl pollock fishery monitor the offload to enumerate salmon bycatch and obtain tissues for genetic analysis from the salmon bycatch. For trips that are delivered to tender vessels and trips outside of the pollock fishery, observers obtain salmon counts and tissue samples from all salmon found within at-sea samples of the total catch.
4. Minimize the number of conditional releases from observer coverage issued. The NMFS aimed not to grant conditional releases or temporary exemptions to vessels subject to observer coverage. It was expected that no conditional releases would be granted in 2018.
5. Randomize deployment of observers into the partial coverage category of fishing activities. This randomization is used to collect observer and EM samples that are representative of the entire fishing fleet (observed and monitored trips are equivalent to unobserved and unmonitored trips within a stratum). Evaluation of this objective is focused on the randomization of observer and EM deployments into primary sampling units, and how departures from a random sample affect data quality.

Observer Deployment Performance Metrics

Performance metrics have been developed to assess whether the trip-selection process (through the implementation of the 2018 ADP) provides a representative sample of fishing trips in the North Pacific in 2018. These metrics reflect four mechanisms that can impact the quality of the data: sample frame discrepancies, non-response, differences in trip characteristics, and sample size.

The performance metrics used in this evaluation are as follows:

1. Deployment rates for each stratum: This is the basic level of evaluation for comparing targeted and achieved sampling rates, where sampling strata are partitions of the entire population about which we want to make inferences (e.g., generate estimates of catch). Implementation challenges can be identified in this step, such as sample frame inadequacy, selection biases, and issues with sample unit definitions. Specifically, this section assesses the following:
    a. Sample rates and number of samples relative to intended values.
    b. Quantification of under- and over-coverage rates (sample frame discrepancies). Over-coverage of a population occurs when the sample frame includes elements that are not part of the target population. When these elements are included in the random sample, effort (time, cost) is expended needlessly. Under-coverage results from having a sample frame that does not include a portion of the target population which can lead to biased data if that portion of the population differs from the population included in the sample frame.
    c. Non-response rates. Non-response occurs when randomly selected elements (trips or vessels) are not actually sampled. If these trips or vessels have different fishing behavior (e.g., catch, areas fished) than the rest of the population, the data collected will not represent the entire fleet (non-response bias).
2. Representativeness of the sample: Randomized sampling is a method used to ensure that the results of sampling reflect the underlying population. Departures from randomization can lead to non-representative data and hence potential bias in estimates of the parameters of interest. A randomized sample design is expected to achieve a rate of observed events that is similar across both space and time. Representativeness of the sample was divided into three separate components:
    a. Temporal representativeness

        i. Effort plots: plots of expected and actual observed effort over time. Areas where these two lines deviate from each other are indicative of periods with differential realized sample rates (and potential temporal bias).

    b. Spatial representativeness

        i. Maps: Maps provide a visual depiction of the spatial distribution of observer coverage relative to effort in each partial coverage stratum, as well as where low or high coverage rates occurred.

        ii. Probability of selecting a sample and observing a fewer or greater number of trips within an area than would be expected given the implemented sample rates. These data are used to identify departures from anticipated sampling rates.

    c. Representativeness of trip characteristics

        i. Consistency of trip characteristics for observed and unobserved portions of the stratum. These metrics are based, in part, on the availability of data for both observed and unobserved fishing activities; for example, data that are reported for all trips on landing reports. Attributes tested in this report include the following:

            – Trip duration (days).
            – Vessel length (feet).
            – The number of NMFS Areas visited during the trip.
            – The amount of landed catch (metric tons).
            – The number of species in the landed catch (also known as species richness).
            – The proportion of the total landed catch that was due to the most prevalent species (pMax, an inverse a measure of species diversity where an increase in pMax indicates a decline in diversity).

3. Adequacy of sample size: A well-designed sampling program will have a sample large enough to reasonably ensure that the characteristics of interest in the entire target population are represented in the data. Whether the sample size collected was adequate was determined through an

examination of the probability of deploying observers at the implemented rate and having no observer coverage in one or more cells (e.g., defined by NMFS Reporting Area and strata).

Although these metrics can identify situations where observed results differ from expectations, it is ultimately a subjective decision as to whether or not these differences are substantial enough to have management implications. This holds true for tests that have associated *p*-values. Additionally, our focus on landed catch is due to the fact that total catch is comprised of retained and discarded portions, and since discarded catch is not available from unobserved trips, landed catch represents the only portion of the catch that is available from all trips.

Changes to This Report from Last Year

At their June 2017 meeting, the Council supported allowing 165 vessels to be monitored with EM, with priority given to longline vessels whose data would be used for inseason management (NMFS 2017b, Appendix A). Ultimately, 141 vessels were allowed into EM (NMFS 2017b): 123 fishing predominantly with longline gear and 18 fishing predominantly with pot gear. Of the 123 longline vessels covered by EM, three were considered zero coverage because they were participants in experimental EM research as opposed to regulated EM methodology.

The strata used in 2018 were largely the same as those used in 2017, with two exceptions: the 2017 *HAL – Tender* and *HAL – No Tender* strata were combined into one *HAL* stratum in 2018 and the *EM HAL* stratum was also added to the regulated partial coverage monitoring program in 2018. The deployment performance of EM was therefore split by gear type, with the performance of *EM HAL* being analyzed to the same degree as observed strata. The performance of the *EM POT - No Tender* and *EM POT - Tender* strata are analyzed to a lesser degree due to their pre-implementation status.
As in 2017, trip-selection was the sole method used for selecting fishing activity for observation or monitoring in 2018. This is in contrast to prior years when vessel-selection was used in conjunction with

trip-selection. No minimum coverage hurdle was employed in 2017, whereas a 15% minimum coverage hurdle was used in 2018.

EVALUATION OF DEPLOYMENT IN 2018

The deployment of observers into the 2018 Federal fisheries in Alaska is evaluated at the level of the deployment stratum because each stratum is defined by a different sampling rate or by a different monitoring method (e.g., observers and EM). In this document, trips in the *EM HAL* stratum are considered successfully monitored if at least some video was reviewed from a trip. The rationale for defining monitored trips this way is that it is most similar to the way in which trips in other strata are considered observed (i.e., irrespective of whether or not haul information or usable species composition data were collected). Since *EM HAL* data was the only EM data used for catch accounting in 2018, EM HAL data are the only EM data analyzed to the same degree as observed strata. For *EM POT - No Tender* and *EM POT - Tender*, which were under pre-implementation, two values are reported: the number of trips for which data were received and the number of trips for which data were reviewed. The goal the NMFS set for *EM POT - No Tender* and *EM POT - Tender* in their pre-implementation status was only to *receive* data for trips at the programmed selection rate, not to *review* data for trips at the programmed selection rate.

Evaluating Effort Predictions

Each year the NMFS sets an annual budget in terms of observer days. Therefore, how close anticipated observed effort is to actual invoiced effort in each ADP is a function of how well the NMFS predicts effort and how well the NMFS achieves its sampling rate. The observer day budget for 2018 was set at 4,470 days in the 2018 ADP (NMFS 2017b). Based on simulations using 2017 fishing data conducted a year in advance of deployment for the 2018 ADP, the FMA predicted it would observe 4,394 fishing days at the end of 2018. In 2018, the FMA paid for 3,206 observer days, which was 27% lower than predicted (Fig. 1). This can partially be explained by the fact that the stratum-specific effort predicted in the 2018 ADP (NMFS 2017b) was higher than actual effort by 59.4% in the *TRW - No Tender* stratum and 51.6% in the *TRW - Tender* stratum, and lower than actual effort by 89.1% in the *POT - No Tender* stratum and 61.5% in the *POT - Tender* stratum (Table 1). The over-prediction of effort in the large *TRW – No Tender* stratum contributed greatly to the over-prediction of total effort.

PERFORMANCE OF THE OBSERVER DECLARE AND DEPLOY SYSTEM IN TRIP-SELECTION

The random selection of trips is made by the ODDS. The ODDS generates a random number according to the pre-determined rates and assigns each logged trip to either "selected to be observed" (selected) or "not selected to be observed" (not selected) categories. The NMFS observer provider has access to all selected trip information necessary to schedule observer logistics. Up to three trips may be logged in advance of fishing to provide industry users with flexibility to accommodate their fishing operations.

Logged trips have different dispositions. When initially logged, they are considered pending and can be either closed or cancelled. Whether these changes can be made by the user (person logging the trip) or must be made by the observer provider (or the NMFS) depends on whether or not the trip is selected to be observed, the stratum the trip belongs to, and the timing of the activity. Trips can be closed (marked as complete) by the ODDS user after the planned trip departure date by either entering the dates of the trip and the port processor of the landing, or by selecting from a list of pre-populated landing reports. For partial coverage strata subject to observation, the observer provider is given 72 hours for an observer to board the vessel prior to the trip start. While a trip may be entered into ODDS that is scheduled to start earlier than 72 hours from the time of entry, if selected for observer coverage, the observer provider can opt to delay the start of the trip up to, but not exceeding 72 hours from the time of trip entry. This helps protect the observer provider from the high cost of deploying an observer with short notice. The vessel operator is protected as well by guaranteeing the assigned observer to the vessel up to 48 hours past the planned start of the fishing trip. This rule helps ensure that an observer is available to the boat in case of unforeseen events such as weather. If, however, the trip start date and time has passed by more than 48 hours, then the observer provider can cancel the trip and release the observer from the vessel and trip, and the vessel would need to log a new trip with a new 72-hour notice in place prior to fishing. These 'forced cancellations' are not present in trips that are not selected for observation since the logging, closing, or cancellation of the trip is entirely under vessel control. The vessel operator may change the dates of a logged trip regardless of selection status prior to, or in lieu of cancellation. However, trips that have not been closed at the end of the calendar year are automatically cancelled by the ODDS to prevent 2018 ODDS trips from affecting the deployment rates set for the 2019 ADP.

The number of trips logged in the ODDS in 2018 and their dispositions are summarized in Tables 2-4. The forced cancellation rate by users and by the ODDS is summarized for selected trips in each stratum (Table 2). Of the 5,734 total trips logged, 1,125 were selected and 179 were cancelled: 2 by ODDS (0.18%) and 177 by users (15.7%). The user cancellation rate for selected trips ranged from 3.8% for *EM HAL* to 37.5% for *TRW - Tender*.

The flexibility offered by the ODDS means that the outcome of random selection is known to the vessel operator for up to three logged trips in advance of fishing. In the case where ODDS users disproportionately cancel selected trips, one would expect observed coverage to be lower than the programmed selection rates. To reduce this potential bias the ODDS is programmed to automatically select the vessel's next logged trip if a previously selected trip was cancelled by the user. Although these "inherited" trips preserve the *number* of selected trips in the year, they cannot prevent the *delay* of selected trips during the year. Therefore, the potential for temporal bias is still present. The percentages of selected trips from either inherits or waivers are found in Table 3. The relative percentage of selected trips that inherited their final selected-status due to a previous cancellation ranged from 5.1% for *EM HAL* to 44.4% for *TRW - Tender* (Table 3). It should be noted that no trips (inherited or otherwise) were selected in the *HAL- Tender* stratum. The number of waived trips (i.e., trips given a “pass” on their required observer coverage by the NMFS) was low, with the highest level occurring in the *HAL* stratum at 2.2% (Table 3).

The extent to which trip-selections were changed from the time they were entered can be determined by comparing the rate of trip observation expected from 1) random selection of all logged trips (initial random selection) and 2) random selection of remaining trips after cancellations, waivers, and inherited trips. In any case, the proportion of trips selected to be observed should fall within what would be expected given the binomial distribution (since each trip was either selected or not selected). The rates obtained (%, with associated *p*-value based on the binomial distribution) in the initial selection process were 16.32% (*p*-value = 0.256) for the *HAL* stratum, 30.15% (p-value = 0.938) for the *EM HAL* stratum, 16.38% (*p*-value = 0.918) for the *POT - No Tender* stratum, 22.45% (*p*-value = 0.346) for the *POT - Tender* stratum, 20.07% (*p*-value = 0.933) for the *TRW - No Tender* stratum, and 22.54% (*p*-value = 0.201) for the *TRW - Tender* stratum (Table 4). This means that there is no evidence that the ODDS was not selecting trips according to the programmed rate. The final selection rate after trips were closed, cancelled, or waived was 17.67% (*p*-value = 0.656) for the *HAL* stratum, 34.20% (*p*-value = 0.018) for the *EM HAL* stratum, 16.45% (*p*-value = 0.869) for the *POT - No Tender* stratum, 32.43% (*p*-value = 0.027) for the *POT - Tender* stratum, 20.88% (*p*-value = 0.462) for the *TRW - No Tender* stratum, and 36.73% (*p*-value = 0.001) for the *TRW - Tender* stratum (Table 4).

Evidence for differences between the initial and final selection rates were present in three of six partial coverage strata in 2018. In 2017, evidence for differences were present in five of six strata (AFSC and AKRO 2018). Once a separation between initial and final selection rates appears in a stratum, it tends to persist throughout the remainder of the year (Fig. 2).

The fact that the final selection rates for most strata were greater than the initial selection rates results from the fact that cancelled trips that were originally selected for coverage were preserved through the inherit process, while cancelled trips that were not originally selected for coverage were not. These patterns are consistent with the hypothesis that trips selected for coverage are being delayed, and cancellation of selected trips resulted in a greater number of selected trips later in the year as the result of the inherit process. Various degrees of separation between the initial and final selection rates have been observed since the implementation of the restructured Observer Program (NMFS 2014, NMFS 2015a, NMFS 2016, NMFS 2017a, AFSC and AKRO 2018).

In addition to the inherit process, the lack of a linkage between the ODDS and *e*Landings contributes to the differences between programmed selection rates in ODDS and trips that were ultimately observed. Currently, ODDS provides users with a list of Report IDs from *e*Landings from which to close their logged trips, and *e*Landings has been updated to prompt the entry of ODDS trip numbers. However, these data were not validated or error checked, making them unreliable in their current state. A linkage between the logged (ODDS) trip (with its selection probability) and associated landing information would be necessary to evaluate potential improvements in deployment efficiency within the partial coverage fleet.

Evaluation of Deployment Rates

This section compares the coverage rate achieved against the expected coverage rates. Data used in this evaluation are stored within the Catch Accounting System (CAS, managed by the AKRO), the Observer Program database NORPAC (managed by the AFSC), and *e*Landings (under joint management by Alaska Department of Fish and Game—ADF&G; the International Pacific Halibut Commission—IPHC; and the NMFS). Separate rate evaluations were conducted depending on whether the unit of observer deployment was at-sea fishing trips or dockside deliveries of pollock.

At-sea Deployments

The 2018 Observer Program had 11 different deployment strata to be evaluated. There was one full coverage stratum; it included trips taken both by vessels that were required to have full coverage (e.g., AFA vessels) and those fishing in the BSAI that opted into full coverage. There were six partial coverage strata: five observed strata defined by gear and tender designation and one regulated EM stratum for hook-and-line vessels. There were two pre-implementation EM strata for pot vessels. There were also two zero coverage strata: one zero coverage EM research stratum and one zero coverage stratum for jig vessels and vessels under 40 ft. length overall.

Evaluations for the full coverage category and zero-selection pool are straightforward - either the coverage achieved was equal to 100% or 0%, respectively, or it was not. The program met expected rates of coverage in all full and zero coverage strata (Table 5). Partial coverage rates are expected to fall 95% percent of the time within a 95% confidence interval computed from the realized coverage rates (under the assumption of a binomial distribution for observed trips). If expected coverage levels were within the 95% confidence intervals, then we conclude that there was no evidence that coverage levels differed from the expected rates. Coverage rates were consistent with expected values in three of the six partial coverage strata but were lower than expected within the *HAL* and *EM HAL* strata and higher than expected in the *TRW - Tender* stratum (Table 5). The coverage rate for EM is based on information provided from the Pacific States Marine Fisheries Commission (PSMFC) that is available to analysts in the AFSC database. In 2018, there was considerable lag-time between the receipt and review of some EM HAL data, with the mean time between receipt and review being 60 days. This is compared to an average of 8 days during pre-implementation in 2016 (NMFS 2017c, Table 3-9). By the end of 2018, the PSMFC had not reviewed 62 selected *EM HAL* trips (Table 6). The PSMFC had received data for 53 of those selected trips. The number of unreviewed *EM HAL* trips differed by month, with the highest numbers of unreviewed trips appearing in September and October (Table 6). None of the 41 selected EM POT trips were reviewed (Table 6), as the goal of EM POT in its pre-implementation phase was to establish a framework for receiving video, not to review video received. In total 4,381 trips (41.2%) and 484 vessels (44.6%) were observed or monitored among all fishing in Federal fisheries of Alaska in 2018 (Table 5). In 2017, 4,220 trips (36.4%) and 407 vessels (36.4%) were observed (AFSC and AKRO 2018, Table 5).

Coverage Rates for Dockside Monitoring

Observers were assigned to monitor deliveries of walleye pollock (*Gadus chalcogrammus*). The objective of this monitoring was to obtain a count of the number of salmon caught as bycatch and to obtain tissue samples for genetic analysis from these fish in each observed pollock delivery. There have been many iterations of the sampling design used to obtain genetic samples from salmon bycatch for the purposes of determining stock of origin (Faunce 2015). The sampling design used for this objective in 2018 remained unchanged from that used since 2011; all deliveries of walleye pollock that are observed at sea were also observed dockside. While all Bering Sea pollock trips and deliveries are observed, this is not the case in the GOA (NMFS 2015b). For this analysis, pollock deliveries were defined as any delivery where the predominant species was pollock in *e*Landings.

Given the design, the level of dockside observation of walleye pollock deliveries should be 100% in the full coverage category. In 2018, 100% of full coverage walleye pollock deliveries were observed in every port (Table 7). This differs from results that were presented in the 2018 Annual Report (AFSC and AKRO 2019), which stated that three pollock deliveries to the port of Dutch Harbor went unobserved. The data for those deliveries were located after the publication of the 2018 Annual Report.

While expectations of the full coverage category are straightforward, evaluations of the partial coverage category are more complex. As a matter of policy, no tender deliveries are observed. While it may seem intuitive that the expected coverage rate for deliveries within the *TRW - No Tender* stratum should be equal to the programmed trip selection rate of 20.18%, this assumption is likely untrue because observers are not deployed into the pollock fishery but into the entire trawl fishery, and the relationship between the number of deliveries and trips is not expected to be constant, especially when measured across ports. Therefore, we present the dockside observation rates for the *TRW - No Tender* stratum (Table 7) but do not include any formal statistical tests.

Bycatch estimates of Chinook salmon in the GOA are estimated using methods described in Cahalan et al. (2015). In the event that a delivery cannot be monitored (e.g., the case in a tendered delivery or non-pollock delivery), then estimation of bycatch comes by applying salmon bycatch rates to landed catch. Estimates of stock of origin from salmon bycatch are produced by the AFSC's Auke Bay Laboratories (e.g., Guthrie et al. 2017).

SAMPLE QUALITY

Temporal Patterns in Trip-Selection

The cumulative number of fishing trips in each stratum was multiplied by the stratum-specific selection rate to obtain the expected number of observed trips. Under the assumption that there was no temporal bias in observer coverage, 2.5% of values should be larger than the upper 95% confidence limit and 2.5% should be smaller than the lower limit. At the end of 2018, the number of observed trips was outside of this expected range in three of the six partial coverage strata: *EM HAL* (expected rate = 0.300, realized rate = 0.220, *p*-value = 0.000), *HAL* (expected rate = 0.173, realized rate = 0.155, *p*-value = 0.041), and *TRW – Tender* (expected rate = 0.167, realized rate = 0.350, *p*-value = 0.005; Fig. 3). Coverage rates were outside of the expected range for 25.8%, 23.8%, and 33.2% of the year for the *EM HAL*, *HAL*, and *TRW - Tender* strata, respectively. Coverage rates were also outside of the expected range for 0.5% of the year for the *POT – Tender* stratum, although this stratum ended the year within its expected range of coverage (expected rate = 0.174, realized rate = 0.290, *p*-value = 0.097). Coverage rates were within their expected ranges for 100% of the year for the *POT – No Tender* (expected rate = 0.162, realized rate = 0.155, *p*-value = 0.665) and *TRW – No Tender* (expected rate = 0.202, realized rate = 0.203, *p*-value = 0.908) strata. Overall, there appeared to be less temporal bias in 2018 than in 2017, when three of six partial coverage strata had coverage rates outside of the expected range for the majority of the year, including the *POT - No Tender* stratum, which had a higher than expected coverage rate for 100% of the year (AFSC and AKRO 2018).

Spatial Patterns in Trip-Selection

Under a strictly random selection of trips and with a large enough sample size, the spatial distribution of observed trips should reflect the spatial distribution of all trips. The hypergeometric distribution can be used to describe the results of sampling from a population of items (fishing trips) with different characteristics (NMFS Area fished). The expected number of trips based on this distribution is the sample rate multiplied by the number of trips that fished in an area (observed and unobserved). Using this method, we compared the expected number of trips and the observed number of trips in each NMFS Reporting Area and stratum combination (Fig. 4). Note that in most cases, the sampling result was close to the expected result; larger differences tend to be associated with lower numbers of trips within a NMFS Area.

Using landings data, we calculated the probability of observing at least the number of trips that we did observe within a stratum and NMFS Area. This calculation uses the sampling rate and the distribution of trips across NMFS Reporting Areas. This evaluation does not test whether the resulting coverage rate in a NMFS Area for a stratum was equal to the stratum selection rate, but instead tests whether the actual coverage rate (realized rate) in a NMFS Area for a stratum was unexpected compared to the stratum-wide realized observation rate. For the purposes of the following discussion, NMFS Areas with an unexpected number of trips (probability of our result is less than 0.05) were considered "low-*p*" areas.

The HAL stratum

Given that there were 19 NMFS Areas fished in *HAL*, we would expect there to be $0.05 \times 19 = 1$ low-*p* area for this stratum. There were three. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 25.5% (median = 8.5%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 5. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *HAL* stratum.

The EM HAL stratum

Given that there were 14 NMFS Areas fished in *EM HAL*, we would expect there to be 0.05 ×14 = 1 low-*p* area for this stratum. There was one. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 50% (median = 23.3%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 6. These results mean that there was no clustering of observed trips among NMFS Areas that was different from expected. No spatial bias appears to have occurred in the *EM HAL* stratum.

The POT - No Tender stratum

Given that there were 14 NMFS Areas fished in *POT - No Tender*, we would expect there to be 0.05 ×14 = 1 low-p area for this stratum. There was one. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 55.6% (median = 15.8%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 7. These results mean that there was no clustering of observed trips among NMFS Areas that was different from expected. No spatial bias appears to have occurred in the *POT - No Tender* stratum.

The TRW - No Tender stratum

Given that there were nine NMFS Areas fished in *TRW - No Tender*, we would expect there to be 0.05 × 9 = 0 low-p areas for this stratum. There were two. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 57.9% (median = 19.9%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 8. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *TRW - No Tender* stratum.

The POT - Tender stratum

Given that there were 6 NMFS Areas fished in *POT - Tender*, we would expect there to be 0.05 × 6 = 0 low-p areas for this stratum. There was one. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 66.7% (median = 34.3%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 9. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *POT - Tender* stratum.

The TRW - Tender stratum

Given that there were four NMFS Areas fished in *TRW - Tender*, we would expect there to be 0.05 × 4 = 0 low-p areas for this stratum. There were two. The percent of trips observed among NMFS Areas in this stratum ranged from 0% to 100% (median = 18.4%). The probability of these coverage rates or rates that deviated further from expected values is depicted in Figure 10. These results mean that there was some clustering of observed trips among NMFS Areas that was different from expected. Some spatial bias appears to have occurred in the *TRW - Tender* stratum.

Trip Metrics

This section is focused on answering one question related to the deployment of observers: were observed trips similar to unobserved trips? A permutation test (a.k.a., randomization test) was used to answer this question. This test evaluates the question "How likely is the difference we found if these two groups had the same distribution (in the metric we are comparing)?" Permutation tests compare the actual difference found between two groups to the distribution of many differences derived by randomizing the labels defining the two groups (e.g., observed and unobserved). Difference values in the permutation test were calculated by subtracting the mean metric value for the "No" condition from the mean metric value for the "Yes" condition. For example, the difference between vessel lengths in a permutation test for an observer effect would be the mean value (i.e., mean vessel length) for unobserved trips subtracted from the mean value for all observed trips, and hence a positive value would indicate that observed trips occurred on larger vessels than unobserved trips. By randomizing group assignments, the distribution of randomized differences represents the sampling distribution under the null hypothesis that the two groups are equal. In this report 1,000 randomized trials are run for each permutation test. The *p*-value from the test is calculated as the number of randomized trials with greater absolute differences than the actual difference divided by the number of randomized trials. Similar to the other statistical tests used in this report, low *p*-values (< 0.05) indicate unlikely events under the hypothesis of equality and are therefore considered evidence against that hypothesis. In an attempt to improve clarity, although five values are calculated in the test; 1) the difference between groups, 2) the mean difference between groups from randomized trials, 3) #1 expressed as a percentage of the mean value of the metric being tested, 4) #2 expressed as a percentage of the mean value of the metric being tested, and 5) the *p*-value of the test, only values one, three and five are presented.

Six trip metrics were examined in the permutation test. These metrics were as follows: the number of NMFS Areas visited in a trip, trip duration (days), the weight of the landed catch (t), the vessel length (ft), the number of species in the landed catch, and the proportion (0 to the most predominant species [pMax]). The metric vessel length is used to help interpret the results from landed weight of catch since fishing power is positively correlated to vessel length. Specifically, differences in weight *and* length were interpreted as a failure to achieve a random sample of vessels of different sizes, whereas differences in weight only lends more evidence that there was an observer effect. The number of species within the landed portion of the catch is a measure of species richness. Our pMax metric follows the concepts behind Hill's diversity number N1 that depicts the number of abundant species (Hill 1973) and is a measure of how "pure" catch is since a value of one would indicate that only the predominant (and presumed desirable) species was landed.

Are observed trips similar to unobserved trips?

This comparison is the basis for examining if there was an observer effect (i.e., differential behavior when observed compared to when not observed) within partial coverage trips. Sample sizes and results for this test are presented in Table 8.

Of the six metrics compared in the *EM HAL* stratum, one had a low *p*-value. Observed trips in this stratum landed 9.7% (0.39) more species than unobserved trips (Table 9). Of the six metrics compared in the *HAL* stratum, two had low *p*-values. Observed trips in this stratum were 14.3% (0.76 days) shorter in duration and landed catch that weighed 15.6% (1.05 metric tons) less than unobserved trips (Table 9). Of the six metrics compared in the *POT - No Tender* stratum, one had a low *p*-value. Observed trips in this stratum landed 14.3% (0.29) more species than unobserved trips (Table 9). Of the six metrics compared in the *POT - Tender* stratum, there were no metrics with low *p*-values (Table 9). Of the six metrics compared in the *TRW - No Tender* stratum, two had low *p*-values. Observed trips in this stratum occurred in 3% (0.03) fewer areas and were 9.4% (0.26 days) shorter in duration than unobserved trips (Table 9). Of the six metrics compared in the *TRW - Tender* stratum, there were no metrics with low *p*-values (Table 9). A visual depiction of individual results of this permutation test for the *HAL*, *POT - No Tender*, and *TRW - No Tender* strata is given in Figure 11 for illustration purposes.

Based on these results, there was some evidence for differences between observed and unobserved trips for *EM HAL*, *HAL*, *POT – No Tender* and *TRW – No Tender* strata, implying that observed trips were not entirely representative of all trips taken for those strata. With 36 statistical tests in total, we would expect $0.05 \times 36 = 1.8$ tests to have *p*-values below 0.05 by chance (assuming independent tests) when there were no differences between observed and unobserved trips – we observed 6 such results. Note that while some of the observed differences in the two POT strata seem quite large, sample sizes were very small (Table 8), making it difficult to detect differences statistically.

Gear, tender, and observed status combinations

Since 2013 permutation tests have been used to analyze the differences between observed trips and unobserved trips within strata. One of the analyses done by the permutation test is to compare trip lengths (in days) between observed and unobserved trips, and determine whether there were significant differences. However, these permutation tests do not visually map the data for observed and tendered states together. To accomplish this, a plot of the trip durations for these states is included as Figure 12. These plots visually display data that complements the results of the permutation test, showing that trip duration differs with observed status most notably in the *HAL* and *TRW – No Tender* strata, in which observed trips tended to be shorter.

ADEQUACY OF THE SAMPLE SIZE

In a well-designed sampling program, the observer coverage rate should be large enough to reasonably ensure that the range of fishing activities and characteristics are represented in the sample data. The Catch Accounting System post-stratifies data into groups of fishing activities with similar trip characteristics such as gear, trip targets, and NMFS Area (Cahalan et al. 2014). At low numbers of trips and low sampling rates, the probability of no observer data within a particular post-stratum is increased and may result in expansions of bycatch rates from one type of fishing activity against landings for a different type of fishing activity. This will result in biased estimates of bycatch. For this reason, it is important to have a large enough sample (observed trips and vessels) to have reasonable expectation of observing all types of fishing.

Over the course of an entire year, some NMFS Areas have low fishing effort and as a result have a relatively high probability of being missed by the simple random sampling represented by observer deployments. The fishing effort data for each stratum and the number of observed trips over the course of 2018 was used to illustrate their combined effect on the probability of a NMFS Area containing observer data using the hypergeometric distribution (Fig. 13). From this figure it can be seen how the likelihood of at least one observation was increased with fishing effort and was also increased with an increase in the selection rate. Given our sampling rates in the 6 partial coverage trip-selection strata, the probability of having no observed trips in a NMFS Reporting Areas increases quickly above 0.05 when there are fewer than 11 trips in the *EM HAL* stratum, 18 trips in the *HAL* stratum, 17 trips in the *POT - No Tender* stratum, 7 trips in the *POT - Tender* stratum, 13 trips in the *TRW - No Tender* stratum, and 6 trips in the *TRW - Tender* stratum in a given area. Including additional factors such as week, gear, and target will decrease the number of trips with the same characteristics and hence increase the probabilities of obtaining no observer data of that character (post-strata of the CAS).

RESPONSE TO COUNCIL AND SSC COMMENTS

The SSC has requested that a specific section with responses to SSC comments be provided in the written report, as is done for SAFE documents. This section addresses comments (in italics) made by the Council and the SSC in response to the presentation of the 2017 Annual Report made at the June 2018 Council meeting.

The Council offered the following comments:

Include an evaluation of observer effects at finer resolution than gear-level strata, so that observer effects in pelagic and non-pelagic trawl can be investigated.

This analysis has been conducted using multiple years of data and constitutes the Appendix. We have included results from multiple years to test the stability of results between years.

Continue to provide details on EM in Chapter 4 and also include information in the report about the number of EM trips selected, the number monitored, and the number reviewed, for clarification.

While the Council specifically mentions Chapter 4, most of the requested data are summarized in Chapter 3. These include the number of EM trips taken, the number of EM trips selected, the number of

EM trips that had an associated hard drive received, and the number of EM trips reviewed. These data are provided for both the regulated EM HAL stratum and the EM POT strata, which were under pre-implementation in 2018.

Add an appendix that describes details of cost calculations for EM and observer days over time.

This was not addressed in Chapter 3 of the 2018 Annual Report (AFSC and AKRO 2019).

The Council also recommends that NMFS communicate with the OAC on the results from the proposed ODDS agency subgroup.

Our response to this is addressed in SSC comments below.

The SSC offered the following recommendations to the Observer Program:

(June) The behavior of the ODDS system with respect to inheritance of trip selection after a trip is cancelled leads to temporally biased sampling of some strata, with many, or most, observed trips coming very late in the season. ***The SSC concurs with the NMFS recommendation that a sub-group be created to evaluate system behavior and identify ways to obtain broader, more representative observer coverage throughout the season.***

(Oct.) The SSC notes that when a trip is designated to be observed via ODDS, but is then canceled, the subsequent trip for that vessel inherits the "observed" status. This systematically shifts sampling effort later in the season, generating temporal bias. Recognizing this, a Trip Inheritance Group has been formed by NMFS. The SSC looks forward to the Trip Inheritance Group's recommendations to resolve this issue.

Staff working on ODDS have first been tasked with documenting the architecture and decision logic of the existing system. Progress on this item can be tracked in the document 'Analytical Tasks Related to the Observer Program', with the task entitled 'Agency ODDS Subgroup'. Once this task is complete, staff will focus on amending (if necessary) trip logging and cancellation rules.

(June) We reiterate that, while the SSC recognizes that development of variances for use in planning of deployments and stock assessment is ongoing, ***we urge the analysts to initiate a comparison of the likely magnitude of bias that has been detected between observed and unobserved trips with the overall magnitude and precision of discard or PSC that is being monitored for compliance by management.*** *The analysts note in the report that further clarification and conversation with the SSC is needed and we look forward to this exchange.*

The differences between observed and unobserved measures of retained catch, NMFS Areas, etc. in Chapter 3 likely do not directly translate to bias in PSC estimates. While it remains unclear to the OSC how such an analysis would be conducted, we note that this task should be included under item #6 for the 'Analytical Tasks Related to the Observer Program' entitled 'Observer Program Performance Metrics'. Staff have yet to be assigned to this issue.

(June) While the SSC supports the NMFS recommendation to use the same observer trip selection strata in 2019 as in 2018, in cases where there are multiple gear types in a stratum (e.g., pelagic and non-pelagic trawls) the SSC recommends analysis of the results by gear type separately in addition to analysis aggregated to the stratum level. Such disaggregation will avoid masking of gear-specific differences in catch composition and other factors that could provide justification for possible further subdivision of strata.

This analysis has been conducted using multiple years of data and constitutes the Appendix. Following Council direction, we have limited our comparisons to pelagic and non-pelagic trawl. We have included results from multiple years to test the stability of results between years.

(June) The SSC remains concerned that performance metrics from EM pre-implementation have not been fully evaluated prior to full implementation of EM in the observer program. We look forward to seeing a full evaluation of this program as soon as is practical, as well as an evaluation of the tradeoffs between use of EM and the existing partially observed coverage category. As the Council considers continued growth of the EM program, it will be important to conduct appropriate cost comparisons,

specifically including video review costs, as well as an evaluation of the ability of EM versus onboard observer data to meet program needs.

The Alaska Fisheries Science Center has embarked on a project to evaluate how loss of biological data resulting from EM expansion has, and will affect its data products including stock assessments. This initial 'scoping' document is planned to be completed in time for the October Council meeting, conditional on the fact that FMA staff are not needed for the draft ADP compilation. Costs and design of the EM program are not planned to be addressed since we feel they are best addressed in the existing ADP process in years when EM is funded from fee revenue.

(June) Compliance and enforcement issues remain a problem within the observer program that are contributing to bias, fluctuate substantially among years, and may be substantially underreported for a variety of social and safety reasons. ***The SSC encourages the training of crew fleet-wide on the necessity of the observing program to proper fishery management and how crew can contribute to the success of the program by interacting appropriately with observers.*** *It is critical that these issues be addressed immediately.*

(Oct.) Another item in the June 2018 SSC minutes that has not yet been addressed was fleet-wide training of crew on the necessity of the observer program. The SSC reiterates that compliance and enforcement issues remain a problem within the observer program that are contributing to bias, fluctuate substantially among years, and may be substantially underreported for a variety of social and safety reasons. This was highlighted as a critical need to be addressed immediately.

We agree with the SSC recommendations above and add that the 2018 Annual Report includes a new appendix in which the FMA has conducted an exploratory analysis of observer statements conditioned by the amount of observer and vessel combinations or the amount of time observers were on a vessel (Appendix D). This standardization of statement summaries into rates is the first of its kind for observer statement data.

OSC Recommendations to Improve Data Quality

Recommendations from the 2017 Annual Deployment Review

The Observer Science Committee made the following recommendations in its 2017 review of observer deployment to be considered in developing the 2019 ADP (NMFS 2018b). Following each italicized recommendation is the outcome of that recommendation.

The Observer Science Committee's Recommendations to improve the 2019 ADP were as follows:

1. *The OSC has three recommendations regarding the ODDS, its relationship to eLandings, and the effect of cancellations on achieved coverage:*
    a. *The OSC reiterates its 4-year recommendation that the NMFS improve the linkages between ODDS and eLandings (OSC recommendation for 2013, 2014, 2015, 2016 version of this Review).*

       While the ability to review *e*Landings associated with a vessel in ODDS is already facilitated, and entry of an ODDS number associated with a landing is already facilitated in *e*Landings, decisions need to be made as to which model would be best to move forward with. Without direction on which option to proceed with, no progress has been made on this issue. Outreach and education of NMFS decision with industry would be necessary before implementation of either option.
    b. *OSC reiterates its 3-year recommendation that the NMFS explore ways to reduce the impact of cancellations on the number of trips selected for observer coverage in the ODDS (OSC recommendation from the 2014, 2015, and 2016 version of this Review). This may be accomplished in a variety of ways that include, but are not limited to the following: reducing the number of trips that can be logged in advance, and/or reducing*

the incentive or ability to cancel trips selected for observer coverage or electronic monitoring, since the ability to change dates is already facilitated.

Most of these proposed solutions are already enabled in ODDS but are only applied in special circumstances. The ability to change dates has been facilitated in ODDS since its inception. For unselected trips, the operator needs only to change the dates in ODDS. Although special rules apply for trips logged within 72 hours of departure, the ability of an operator to change dates for selected trips is still facilitated but must be made through coordination with the observer provider. New restrictions on the number of trips a vessel operator can log into ODDS was implemented with the addition of EM vessels to ODDS in 2018, although they only apply to vessels that have special VMP or Appeal status changes. Should FMA desire to change the business rules to ODDS for the upcoming year, they could be enacted quickly by FMA technical staff. However, outreach and education of NMFS decision with industry would be necessary before implementation.

c. *This is the first year in which the OSC recommends that NMFS form an agency sub-group to document the way in which the ODDS currently operates and to describe alternatives for how it can be improved, particularly in regards to points a and b and whether technical improvements to ODDs could address these issues.*

In 2019 funds were received and a person hired to conduct this project. At the time of this writing three of six ODDS modules have been documented, and a list of recommended improvements is ongoing. This project is scheduled to be completed by the end of the calendar year.

2. *The OSC has two recommendations concerning stratification:*

   a. *The OSC recommends that the strata be kept the same between the 2018 and 2019 ADPs. These strata are as they were in 2017, with the exception of combining the HAL - No Tender and HAL - Tender strata into one HAL stratum. The OSC makes this recommendation both to preserve stability in methods across years, and because further stratification would likely decrease sample size within some strata to undesirably small sizes, as was seen with the HAL - Tender stratum in 2017.*

This recommendation was implemented in the final 2019 ADP.

b. *The OSC provided evaluation of the Council's request to explore differences between NPT and PTR gear. Based on this evaluation, which considers factors pertinent to stratification, the OSC recommends against stratifying trawl trips by pelagic and non-pelagic gear types. The supporting analysis for this recommendation can be found in Appendix A [of the 2017 Annual Report; AFSC and AKRO 2018].*

Trawl gear was not split by NPT or PTR gear in the 2019 ADP.

3. *The OSC has two recommendations concerning future at-sea coverage rates for observers (and potentially monitoring):*

a. *We reiterate our recommendation from last year that sampling rates in future ADPs be high enough in each stratum to maximize the probability of achieving three observed trips in each of the NMFS Areas.*

The 2019 ADP established a base-coverage rate of 15% above which available sea-days are allocated according to a blended optimization routine focused on PSC fishes (Council intent). While the choice of this base coverage rate among all strata has been debated by the FMAC, numerous analyses performed by members of this group for the Council and NMFS have demonstrated that the likelihood of observing at least one or three trips within a stratum and area domain are increased dramatically up to 15% coverage above which the magnitude of additional gains are reduced.

b. *The OSC recommends that future ADPs include, as one option, a sample design in which strata are selected at the same rate. Although this design could be considered a baseline used for making comparisons to other proposed designs, under some scenarios, this option may be recommended.*

This design was included in the draft 2019 ADP.

4. *The OSC recommends that the performance standards used to evaluate observer effects in the Annual Report be reassessed by the OSC. The performance standards were developed in 2013 with the restructuring of the Observer Program and have yet to be reviewed. The original*

purpose of this set of indicators was to evaluate the differences between the unobserved and observed population of trips using available information for the two groups; information that can be directly measured in both groups (e.g., total weight of landed catch). These metrics have been useful for evaluating whether the deployment of observers into the sampling strata has resulted in a representative sample of trips. However, an evaluation has not been conducted that relates these metrics to at-sea information. Additionally, the magnitude of the differences (the effect size) has not been evaluated relative to whether differences seen between the two groups are meaningful in the context of the overall data. We recommend evaluating the suite of metrics in context with how they relate to at-sea data collections and, to the extent feasible, provide additional information regarding interpretation of effect sizes and p-values (e.g., consideration of sample sizes).

In recognition of the fact that the current methods have the drawback of multiple comparisons, we continue to focus on family-wide error rates when interpreting *p*-values. However, we acknowledge that in the public arena any one 'significant' test result can be misinterpreted. OSC members met during March and April 2019 to discuss this issue, but were unable to come to a conclusion about how best to proceed. No change to the performance metrics were made for the 2018 Annual Report. Model-based approaches are being considered as an alternative because of the ability to focus on a single factor (observer effects) while controlling for other effects (strata, vessel size, gear). 'Plans to explore alternative approaches to evaluating observer effects' should be added to the list of analytical priorities related to the observer program maintained for the Council by the Alaska Regional Office.

Recommendations to Improve Data Quality and Guide the 2020 ADP

The OSC continues to be concerned about the consistently high cancellation rates for selected trips in tendering stratum, and notes that most issues reported to OLE by FMA related to ODDS now focus on tendering strata definitions.

We recommend that the ODDS trip logging and cancellation rules be re-evaluated and communicated to the Council and industry as soon as possible.

We recommend that draft 2020 ADP stratification designs include a re-examination of tendering strata. This may be accomplished in a variety of ways not limited to eliminating tendering stratum altogether, or holding selection rates the same between tendered and non-tendered stratum within a gear type.

There has been continued interest by the Council in the evaluation of observer effects between pelagic (PTR) and non-pelagic (NPT) types of trawl gear. In response to the Council's June 2017 request, a detailed analysis of the issue was provided in the 2017 Annual Report (AFSC and AKRO 2018, Appendix A). Despite this analysis, the Council again asked for another analysis of this issue in their June 2018 review of the Annual Report. We have included the raw results of permutation tests in response to the Council's continued request. We continue to hold concerns with this analysis and defer to our conclusion contained within Appendix A of the 2017 Annual Report: **the OSC do not recommend stratification by type of trawl gear (i.e., NPT and PTR strata).** Briefly, this analysis is complicated by the fact that trips occur with both sub-gear types and results differ between years. This exposes the analyses to the Simpsons' paradox, wherein the results of numerous tests of different aggregations contradict each other. While the number of low probability tests exceeded the number of tests expected by a greater amount for non-pelagic than for pelagic trawl, we point out that both sub-gear types exhibited evidence that data from the observed trips are not likely to be representative of total trips for some metrics between years. We strongly recommend against the creation of separate ADP strata for these sub-gear types for the following reasons. First, vessels may carry both gear types on a trip, and which of these will be fished is not necessarily known before a trip begins. This is important because definitions of which stratum a trip belongs to is necessary for the correct selection rate to be applied to the trip. Second, post-stratification methods in the Catch Accounting System already correctly account for differences in sampling effort between NPT and PTR trips. Third, we must look to our past and draw from our experience with creating tendering stratum for each gear type. This effort has been marginally successful.

Although the selection rate for those trips is known, there are now so few trips in these tendering strata that evaluating whether or not observer deployments were successful in gathering representative data has become difficult. It is at least worth exploring whether current sample sizes of the observer program can support such a split between sub gear types for the purposes of creating new ADP strata and whether the uncertainty in gear type known before the trip can be overcome.

We recommend continuation of the baseline + optimization approach for determining coverage levels among strata. Based on the results of analyses presented at the October 2018 Council meeting, we do not see strong evidence to reduce gear-specific baseline coverage levels below 15%.

The lack of coverage for the end of 2018 in the EM HAL stratum was apparently due to a lack of EM review. **We recommend that EM review rates be set to ensure that the entire year is sampled and review is timely enough so that data from EM can be used for catch accounting and fisheries monitoring as envisioned by the Council.** EM selection and EM review rates in the 2020 ADP may be calculated using existing ADP protocols as long as costs for EM are available.

Citations

Alaska Fisheries Science Center (AFSC) and Alaska Regional Office (AKRO). 2019. North Pacific Observer Program 2018 Annual Report. AFSC Processed Rep. 2019-04, 148 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Alaska Fisheries Science Center and Alaska Regional Office. 2018. North Pacific Observer Program 2017 Annual Report. AFSC Processed Rep. 2018-02, 136 p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

AFSC (Alaska Fisheries Science Center). 2016. 2017 Observer sampling manual. Fisheries Monitoring and Analysis Division, North Pacific Groundfish Observer Program. AFSC, 7600 Sand Point Way N.E., Seattle, Washington, 98115. Available online at https://www.afsc.noaa.gov/FMA/Manual_pages/MANUAL_pdfs/manual2017.pdf.

Cahalan, J., J. Mondragon, and J. Gasper. 2014. Catch sampling and estimation in the Federal groundfish fisheries off Alaska: 2015 Edition. NOAA Tech. Memo. NMFS-AFSC-286, 46 p.

Cahalan J., J. Gasper, and J. Mondragon. 2015. Catch estimation in the federal trawl fisheries off Alaska: a Simulation approach to compare the statistical properties of three trip-specific catch estimators. Can J. Fish. 72:1024:1036.

Faunce, C. H. 2015. Evolution of observer methods to obtain genetic material from Chinook salmon bycatch in the Alaska pollock fishery. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-288, 28 p.

Faunce, C. H., and Barbeaux, S. J. 2011. The frequency and quantity of Alaskan groundfish catcher-vessel landings made with and without an observer. ICES J. Mar. Sci. 68:1757-1763.

Guthrie, C. M. III, Hv. T. Nguyen, A. E. Thomson, and J. R. Guyon. 2017. Genetic stock composition analysis of Chinook salmon bycatch samples from the 2015 Gulf of Alaska trawl fisheries. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-343, 33 p.

Hill, M.O. 1973. Diversity and evenness: A unifying notation and its consequences. Ecology 61: 225-236.

Nelson Jr., R., R. French, and J. Wall. 1981. Sampling by U.S. observers on foreign fishing vessels in the eastern Bering Sea and Aleutian Island region, 1977-78. Mar. Fish. Rev. 43:1-19.

NMFS (National Marine Fisheries Service). 2017a. North Pacific Observer Program 2016 Annual Report. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available online at https://www.afsc.noaa.gov/Publications/ProcRpt/PR2017-07.pdf.

NMFS. 2017b. 2018 Annual deployment plan for observers and electronic monitoring in the groundfish and halibut fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available online at https://alaskafisheries.noaa.gov/sites/default/files/final_2018_adp.pdf.

NMFS. 2016. North Pacific Groundfish and Halibut Observer Program 2015 annual report. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available online at https://alaskafisheries.noaa.gov/sites/default/files/2015observerprogramannualreport.pdf.

NMFS. 2015a. North Pacific Groundfish and Halibut Observer Program 2014 annual report. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available online at https://alaskafisheries.noaa.gov/sites/default/files/annualrpt2014.pdf.

NMFS. 2015b. Draft supplement to the Environmental Assessment for restructuring the program for observer procurement and deployment in the North Pacific. NMFS, Alaska Regional Office, Juneau. May 2015. Available online at https://alaskafisheries.noaa.gov/sites/default/files/analyses/finalea_restructuring0915.pdf.

NMFS. 2014. North Pacific Groundfish and Halibut Observer Program 2013 annual report. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available online at https://alaskafisheries.noaa.gov/sites/default/files/annualrpt2013.pdf.

NMFS. 2013. 2014 Annual deployment plan for observers in the groundfish and halibut fisheries off Alaska. National Oceanic and Atmospheric Administration, 709 West 9th Street. Juneau, Alaska 99802. Available online at https://alaskafisheries.noaa.gov/sites/default/files/adp2014.pdf.

Wall, J., R. French, and R. Nelson Jr. 1981. Foreign fisheries in the Gulf of Alaska, 1977-78. Mar. Fish. Rev. 43:20-35.

Table 1. -- Comparison between predicted and actual trip days for partial coverage strata in 2018. Predicted values come from the 2018 Annual Deployment Plan (ADP).

Strata	Predicted number of trip days in ADP	Actual number of trip days	% Difference from predicted
HAL	9,736	10,608	9.0
POT - No Tender	1,470	2,780	89.1
POT - Tender	169	273	61.5
TRW - No Tender	11,667	4,742	-59.4
TRW - Tender	552	267	-51.6
Total	23,594	18,670	-20.9

Table 2. -- Trip cancellation rates in the ODDS for 2018. A trip is cancelled by the system if the user did not identify whether fishing had occurred by the end of the year. "Paper" indicates that a trip was logged when the ODDS was not available.

Strata	**Random number outcomes**	**Logged (*a*)**	**Cancelled by system (*b*)**	**Trips remaining (*c* = *a*-*b*)**	**Cancelled by user (*d*)**	**Paper**	**% User cancellation (*d*/*c* * 100)**
HAL	Not Selected	1,820				0	
HAL	Selected	355	2	353	84	0	23.8
EM HAL	Not Selected	542				0	
EM HAL	Selected	234	0	234	9	0	3.8
POT - No Tender	Not Selected	577				0	
POT - No Tender	Selected	113	0	113	21	0	18.6
POT - Tender	Not Selected	38				0	
POT - Tender	Selected	11	0	11	3	0	27.3
TRW - No Tender	Not Selected	1,577				0	
TRW - No Tender	Selected	396	0	396	54	0	13.6
TRW - Tender	Not Selected	55				0	
TRW - Tender	Selected	16	0	16	6	0	37.5
Total	Not Selected	4,609				0	
Total	Selected	1,125	2	1,123	177	0	15.8

Table 3. -- Number of remaining trips after cancellation in each trip-selection strata (*HAL*, *POT - No Tender*, *POT - Tender*, *TRW - No Tender*, *TRW - Tender*, and *EM HAL*) that were selected using the initial random number generator (Random Number Selection) and those that remained after user manipulation (Total Final Selected). The relative impact of waivers in trip-selection is also shown (% Reduction of Selected Trips due to Waivers). **Not from random numbers.

Strata	**Total trips**	**Random number selection (*r*)**	**Inherited selection** (*i*)**	**Randomly selected but waived (*w*)**	**Total final selected (*T*=*r*+*i*-*w*)**	**% Selected from inherits ((*i*/*T*)*100)**	**% Reduction of selected trips due to waivers (*w*/(*T*+*w*)*100)**
HAL	1,726	269	43	7	305	14.1	2.2
EM HAL	692	225	12	2	236	5.1	0.8
POT - No Tender	608	92	10	2	100	10.0	2.0
POT - Tender	37	8	4	0	12	33.3	0.0
TRW - No Tender	1,796	342	33	0	375	8.8	0.0
TRW - Tender	49	10	8	0	18	44.4	0.0

Table 4. -- Number of logged trips in each partial coverage stratum (*HAL*, *POT - No Tender*, *POT - Tender*, *TRW - No Tender*, *TRW - Tender*, and *EM HAL*) that were selected using the initial random number generator (Random Selection Only) and those that remained after user manipulation (Final Expected). The relative impact of waivers in trip-selection is also shown (No Waivers).

Strata	Trip disposition	Selected trips	Total trips	Actual selection (%)	Programmed selection (%)	*p*-value (H0: Actual = Programmed)
HAL	Initial Random Selection, *a*	355	2,175	16.32	17.26	0.256
	After Cancellations, *b* (*a-b*)	269	1,726	15.59	17.26	0.069
	With Inherits, *c* (*a-b*+*c*)	312	1,726	18.08	17.26	0.372
	After Waivers, *d* (*a-b*+*c-d*)	305	1,726	17.67	17.26	0.656
EM HAL	Initial Random Selection, *a*	234	776	30.15	30.00	0.938
	After Cancellations, *b* (*a-b*)	225	692	32.51	30.00	0.158
	With Inherits, *c* (*a-b*+*c*)	238	692	34.39	30.00	0.013
	After Waivers, *d* (*a-b*+*c-d*)	236	692	34.10	30.00	0.020
POT - No Tender	Initial Random Selection, *a*	113	690	16.38	16.21	0.918
	After Cancellations, *b* (*a-b*)	92	608	15.13	16.21	0.509
	With Inherits, *c* (*a-b*+*c*)	102	608	16.78	16.21	0.700
	After Waivers, *d* (*a-b*+*c-d*)	100	608	16.45	16.21	0.869
POT - Tender	Initial Random Selection, *a*	11	49	22.45	17.39	0.346
	After Cancellations, *b* (*a-b*)	8	37	21.62	17.39	0.514
	With Inherits, *c* (*a-b*+*c*)	12	37	32.43	17.39	0.027
	After Waivers, *d* (*a-b*+*c-d*)	12	37	32.43	17.39	0.027
TRW - No Tender	Initial Random Selection, *a*	396	1,973	20.07	20.18	0.933
	After Cancellations, *b* (*a-b*)	342	1,796	19.04	20.18	0.240
	With Inherits, *c* (*a-b*+*c*)	375	1,796	20.88	20.18	0.462
	After Waivers, *d* (*a-b*+*c-d*)	375	1,796	20.88	20.18	0.462
TRW - Tender	Initial Random Selection, *a*	16	71	22.54	16.67	0.201
	After Cancellations, *b* (*a-b*)	10	49	20.41	16.67	0.446
	With Inherits, *c* (*a-b*+*c*)	18	49	36.73	16.67	0.001
	After Waivers, *d* (*a-b*+*c-d*)	18	49	36.73	16.67	0.001

Table 5. -- Number of total vessels (V), sampled vessels (v), total trips (N), sampled trips (n) for each stratum and observer deployment method (vessel and trip-selection) in 2018. The expected coverage and 95% confidence interval columns are expressed as percentages of the total number of trips taken within each stratum. Fleet totals are reported with and without Electronic Monitoring (EM) data since EM were not used for catch estimation in 2018. *Values for sampled trips and realized coverage for EM POT strata are based on EM hard drives received, not actual data reviewed. See Table 6 for review data.

Coverage	**Strata**	***V***	***v***	***N***	***n***	**Realized coverage**	**Expected coverage**	**95% confidence interval lower limit**	**95% confidence interval upper limit**	**Realized meets expected?**
Full	FULL	159	159	3,400	3,400	100.0	100.0			Yes
Partial	HAL	364	176	1,990	309	15.5	17.3	14.0	17.2	No
Partial	EM HAL	120	81	767	174	22.7	30.0	19.8	25.8	No
Partial	POT - No Tender	73	53	626	97	15.5	16.2	12.7	18.6	Yes
Partial	POT - Tender	15	7	31	9	29.0	17.4	14.2	48.0	Yes
Partial	TRW - No Tender	76	67	1,864	378	20.3	20.2	18.5	22.2	Yes
Partial	TRW - Tender	18	11	40	14	35.0	16.7	20.6	51.7	No
Gear-based Total		602	364	5,318	981	18.4				
Partial	EM POT - No Tender	17	12*	163	41*	25.2*	30.0	18.7	32.5	Yes
Partial	EM POT - Tender	1	1*	1	1*	100.0*	30.0	2.5	100.0	Yes
Partial	Zero Coverage	361	0	1,725	0	0.0	0.0			Yes
Partial	Zero EM Research	3	0	23	0	0.0	0.0			Yes
Total Fleet (without EM POT)	Total	1084	484	10,630	4,381	41.2% Trips; 44.6% Vessels				

Table 6. -- The number of EM hard drives received and reviewed by gear type and month. Totals may differ from Table 5 since trip start date was used to define trips here, rather than the landing date used to define trips in Table 5.

Gear	Data reviewed?	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Total
EM HAL	Yes	3	5	19	42	41	21	10	17	16	0	0	0	174
EM HAL	No	0	0	0	3	2	4	3	5	14	29	2	0	62
EM POT	No	19	2	0	0	1	2	1	1	8	2	2	3	41

Table 7. -- The number of *TRW - No Tender* pollock deliveries by port and coverage category. IFP: Inshore Floating Processor, Hbr: Harbor. * These values have been corrected from what is published in the 2018 Annual Report (AFSC and AKRO 2019) after missing data were found.

FMP	Coverage category	Port	Total deliveries (*N*)	Observed deliveries (*n*)	% Observed
Bering Sea	Full	Akutan	817	817	100.0
Bering Sea	Full	Dutch Hbr.	1,121	1,121*	100.0*
Bering Sea	Full	IFP	2	2	100.0
Bering Sea	Full	King Cove	81	81	100.0
Bering Sea	Full	Sand Point	12	12	100.0
Total	Full		2,033	2,030	99.9
Gulf of Alaska	Partial	Akutan	78	18	23.1
Gulf of Alaska	Partial	King Cove	1	0	0.0
Gulf of Alaska	Partial	Kodiak	1,087	216	19.9
Gulf of Alaska	Partial	Sand Point	273	46	16.8
Total	Partial		1,439	280	19.5

Table 8. -- Number of trips by observation status in the 2018 trip-selection strata.

Strata	Observed	Unobserved
HAL	309	1,681
EM HAL	174	593
POT - No Tender	97	529
TRW - No Tender	378	1,486
POT - Tender	9	22
TRW - Tender	14	26

Table 9. -- Results of permutation tests between observed and unobserved trips in the 2018 trip-selection strata. OD: Observed difference (Observed - Unobserved).

Strata	Metric	NMFS areas	Days fished	Vessel length (ft)	Species landed	pMax species	Landed catch (t)
EM HAL	Observed difference	0.003	-0.113	0.361	0.387	-0.007	0.143
	OD (%)	0.267	-2.179	0.684	9.700	-0.824	2.108
	p-value	1.000	0.557	0.660	0.022	0.558	0.727
HAL	Observed difference	0.024	-0.760	-0.568	0.049	0.008	-1.048
	OD (%)	2.158	-14.345	-1.037	1.372	0.946	-15.593
	p-value	0.267	< 0.001	0.439	0.694	0.301	0.004
POT - No Tender	Observed difference	-0.019	-0.103	2.098	0.288	0.007	1.842
	OD (%)	-1.821	-2.337	2.732	14.296	0.743	5.632
	p-value	0.523	0.663	0.481	0.024	0.087	0.581
POT - Tender	Observed difference	0.131	-0.510	12.449	0.061	-0.004	49.863
	OD (%)	11.631	-5.792	14.580	2.646	-0.424	29.994
	p-value	0.688	0.891	0.214	1.000	0.294	0.647
TRW - No Tender	Observed difference	-0.032	-0.256	-1.480	-0.096	0.015	-4.352
	OD (%)	-3.040	-9.403	-1.750	-1.657	1.590	-4.549
	p-value	0.024	0.042	0.124	0.676	0.073	0.070
TRW - Tender	Observed difference	0.071	1.819	3.593	0.308	-0.004	122.714
	OD (%)	6.969	27.262	5.800	5.806	-0.407	51.755
	p-value	0.359	0.125	0.299	0.599	0.897	0.197

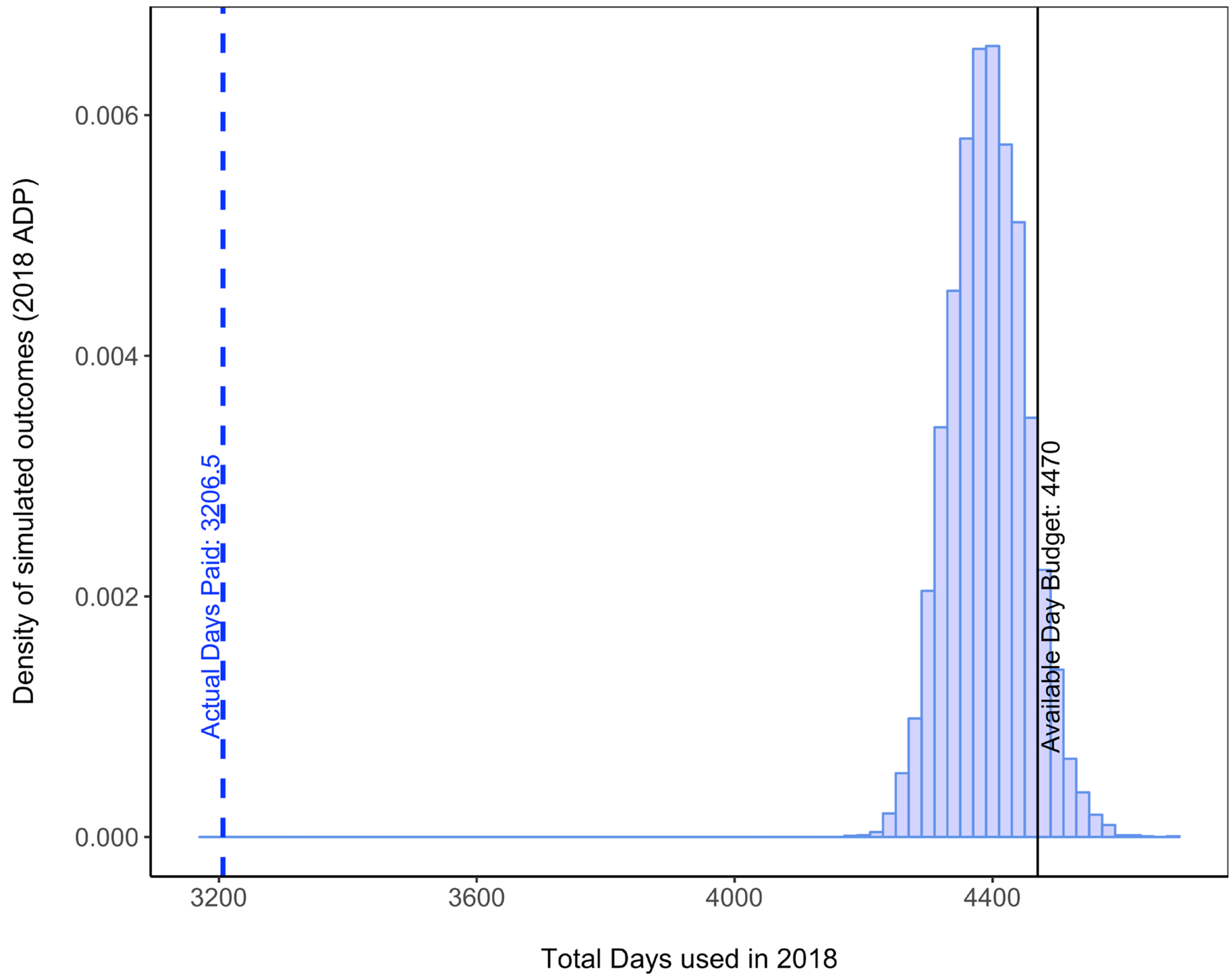

Figure 1. -- Actual paid sea-days in 2018 (dotted line) in relation to the range of potential budgetary outcomes estimated in December 2017 for the Final 2018 Annual Deployment Plan (vertical bars).

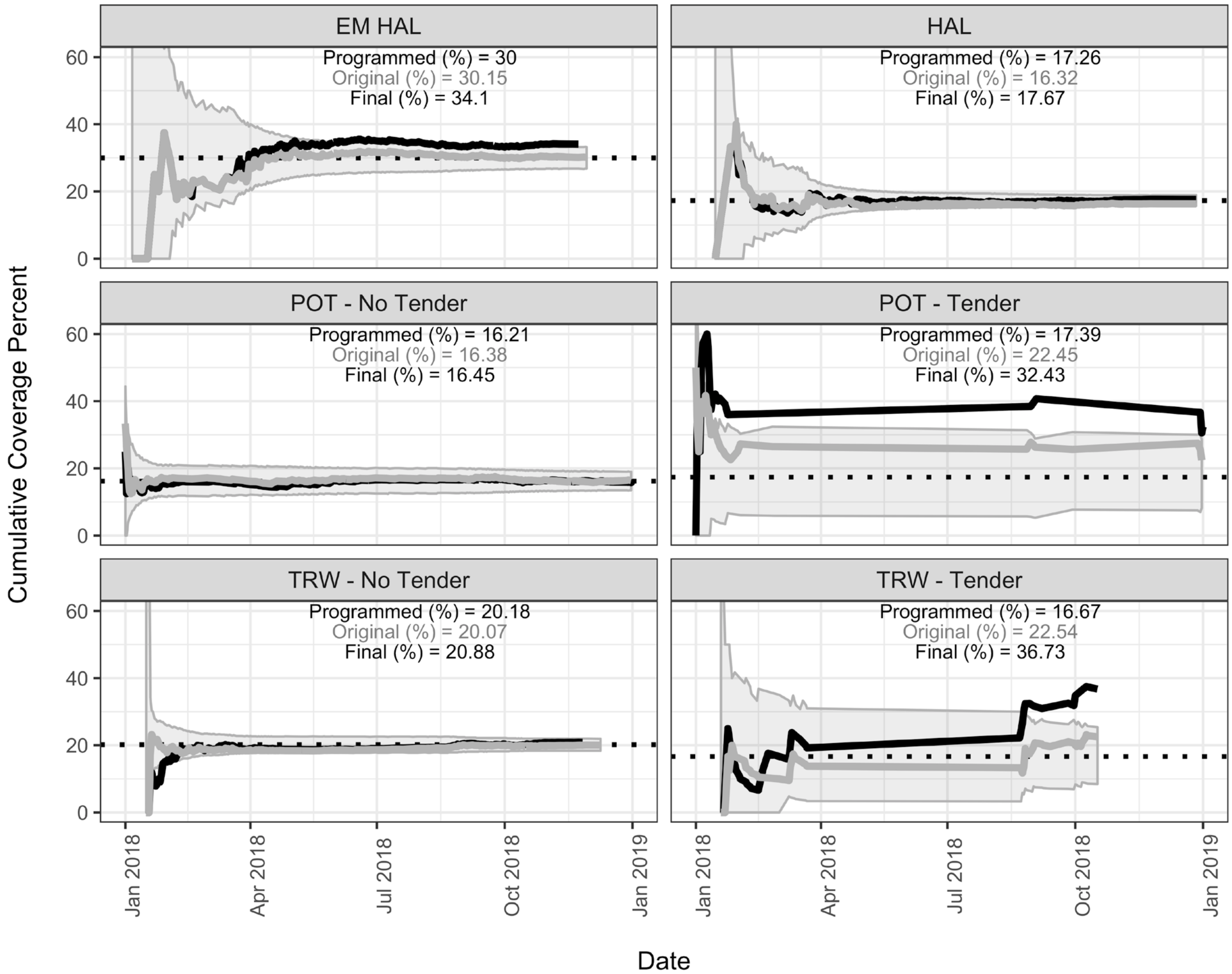

Figure 2. -- Rate of selected trips logged into ODDS organized by original date entered for all trips (grey line and grey text), and final date considering only non-cancelled trips (black line and black text). The programmed selection rate is depicted as the dotted line. Grey shaded areas denote the range of coverage rate corresponding to the 95% confidence intervals expected from the binomial distribution. The final coverage rates were higher than if trip dates had not been altered and/or cancelled.

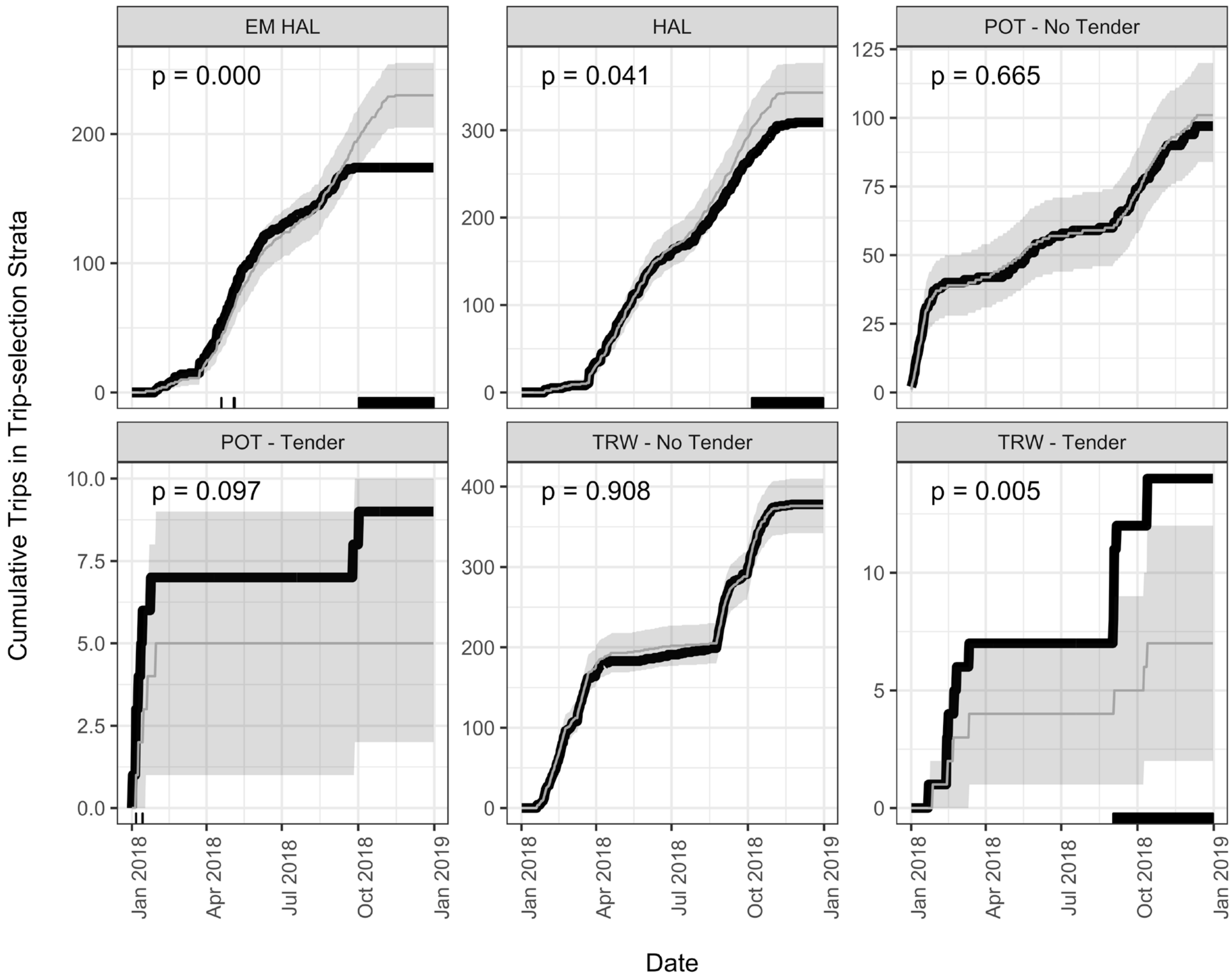

Figure 3. -- Cumulative number of trips observed during 2018 (black line) compared to the expected range of observed trips (shaded area) given fishing effort and sampling rates. Dates where the observed number of trips is outside of expected (less or more than the range; OOE) are depicted as tick marks on the horizontal x-axis. The results of tests that the observed rate derived from a binomial distribution sampled at the selection rate are denoted as *p*-values.

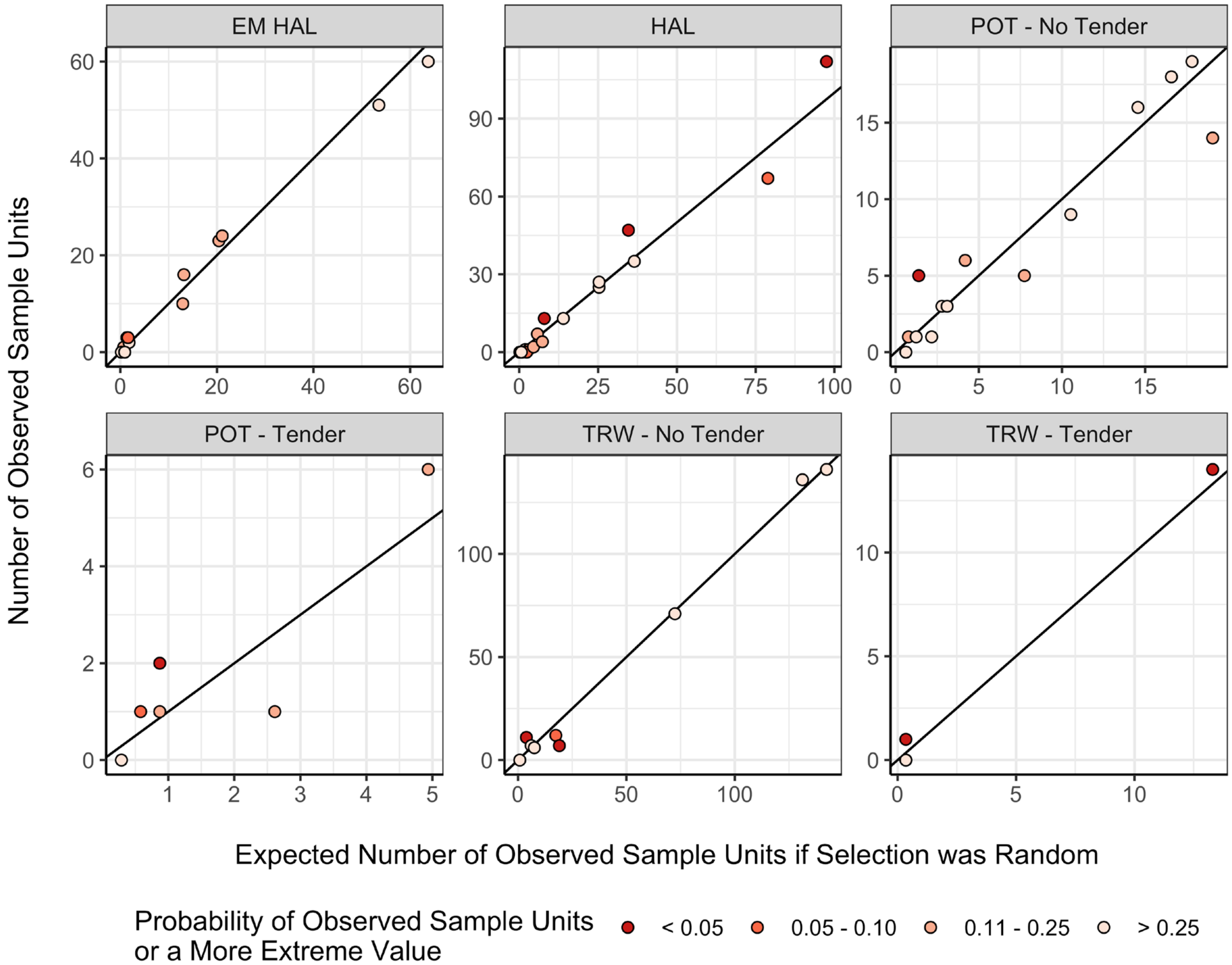

Figure 4. -- Comparison plots depicting the number of observed sample units compared to the number of expected observed sample units for each partial coverage stratum. Each point on a plot represents a NMFS Area. The darker the point, the more unusual the result.

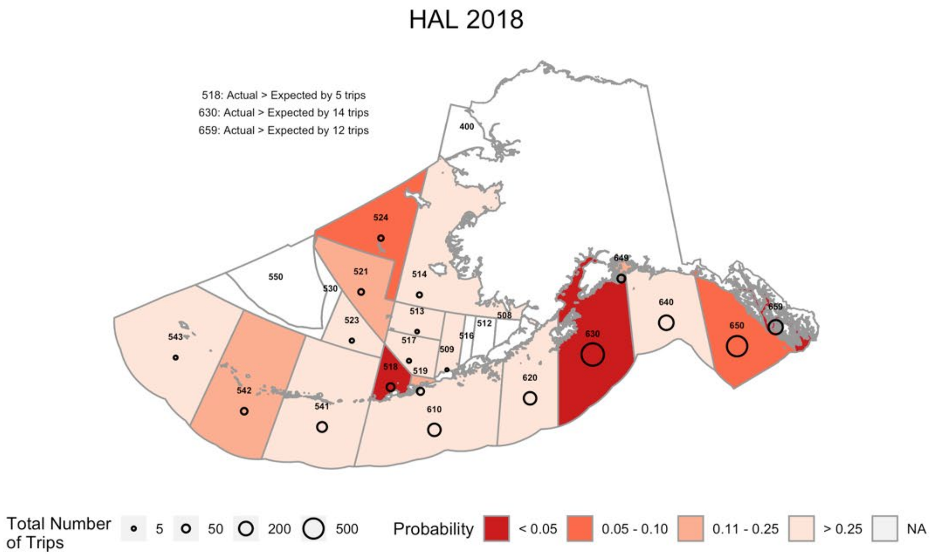

Figure 5. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *HAL* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

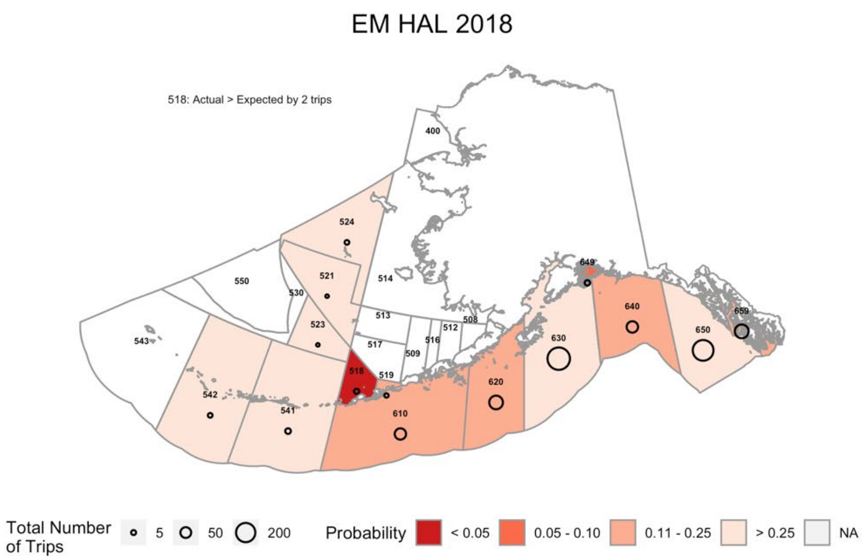

Figure 6. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *EM HAL* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

POT - No Tender 2018

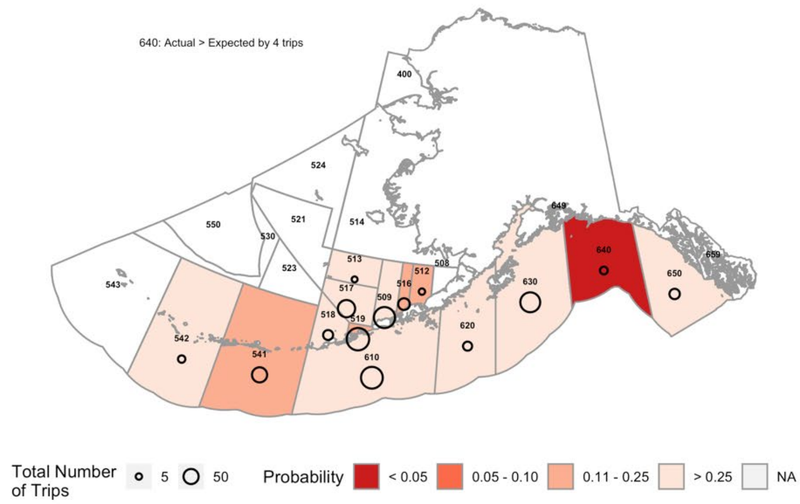

Figure 7. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *POT - No Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

TRW - No Tender 2018

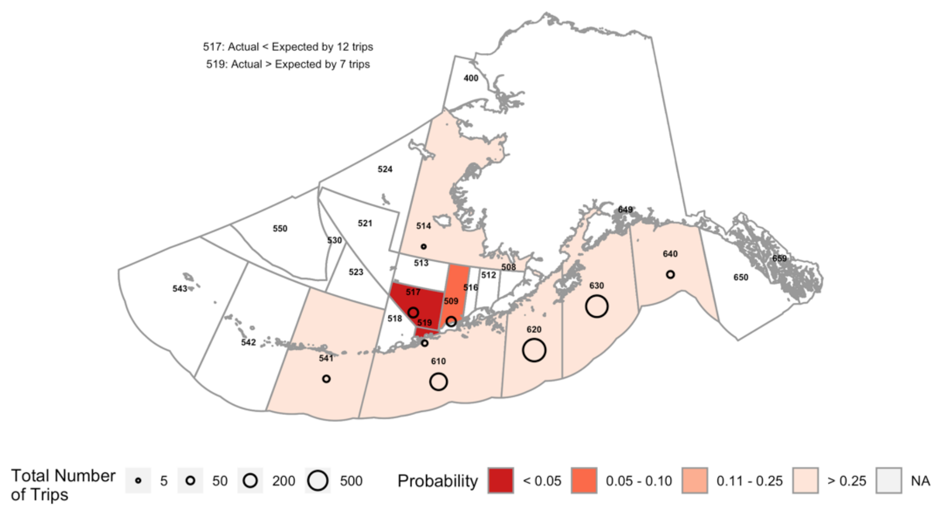

Figure 8. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *TRW - No Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

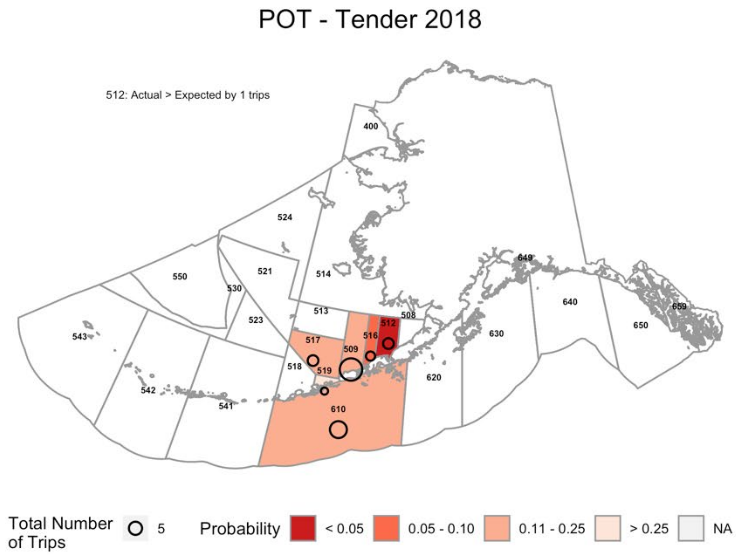

Figure 9. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *POT - Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

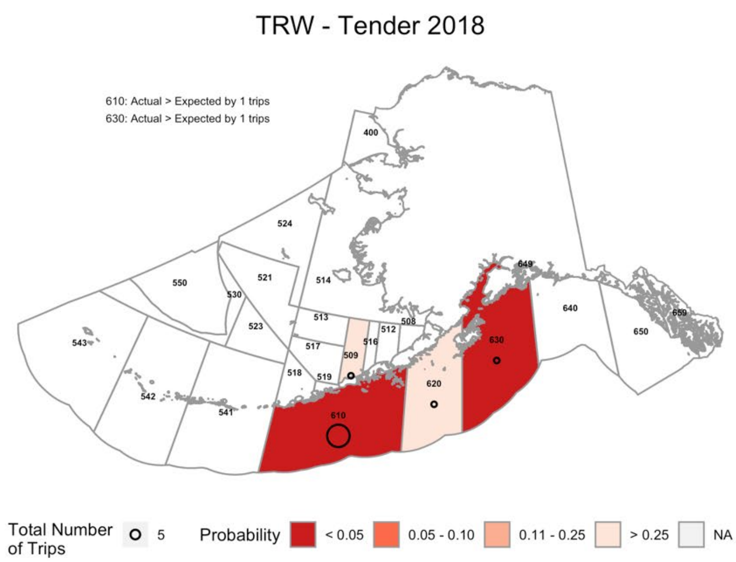

Figure 10. -- Probability of observing the realized or more extreme outcome (coverage rate) in a NMFS Reporting Area in the *TRW - Tender* stratum. Reporting Areas where unlikely outcomes occurred are shaded in darker colors.

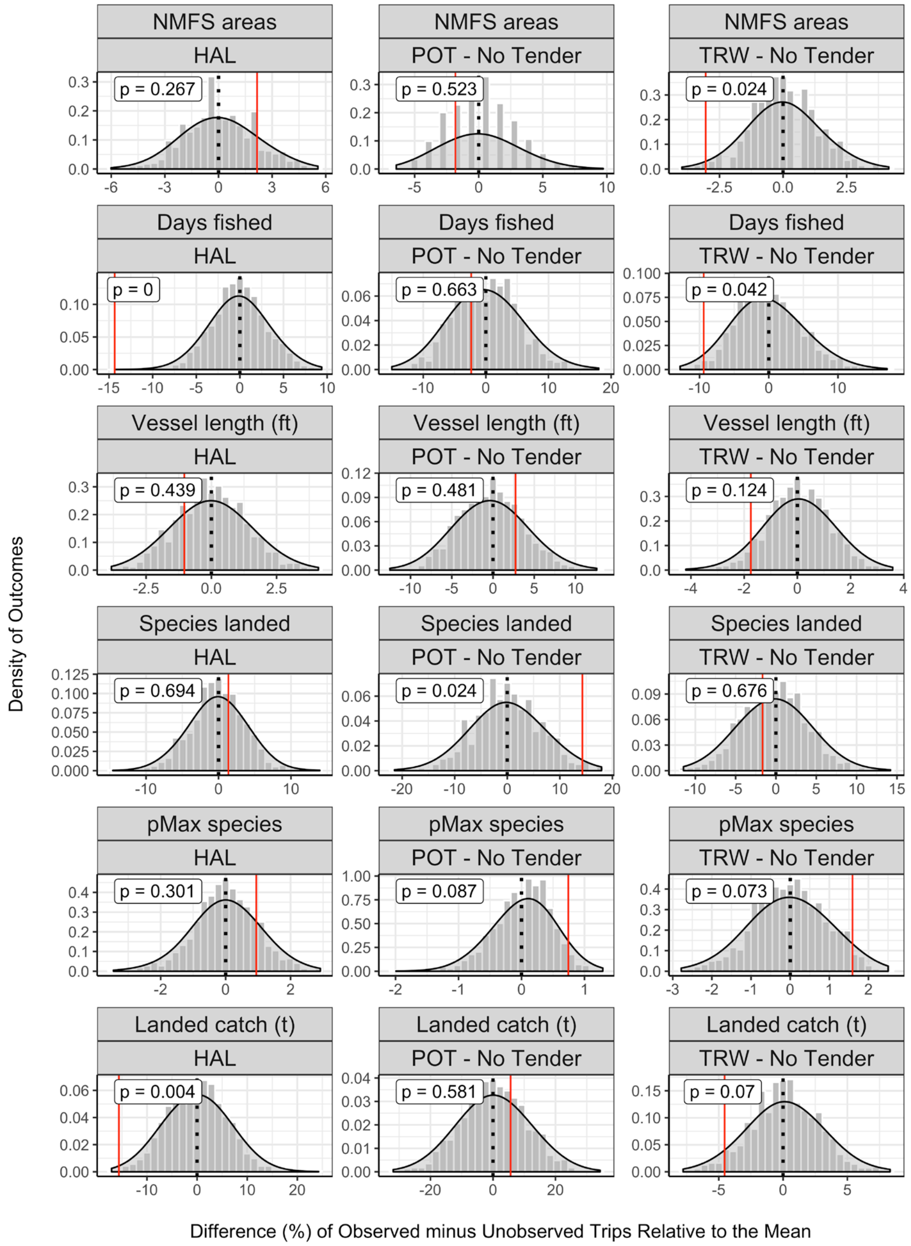

Figure 11. -- Example of results from permutation tests depicting percent differences between observed and unobserved trips for each strata in the partial coverage category. Grey bars depict the distribution of differences between observed and unobserved trips where the assignment of observed status has been randomized (this represents the sampling distribution under the null hypothesis that observed and unobserved trips are the same). The vertical line denotes the actual difference between observed and unobserved trips. Values on the x-axis have been scaled to reflect the relative (%) differences in each metric. The *p*-value for each test is denoted in the upper left corner. Low *p*-values are reason to reject the null hypothesis and conclude that there is an observer effect. Results from all permutation tests can be found in the Tables section of this report.

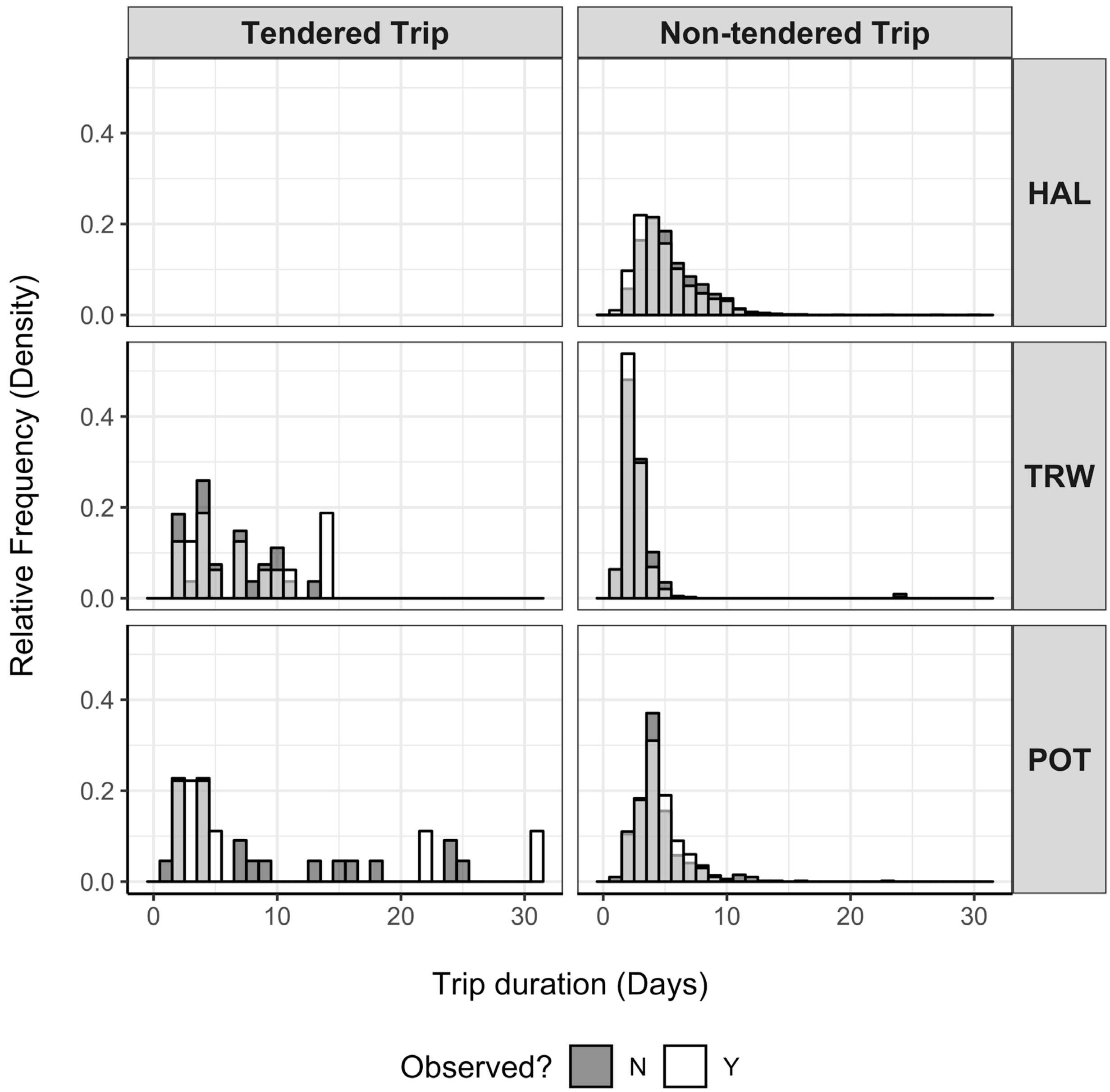

Figure 12. -- Distribution of trip durations for vessels in the partial coverage category by gear and observation status. Observed trips are depicted as transparent white bars overtop of solid black bars for unobserved trips. Trip durations where both observed and unobserved status exist are depicted in gray (This is not the same as a stacked bar chart, in which the height of the bar would reflect observed and unobserved on top of one another- this plot has each observation status in front of the other).

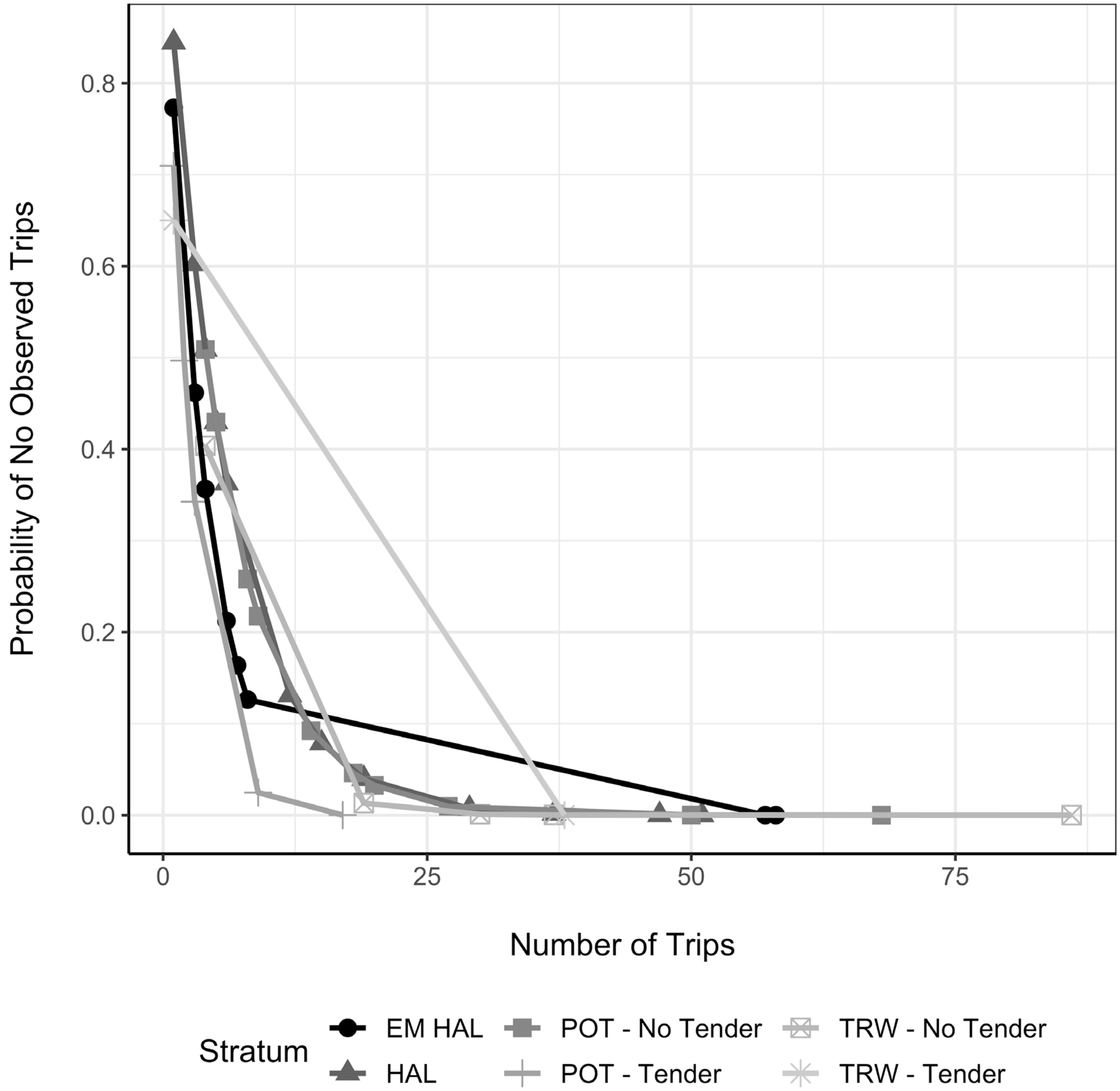

Figure 13. -- Probability of observing no trips in a NMFS Area and stratum given fishing effort and sampling rate. The x-axis has been truncated to increase resolution at low levels of fishing effort. The likelihood of having no observer data decreases with increasing total fishing effort and selection rate. The selection rate is 17.26% in the *HAL* stratum, 16.21% in the *POT - No Tender* stratum, 17.39% in the *POT - Tender* stratum, 20.18% in the *TRW - No Tender* stratum, 16.67% in the *TRW - Tender* stratum, and 30% in the *EM HAL* stratum.

Appendix - Evaluation of Pelagic and Non-pelagic Trawl Trips

Introduction

At its June 2017 meeting, the North Pacific Fishery Management Council requested that NMFS evaluate whether there is evidence of an observer effect in either pelagic trawl (PTR) or non-pelagic trawl (NPT) gear fished by partial coverage vessels. These two gear types are typically used for different styles of fishing, with NPT gear associated with bottom contact and PTR gear typically fished in the water column. While this is often the case, both gear types can be fished on the bottom.

The recommendation followed an FMAC request for the evaluation, including a discussion about the “pros and cons” of separate observer deployment strata for those two gear types. One concern identified is vessels selected for observer coverage being directed to fish for pollock in order to avoid the at-sea sampling of salmon PSC that is done on non-pollock trips in the Gulf of Alaska. This type of activity can only occur when pollock is open for directed fishing but would result in a vessel avoiding an at-sea sample for salmon by taking a pollock trip when observed. Such behavior would result in higher observer coverage in PTR gear since it is used to target pollock. Whereas salmon accounting for observed vessels fishing with NPT gear is based on highly variable at-sea samples, observed vessels fishing for pollock (usually using PTR gear) typically have salmon accounted for during the offload at the shoreside processing plant. In management situations where salmon PSC caps are a concern, industry may choose to fish such that their offload is primarily pollock on observed trips, thus obtaining a shoreside count of salmon PSC. Conversely, if halibut PSC limits are a management concern, industry may direct more observed vessels to fish with NPT gear to obtain a larger sample of fishing activity with that gear type.

The 2018 Annual Deployment Plan separates trawl strata by tender status, not by whether the gear being used is pelagic or non-pelagic. The catch accounting system (CAS) post-stratifies observer and landings data based on whether the trip is recorded as NPT or PTR on the landing report (“fish tickets”) or in the observer data. In both cases, the vessel operator is reporting the gear type being used to the observer (usually through the logbook) or through *e*Landings. Although the gear information is "self-reported", regulations at, 50 CFR 679.2 (definitions) define pelagic and non-pelagic trawl gear to be of certain configurations (e.g., floats, mesh configurations, line configurations).

The two gear types are also associated with differing fishery management issues, with salmon bycatch being the primary issue for the pollock pelagic trawl fishery and halibut PSC being of concern for the non-pelagic trawl fishery. Being a relatively rare species, salmon are accounted for shoreside when an observer is onboard and the vessel is not delivering to a tender. These counts are extrapolated to unobserved trips. In contrast, halibut discard estimates are only based on data collected by observers at sea, and extrapolated from observed to unobserved trips.

The OSC responded to the Council's request with an Appendix in the 2017 Annual Report (AFSC and AKRO 2018, Appendix A). The OSC chose not to include a permutation test in that analysis after concluding that stratification by these sub-gear types is not feasible due to the fact that both PTR and NPT gear is used in some trips (AFSC and AKRO 2018, Appendix Table A-1). At its June 2018 meeting, the Council again requested that NMFS evaluate whether there is evidence of an observer effect in either pelagic trawl or non-pelagic trawl gear fished by partial coverage vessels. That analysis is provided here, with permutation test results included. All analyses consider only non-tendered trawl trips.

Results

Since 2016, 99.7% of the partial coverage category PTR landings targeted pollock (Appendix Table 1). Of these 4,383 pollock trips, 97.3% had a catch composition of at least 95% pollock, which falls into the CAS "pelagic" pollock target (suggesting mid water tows). The remaining pollock landings were in the "bottom" pollock target category, which is based on the pollock being the predominant species retained (but less than 95% of the retained catch). The predominant targets for vessels fishing NPT gear were Pacific cod (54.3% of trips) and arrowtooth flounder (31.1% of trips), followed by pollock (8.8% of trips; Appendix Table 1).

Observation rates for PTR gear were consistently higher than observation rates for NPT gear (Appendix Table 2). This supports the theory that vessels may be opting to fish PTR gear when chosen for observer coverage. Also of note is that mixed-gear trips, where the vessel fishes both pelagic and non-pelagic gear during a trip, are not uncommon (Appendix Table 2).

The results from the permutation test are mixed. Between 2016 and 2018, 11 of 18 metric/year combinations showed evidence of an observer effect in NPT gear (Table 3). This is in comparison to 4 of 18 metric/year combinations showing evidence of an observer effect in PTR gear (Table 3). Within each gear type, we would expect $18 \times 0.05 \approx 1$ metric/year combinations to show evidence of an observer effect by random chance. This means that while both gear types showed evidence of an observer effect, there was more evidence of an observer effect in NPT gear.

Discussion

The results of this analysis do not conflict with the conclusions made in Appendix A of the 2017 Annual Report (AFSC and AKRO 2018). This analysis simply provides more nuanced information requested by the Council that relates to the observation of trawl trips within the PTR and NPT gear types. The OSC continues to recommend against stratification by the NPT and PTR gear types. The reasons for this recommendation can be found in Appendix A of the 2017 Annual Report (AFSC and AKRO 2018).

Appendix Table 1. -- Number of trips by target species for NPT and PTR gear types between 2016 and 2018. For the purpose of this table, mixed-gear trips are excluded.

Gear	Target	N
NPT	Pacific cod	1,020
	Arrowtooth flounder	584
	Pollock	166
	Flatfish (shallow water)	70
	Flathead sole	13
	Rex sole	7
	Sablefish	5
	Atka mackerel	4
	Rockfish	4
	Yellowfin sole	4
	Other	3
NPT Total		1,880
PTR	Pollock	4,383
	Arrowtooth flounder	6
	Flatfish (shallow water)	3
	Pacific cod	2
	Atka mackerel	1
	Rockfish	1
PTR Total		4,396

Appendix Table 2. -- Number of total trips (N) and sampled trips (n) by trawl gear type. For the purpose of this table, each mixed-gear trip is counted three times: as a PTR trip, as an NPT trip, and as an NPT & PTR trip.

Gear	N	n	% Observed
2016			
PTR	1,560	421	27.0
NPT	844	205	24.3
NPT & PTR	62	19	30.6
2017			
PTR	1,544	350	22.7
NPT	508	82	16.1
NPT & PTR	38	1	2.6
2018			
PTR	1,292	272	21.1
NPT	528	92	17.4
NPT & PTR	44	14	31.8

Appendix Table 3. -- Results of permutation tests between observed and unobserved trips within the NPT and PTR gear types between 2016 and 2018. For the purpose of these tests, mixed-gear trips were excluded.

Gear	Metric	NMFS areas	Days fished	Vessel length (ft)	Species landed	pMax species	Landed catch (t)
2016							
NPT	Observed difference	-0.037	-0.493	-1.044	-1.764	0.049	-23.684
	OD (%)	-2.925	-14.584	-1.188	-26.854	6.022	-39.879
	p-value	0.335	< 0.001	0.373	< 0.001	0.008	< 0.001
PTR	Observed difference	-0.012	0.246	4.884	-0.185	-0.001	7.952
	OD (%)	-1.181	11.287	6.024	-3.833	-0.112	8.484
	p-value	0.195	0.010	< 0.001	0.128	0.395	< 0.001
2017							
NPT	Observed difference	0.063	-0.263	1.504	-1.521	0.056	-16.168
	OD (%)	5.051	-7.757	1.689	-21.083	6.765	-19.774
	p-value	0.278	0.055	0.354	0.008	0.016	0.007
PTR	Observed difference	-0.012	-0.115	-1.437	-0.224	-0.002	-3.072
	OD (%)	-1.178	-5.353	-1.698	-5.200	-0.169	-2.857
	p-value	0.193	0.285	0.186	0.109	0.002	0.150
2018							
NPT	Observed difference	-0.089	-0.290	-4.309	-0.360	0.032	-18.648
	OD (%)	-7.746	-8.889	-5.059	-4.035	4.142	-26.359
	p-value	0.046	0.011	0.019	0.572	0.168	< 0.001
PTR	Observed difference	-0.001	-0.203	-0.644	0.195	-0.002	-2.076
	OD (%)	-0.144	-8.149	-0.765	4.410	-0.158	-1.986
	p-value	1.000	0.193	0.611	0.178	0.441	0.387

U.S. Secretary of Commerce
Wilbur L. Ross, Jr.

Acting Under Secretary of Commerce for Oceans and Atmosphere
Dr. Neil Jacobs

Assistant Administrator for Fisheries
Chris Oliver

September 2019

www.nmfs.noaa.gov

OFFICIAL BUSINESS

National Marine Fisheries Service
Alaska Fisheries Science Center
7600 Sand Point Way N.E.
Seattle, WA 98115-6349