

This information is distributed solely for the purpose of predissemination peer review. It has not been formally disseminated by NOAA. It does not represent any final agency determination or policy.

**A Description of the Allocation Procedure
applied to the 1994 to present Commercial Landings Data**

by

S.E. Wigley, P. Hersey and J.E. Palmer

Working Paper in support of Terms of Reference A

Groundfish Assessment Review Meeting
GARM Data Review Meeting
Woods Hole, MA

October 29 – November 2, 2007

Table of Contents

Executive Summary

INTRODUCTION

Background

Data sources

METHODS

Allocation design

Allocation levels

Meta fields

Level A

Level B, C, and D

Area probability density functions

Effort estimation

Allocation checks and constraints

Allocation Assumptions

PREPARATION of DEALER DATA

Consolidated trips

Dealer Trip Identifier

Non-allocated and Allocated data sets

PREPARATION of VESSEL TRIP REPORT DATA

VTR matching data sets

Level B, C and D area data sets

Level B, C and d effort data sets

MATCHING: Creation of Header and Species detail records

Header records

Species records

COMPARISON OF ALLOCATION INPUT DATA SETS

ALLOCATED DATA

Results of matching

Biological samples

Evaluation of the random component of the allocation procedure

CONCLUSIONS/DISCUSSION

Acknowledgments

Literature Cited

DO NOT CITE OR DISTRIBUTE

EXECUTIVE SUMMARY

The purpose of the allocation is to supplement the mandatory commercial landing data (1994 onward) with area fished and effort information using the Vessel Trip Reports (VTR). The goal is to eliminate the need for single species allocation for each analysis conducted and to maintain a consistent, comprehensive, commercial landings database from the 1963 -present containing the information need to address management questions, conduct stock assessments and ecosystem research.

The multi-tier trip-based allocation is designed to combine each mandatory reporting dealer (Dealer) trip with a vessel trip report (VTR) trip or a group of VTR trips of similar characteristics to obtain area fished and effort associated with the Dealer trip. Although the trip-based allocation and the single species proration yield similar results with regard to stock landings (Wigley et al. 2007b), the trip-based allocation is an improvement over the single-species proration because it provides area fished at a fine level of resolution (statistical area rather than stock level) for all species. It also estimates effort associated with these landings. The trip-based allocation represents a comprehensive approach to determining area fished and effort in Northeast region's commercial landings in order to meet scientific and fishery management needs as well as commercial data reporting requirements to Northwest Atlantic Fisheries Organization.

The multi-tier trip-based allocation has been developed to augment commercial landings data with area fished and effort; however, trip characteristics, species landings and price information will not change. All species on a given trip/subtrip will be assigned the same area and effort. The multi-tier trip-based allocation utilizes Vessel Trip Report data that has been aggregated into four levels: Level A, Level B, Level C and Level D. At Level A, Dealer and VTR trips are matched one to one. At Levels B, C and D, VTR trips are grouped together to form a pool of trips with similar characteristic which define the stratification cell within the level of aggregation.

A Dealer trip seeks an area match at Level A, and progresses through the increasing levels of aggregated VTR data until a match occurs. Area is obtained first, then effort is obtained .

For each area level and stratification cell, a discrete probability distribution function is formed representing the proportion of trips which fished in a unique statistical area. A discrete cumulative distribution is formed using the statistical area probabilities. Each unique statistical area within the VTR group will have a cumulative probability associated with it. Before the allocation begins, every Dealer trip is assigned a random number between 0 and 1. The random number is compared with the cumulative probability associated with each area. The cumulative probabilities are in ascending order; when the random number is greater than or equal to the cumulative probability value, the statistical area associated with the cumulative probability is assigned to the Dealer trip. Thus, a single area fished is assigned to a Dealer trip on a probabilistic basis by sampling (with replacement) the distribution of VTR areas within the group.

Total effort is not known in the Dealer data; each Dealer trip will be supplemented with effort (number of trips, *ntrips*, days fished and days absent) taken directly from a VTR trip or estimated from the pool of VTR trips with similar characteristics. When a match occurs at Level A, days fished and days absent are transferred to the Dealer trip only when both effort metrics have values (both must be not null). If available, the number of hauls, haul duration, crew size, gear quantity, and gear size are also transferred. If a match occurs at Level B, C or D, then an estimate of days fished (DF) per trip and an estimate of days absent (DA) per trip are assigned to the Dealer trip. Both days fished and days absent are estimated by the median of their distributions, respectively, within the cell. The median was selected as the simplest statistic of central tendency for distributions of various shapes.

The estimates of days fished and days absent are assigned to an entire trip. Therefore, all VTR trips within the group must be converted into 'whole trips'. The days fished and the days absent are each multiplied by the inverse of *ntrips*, the portion of the trip. This calculation is non-consequential for non-split trips; for split trips, the effort is expanded to represent a whole trip. For example, if a subtrip had an *ntrips* = 0.3 then, to convert this partial trip to a whole trip, the converted DF = DF * 1/0.3 and the converted DA = DA * 1/0.3.

In addition to estimating the median (second quartile), the first and third quartiles are also derived to provide a measure of dispersion. The quartile deviation can be calculated as $(q_3 - q_1)/2$. The first and third quartiles of DF and DA will be transferred into the Dealer trip record for analysts to use, if desired.

The allocation assumes the follow: 1) Dealer landings are a census of total landings; 2) vessels land only once per trip; 3) each Dealer trip that enters the allocation represents one trip; and 4) VTR data set is a representative subset of the Dealer set

The proportion of Dealer landings entering the allocation range between 19% and 32%. Between 51% and 73% of the landings that enter the allocation to find area fished match at Level A (a one to one match of Dealer and VTR trips). Total commercial landings changed very slightly (< 1 mt) due to rounding of whole species pounds on split trips. An evaluation of input data for allocation revealed the VTR subset generally reflected Dealer data. An evaluation of the random component of the allocation indicated that the random component did not contribute to a wide spread in stock landings, indicating that the random component is not a large source of stock landings variability. Although some statistical areas on the biological samples associated with allocated trips changed, the majority of samples remained unchanged.

The trip-based allocation is an improvement over the single-species proration.

INTRODUCTION

Commercial landings data are used to address management questions, to conduct stock assessments and to meet reporting requirements for fishery resources off the east coast of the United States. Beginning in June 1994, the National Marine Fisheries Service/Northeast Region's data collection system was changed from a voluntary to a mandatory reporting system for USA fishermen and dealers who catch and buy/sell groundfish species regulated by the Northeast Multi-species Fishery Management Plan. The mandatory reporting system consists of two components: 1) dealer reporting and 2) vessel trip reporting. Each component contains information needed for fishery management and stock assessment analyses: the dealer reports contain total landings by market category, while the vessel trip reports contain information on area fished, kept and discarded portions of the catch, and fishing effort. There is no unique identifier to link these two components into one database. A multi-tier trip-based allocation scheme has been developed to combine information for these two components into a single database which is consistent with commercial landings data prior to 1994. This comprehensive trip-based allocation will eliminate the need for single species proration. This project was undertaken by the staffs of the Northeast Fisheries Science Center (NEFSC) and the Northeast Regional Office (NER).

The purpose of this document is to describe the multi-tier trip-based allocation designed to combine each mandatory reporting dealer (Dealer) trip with a vessel trip report (VTR) trip (or a group of VTR trips of similar characteristics) to obtain area fished and effort associated with the Dealer trip. The document describes: 1) allocation design; 2) the qualitative approach used to evaluate if the VTR data set is a representative subset of the Dealer data set; 3) the results of matching within the allocation; 4) an evaluation of the random component of the allocation; and 5) the changes to statistical area previously assigned to biological samples. A detailed technical in-house manual documenting the computer programs has been created.

Background

An evaluation of the 1994 vessel trip report data collected under the mandatory system was undertaken in spring 1996 by the Northern Demersal and Coastal/Pelagic Working Groups of the Stock Assessment Workshop (SAW). Findings were reported to the 22nd SAW (NEFSC 1996). The Stock Assessment Review Committee (SARC) recommended that: 1) the data needed further auditing; 2) use of existing data for provisional assessment calculations should be "performed with extreme caution and full awareness of the problems in the database"; 3) analysis and design of the mandatory data collection system should be completed and implemented with consideration given to the following features: a) an unambiguous linking criterion for dealer, VTR, sea sampling and effort monitoring databases; b) pre-audits of all submitted data to eliminate ambiguities and preserve the original integrity of the VTR information; and c) create user-friendly data collection forms with clear instructions for recording information; and 4) until long-range problems are resolved, immediate steps should be taken to improve the existing data collection process (NEFSC 1996).

Subsequent to the 1996 VTR data evaluation, further auditing of the VTR data has continued at

NER. Since 1997, single species proration of landings data have been cautiously performed on an ad-hoc basis to meet stock assessment and management needs. The single species proration is narrow in scope, determining stock areas landings (comprising several statistical areas) by calendar quarter (Wigley et al. 1998) and does not estimate effort. The multi-tier trip-based allocation expands upon these limitations and derives area fished for landings of all species caught on a trip to statistical area (Figure 1) and estimates effort, while maintaining the Dealer's original temporal resolution of month and day. Similar to the single species proration, the multi-tier trip-based allocation does not alter species landings, but augments the Dealer landings data record with area fished and effort.

Other NEFSC analysts have also prorated dealer data by time and area for protected species by port groups (Bisack 2003).

Data Sources

Dealer data

The Dealer data used in the allocation originated from the Commercial Fisheries Database System (CFDBS) Oracle tables maintained by Northeast Fisheries Science Center. Landings data that were not part of the mandatory reporting system did not enter the allocation, and accordingly, these landings were held aside until the allocation was complete, and then re-joined for a complete commercial landings data set (Figure 2). The mandatory Dealer data contain species landed and live pounds by market category, date landed, vessel permit, gear type, ton class, port landed and price. The mandatory Dealer data do not include area fished, gear characteristics (mesh size, gear quantity and gear size) or effort (crew size, number of hauls, haul duration, days fished or days absent); this information will be supplied from the VTR data during the allocation. The steps taken to identify which Dealer trips will enter the allocation process and the procedures developed to prepare the Dealer data are described in a subsequent section.

In May 1, 2004 Dealer Electronic Reporting (DER) was implemented as part of Amendment 13 of the Multispecies FMP¹. There are no requirements for Dealers to submit gear information through DER, however, many of the Dealers do so. This aspect will need further investigation as the 2004 to 2006 data are evaluated. *Author's note: the 2004 to 2006 commercial data are being processed through the allocation procedure.*

Vessel Trip Report data

Northeast multi-species VTR data used in the allocation originated from Oracle tables (*DOCUMENT*, *CATCH* and *IMAGES*) maintained by the Northeast Regional Office. These data were used to populate Oracle tables (*VESLOGyyyyT*, *VESLOGyyyyG*, *VESLOGyyyyS*, where yyyy is 4-digit year) created by Northeast Fisheries Science Center (Appendix Figure 1). The VTR data tables used to determine year were based on the date landed or date sold as these dates would most closely correspond to the date in the Dealer data. These data contain logbooks from

¹ http://www.nero.noaa.gov/dealer_er/highlights/04edrfr.pdf

charter, party and commercial trips, as well as logbooks which document that no fishing took place during a given month. Only commercial trips which fished and had kept catch were used in the allocation. The VTR data contain information on area fished, kept and discarded species pounds, gear type (gear size, gear quantity, mesh size), and effort (number of hauls, haul duration and crew size). Extensive data summaries and analyses revealed the VTR data were in 'raw' form and that procedures were needed to further audit the 1994-2001 VTR data before the data could be used in the allocation scheme. VTRs which did not contain fishing area location data (e.g. statistical area, latitude/longitude or loran) were eliminated for the data set. The VTR data used in the allocation procedure are described in a subsequent section.

Southeast pelagic vessel trip reports contain area and effort data needed to supplement the large pelagic landings data in the Northeast Region. The Southeast pelagic VTRs will not be incorporated into the allocation at this time; however, the allocation scheme could accommodate multiple but separate sources of VTR data.

Data constraints

Several data constraints precluded a simple direct match of each Dealer trip with a VTR trip. These data constraints include: 1) the lack of a unique identifier between the two data sets; 2) not every Dealer trip has a corresponding VTR trip due to the lack of 100% compliance by fishermen (i.e. fishermen are not submitting a logbook for each and every fishing trip); 3) incomplete logbooks (missing data within logbook); and 4) data collection inconsistencies between the two data sets.

In lieu of a unique identifier², data elements common to each data set are essential in establishing an indirect link to match trips between the two data sets. Common elements that uniquely describe a trip or characterize a trip's fleet include: vessel permit, ton class, month, day, gear type, and port. The use of day, port and gear as common elements have some associated caveats which are described later in this section.

Although mesh size is a key factor identifying sub-fleets, mesh size is not a common element in the two data sets and therefore could not be used. However, species caught during a trip is an indirect indicator of the mesh size used. As a surrogate for mesh size, VTR and Dealer trips were categorized into 12 main species groups based on the species kept (VTR) or species landed (Dealer) for a given trip. Main species groups are useful in differentiating sub-fleet sectors that are spatially distinct such as the long-line monkfish trips and long-line tilefish trips as well as the large-mesh and small-mesh otter trawl fisheries. Details on these main species groups are discussed later in this section.

There is not a one-to-one correspondence of trips between the Dealer and VTR data sets due to less than 100% compliance for VTR submission and incomplete or unuseable VTRs; hence, the VTR data are a subset of Dealer landings. Without a one-to-one correspondence between data sets, it was necessary to develop a multi-tier allocation scheme that would allow both one-to-one

² In May 2004, a unique trip identifier was established; due to limited QA/QC procedures, it was not possible to utilize the unique trip identifier in the allocation for 2004 to 2006. It will be possible to incorporate the trip identifier into the allocation when this field is fully audited.

matches as well as one-to-many VTR trips of similar characteristics in order to determine area fished and effort associated with the Dealer trip. The VTR data are examined to determine if these data are a representative sample of the Dealer data. A qualitative evaluation to identify potential bias is described in a subsequent section.

During the 1994-2001 period, incomplete VTR logbooks were submitted. Incomplete logbooks, in this case, consisted of VTRs that did not contain chart (statistical) area fished and/or VTRs which did not report the number of hauls and/or haul duration, the information needed to derive effort in terms of days fished. For those VTRs that did not report a statistical area fished but did report a latitude/longitude or *Loran*, those data are used to derive statistical area, thus increasing the number of VTRs with statistical area. Any VTR trip that did not report statistical area fished or for which an area could not be derived from the point location was eliminated from the VTR data set. Intermittently since 2001, incomplete VTRs have been returned to the fisherman for re-submission of a complete VTR.

Regarding area and effort information, all combinations of logbook completeness existed: area and effort reported; area and no effort reported, effort and no area reported; and no area or effort reported. Given the number of logbooks with incomplete effort (missing number of hauls and/or haul duration) in 1994, the allocation scheme separated the determination of area from the determination of effort in order to maximize the number of VTRs used in the allocation. The allocation consisted of two phases: in the first phase, area fished was determined for each Dealer trip using VTRs with area fished regardless of missing effort. However, if a VTR trip reported subtrips (i.e., a split trip), and effort is missing on one or more of the subtrips, then the entire trip is eliminated from VTR set because subtrip effort is used to partition the proportion of ntrips (number or fraction of a trip) to each subtrip. In the second phase, effort (days fished and days absent) is determined using VTR trips that report both area fished and effort (number of hauls and haul duration).

Data collection and coding inconsistencies between the Dealer and VTR data collection systems necessitated the grouping of similar gear types and similar ports. These data groups are used in the characterization and stratification of trips for the allocation matching; however, the original Dealer data is not replaced with these data groups. The data groups are created as additional fields on the data record.

Gear groups: The formation of gear groups is necessary because not all gear codes in the Dealer data have a corresponding VTR gearcode. The gear groups are generally based upon the CFDBS *negear2* code, with some exceptions such as *negear2* = 05 where otter trawl gear types are separated out into three unique gear group codes (i.e. scallop trawl, shrimp trawl and otter trawl). On the other hand, where distinct *negear2* codes represent similar gear types such as hoes and shovels and rakes, these gear types are coded into one group. The gear groups are formed in an ad-hoc fashion based on general gear knowledge as well as assistance from the Regional Office staff. VTR trips are assigned a gear group for each gear used on a trip. For a Dealer trip, if multiple gear types are reported, then the gear type associated with the plurality of the catch is used.

Port groups: Port groups are formed to facilitate the aggregation of VTR data to capture fleet behavior patterns. Port groups are defined by concatenating the state and county codes (the first two and the last two digits of the 6 digit CFDBS port code), with a fifth column appended at the end, i.e. `statecd||countycd||'0'`. Qualitative analyses³ of the gillnet fishery revealed that some ports within a county should not be grouped together due to different spatial fishing patterns (i.e. fishing in different statistical areas) by each port within the county. To capture these port specific spatial patterns within a state/county group, a county is subdivided when a statistical area boundary bisected the county. The 5th column of the port group code is utilized to indicate which counties had been sub-divided and which ports within the split county are grouped together. A zero in the 5th column indicate the county was not split; a value greater than 0 indicate the county was split, and indicate which ports belonged within the sub-county. Those port codes used to represent 'other county' ports, e.g. 'Other Barnstable', have been re-assigned a port code representing 'other state', e.g. 'Other MA', because it is not known which sub-division of the county 'Other Barnstable' should be assigned. Thus, the port is 'bumped up' to the state level. Each of the five counties listed below have their 'other county' port codes re-assigned to the corresponding 'other state' port code.

Five counties along the east coast identified as containing ports which may have different spatial patterns are:

- 1) Barnstable County in MA where areas 514, 521, and 538 trisect;
- 2) Suffolk County in NY (Long Island) where areas 611,612 and 613 trisect;
- 3) Washington County in RI where areas 538 and 539 bisect;
- 4) Ocean County in NJ where areas 612 and 614 bisect; and
- 5) Cape May County in NJ where area 614 and 621 bisect.

Barnstable County has two sub-divisions representing north and south of Cape Cod. Suffolk County has two sub-divisions representing east and west on Long Island's south coast (there was insufficient gillnet data to separate Suffolk County trips fishing in Long Island sound from those fishing south of Long Island; the lobster fishery was not examined).

Not all ports reported in the VTR data have a corresponding CFDBS port code, thus some of the detailed port information obtained in the VTR data can not be fully utilized. For example, Moriches is a port whose fleet fishes primarily in statistical area 613, however, this port does not have a unique CFDBS port code; therefore this port is assigned 'Other Suffolk' port, but, is re-assigned again to the state level because the county is a sub-divided county. Consequently, the trips from this port were grouped with trips from 'Other NY', which may have a broad range of statistical areas than the single port of Moriches. Expanding the list of CFDBS port codes could be useful to capture the spatial patterns of small ports within the allocation.

Throughout the time series, there are some Dealer trips reporting port as 'other state' indicating that the state is known but the specific port is not known. To accommodate these Dealer trips within the allocation, an additional VTR port group was formed by combining all VTR data for a

³ We thank M. Rossman for the plots depicting the gillnet fishery spatial fishing patterns by port which revealed port-specific fishing areas.

given state into one port group for that state.

Another issue relating to port groups involved the distinction between ‘port landed’ and ‘port sold’. In the VTR data, the port landed is the location where the fish were taken off the vessel. In the Dealer data, the port may or may not be the port where the fish are landed since fish product can be trucked and sold in other locations. The Dealer ports most effected are: Portland, Gloucester, and New Bedford where auction houses exists and attract fish product from surrounding ports. Port agents often know when Dealer transactions involve an ‘out-of-town’ vessel where the fish have been trucked and the port agent will either send the weighout slip to that port, or assign the appropriate port landed to that transaction (pers. comm. Scott McNamara, NER port agent in Portland, ME). Of course, it is unrealistic to expect the port agents to track all vessel transactions.

It may be possible to ascertain port sold from the VTR using the dealer permit number, however, limitations occur because not all VTRs report a dealer permit number, and some VTRs report two different dealers in two or more cities or states. Given these limitations and the fact that the dealer permit numbers in the VTR data are not audited, this aspect is not incorporated into the allocation scheme at this time.

For simplicity, the allocation scheme used port landed in the VTR as a corresponding element to port in the Dealer data. Recognition of a potential mis-match is acknowledged for two of the four matching allocation levels (Level C and D utilize port group in the stratification, as described in a subsequent section). The potential mis-match is expected to be minor as preliminary analyses indicated only a small portion of Dealer trips enter these levels. Additionally, it is expected that the number of Dealer trips affected by this potential mis-match will decline over the time series as more and more Dealer trips have a direct match with a VTR trip, reducing the need for fleet characteristics such as port group.

Main species groups: Dealer and VTR trips are assigned a main species group for each trip and gear group used on that trip. Main species groups are used as a surrogate for mesh size to further sub-divide the fleet definition of ton class, port, gear, and month. Twelve main species groups were formed with the intent to generally capture major sub-fisheries within a fleet. The species groups were defined in an ad-hoc fashion. An exploratory analysis of data grouped by the 12 species groups detected differences in spatial distribution patterns by main species. These analyses were conducted on longline, gillnet and otter trawl gears. There were four gear types (scallop dredge, lobster pot, shrimp trawl, scallop trawl) which were identified as gears having a single primary species associated with the gear, hence, these four gear types are assigned a single main species group. The species groups were formed based on the reported kept quantity in the VTR and the landed pounds in the Dealer data. To derive main species group, each species is assigned to a main species group, then the species weights are summed by main species group for each trip. The main species group is the group with the maximum species weight.

The main species group allows the category defined by ton class, gear group and port to be further sub-divided to capture species-specific spatial patterns. For example: Ton class 3 longline boats from Montauk were fishing in spatially distinct areas depending on which species

they were targeting (Statistical Area 537 for monkfish, Statistical Area 539 for tilefish⁴). Main species group and species group are synonymous.

Other caveats: Another issue related to the input data involves the VTR date fields. During the 1994 - 1996 data entry, date sailed was a required field for data entry and must be reported at the time the VTR was submitted; however, date landed and date sold were not required fields for submission (i.e. logbooks were accepted with this information missing). At data entry, if logbooks were missing date sold, then date landed was used. If date landed was missing, date sold was used and if date sold was missing then date sailed was used. No indicator or flag was used to identify these trips where a 'substitute' date was supplied at data entry, hence no way to identify actual days absent from estimated days absent [days absent is calculated in hundredths of days as (date landed - date sailed + 1), no time component was used in the 1994 allocation]. For those trips that do not report date landed or date sold, days absent will be underestimated; these trips will be incorrectly categorized as day trips. Given the uncertainty in the date landed and date sold, and the lax submission requirements for these fields, it was decided to use the date landed (the field most often reported) as the date to derive year, month and day for the VTR data. More importantly, however, the distinction between date sold and date landed is that date sold represents a transaction at the trip-species level (there may be multiple sold dates) while date landed is a trip-level variable. Date landed was selected because this is a trip-based allocation. It is recognized that fish from one trip may be sold on multiple days or fish from one trip sold on a day that is different from the date of landing. To account for these situations and to bridge this apparent disparity in date landed vs date sold, the VTR sold date was utilized to the extent possible given the aforementioned issues. The steps taken to utilize date sold and date landed are further described in a subsequent section.

Another data issue in the VTR was the unit of measure of the species quantity kept. Because of the uncertainty of the unit of measure of the species quantity kept, it was presumed that most species weights were in pounds, live weight; however, landed pounds may be have reported for some species. Species codes were reviewed and using Northeast Conversion Factors (established for CFDBS), species kept quantities that were reported in bags, trays, bushels, etc were converted to live pounds. Additionally, some of the reported quantities were questionable (ie. amounts large enough to sink a vessel); given this, species proportions are used in allocation. The VTR species weights are not used to replace the Dealer landings.

Given the data constraints and the goal of creating a consistent, comprehensive and compatible data set with commercial landings data prior to 1994, a multi-tier trip-based allocation scheme has been developed to match a Dealer trip to a corresponding VTR trip or a group of VTRs based upon fields which are in common to both data sets.

METHODS

Allocation Design

The multi-tier trip-based allocation scheme is designed to resemble, as best as possible, the

⁴ We thank Paul Nitschke for the tilefish fishery analyses.

methods utilized during the voluntary data collection system where port agents collected and/or estimated area fished and effort based upon knowledge of individual vessel and fleet behaviors. Prior to 1994, NMFS port agents would interview vessel captains to obtain area fished (a ten minute square location within a statistical area) and effort information such as days absent from port, number of hauls, haul duration, crew size, and the quantity and size of the gear used on a given trip. However, not every fishing trip was interviewed. For non-interviewed trips, the port agent used knowledge gained through prior interviews of the vessel and the fleet to assign a statistical area and estimate days fished (time the gear was actively fishing). For non-interviewed trips, the resolution of area fished was not as fine as for interviewed trips, and similarly, detailed effort information was not obtained; however, days fished and days absent were estimated. In the multi-tier trip-based allocation, the VTR trips are considered a sample of the commercial trips under the mandatory system, and thus provide the information previously collected during a port agent's 'interview', and accordingly, these VTR trips pooled into vessel and fleet groups form the informational base for the 'non-interviewed' trips.

Total commercial landings in the Dealer data are assumed to be known, but the spatial pattern of these landings is not known. The allocation determines an area fished for the landings based upon the spatial patterns observed in the VTR data. Dealer landings (pounds and value) are not altered during the allocation; area fished (statistical area) is added to the Dealer data record. Total effort in the Dealer data is not known; the allocation determines effort based upon the effort reported in the VTR data. The allocation is trip-based, hence a trip's area fished and effort will be associated with all the species landings from that trip. The allocation determines area fished first, and then effort is determined. The word 'determine' is used because in some cases area fished and effort information come directly from a VTR and in other cases, area fished and effort have been estimated based on a group for VTRs.

In this allocation, a trip is defined as a group of data records with the same year, month, day, and vessel permit in both the Dealer and VTR data sets. A split trip is defined as a trip which used either multiple gear types, multiple mesh sizes or fished in multiple statistical areas.

Allocation Levels

The VTR data are aggregated into groups containing VTR trips of similar characteristics. Four groups (Levels A, B, C and D) of increasingly aggregated VTR data are created, stored as Oracle tables, and used in the allocation. Two levels (Levels A and B) represent vessel-oriented data and two levels (Levels C and D) represent fleet-oriented data (Figure 3). Level A is comprised of audited VTR trips that have not been grouped. Level B is comprised of VTR trips from Level A that have been pooled by vessel permit, gear group, main species group, and month. Level C is comprised of VTR trips from Level A that have been pooled by ton class, port group, gear group, main species group, and calendar quarter. Level D is comprised of VTR trips from Level A that have been grouped by port group.

Every attempt was made to keep the allocation scheme at a monthly resolution; however, for the fleet-oriented data (Level C), it necessitated using quarter-year to ensure sufficient sample sizes within each stratification cell. Thus, Level A uses the month and day, Level B combined trips of the same vessel over the month, Level C combines trips with similar fleet characteristics over the

quarter, and Level D combined trips with similar port characteristics. Level D is intentionally a broad group to capture all trips that did not find a match at a previous level. A total of seven data sets were formed: one table for Level A containing area and effort information; and an area determination table and an effort determination table for Levels B, C and D (Figure 3).

If a Dealer trip has a corresponding VTR (i.e. a one-to-one match on vessel permit, month and day between the two data sets), then the area fished and the effort information, if present, is transferred directly onto the Dealer trip and the Dealer trip data record is complete. The Level A match corresponds to a pre-1994 port agent's interview. However, if a Dealer trip does not have a corresponding a VTR trip, then the Dealer trip is matched to a group of VTR trips that have similar trip characteristics. If a match occurs with a group of VTR trips (from Levels B, C or D), then a single area will be assigned to the Dealer trip on a probabilistic basis by resampling (with replacement) the distribution of VTR trips within the group. Days fished and days absent will be assigned to the Dealer trip based upon the median days fished per trip and median days absent, respectively, from trips within the pooled VTR data for that given area fished. The increasing levels (Levels B, C, D) of pooled VTR trips form an information base similar to the pre-1994 data collection system, where the port agent estimated area fished and effort for a non-interviewed trip based upon either previous vessel interviews and/or based on fleet patterns.

The allocation sequentially searches each of the four VTR data levels until a matching VTR trip (or group of trips) has been obtained for a Dealer trip (Figure 2). The first objective is to find an area fished; once area has been determined, then effort can be determined. Since area and effort are acquired sequentially in the allocation, area may be determined at one level, and effort may be determined at the same or higher level. For Dealer trips which do not find a match in one of the four levels, the area and/or effort fields are assigned the CFDBS default values (days fished and days absent are set to null).

Meta Fields

Two meta fields have been created and appended to the Dealer data to document which VTR data aggregation level was used to obtain the area and effort information. The area and effort meta fields are independent of each other. The meta fields can guide users to which Dealer data may be appropriate for certain analyses and which may not be appropriate. For example, catch per unit effort analyses would utilize data from Level A only (for which effort has not been estimated). The user will employ these meta fields in a similar fashion as they used the *interview_indicator* in pre-1994 data to discriminate between actual versus estimated area and effort data.

The meta field, *Alevel*, will have character values A, B, C, D indicating that area fished was obtained from either Level A, B, C, or D, respectively. A value of X indicates that the Dealer trip entered the allocation, but did not find a match at any of the four levels. Null indicates that the Dealer trip did not enter the allocation. All trips that enter the allocation are expected to find a match, Level D has been designed to capture all trips.

The meta field, *Elevel*, will have character values A, B, C, D indicating that effort was obtained from either Level A, B, C, or D, respectively. A value of X indicates the Dealer trip entered the

allocation, but did not find a match at any of the four levels. Null indicates that the Dealer trip did not enter the allocation.

Level A: Dealer trip matches a VTR trip

Dealer trips that matched at Level A are augmented with as much VTR information as available including: mesh size, depth, latitude, longitude, ten minute square, quarter degree square, statistical area, crew size, gear quantity, gear size, number of hauls, haul duration, days fished, days absent. For a Level A area match to occur, statistical area must be present on the VTR; all other area fields may or may not be present. If the VTR trip does not contain information on the aforementioned area fields; these fields will be assigned the CFDBS default value. For a Level A effort match to occur, days fished and days absent must be present in the VTR; all other effort fields may be null and will be assigned the CFDBS default value. At Level A, area fished and effort, if available, are assumed known from the VTR; no estimation is performed in the allocation for Level A matched trips. The VTR *tripid* and *gearid*⁵ is added to the Dealer data record for documentation purposes. The meta fields for area and effort are set to 'A'.

If the Dealer trip matched a VTR trip which does not contain days fished and day absent, then the VTR area information is used to augment the Dealer trip at Level A. The search for an effort match continues through Level B, C and D effort tables to determine days fished and days absent (Figure 2). The VTR *tripid* is added to the Dealer data record, the area meta field is set to 'A' and the effort meta field is set to the level where effort was determined.

A Dealer trip matched at Level A can result in a split trip when the matching VTR indicates a split trip (i.e., the VTR trip fished in multiple areas, or used multiple gears or mesh sizes). The Dealer trip landings (species by market category) and the value (dollar amount) associated with species landings are partitioned into subtrip components. The process of partitioning landings and price among subtrips is described in a subsequent section.

Levels B, C and D: Dealer trip matches a pool of VTR trips

Dealer trips that matched a pool of VTR trips at Level B, C or D are augmented with an estimate of area fished and an estimate of days fished and days absent. No VTR *tripid* and *gearid* are assigned to Dealer trips which match at Levels B, C or D. Fine scale area and effort information, such as latitude, longitude, quarter degree square, ten minute square, crew size, depth, mesh size, gear quantity, gear size, number of hauls, and haul duration are not estimated; these fields are assigned the CFDBS default value of null. No split trips will result from Levels B, C or D; a single area is estimated for the entire trip and days fished and days absent are determined on a 'per-trip' basis. Estimates of area fished and effort are described below.

Area probability distribution functions

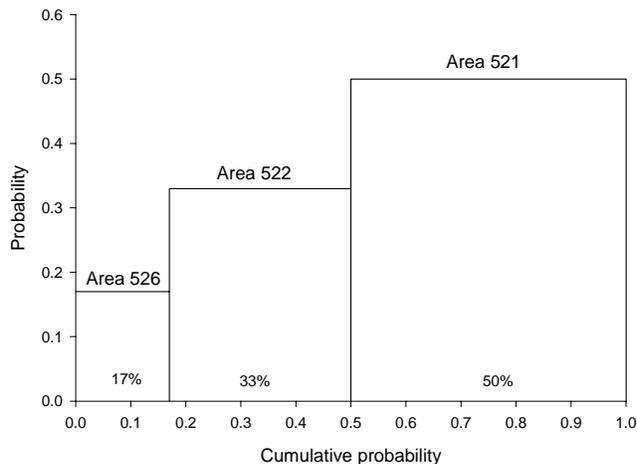
At Levels B, C and D, VTR trips are grouped based upon the stratification criteria for each level.

⁵ VTR *tripid* and *gearid* are computer-generated numbers that uniquely identify each logbook sheet based on permit, and date/time sailed.

For each level and stratification cell, a discrete probability distribution function is formed representing the proportion of trips which fished in each unique statistical area. A discrete cumulative distribution is then formed using the statistical area probabilities. The number of trips within each cell and the number of trips within each unique statistical area are also stored⁶.

Each Dealer trip is assigned a random number between 0 and 1 that has been generated using a large, odd number as the seed; this seed is stored in the software. When a Dealer trip matches a stratification cell, a single area fished (statistical area) is assigned to the Dealer trip on a probabilistic basis by resampling (with replacement) the distribution of statistical areas with the cell. The random number value is compared with each discrete cumulative probability (in ascending order) associated with an unique statistical area. When the random number is less than or equal to the cumulative probability value, the statistical area associated with the cumulative probability is assigned to the Dealer trip.

The following example is given for illustration:



For a given Level and stratification cell, there are 6 VTR trips in the cell which fished in 3 unique statistical areas. Three trips fished in Area 521, two trips fished in Area 522 and 1 trip fished in Area 526. For this cell, the probability of fishing in Area 521 is 0.50 (3 trips / 6 trips), 0.33 (2 trips / 6 trips) for Area 522, and 0.17 (1 trip / 6 trips) for Area 526. When ordered by ascending probability, the cumulative probabilities for the three areas 526, 522 and 521 are 0.17, 0.50, and 1.0, respectively. Each Dealer trip is randomly assigned a number between 0 and 1. This number is compared with the cumulative probabilities to determine a single area fished. In this example, if a given Dealer trip with a randomly assigned value of 0.75 matched this cell, the Dealer trip would be assigned Area 521. On average, for Dealer trips which match this Level and cell, 50% of matches would be assigned Area 521, 33% of the trips would be assigned Area 522, and 17% of the trips would be assigned Area 526. This example illustrates whole trips in each statistical area; however, this algorithm also works with a mix of whole and partial trips.

⁶ This information can later be used to calculate the multinomial probability to capture the uncertainty associated with statistical area landings determined at Levels B, C and D.

Effort estimation

Effort information is collected only in the VTR component of the mandatory data collection system; therefore, total effort is not known in the Dealer data. Effort can not be distributed in a similar fashion as the landings; each Dealer trip's effort must be estimated from the VTR data. Dealer trips acquire effort directly from a corresponding VTR trip at Level A or effort is estimated from a group of VTR possessing similar trip characteristics at Level B, C or D. Exploratory analyses indicate that days fished (DF) and days absent (DA) can be a function of statistical area (i.e. longer trips were made to statistical areas farther from home port). At Levels B, C and D, DF and DA are estimated using the median days fished and median days absent from the VTR trips within the stratification cell for a given statistical area. Also, the first and third quartiles are derived so the semi-interquartile range $[(Q3 - Q1) / 2]$ may be used as a measure of dispersion. In addition to the median, another measure of central tendency, $(Q1+Q3)/2$, may be also be computed from the quartiles. If needed, the interquartile range could be used as a diagnostic for homogeneity of effort within the cell during the input data set evaluation.

Days fished and days absent, although correlated, are independent of each other. Days fished measure the time (in tenths of days) the gear was actively fishing while days absent measure the time (in hundredths of days) the vessel was away from port. Only VTR trips that contain both DF and DA are used to create the data tables at Level B, C, and D.

To estimate the median value of a distribution of days fished per trip and days absent per trip, effort associated with split trips is multiplied by the inverse of *ntrips* (or the fraction of the trip which is associated to the subtrip) to convert effort from a partial trip basis to a 'per-trip' basis. This allowed all VTR trips and subtrips within a stratification cell to be combined into one distribution and the median value would represent effort on a 'per-trip' basis.

Probability density functions (PDF) to determine DF and DA (as used for area fished) were not appropriate because DF and DA may be correlated on a given trip (especially for mobile gear types) and using a separate PDF for each effort measure may result in a mis-match of the two due to the random nature. A joint PDF could have been used, but the objective was to keep the estimation of effort as simple as possible given all the data constraints. The median DF and median DA are selected as the simplest statistic of central tendency for various shaped distributions.

When a match occurs at Levels B, C or D, the meta fields for area (*Alevel*) and effort (*Elevel*) are set to the corresponding letter representing the Level. When no match occurs for effort, effort fields will be assigned their CFDBS default values (null). In addition to meta field *Elevel*, the effort indicator field used in pre-1994 data, *effind*, is assigned as given below:

Effort indicator	Criteria
4	<i>Alevel = A and Elevel = A</i>
3	<i>Alevel = A, B, C or D and Elevel = B, C, or D</i>
2	<i>Alevel = A, B, C or D and Elevel = X</i>

Allocation checks

Two diagnostic fields were created to monitor the matching of Dealer and VTR trips for area and effort. Each time the VTR data are used in a match, a counter is incremented. There is a counter for area and a counter for effort. The counters may be used to evaluate the frequency of cell usage in estimating area and effort. Evaluating how many times a given cell was used provides feedback on allocation, the Levels, and the stratification. The area and effort counters at Level A are carefully reviewed to ensure that only a VTR trip was used only once at this level.

Allocation assumptions

- Assume Dealer landings is a census of total landings;
- Vessels land only once per day;
- Each trip (permit-month-day) in the Dealer data set represents only one trip (consolidated trips are special cases and handled accordingly);
- VTR data are representative subset of the Dealer data.

It is recognized that a trip may sell its catch over several days; if this is not accounted for, then the number of trips will be over-estimated. We have addressed this issue to the extent possible when identifying unique Dealer trips. We have established a unique Dealer trip identifier, *dlrtripid*. This trip identifier links together all transactions that are associated with a trip.

Although landings data are collected in both dealer and vessel components of the mandatory reporting system: 'kept' pounds are recorded in the VTR and 'landed' pounds are recorded in the dealer report. It is assumed for the purposes of these analyses that the dealer data contain the most complete record of total landings, and that the VTR data are an unbiased subset of the commercial data set.

In 1994, an exploratory analysis revealed that there were potentially 74 trips that reported the same permit, month, day and had more than one time sailed. On 29 of the 74 trips, one of the multiple time sailed was '0000'; and on another 17 trips, the two time sailed values were within one hour of each other; thus over half the apparent two-trips-per-day trips had data errors. Thus, a potential 36 trips out of ~50,000 plus trips were incorrectly combined with another trip made by the same vessel.

We have decided to "ignore" multiple trips per day because we can not distinguish the trips that land multiple times per day from those that have mis-reported, i.e. do not fill out the logbook

correctly. We recognize that day boats may make multiple trips per day; in this case, the number of trips will be an underestimate. In the near future, when electronic VTRs are implemented and/or the VTR unique trip identifier is fully in place, this issue will diminish.

PREPARATION OF DEALER INPUT DATA

Consolidated Trips

The Dealer data contains consolidated trips; consolidate trips have landings from multiple trips for the same vessel. Consolidated Dealer trips have day = '00'. These trips do not enter the area matching phase at Level A; these trips are only matched at Level B, C or D to estimate a single area. To assign an area based on vessel or fleet characteristics seemed more appropriate than to assign a single VTR to multiple trips. Consolidated trips do not enter the effort matching phase; effort (days fished and days absent) are assigned the CFDBS default value. Since a consolidated trip represents an unknown number of trips, it is inappropriate to apply the estimated effort that represents one trip. The effort fields of consolidated Dealer trips are assigned the CFDBS default value (null). To estimate the number of trips the consolidated Dealer trip represents, the number of unique month||docn is used as a surrogate for the number of unique trips.

It should be noted that records within the Dealer data represent 'transactions' for which single or multiple transactions may comprise a trip. Multiple transactions may occur on the same day or over several days. Since the allocation is 'trip-based', it is necessary to identify all transactions that are associated with a given trip. All transactions for a given trip are assigned a unique Dealer trip identifier.

Dealer Trip Identifier

Three types of multiple Dealer transactions exist: 1) multiple transactions the same day from one trip; 2) multiple transactions on different days from one trip; and 3) multiple transactions on the same day and multiple transactions on different days from one trip.

Type 1 transactions can be identified using only the Dealer data based on the permit, month and day; all transactions for a trip will be tagged with the same *dltrtpid*.

Type 2 and Type 3 transactions can not be identified solely with Dealer data, VTR data are needed to identify these transactions.

Type 1 min sold = max sold = date landed	Trips that sold on a single date, the landed date
Type 2 (min sold = max sold) <> date landed	Trips that sold on a single day but not the landed date
Type 3 (min sold <> max sold) = or <> date landed	Trips that sold on multiple days

In the allocation, the Dealer uses *month* and *day* (representing sold date) while the VTR uses *month* and *day* based on date of landing because this is a trip-level data element and this is a trip-based allocation. It is recognized that fish from one trip may be sold on multiple days or fish from one trip sold on a day that is different from the date of landing. To account for these situations and to bridge this apparent disparity in date landed vs date sold, the following steps have been developed to: 1) maintain the original date in the Dealer data, and 2) use the VTR landing date for matching purposes only for multi-day (trip-boats) trips which sold on multiple days or the sold date is different from the landed date.

A set of VTR trips are identified that have days absent greater than 1 day (trip-boat trips) and sold on multiple days or sold on a day different from date landed. This set does not include the following: 1) trips with erroneous maximum sold dates (if *maxsold - dateInd1* > 10 then delete); 2) trips with more than 3 different sold dates (there are many VTRs where fishermen had reported many trips on one log sheet); 3) trips where minimum date sold is less than the date landed; and 4) overlapping VTR trips with overlapping dates for the same permit. This set also excludes day trips as it is unlikely that they would sell over multiple days if they were a day-boat, with the exception of fishermen who pound/carr (i.e., hold in a cage).

Dealer transactions and the subset of VTR trips (Type 2 and 3 above) are merged based on permit and where the Dealer date landed is between the VTR minimum date sold and maximum date sold. All Dealer transactions associated with the VTR trip from the subset will be assigned the same *dlrtrpid* and the *month* and *day* based on the VTR date land. The original Dealer month and day values are stored and will be used to re-populate the month and day fields after the allocation.

A Dealer trip identifier, *dlrtrpid*⁷, will be assigned to all transactions in the Dealer data and is defined as the concatenation of month and document number, *month||docn*. A trip is defined as a unique *permit-month-day*. Single transaction trips will have the *dlrtrpid* equal to *month||docn*. A trip with multiple transactions will have one *dlrtrpid* for all transactions, and the *dlrtrpid* will be based on the *month||docn* of the transaction with the most landed pounds.

⁷ *dlrtrpid* is similar to *link* in CFDETT/S (pre-1994 data), however, differs when multiple transactions for a given trip occur.

Non-allocated and allocated Dealer data

As noted earlier, not all Dealer trips will enter the allocation; the Dealer data are partitioned into two data sets: allocated and non-allocated.

Dealer data entering the allocation have the following criteria: 1) *source* = '07' (i.e. mandatory reporting); 2) month between 01 and 12 (except for 1994, where only data between April – December only); 3) vessel with a unique permit represent a single vessel.

Dealer data not entering the allocation have the following criteria: 1) Dealer data with *source* != '07' including non-mandatory reporting data such as state data, bluefin tuna and other highly migratory species data, surfclam and ocean quahog fishery trips, landings data from Foreign Sea Sampling and the Potomac River Fish Commission (as identified by the letters E, Q to Z in the document number); 2) Dealer data between January to March 1994 (data prior to mandatory reporting) or data with month = '00'; 3) vessels with non-unique permits such as permits in (000000, 190998, 390998); 4) data where gear is in (400, 040, 115, 999) or *ngear2* in = (03, 17), as these gear type have other data reporting systems other than the VTR system.

Trips with transactions that fall in both the non-allocated and allocated sets are identified and all transactions associated with these trips are relocated to the non-allocated set. All transactions associated with a trip are either in the allocated set or the non-allocated set.

PREPARATION OF VESSEL TRIP REPORT DATA

Not all VTR data are used in the allocation. Vessel trip reports from the party and charter industry will not be used to allocate commercial trips. Only commercial trips (*tripcatg* = 1) that fished and reported an area fished (either a statistical area or a latitude and longitude or *Loran* from which a statistical area could be derived) will enter the allocation. Although there were several different audit procedures conducted on the 1994 data (Regional Office audits and the NEFSC side-by-side audits) the data needed more auditing before the data could be used in the allocation. Thus, a suite of programs were developed to further audit the VTR data. The allocation audits do not screen every field in the VTR data, only a limited number of fields pertaining to the allocation itself (especially the fields used to match the dealer data with the VTR data) were considered. Other fields were corrected on an 'ad-hoc basis', i.e. if an error was discovered, then it was corrected; however, a thorough screening of all fields (such as crew size and depth) was not undertaken at this time. Improperly submitted VTRs or trips with no catch were excluded from the allocation.

All records with area recorded as 551, 552 (Canadian territory) were changed to 561 or 562 after verifying the ten minute square was on the Hague Line between Divisions 55 and 56. All records with area recorded as 523 & 524 were re-assigned to 561 & 562 respectively (Areas 561 and 562 were formerly 523 and 524 before the USA and Canadian boundary line was established). Area is a required field for a VTR to be used in the allocation scheme. Any records for which area was unresolved were not used.

VTR trips that reported the following statistical areas were not used in the allocation: 110, 100, 500, 510, 520, 528, 530, 540, 550, 551, 552, 560, 600, 610, 620, 630, 799, 800, 899. This list comprises statistical areas that are 1) beyond the range of the fishing activity in the Northeast region and/or 2) represents 'generic' statistical area that represent a group of statistical areas. For example, 510 represents the collection of statistical areas from 511 to 515, a group of 5 statistical areas.

A trip should be split into subtrips if either area, gear or mesh changed during the trip. A trip is defined a group of VTR records with the same year, month, day, and permit. Year, month and day are based on date landed. A subtrip is an integer assigned to a record or group of records which comprise part of a trip. This number starts with one and is incremented when the gear, area or mesh size changes within a given permit, year, month, day. The number of subtrips, *nsubtrip* is an integer value indicating the number of subtrips for a given trip, $nsubtrips = \max(\text{subtrip})$.

To identify trips which may be artificially split when mesh is either not reported on one of the subtrips, reported incorrectly or entered incorrectly at data entry, additional screening was conducted for mesh.

Other data preparation of the VTR data included: 1) converting species weights, based on species codes, to pounds when quantity kept was reported in bushels, trays, gallons, barrels, etc. ; and 2) limited auditing of days fished and days absent was performed to identify and remove outliers (unrealistic values).

VTR data set

VTR trips with the following criteria are used in the allocation: 1) statistical area (derived from *cnemarea*) is not null or 0; 2) *tripcatg* = 1 (omit charter and party trips); 3) *not_fished* = 0 or is null (omit trips which did not fish); and 4) vessels that landed in ME, NH, MA, RI, CT, NY, NJ, DE, MD, VA, NC (omit NC for 1994-1996 because NC landings were not included in the Dealer landings for these years);

VTR trips excluded from the allocation include:

1) trips where statistical area could not be derived; 2) trips for which date landed was less than date sailed; 3) trips from vessels with permits in the 800 series representing NY state vessels that are not federally permitted vessels; 4) trips that used the following gear codes (drc, llp, hrp, ptm, gnd, mix, oth, null) and lobster pot gear fishing in zones 0 ,1 or null⁸; 5) trips with more than one subtrip that have any excluded subtrip; and 6) trips with more than one subtrip for which one or more of the subtrips has no effort⁹.

⁸ inshore lobster pot gear will not be included in the allocation because a) it is expected that inshore fisherman would not have a federal permit to fish in federal waters, b) gear code 'PTL' does not distinguish between offshore and inshore pots (200 and 210)

⁹ if the trip is split, days fished must be present on all subtrips of the trip to calculate the *ntrip* for each subtrip; if *ntrip* can not be determined, then the trip can not be used, not even for area, because the *ntrip* on the subtrip is utilized in the area probability density function.

Seven VTR data sets are created. A base data set is created containing all useable individual VTRs trips, this forms Level A. From this data set, six additional data sets are created, an area data set and effort data set for Levels B, C, and D described below.

Level B, C and D Area sets

Level B area data set contains VTR trips from Level A grouped into cells stratified by **permit, month, gear group, and species group**. Trips for which the species group is null are excluded.

Level C area data set contains VTR trips from Level A grouped into cells stratified by **ton class, quarter, port group, gear group, species group**. Trips for which species group is null are excluded as well as trips that fished in the Grand Banks statistical areas in (330, 340, 350).

For Level D area data set contains VTR trips from Level A grouped into cells stratified by **port group**. Trips fishing on the Grand Banks (area 330, 340, 350) were excluded.

The unique statistical areas fished by these trips are determined within each Level and cell, and their associated probability and cumulative probability are calculated. These data sets contains the following variables for each cell: a) probability; b) cumulative probability; c) number of trips (*ntrips*) in the cell; d) given an area, the number of trips in the cell; e) count of the number of trips or subtrips which formed this cell; f) area: an unique statistical area within the cell; g) a counter of dealer matches within the cell.

This information can later be used to calculate multinomial probability to capture the uncertainty associated with statistical area landings at Levels B, C and D.

Level B, C and D Effort sets

For Level B effort data, the VTR trips in Level A are grouped into cells stratified by **permit, month, gear_group, species group, and area**. Trips for which species group is null are excluded as well as trips for which days fished or days absent or null.

For Level C effort data, the VTR trips are grouped into cells stratified by **ton class, quarter, port group, gear group, species group and area**. Trips for which species group is null are excluded as well as trips that fished in the Grand Banks statistical areas in (330, 340, 350) and trips with no effort.

For Level D Effort data set, the VTR trips from Level A are grouped into cells stratified by **port group and area**. Trips fishing on the Grand Banks (statistical areas 330, 340, 350) were excluded as well as trips with no effort.

IX. MATCHING: Creation of Header records and Speices detailed records

The Dealer data are comprised of 1) 'header' records that contain trip landings, trip value and effort for each trip/subtrip, and 2) detailed species records that contain species, market category

(grade), weight and value for each species-market category and trip/subtrip.

Header records

When a Dealer trip matches a VTR trip that has subtrips (Level A only), multiple headers will be created. The match will return the number of headers records equivalent to the number of subtrips. The additional header records will have the same Dealer trip identifier, and the subtrips will be sequential; this information will come directly from the VTR trip. The landings and value for the trip are partitioned among the additional headers based on species area proportions observed in the VTR data if available or based on the effort (*ntrip*).

If a Dealer trip has multiple transactions and the Dealer trip matches a VTR (split or non-split), then effort (*ntrips*, days fished and days absent) must be partitioned evenly among the subtrips. In all other circumstances, *ntrip* is used as the basis to partition effort from the VTR. For non-split trips, landings and value will remain the same; for split trips, the landings and value will be partitioned among the subtrips based on *ntrip*. Details on the additional headers created and how effort (*ntrips*, *df* and *da*) is partitioned among the subtrips are given in Appendix Tables 1 and 2 for single and multi-transaction trips, respectively.

Species detailed records

For non-split trips, statistical area will be transferred onto the each Dealer species records for a given matched trip. For split trips, Dealer species-market category landings and value will be distributed to the subtrips based on a species-specific area proportion¹⁰ derived from the VTR, if species information is available; otherwise effort¹¹ (*ntrips*) from the VTR trip is used to partition landings and value when species information is not present.

Combined Dealer data sets

The non-mandatory data and allocated data are combined into one data set. An audit is run of the entire data set using a modified version of a master audit.

Due to poor VTR logbook instructions, gear size and gear quantity for some fixed gear (lobster pots and crab pots) must be nulled out when *Alevel* = 'A' and *source* = '07'. Additionally, gear size and gear quantity are set to null when *negear* in (200, 210, 300) and *Alevel* = 'A' and *source* = '07'.

Master Oracle tables, CFDETSyyyyyAA CFDETTyyyyAA are created for use (Appendix Figure

¹⁰ It is assumed that the VTR species pounds are reported by subtrip accurately. Using proportions guards against some of the reporting difficulties encountered.

¹¹ This assumes that the probability of catching this species is based on the amount of time the vessel fished in an area (and not based on other species catch amounts). It assumes that some of the catch may be mis-assigned to an area, however this is less 'evil' than picking a single area and wrongly assigning the species catch to one incorrect area.

1). Biological samples stored in CFLENyyyy and CFAGEyyyy were updated with the allocated area assigned to the trips from which the sample was taken. If a biological sample was taken from a split-trip, the sample was assigned area = '000'. This prevents the mis-use of these samples for multi-stock species while allowing their use for single stock species.

COMPARISON OF ALLOCATION INPUT DATA SETS

The allocation assumes the VTR data are a representative subset of the mandatory reporting Dealer data. The VTR and Dealer data are compared to identify any potential bias in the VTR data which may exist due to reporting compliance. The comparisons were performed at the same level of resolution at which the allocation would be conducted, i.e. month, quarter, port group, gear group, ton class, and species group. Annual comparisons were qualitatively evaluated based upon the percent distribution of trips in the VTR and Dealer sets by the stratification variables: month, quarter, state, port group, ton class, gear group, and species group. An illustrative example, using data from 2000 was selected to display the percent distributions for each stratification variable (Figures 4a to 4e). To summarize the percent distributions for all years, and categories within each stratification variable, the differences between Dealer percentage and VTR percentage were calculated and plotted (Figures 4a - 4e).

The differences in percentages between the Dealer and VTR general range between +/- 5 percent for most years except 2000, where gear groups and species group differences range between +/- 10 percent. In general, there is close agreement between the overall percent distributions of the VTR and Dealer data, indicating that the VTR input data general reflects the Dealer data, with the possible exception of data for ton class 1 (under tonnage) vessels caused by the exclusion of these vessel in the Dealer data.

Summary of the number of VTR trips and trips with subtrips are given in Table 1. As shown in Palmer and Wigley (in review), some VTR multiple-subtrip trips underreport the number of statistical area fished, resulting in fewer split trips than expected. However, as Palmer and Wigley (in review) show, this does not have serious implications for the overall use of the allocation procedure.

ALLOCATED DATA

Matching results

Summary statistics of the number of metric tons landed in the Dealer data, the proportion of landings that entered the allocation and the proportion of landings that matched at each allocation level for area and effort are given in Table 2 for 1994 to 2003. In the allocation, there is a very small amount (< 1 mt) of increased landings that result from the rounding of species pounds in trips that have subtrips. The proportion of Dealer landings entering the allocation range between 19% and 32%. Between 51% and 73% of the landings that enter the allocation to find area fished match at Level A. The percent of total landings subject to the random component of the allocation ranged between 9% and 12%.

Annual species landings and percent landings, by non-allocated and allocation level for area fished, are given in Tables 3 and 4 and Figure 5. The species summarized here are the eight species with multi-stock components and as well as all species combined. Except for 1994, generally more than 90% of the total species landings entered the allocation for all but two of multi-stock species. Since first quarter of 1994 did not enter the allocation, it is expected to have a lower percentage of total landings entering the allocation. Over all years, both red hake and silver hake have a higher percent on non-allocated landings than the other multi-stock species.

Biological Samples

Biological samples (lengths and age structures) taken for species landed by trips that entered the allocation will acquire the allocated trip area. For samples taken from split trips, the area is assigned '000'. If a trip did not enter the allocation, the original area, if present, remained on the sample.

There are some samples for which statistical areas changed. There are various reasons for the changes, including: the sample did not have an area and acquired an area via the allocation, or the statistical area changed due to internal consistency checks performed on the VTR after the VTR was used to assign statistical area to the sample (this represents about 2% of the samples with areas that changed); the allocated Dealer trip matched at Level B, C or D and an estimated area was obtained for the sample.

A summary comparison of the original area and the allocated area for the samples collected during 1994 to 2003 is given in Tables 5 and 6 for lengths and ages, respectively. Comparison of original area and allocated area for individual species and stocks are given in Wigley et al. (in review).

Evaluation of random component (1,000 realizations)

To evaluate the random component of the allocation, the 1994 Dealer data was run through the allocation procedure 1,000 times, each time using a different seed to generate a series of random numbers that were assigned to each Dealer trip. The 1994 Dealer data were selected because it had expected to have the largest proportion (49% of allocated data) of Dealer landings that matched at Levels B, C and D where area was assigned on a probabilistic basis.

There are eight species in the Northeast with multiple stock components: cod, haddock, yellowtail flounder, winter flounder, windowpane, monkfish, red hake and silver hake (Table 7, Figure 1). Stock area landings were summed for each species and stock from each of the 1,000 runs. A frequency distribution and the 80% CI were calculated (Figure 6). The range of stock landings by species varied: haddock had the smallest range (5 mt) of stock landings while silver hake had the largest range (325 mt). The percent spread, calculated as the range / mean, varied between 1.2% (GB COD) and 32.1% (SNE YT). For most species, the 1994 point estimate from the base run was within the 80% CI of the 1,000 realizations, with the exception of silver hake. For both windowpane flounder and winter flounder, the 1994 point estimate was at the boundary of the confidence interval. For most species, the random component did not contribute

to wide spread in stock landings, indicating that the random component is not a large source of stock landings variability.

Conclusions/Discussion

- There is a high percent of landings that match at the vessel level (Levels A and B) to obtain area fished and a small percentage of landings where area fished is estimated from fleet patterns. This is an improvement over the pre-1994 landings data where approximately less than 10% of the trips were interviewed by port agents.
- There are trade-offs between using a trip-based allocation procedure and other methods, however, the need to link biological samples to individual trips to obtain area for the sample was a necessary element.
- An evaluation of the random component of the allocation indicated that the random component did not contribute to wide spread in stock landings, indicating that the random component is not a large source of stock landings variability.
- The allocation is predicated upon using clean VTR data. To the extent possible, VTR data was audited for use in the allocation. Continued efforts to expand the routine auditing of VTR as soon as logbooks are submit are encouraged to improve data quality and accuracy.
- In the future, the allocation can be expanded to include vessels that make multiple trips per day as well as the incorporation of the unique trip identifier established in 2004 to link the Dealer and VTR databases. It is premature to use this identifier until further data quality and accuracy procedures are established.
- Due to implementation of Dealer Electronic Reporting in May 2004, further evaluation of 2004 to 2006 allocated data will be needed.
- Examination of effort (number of trips, days fished, day absents) in the 1994 allocated data compared favorably with 1993; however, examination of effort over the entire allocated time series is needed.
- Evaluation of the allocated data at the stock level is provided in Wigley et al., in review.

Acknowledgments

We wish to thank R. Mayo for his guidance throughout this project; M. Rossman and P. Nitschke for their analyses of VTR data in support of allocation stratification; M. Palmer for his analysis of the impacts of the assumptions; Data Management Systems staff, including Barbara North, Heidi Marrota and David Hiltz for their technical assistance and we thank our reviewers for their constructive comments.

Literature Cited

- Bisack, K. 2003. Estimates of marine mammal bycatch in the Northeast (New England) multispecies sink gillnet fishery in 1996. *Northeast Fish. Sci. Cent. Ref. Doc.* 03-18; 21 p. Available on-line:
<http://www.nefsc.noaa.gov/nefsc/publications/crd/crd0318/crd0318.pdf>
- NEFSC (Northeast Fisheries Science Center). 1996. Report of the 22nd Northeast Regional Stock Assessment Workshop (22nd SAW): Public Review Workshop. *Northeast Fish. Sci. Cent. Ref. Doc.* 96-16; 45 p.
- Palmer M.C. and S.E.Wigley (in review). Validating the stock apportionment of commercial fisheries landings using positional data from Vessel Monitoring System (VMS). *Northeast Fish. Sci. Cent. Ref. Doc.* 07-xx. (also a GARM Data Review Meeting working paper).
- Wigley, S. M. Terceiro, A. DeLong, and K. Sosebee. 1998. Proration of the 1994-96 USA Commercial Landings of Atlantic cod, Haddock and Yellowtail Flounder to Unit Stock Areas. *Northeast Fish. Sci. Cent. Ref. Doc.* 98-02; 32 p.
- Wigley, S.E., C. Legault, E. Brooks, S. Cadrin, L. Col, L. Hendrickson, R. Mayo, P. Nitschke, M. Palmer, K. Sosebee and M. Terceiro. GARM working paper in review. Annual comparisons of the trip-based allocated and single-species prorated commercial landings, biological samples and numbers of landed fish at age. 67 p.

Table 1. Summary of the number of VTR subtrips used in the allocation, by year and subtrip;

VTR Subtrips	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
1	51,488	88,198	91,297	88,525	90,502	84,488	86,747	89,617	84,190	79,193	81,226	86,639	86,478
2	626	1,040	1,064	790	789	723	1,019	1,373	1,553	1,445	1,601	1,870	1,721
3	70	68	103	88	81	93	120	107	92	136	127	159	154
4	15	10	7	14	19	15	16	16	16	17	22	11	19
5	3	1	2	5	1		5	4	3	5	1		3
6	1	1		1			1				1		
7													
Total trips	52,203	89,318	92,473	89,423	91,392	85,319	87,908	91,117	85,854	80,796	82,978	88,679	88,375
Trips with subtrips	715	1,120	1,176	898	890	831	1,161	1,500	1,664	1,603	1,752	2,040	1,897
	1.4%	1.3%	1.3%	1.0%	1.0%	1.0%	1.3%	1.6%	1.9%	2.0%	2.1%	2.3%	2.1%

Table 2. Summary statistics of allocation procedure for commercial landings data (Dealer data) by year. *Note: 2004 to 2006 data are being processed.*

	1994			1995			1996		
	mt	% mt	% allocated	mt	% mt	% allocated	mt	% mt	% allocated
Dealer Data									
before Allocation	654527.7			757964.9			726312.1		
after Allocation	654527.8			757965.3			726312.5		
Difference	0.1			0.4			0.4		
Non-Allocated data	532571.5	81.4%		572974.6	75.6%		521473.5	71.8%	
Allocated data	121956.3	18.6%		184990.7	24.4%		204839.0	28.2%	
AREA match									
Level A	62236.2	9.5%	51.0%	106473.0	14.0%	57.6%	136592.4	18.8%	66.7%
Level B	18724.1	2.9%	15.4%	39078.9	5.2%	21.1%	36777.9	5.1%	18.0%
Level C	30501.5	4.7%	25.0%	31730.3	4.2%	17.2%	25731.8	3.5%	12.6%
Level D	10494.5	1.6%	8.6%	7708.5	1.0%	4.2%	5736.9	0.8%	2.8%
Allocated data	121956.3	18.6%	100.0%	184990.7	24.4%	100.0%	204839.0	28.2%	100.0%
subject to random component		9.1%	49.0%		10.4%	42.4%		9.4%	33.3%
EFFORT match									
Level A	56430.3		46.3%	91930.4		49.7%	120880.2		59.0%
Level B	18558.4		15.2%	35391.5		19.1%	34354.4		16.8%
Level C	16153.2		13.2%	21887.3		11.8%	19491.3		9.5%
Level D	20682.6		17.0%	28721.8		15.5%	25429.4		12.4%
Level X	10131.8		8.3%	7059.7		3.8%	4683.7		2.3%
Allocated data	121956.3		100.0%	184990.7		100.0%	204839.0		100.0%

Table 2 continued.

Dealer Data	1997			1998			1999		
	% allocated			% allocated			% allocated		
	mt	% mt	mt	mt	% mt	mt	mt	% mt	mt
before Allocation	709673.1			696894.9			615488.35		
after Allocation	709673.4			696895.2			615488.57		
Difference	0.3			0.285			0.218		
Non-Allocated data	509288.4	71.8%		471243.6	67.6%		447878.63	72.8%	
Allocated data	200385.0	28.2%		225651.5	32.4%		167609.94	27.2%	
AREA match									
Level A	129144.1	18.2%	64.4%	139152	20.0%	61.7%	112428.62	18.3%	67.1%
Level B	42440.9	6.0%	21.2%	47216.77	6.8%	20.9%	33883.188	5.5%	20.2%
Level C	24131.1	3.4%	12.0%	29541.56	4.2%	13.1%	16834.022	2.7%	10.0%
Level D	4668.9	0.7%	2.3%	9741.213	1.4%	4.3%	4464.1119	0.7%	2.7%
Allocated data	200385.0	28.2%	100.0%	225651.5	32.4%	100.0%	167609.9	27.2%	100.0%
subject to random component		10.0%	35.6%		12.4%	38.3%		9.0%	32.9%
EFFORT match									
Level A	112856.2		56.3%	122574.9		54.3%	98711.903		58.9%
Level B	40169.2		20.0%	46795.95		20.7%	33034.842		19.7%
Level C	13345.5		6.7%	15030.53		6.7%	11006.804		6.6%
Level D	27615.0		13.8%	35774.2		15.9%	21830.79		13.0%
Level X	6399.1		3.2%	5475.922		2.4%	3025.5976		1.8%
Allocated data	200385.0		100.0%	225651.5		100.0%	167609.9		100.0%

Table 2 continued.

Dealer Data	2000			2001			2002		
	% allocated			% allocated			% allocated		
	mt	% mt	mt	mt	% mt	mt	mt	% mt	mt
before Allocation	601354.2			672812			596941.3		
after Allocation	601354.6			672812.7			596942.1		
Difference	0.418			0.714			0.815		
Non-Allocated data	440934.6	73.3%		509009.7	75.7%		445862.6	74.7%	
Allocated data	160420.0	26.7%		163803	24.3%		151079.5	25.3%	
AREA match									
Level A	109866.3	18.3%	68.5%	118937.6	17.7%	72.6%	106470.2	17.8%	70.5%
Level B	30432.7	5.1%	19.0%	30194.01	4.5%	18.4%	29765.88	5.0%	19.7%
Level C	14421.1	2.4%	9.0%	13223.88	2.0%	8.1%	12701.14	2.1%	8.4%
Level D	5699.9	0.9%	3.6%	1447.471	0.2%	0.9%	2142.28	0.4%	1.4%
Allocated data	160420.0	26.7%	100.0%	163803.0		100.0%	151079.5	25.3%	100.0%
subject to random component		8.4%	31.5%		6.7%	27.4%		7.5%	29.5%
EFFORT match									
Level A	105015.7		65.5%	117555.3		71.8%	104882.1		69.4%
Level B	32140.8		20.0%	30706.99		18.7%	29810.83		19.7%
Level C	8099.2		5.0%	6077.016		3.7%	6529.462		4.3%
Level D	14375.3		9.0%	9072.804		5.5%	8141.667		5.4%
Level X	789.0		0.5%	390.846		0.2%	1715.504		1.1%
Allocated data	160420.0		100.0%	163803.0		100.0%	151079.5		100.0%

Table 2 continued.

2003			
Dealer Data	mt	% allocated	
		% mt	mt
before Allocation	622149.9		
after Allocation	622150.8		
Difference	0.925		
Non-Allocated data	448153.6	72.0%	
Allocated data	173997.2	28.0%	
AREA match			
Level A	108891.5	17.5%	62.6%
Level B	33895.08	5.4%	19.5%
Level C	20018.26	3.2%	11.5%
Level D	11192.32	1.8%	6.4%
Allocated data	173997.2	28.0%	100.0%
subject to random component		10.5%	37.4%
EFFORT match			
Level A	107433.1		61.7%
Level B	34017.51		19.6%
Level C	6331.127		3.6%
Level D	23666.68		13.6%
Level X	2548.746		1.5%
Allocated data	173997.2		100.0%

Table 3. Summary of landings (live, mt) for selected species and all species (area fished).

Monkfish	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Allocated Level A	3088.1	6523.1	6560.6	6723.3	6799.078	7802.8	7837.3	8624.4	8912.4	9228.7
Allocated Level B	1238.2	2485.5	2214.8	2010.7	1704.032	2606.1	2183.1	1981.9	2127.9	2468.3
Allocated Level C	1306.1	1683.5	1253.3	1387.8	1426.553	1402.6	1363.3	1246.1	1215.7	1025.3
Allocated Level D	390.9	495.8	600.6	334.7	249.6541	128.6	76.4	73.3	91.2	112.3
Non-allocated	3716.5	904.3	945.0	1309.8	1206.169	816.8	757.8	638.7	582.9	792.7
Total	9739.8	12092.2	11574.2	11766.3	11385.5	12757.0	12217.9	12564.5	12930.1	13627.4
Cod										
Allocated Level A	4896.6	7937.5	8676.4	7916.7	6827.4	6152.3	6915.1	9608.0	7920.5	6553.1
Allocated Level B	1692.6	2268.3	2223.0	1920.5	1567.6	1386.6	1775.6	2168.2	1976.2	1478.9
Allocated Level C	1429.2	1060.2	963.4	936.0	814.2	531.1	763.9	737.0	862.3	739.1
Allocated Level D	52.9	14.9	13.6	35.2	64.4	44.3	27.6	32.5	23.8	14.4
Non-allocated	6955.2	200.9	154.4	129.7	124.2	77.0	141.3	269.4	342.3	317.3
Total	15026.5	11481.8	12030.8	10938.0	9397.8	8191.2	9623.5	12815.1	11125.2	9102.9
Winter flounder										
Allocated Level A	1608.8	2558.1	3087.4	3353.6	3241.9	3068.0	4029.6	4764.3	3720.3	4164.3
Allocated Level B	632.5	915.7	1002.6	1195.8	1008.6	1046.7	1195.9	1456.2	1473.2	1031.6
Allocated Level C	408.0	303.1	417.3	496.2	589.5	313.1	340.8	383.3	432.2	472.2
Allocated Level D	12.1	4.9	10.2	6.9	3.9	3.0	17.8	34.7	12.8	17.0
Non-allocated	941.6	247.4	242.6	290.9	245.1	206.3	258.4	291.9	239.0	206.9
Total	3603.0	4029.3	4760.1	5343.4	5089.0	4637.1	5842.6	6930.4	5877.5	5891.9
Yellowtail flounder										
Allocated Level A	1544.5	1272.4	1572.2	1919.8	2366.0	3052.6	4495.8	4814.6	3202.5	3612.1
Allocated Level B	571.5	478.5	565.9	622.7	871.3	1001.9	1970.9	1799.8	1491.6	1494.6
Allocated Level C	350.6	140.3	183.4	236.1	283.8	303.5	339.3	456.8	397.6	415.2
Allocated Level D	11.5	1.5	4.4	6.9	3.8	3.8	9.7	5.2	2.2	2.9
Non-allocated	618.7	35.3	70.7	86.0	94.8	66.2	118.7	212.9	231.3	40.8
Total	3096.8	1927.9	2396.6	2871.5	3619.6	4428.0	6934.3	7289.3	5325.2	5565.6
Windowpane										
Allocated Level A	221.4	431.4	666.7	366.7	396.3	118.1	189.2	127.1	65.4	40.7
Allocated Level B	108.9	247.4	212.8	92.2	78.1	35.2	52.2	36.1	20.5	15.5
Allocated Level C	38.2	95.2	73.3	54.0	39.4	7.1	13.7	5.9	3.9	3.5
Allocated Level D	0.5	0.2	0.4	0.3	0.3	0.8	1.4	0.4	0.1	0.5
Non-allocated	155.9	13.2	10.9	19.0	5.7	5.0	15.7	7.7	7.6	3.5
Total	524.9	787.4	964.0	532.2	519.7	166.2	272.2	177.2	97.5	63.8
Haddock										
Allocated Level A	121.4	249.6	359.3	936.5	1777.6	2035.6	2480.6	3775.3	4577.3	4299.2
Allocated Level B	39.8	68.1	93.9	272.2	485.6	587.5	835.0	1119.7	1621.0	1185.5
Allocated Level C	27.1	38.4	46.0	107.7	209.5	112.2	186.7	168.0	362.3	433.0
Allocated Level D	0.1	1.1	1.6	1.3	3.1	0.9	0.6	0.7	1.2	6.5
Non-allocated	99.9	2.2	2.9	1.4	13.3	21.2	7.6	47.5	54.2	28.2
Total	288.3	359.3	503.7	1319.2	2489.0	2757.3	3510.6	5111.1	6616.0	5952.4
Red Hake										
Allocated Level A	742.5	714.2	650.2	766.9	836.3	993.4	1021.5	1055.0	511.8	387.7
Allocated Level B	144.4	221.3	235.0	246.2	281.3	317.1	330.6	354.9	178.3	164.3
Allocated Level C	262.8	126.2	33.0	71.5	80.0	71.2	53.1	61.3	38.5	60.7
Allocated Level D	9.7	2.1	3.0	4.5	2.4	2.5	7.7	7.6	0.8	4.3
Non-allocated	541.7	535.8	172.9	231.8	127.1	173.0	176.3	193.3	178.1	191.3
Total	1701.2	1599.5	1094.1	1321.0	1327.0	1557.2	1589.2	1672.1	907.6	808.3
Silver Hake										
Allocated Level A	5670.8	7105.7	8217.8	7534.5	7561.0	7187.9	6274.6	6756.5	4471.9	4806.3
Allocated Level B	1475.2	3069.9	3861.2	4352.1	3827.1	3012.8	2630.0	3293.3	1869.3	2142.6
Allocated Level C	1442.4	1082.7	458.2	1044.3	1488.4	821.5	604.2	395.0	293.1	543.4
Allocated Level D	26.6	33.9	15.0	20.1	14.5	24.0	25.6	21.4	7.3	26.4
Non-allocated	7442.2	3403.1	3627.6	2613.7	1976.0	2974.8	2827.0	2442.1	1295.7	1123.9
Total	16057.2	14695.3	16179.8	15564.7	14866.9	14021.0	12361.5	12908.2	7937.3	8642.6
All Species										
Allocated Level A	62236.2	106473	136592.4	129144.1	139152.0	112428.6	109866.3	118937.6	106470.2	108891.5
Allocated Level B	18724.1	39078.9	36777.9	42440.93	47216.8	33883.19	30432.7	30194.01	29765.88	33895.08
Allocated Level C	30501.5	31730.3	25731.8	24131.05	29541.6	16834.02	14421.1	13223.88	12701.14	20018.26
Allocated Level D	10494.5	7708.48	5736.9	4668.942	9741.2	4464.112	5699.9	1447.47	2142.28	11192.32
Non-allocated	532572	572975	521473.5	509288.4	471243.6	447878.6	440934.6	509009.7	445862.6	448153.6
Total	654527.8	757965.3	726312.5	709673.4	696895.2	615488.6	601354.6	672812.7	596942.1	622150.8

Table 4. Summary of percent landings for selected species and all species (area fished).

Monkfish	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Allocated Level A	31.7%	53.9%	56.7%	57.1%	59.7%	61.2%	64.1%	68.6%	68.9%	67.7%
Allocated Level B	12.7%	20.6%	19.1%	17.1%	15.0%	20.4%	17.9%	15.8%	16.5%	18.1%
Allocated Level C	13.4%	13.9%	10.8%	11.8%	12.5%	11.0%	11.2%	9.9%	9.4%	7.5%
Allocated Level D	4.0%	4.1%	5.2%	2.8%	2.2%	1.0%	0.6%	0.6%	0.7%	0.8%
Non-allocated	38.2%	7.5%	8.2%	11.1%	10.6%	6.4%	6.2%	5.1%	4.5%	5.8%
Total	100.0%									
Cod										
Allocated Level A	32.6%	69.1%	72.1%	72.4%	72.6%	75.1%	71.9%	75.0%	71.2%	72.0%
Allocated Level B	11.3%	19.8%	18.5%	17.6%	16.7%	16.9%	18.5%	16.9%	17.8%	16.2%
Allocated Level C	9.5%	9.2%	8.0%	8.6%	8.7%	6.5%	7.9%	5.8%	7.8%	8.1%
Allocated Level D	0.4%	0.1%	0.1%	0.3%	0.7%	0.5%	0.3%	0.3%	0.2%	0.2%
Non-allocated	46.3%	1.7%	1.3%	1.2%	1.3%	0.9%	1.5%	2.1%	3.1%	3.5%
Total	100.0%									
Winter flounder										
Allocated Level A	44.7%	63.5%	64.9%	62.8%	63.7%	66.2%	69.0%	68.7%	63.3%	70.7%
Allocated Level B	17.6%	22.7%	21.1%	22.4%	19.8%	22.6%	20.5%	21.0%	25.1%	17.5%
Allocated Level C	11.3%	7.5%	8.8%	9.3%	11.6%	6.8%	5.8%	5.5%	7.4%	8.0%
Allocated Level D	0.3%	0.1%	0.2%	0.1%	0.1%	0.1%	0.3%	0.5%	0.2%	0.3%
Non-allocated	26.1%	6.1%	5.1%	5.4%	4.8%	4.4%	4.4%	4.2%	4.1%	3.5%
Total	100.0%									
Yellowtail flounder										
Allocated Level A	49.9%	66.0%	65.6%	66.9%	65.4%	68.9%	64.8%	66.1%	60.1%	64.9%
Allocated Level B	18.5%	24.8%	23.6%	21.7%	24.1%	22.6%	28.4%	24.7%	28.0%	26.9%
Allocated Level C	11.3%	7.3%	7.7%	8.2%	7.8%	6.9%	4.9%	6.3%	7.5%	7.5%
Allocated Level D	0.4%	0.1%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%
Non-allocated	20.0%	1.8%	2.9%	3.0%	2.6%	1.5%	1.7%	2.9%	4.3%	0.7%
Total	100.0%									
Windowpane										
Allocated Level A	42.2%	54.8%	69.2%	68.9%	76.2%	71.1%	69.5%	71.7%	67.1%	63.8%
Allocated Level B	20.7%	31.4%	22.1%	17.3%	15.0%	21.2%	19.2%	20.4%	21.0%	24.4%
Allocated Level C	7.3%	12.1%	7.6%	10.1%	7.6%	4.3%	5.0%	3.3%	4.0%	5.5%
Allocated Level D	0.1%	0.0%	0.0%	0.1%	0.1%	0.5%	0.5%	0.2%	0.1%	0.8%
Non-allocated	29.7%	1.7%	1.1%	3.6%	1.1%	3.0%	5.8%	4.3%	7.8%	5.5%
Total	100.0%									
Haddock										
Allocated Level A	42.1%	69.5%	71.3%	71.0%	71.4%	73.8%	70.7%	73.9%	69.2%	72.2%
Allocated Level B	13.8%	19.0%	18.6%	20.6%	19.5%	21.3%	23.8%	21.9%	24.5%	19.9%
Allocated Level C	9.4%	10.7%	9.1%	8.2%	8.4%	4.1%	5.3%	3.3%	5.5%	7.3%
Allocated Level D	0.0%	0.3%	0.3%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%
Non-allocated	34.7%	0.6%	0.6%	0.1%	0.5%	0.8%	0.2%	0.9%	0.8%	0.5%
Total	100.0%									
Red Hake										
Allocated Level A	43.6%	44.6%	59.4%	58.1%	63.0%	63.8%	64.3%	63.1%	56.4%	48.0%
Allocated Level B	8.5%	13.8%	21.5%	18.6%	21.2%	20.4%	20.8%	21.2%	19.7%	20.3%
Allocated Level C	15.4%	7.9%	3.0%	5.4%	6.0%	4.6%	3.3%	3.7%	4.2%	7.5%
Allocated Level D	0.6%	0.1%	0.3%	0.3%	0.2%	0.2%	0.5%	0.5%	0.1%	0.5%
Non-allocated	31.8%	33.5%	15.8%	17.6%	9.6%	11.1%	11.1%	11.6%	19.6%	23.7%
Total	100.0%									
Silver Hake										
Allocated Level A	35.3%	48.4%	50.8%	48.4%	50.9%	51.3%	50.8%	52.3%	56.3%	55.6%
Allocated Level B	9.2%	20.9%	23.9%	28.0%	25.7%	21.5%	21.3%	25.5%	23.6%	24.8%
Allocated Level C	9.0%	7.4%	2.8%	6.7%	10.0%	5.9%	4.9%	3.1%	3.7%	6.3%
Allocated Level D	0.2%	0.2%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.1%	0.3%
Non-allocated	46.3%	23.2%	22.4%	16.8%	13.3%	21.2%	22.9%	18.9%	16.3%	13.0%
Total	100.0%									
All Species										
Allocated Level A	9.5%	14.0%	18.8%	18.2%	20.0%	18.3%	18.3%	17.7%	17.8%	17.5%
Allocated Level B	2.9%	5.2%	5.1%	6.0%	6.8%	5.5%	5.1%	4.5%	5.0%	5.4%
Allocated Level C	4.7%	4.2%	3.5%	3.4%	4.2%	2.7%	2.4%	2.0%	2.1%	3.2%
Allocated Level D	1.6%	1.0%	0.8%	0.7%	1.4%	0.7%	0.9%	0.2%	0.4%	1.8%
Non-allocated	81.4%	75.6%	71.8%	71.8%	67.6%	72.8%	73.3%	75.7%	74.7%	72.0%
Total	100.0%									

Table 5. Comparison of the number of length samples by original area and allocated area, allocation level and whether a change in statistical area occurred for split and non split trips. Yellow shade indicated the number of samples there will be assigned '000' due to a split trip.

All species									
CFLEnyyyyAA			Allocated Area						
Number of samples			Level A		Level B, C, D		Non-allocated		Total
YEAR	Orig Area	Split Trip	Differs	Same	Differs	Same	Differs	Same	
1994	area	no	3		4	3	1	159	170
		yes	3						3
	no-area	no	313		201		38		552
yes		37				4		41	
1994 Total			356		205	3	43	159	766
1995	area	no	44	207	8				259
		yes	10						10
	no-area	no	103		143		9	1	256
yes		18				2		20	
1995 Total			175	207	151		11	1	545
1996	area	no	99	292	88	76		9	564
		yes	13						13
	no-area	no	316		142		3		461
yes		26						26	
1996 Total			454	292	230	76	3	9	1064
1997	area	no	241	747	180	158	1	13	1340
		yes	51						51
	no-area	no	78		50				128
yes		7				16		23	
1997 Total			377	747	230	158	17	13	1542
1998	area	no	195	556	121	156	3	14	1045
		yes	27						27
	no-area	no	35		33		2		70
yes		1				1		2	
1998 Total			258	556	154	156	6	14	1144
1999	area	no	239	569	186	162	2	51	1209
		yes	20						20
	no-area	no	56		50		6		112
yes		1				1		2	
1999 Total			316	569	236	162	9	51	1343
2000	area	no	234	813	206	220	4	32	1509
		yes	43						43
	no-area	no	82		41		6		129
yes		8				7		15	
2000 Total			367	813	247	220	17	32	1696
2001	area	no	282	829	236	212	4	39	1602
		yes	111						111
	no-area	no	63		22		2		87
yes		5				1		6	
2001 Total			461	829	258	212	7	39	1806
2002	area	no	258	748	203	238		109	1556
		yes	63						63
	no-area	no	21		24				45
yes						3		3	
2002 Total			342	748	227	238	3	109	1667
2003	area	no	373	1152	339	333	5	164	2366
		yes	113						113
	no-area	no	12		45		6		63
yes		2				12		14	
2003 Total			500	1152	384	333	23	164	2556

Table 6. Comparison of the number of age samples, by original area and allocated area, allocation level and whether a change in statistical area occurred for split and non split trips. Yellow shade indicated the number of samples there will be assigned '000' due to a split trip.

All species

CFAGEyyyyAA			Allocated Area						Total
Number of samples			Level A		Level B, C, D		Non-allocated		
YEAR	Orig Area	Split Trip	Differs	Same	Differs	Same	Differs	Same	
1994	area	no	73	90	67	59	20	128	437
		yes	30						30
1994 Total			103	90	67	59	20	128	467
1995	area	no	96	139	63	34	1	19	352
		yes	8						8
	no-area	no			4				4
1995 Total			104	139	67	34	1	19	364
1996	area	no	114	335	97	85		37	668
		yes	12						12
1996 Total			126	335	97	85		37	680
1997	area	no	141	452	101	102		28	824
		yes	23						23
	no-area	no	1						1
1997 Total			165	452	101	102		28	848
1998	area	no	88	356	51	72	1	23	591
		yes	11						11
1998 Total			99	356	51	72	1	23	602
1999	area	no	127	319	74	66		29	615
		yes	6						6
1999 Total			133	319	74	66		29	621
2000	area	no	149	593	134	168		43	1087
		yes	26						26
2000 Total			175	593	134	168		43	1113
2001	area	no	134	461	120	108	2	25	850
		yes	42						42
2001 Total			176	461	120	108	2	25	892
2002	area	no	114	396	83	135		63	791
		yes	17						17
2002 Total			131	396	83	135		63	808
2003	area	no	186	564	150	176	1	50	1127
		yes	36					1	37
2003 Total			222	564	150	176	1	51	1164

Table 7. Statistical areas associated with the eight species and stocks examined.

Species	Stock	Statistical Areas
Atlantic cod	Georges Bank	521, 522, 525, 526, 533, 534, 537 - 539, 541 - 543, 551, 552, 561, 562, 611 - 616, 621 - 629, 631 - 639
	Gulf of Maine	464, 465, 511 - 515
Haddock	Georges Bank	521, 522, 525, 526, 533, 534, 537 - 539, 541 - 543, 551, 552, 561, 562, 611 - 616, 621 - 629, 631 - 639
	Gulf of Maine	464, 465, 511 - 515
Yellowtail flounder	Georges Bank	522, 525, 551, 552, 561, 562
	Cape Cod/Gulf of Maine	464, 465, 511, 512, 513, 514, 515, 521
	SNE/Mid-Atlantic	526, 533, 534, 537 - 539, 541 - 543, 611 - 616, 621 - 629, 631 - 639
Winter flounder	Georges Bank	522, 525, 551, 552, 561, 562
	Gulf of Maine	464, 465, 511, 512, 513, 514, 515
	SNE/Mid-Atlantic	521, 526, 533, 534, 537 - 539, 541 - 543, 611 - 616, 621 - 629, 631 - 639
Windowpane flounder	North	464, 465, 511 - 515, 521, 522, 525, 542, 543, 551, 552, 561, 562
	South	526, 533, 534, 537 - 539, 541, 611 - 616, 621 - 629, 631 - 639
Goosefish	North	464, 465, 511 - 515, 521, 522, 551, 561
	South	525, 526, 533, 534, 537 - 539, 541 - 543, 552, 562, 611 - 616, 621 - 629, 631 - 639
Silver hake	North	464, 465, 511 - 515, 521, 522, 551, 561
	South	525, 526, 533, 534, 537 - 539, 541 - 543, 552, 562, 611 - 616, 621 - 629, 631 - 639
Red hake	North	464, 465, 511 - 515, 521, 522, 551, 561
	South	525, 526, 533, 534, 537 - 539, 541 - 543, 552, 562, 611 - 616, 621 - 629, 631 - 639

Table 8. Summary statistics from 1,000 realizations of 1994 Dealer data for the eight species examined.

Species/Stock	1,000 Realizations							1994	1994 %tile	pt/mean
	Min	Mean	Max	Range	%spread	p10	p90			
MONKFISH-N	4242	4344	4485	243	5.6%	4304	4385	4352	60.9	1.002
MONKFISH-S	5066	5204	5305	239	4.6%	5163	5245	5196	41	0.998
<i>MONKFISH-Oth</i>	189	193	209	20						
COD-GB	8057	8106	8156	99	1.2%	8085	8128	8099	35.1	0.999
COD-GM	6776	6825	6883	107	1.6%	6804	6846	6836	76.8	1.002
<i>COD-Oth</i>	87	95	110	23						
WINTER FLD-GB	869	913	983	114	12.5%	897	929	929	90.3	1.017
WINTER FLD-GM	521	534	558	37	6.9%	527	541	528	16.2	0.990
WINTER FLD-SNE	2047	2124	2169	122	5.7%	2107	2140	2114	23.3	0.995
<i>WINTER FLD - Oth</i>	32	32	37	5						
YELLOWTAIL FLD-GB	1360	1437	1523	163	11.3%	1401	1473	1429	39.5	0.995
YELLOWTAIL FLD-GM	1066	1132	1214	148	13.1%	1098	1167	1139	61.8	1.006
YELLOWTAIL FLD-SNE	333	374	453	120	32.1%	348	403	367	41.5	0.980
<i>YELLOWTAIL FLD-Oth</i>	145	154	168	23						
WINDOWPANE-N	324	334	341	17	5.1%	330	337	338	96.8	1.013
WINDOWPANE-S	183	190	200	17	8.9%	186	194	186	10.5	0.977
<i>WINDOWPANE - Oth</i>	1	1	2	1						
HADDOCK-GB	179	181	184	5	2.8%	180	182	180	16.5	0.995
HADDOCK-GM	103	106	108	5	4.7%	105	107	107	99.5	1.010
<i>HADDOCK-Oth</i>	1	1	2	1						
RED HAKE-N	707	714	723	16	2.2%	712	717	714	54.9	1.000
RED HAKE-S	971	980	987	16	1.6%	977	982	980	64.1	1.000
<i>RED HAKE -Oth</i>	7	7	7	0						
SILVER HAKE-N	3820	3999	4144	324	8.1%	3942	4059	3887	0.4	0.972
SILVER HAKE-S	11881	12027	12206	325	2.7%	11967	12084	12139	99.3	1.009
<i>SILVER HAKE - Oth</i>	32	32	32	0						

Figure 1. Statistical areas off the northeast coast of the United States.

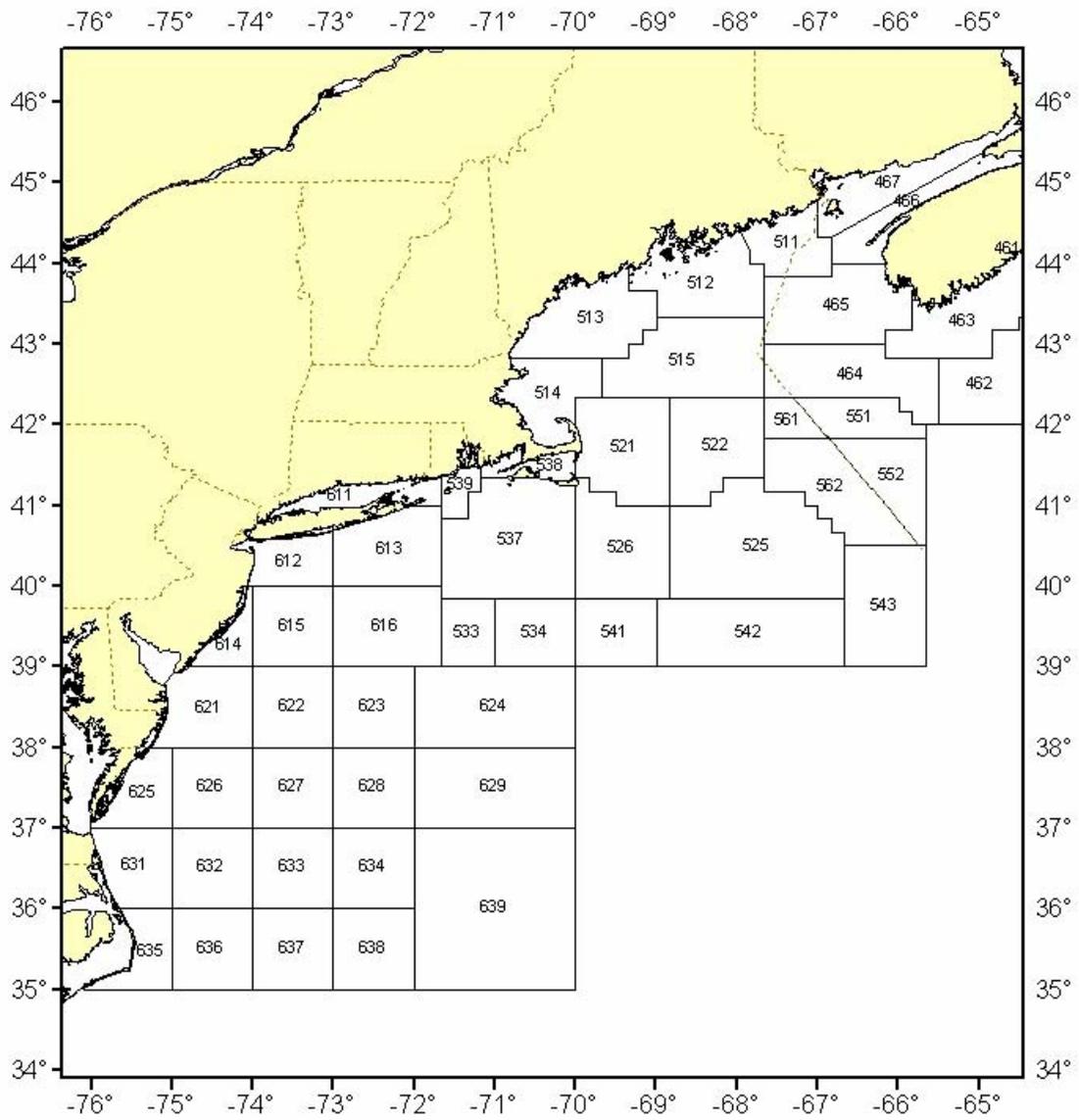


Figure 2. Diagram of the data flow through the four tiers of allocation. Only mandatory Dealer data enters the allocation. Non-mandatory and mandatory data are later combined into one data set

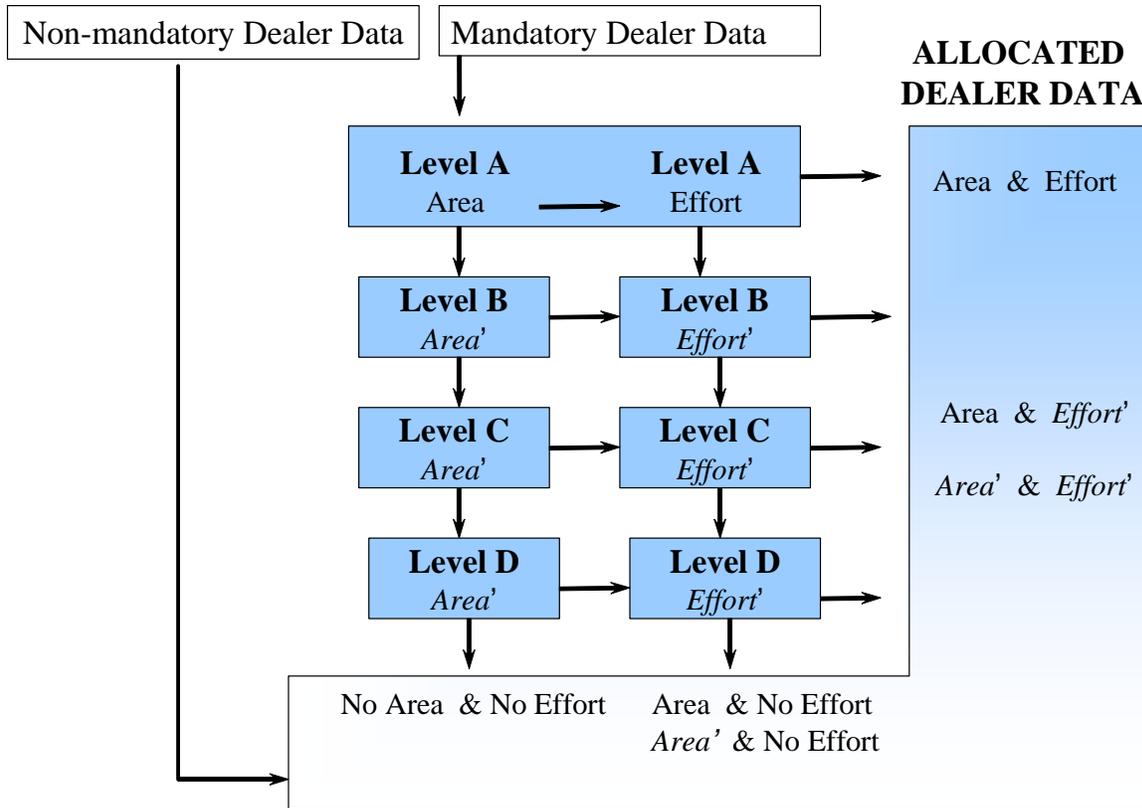


Figure 3. Diagram of the seven VTR data sets used in the allocation and the stratification variables used at each level.

VESSEL TRIP REPORT DATA:

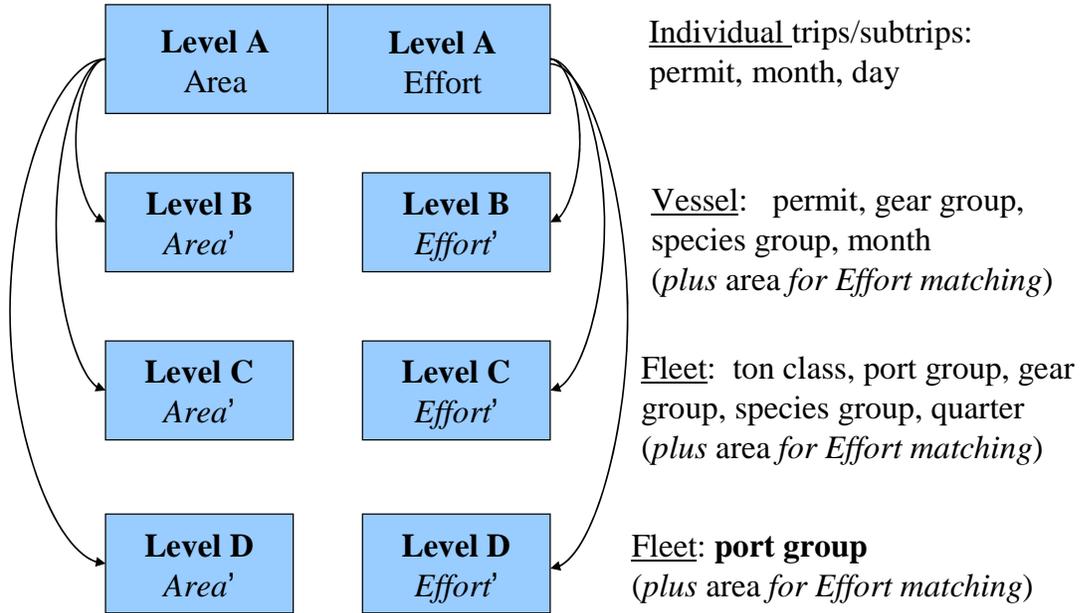


Figure 4a. Percent distribution of trips, by month, for Dealer and VTR in 2000 (top) and difference between the Dealer and VTR percentages for each month, 1994 – 2003 (bottom).

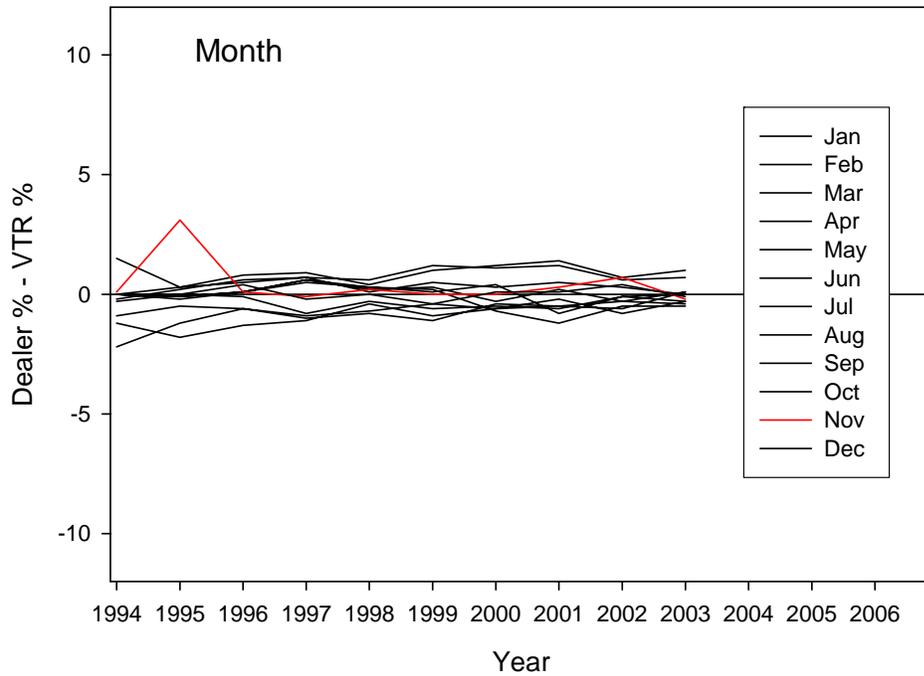
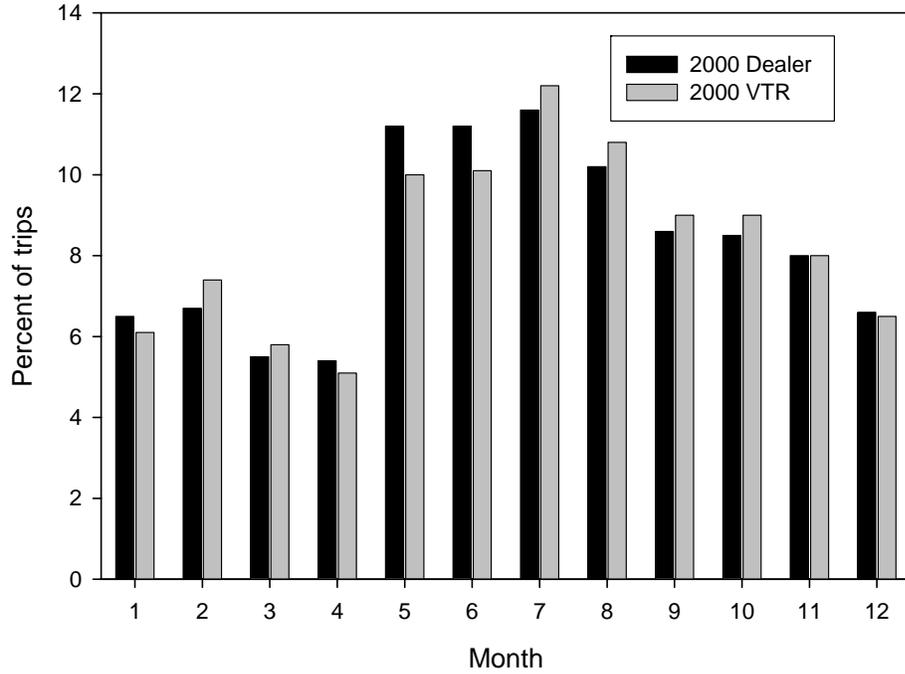


Figure 4b. Percent distribution of trips, by ton class, of Dealer and VTR in 2000 (top) and difference between the Dealer and VTR percentage, for each ton class, 1994 – 2003 (bottom).

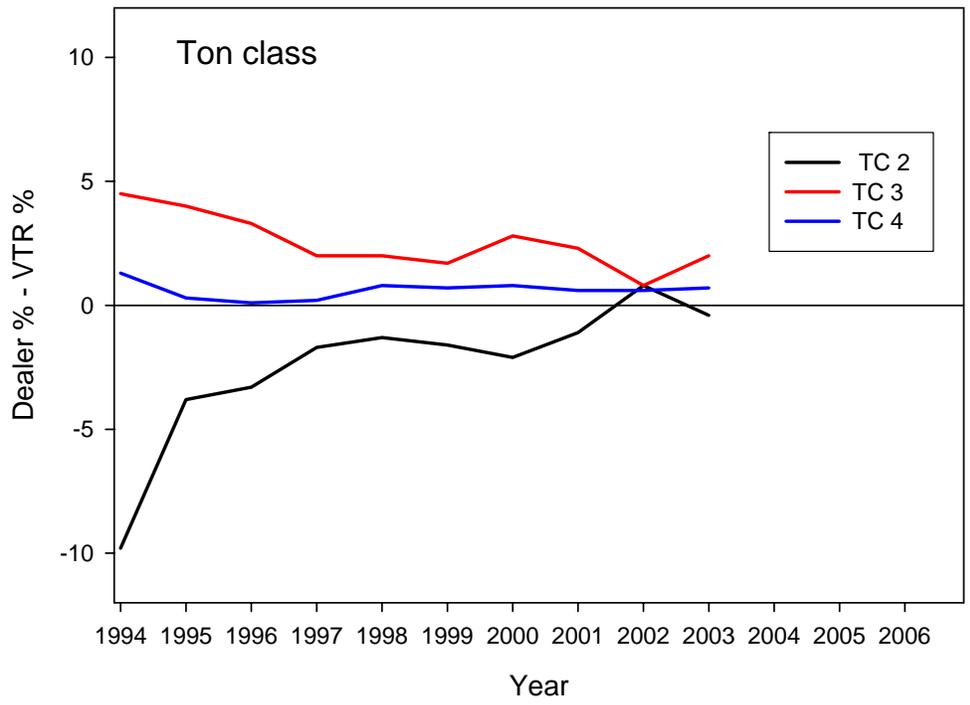
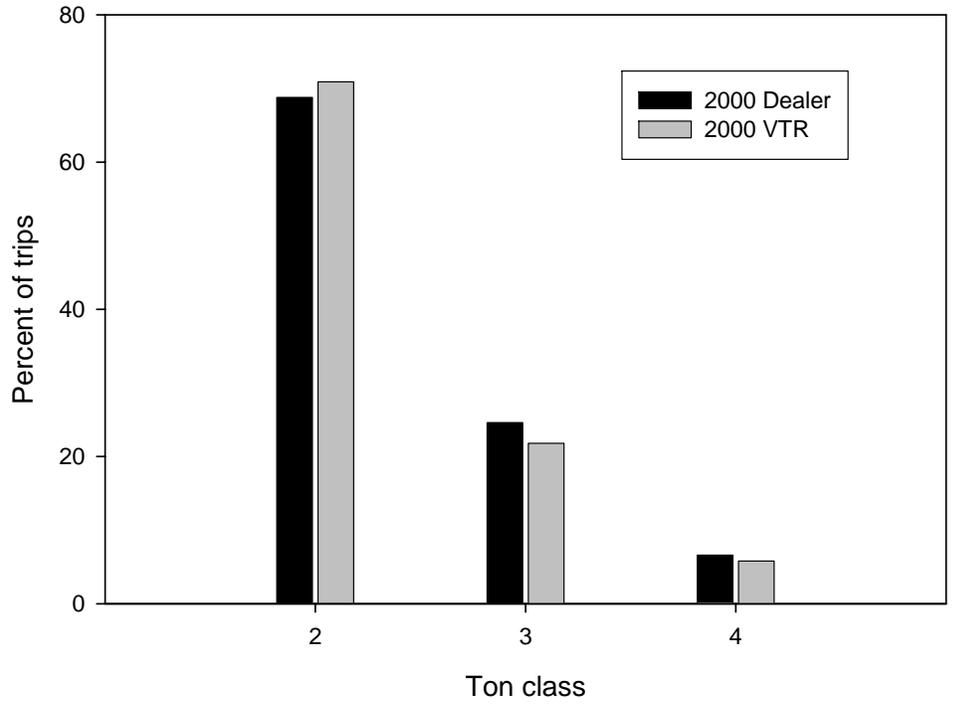


Figure 4c. Percent distribution of trips, by gear group, of Dealer and VTR in 2000 (top) and difference between the Dealer and VTR percentage for each gear group, 1994 – 2003 (bottom).

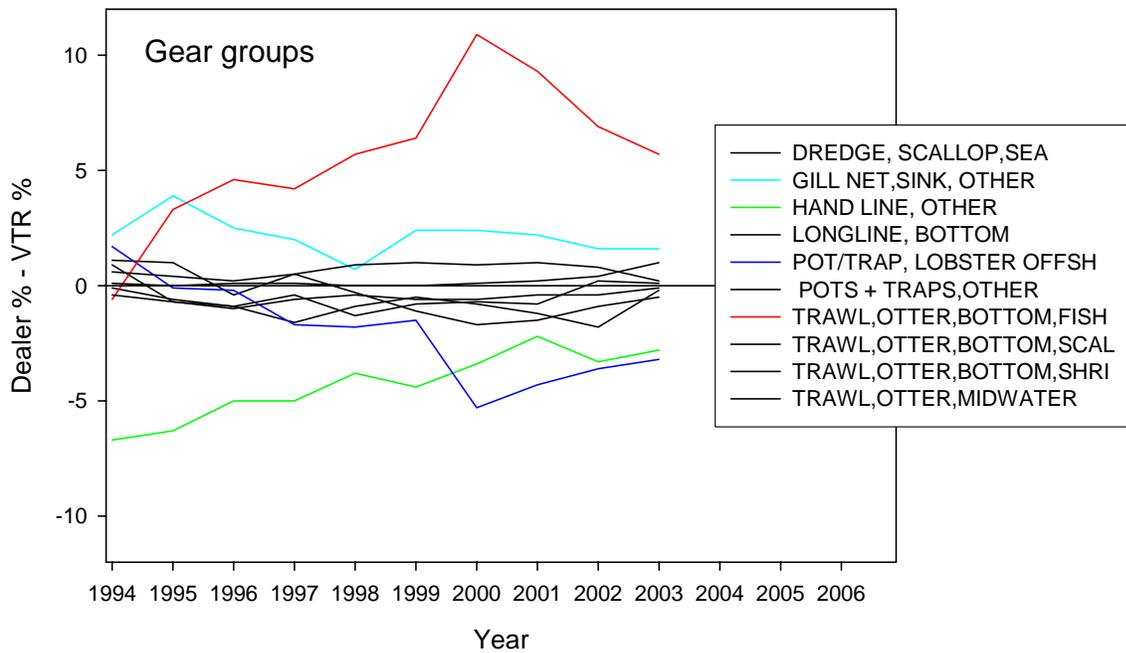
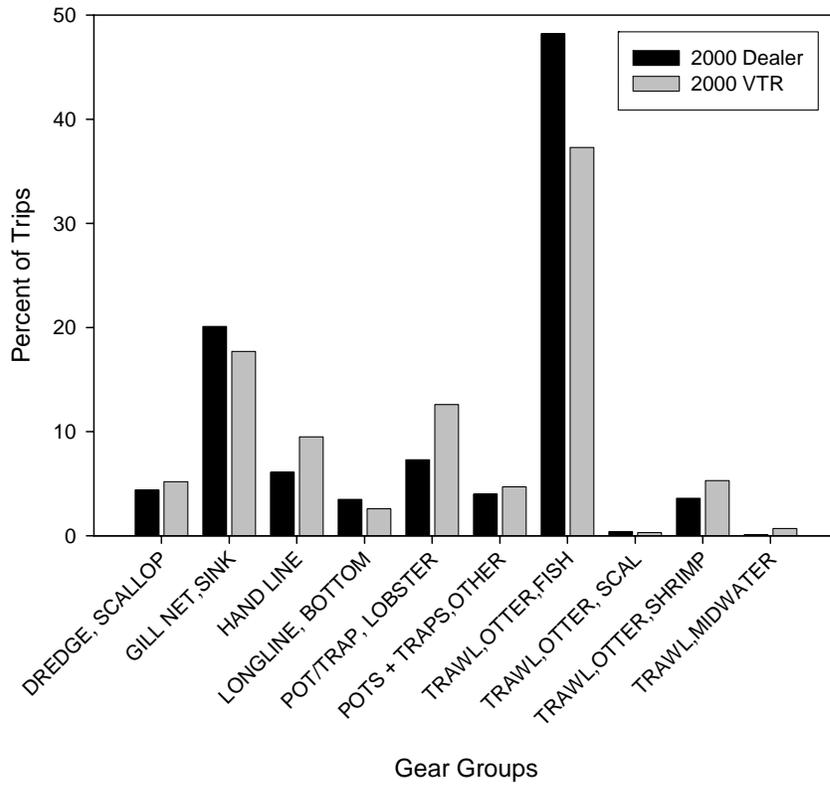


Figure 4d. Percent distribution of trips, by species group, of Dealer and VTR in 2000 (top) and difference between the Dealer and VTR percentage for each species groups, 1994 – 2003 (bottom).

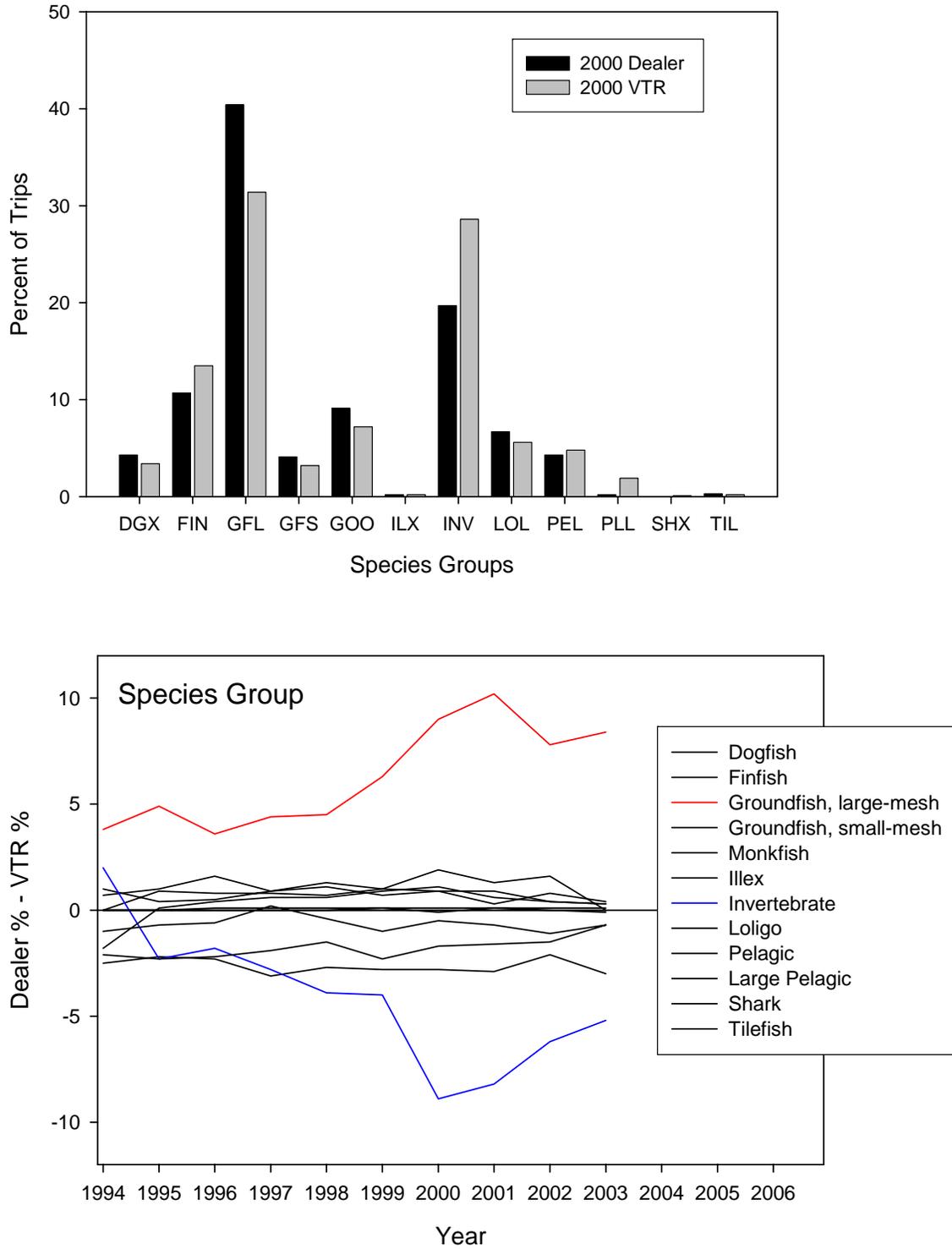


Figure 4e. Percent distribution of trips, by port group, for Dealer and VTR in 2000 (top) and difference between the Dealer and VTR percent for each port group, 1994 – 2003 (bottom).

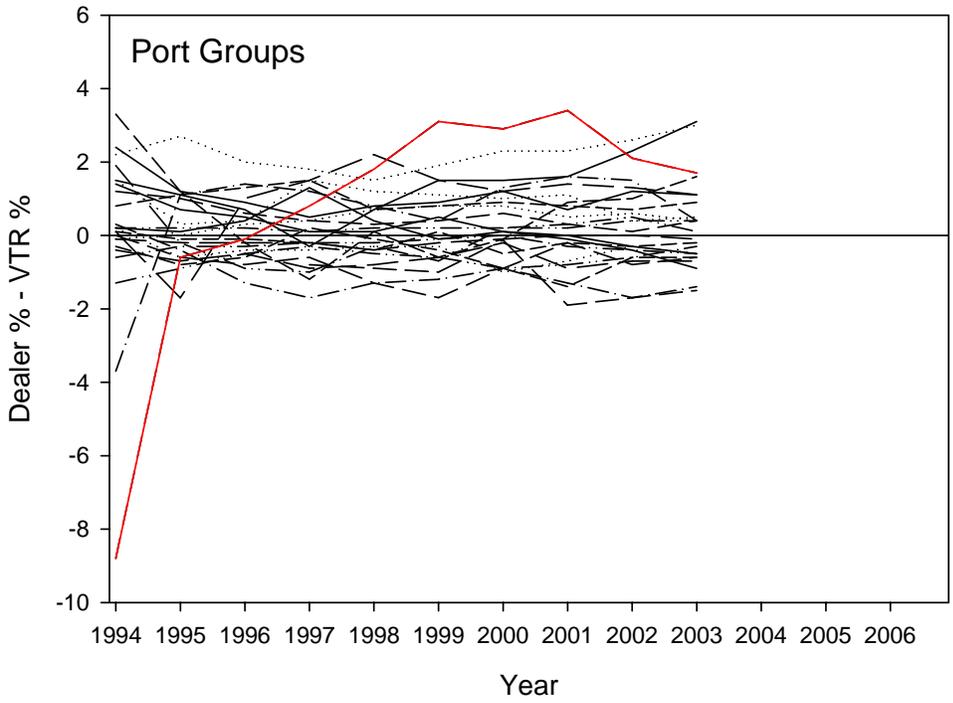
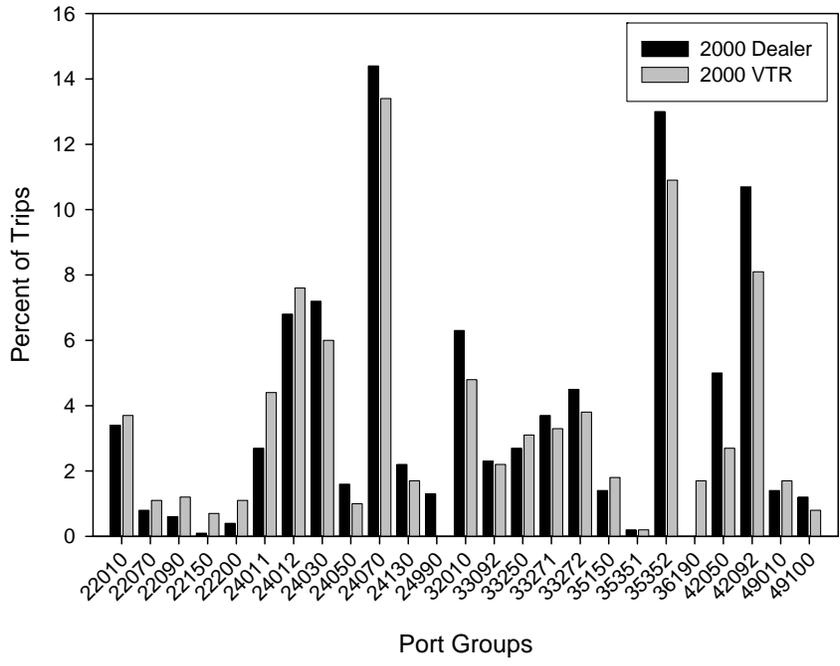
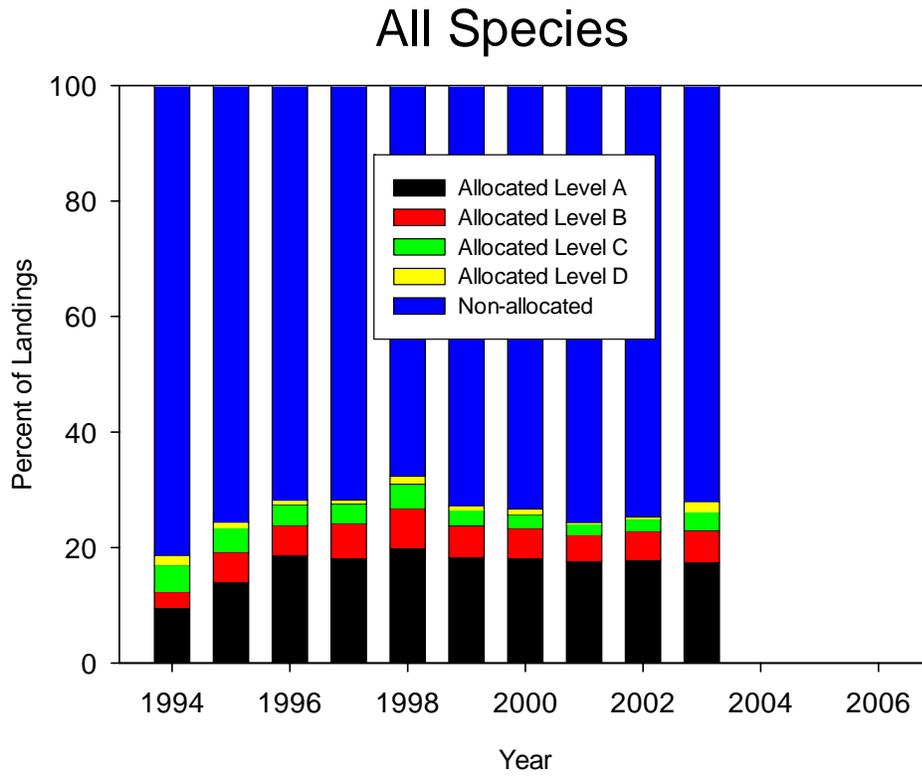
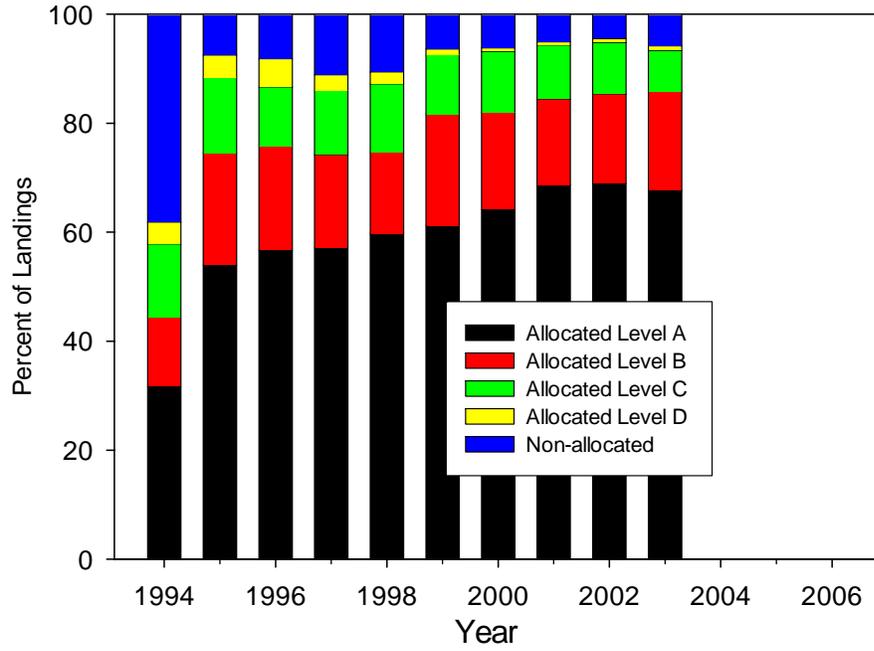


Figure 5. Percent of landings, by allocation level, for 1994 to 2003.



MONKFISH



COD

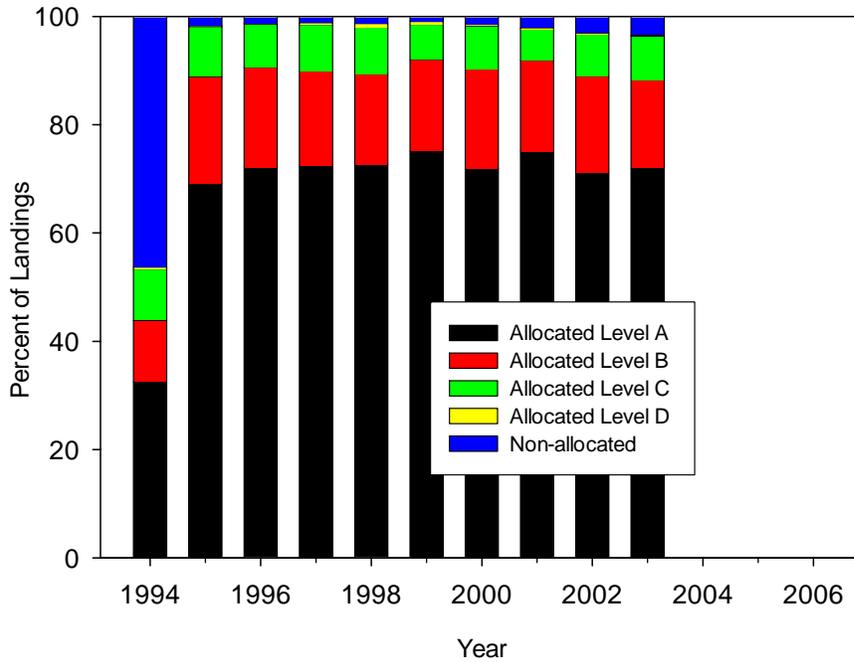
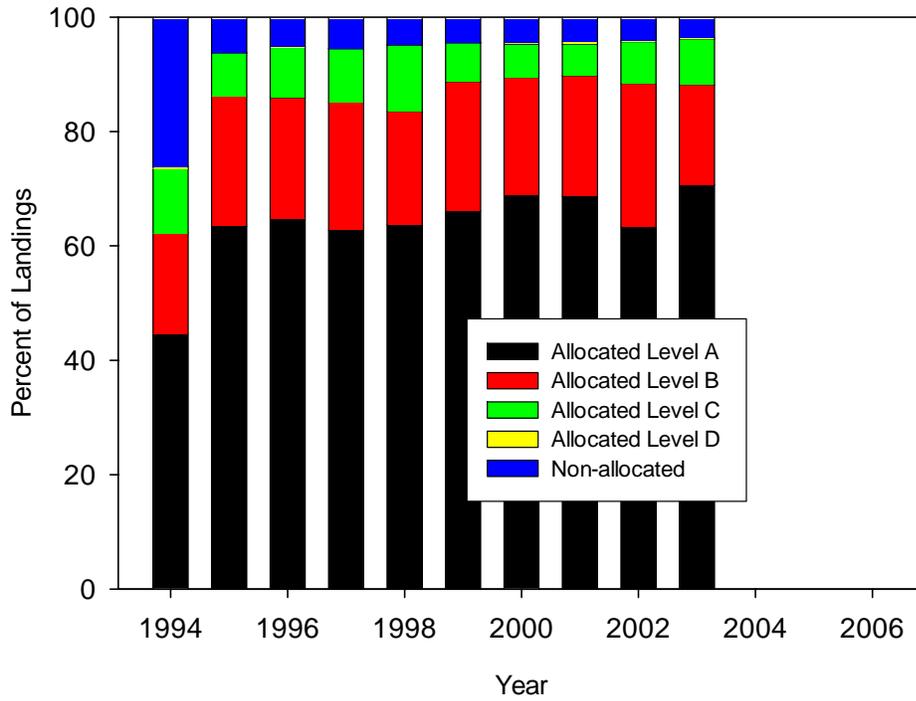


Figure 5 continued.

WINTER FLD



YELLOWTAIL FLD

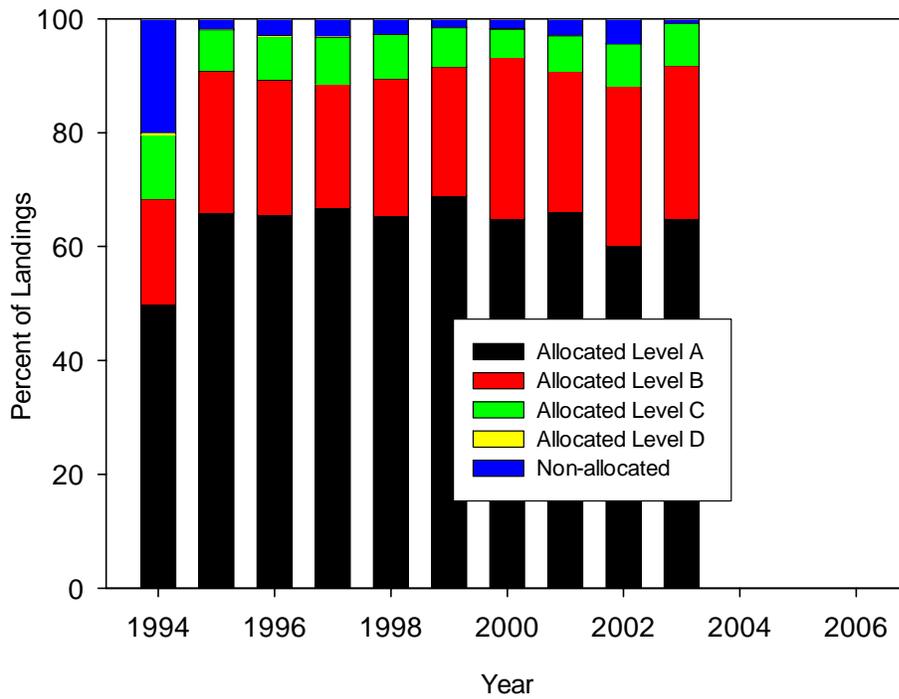
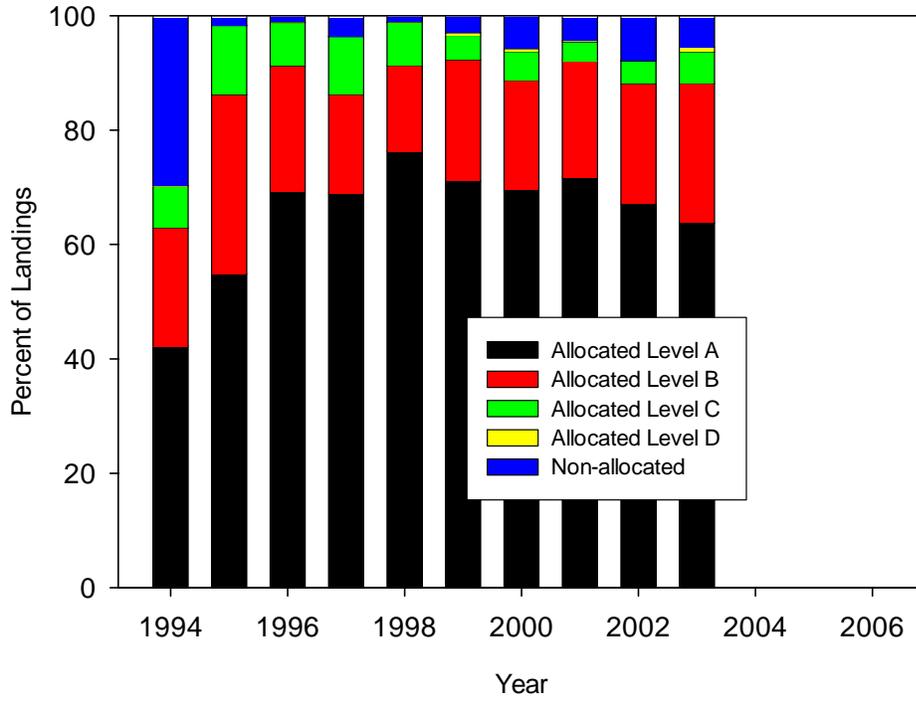


Figure 5 continued.

WINDOWPANE



HADDOCK

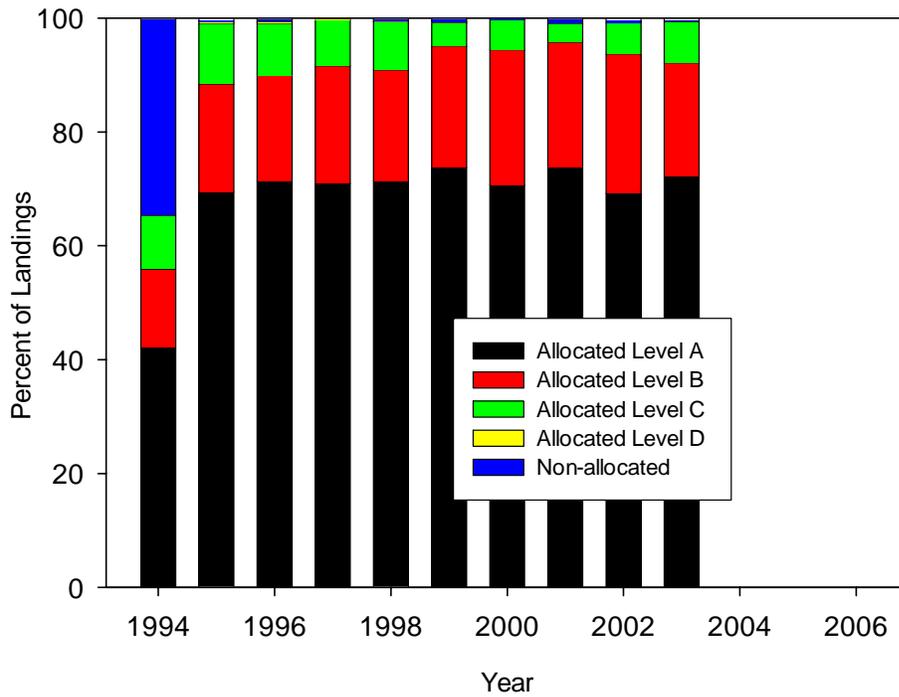


Figure 5. continued.

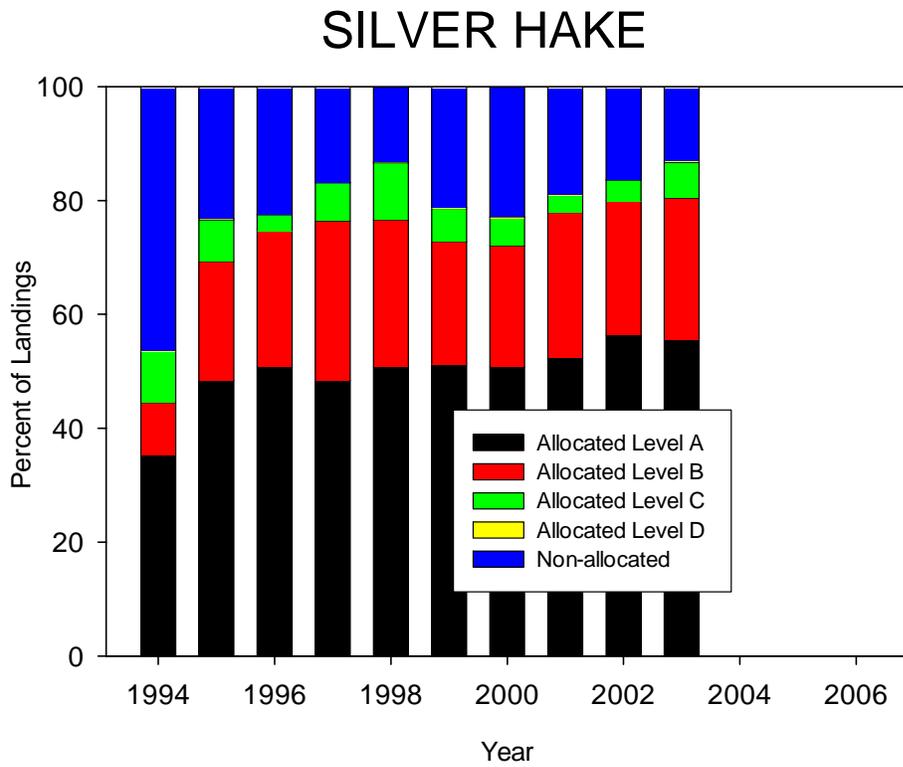
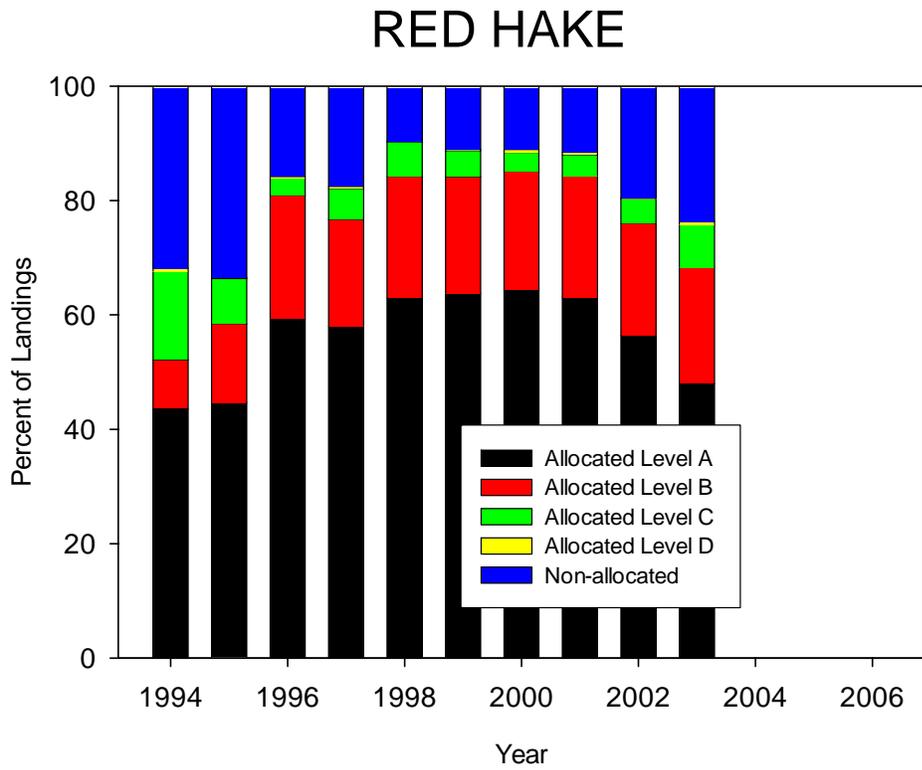


Figure 6. Frequency distribution of stock landings (mt-land wt) from 1,000 realizations of 1994 Dealer data, by species; vertical red lines indicate 80% confidence interval, circle represents 1994 base run.

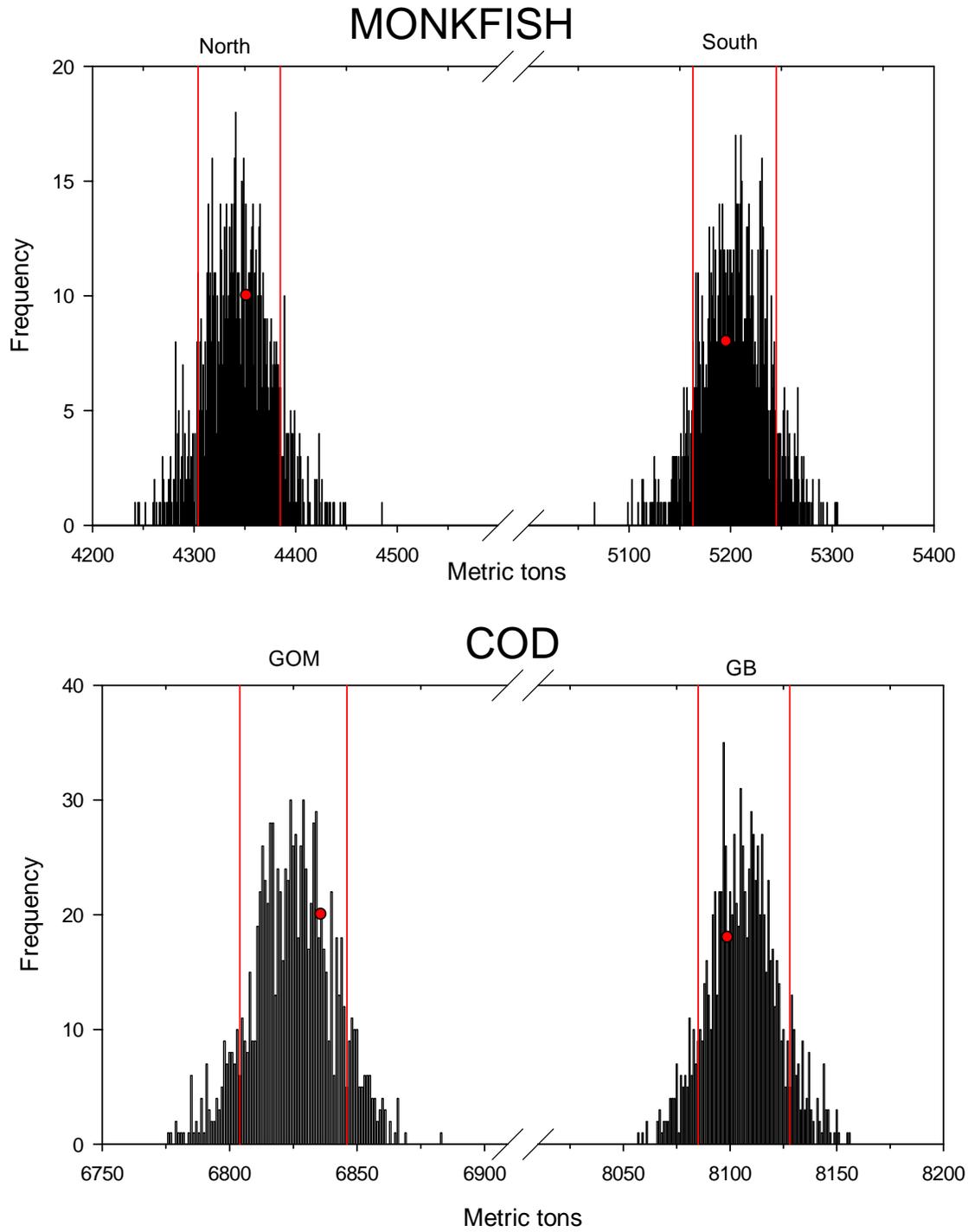


Figure 6 continued

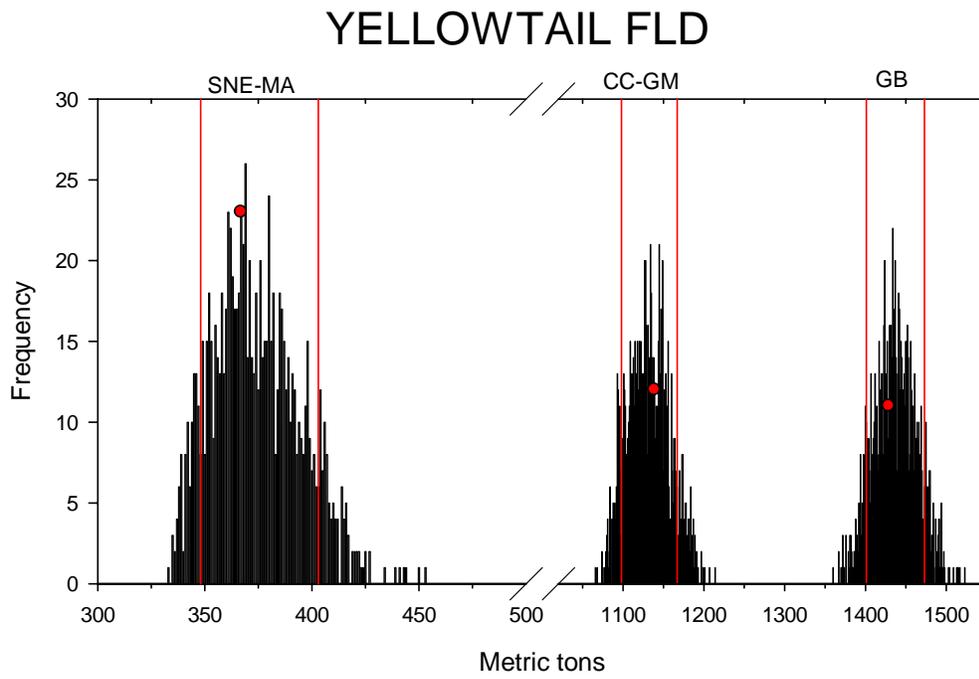
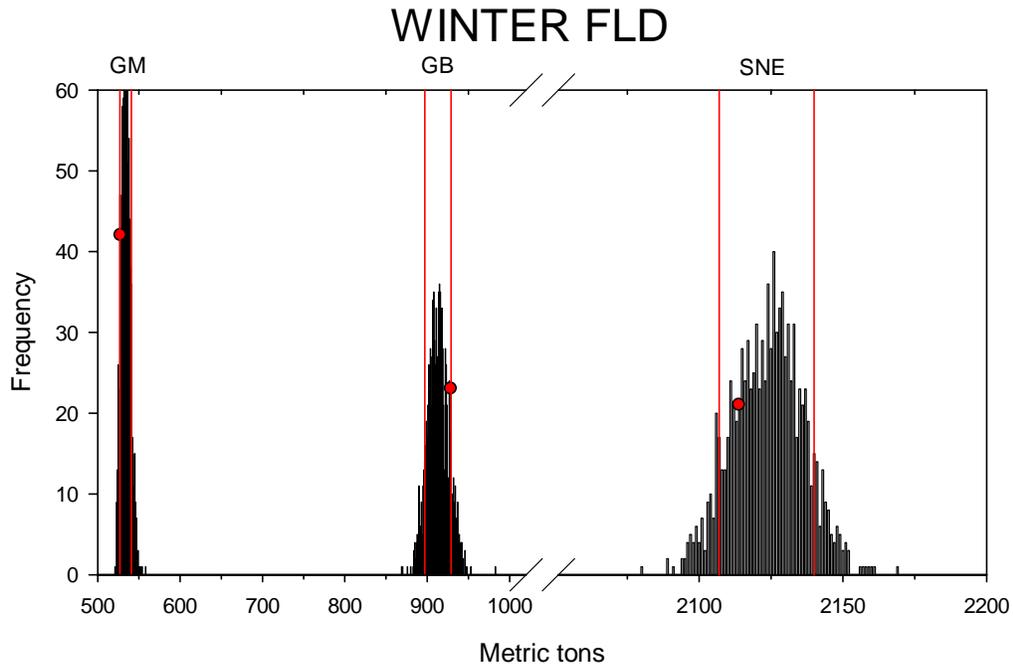


Figure 6. continued.

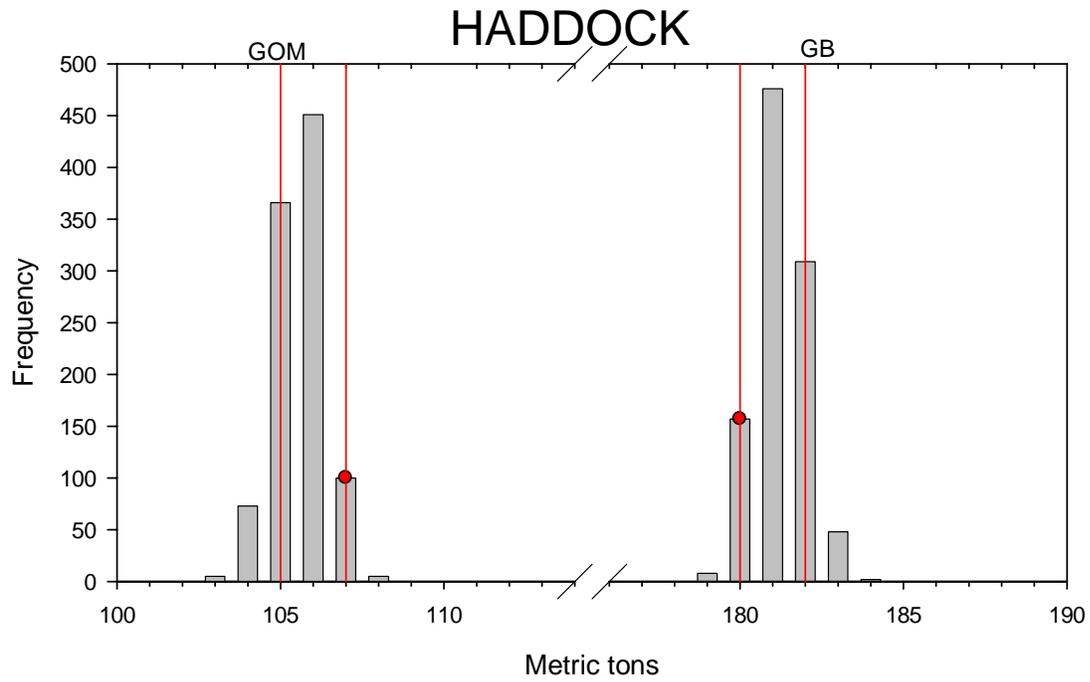
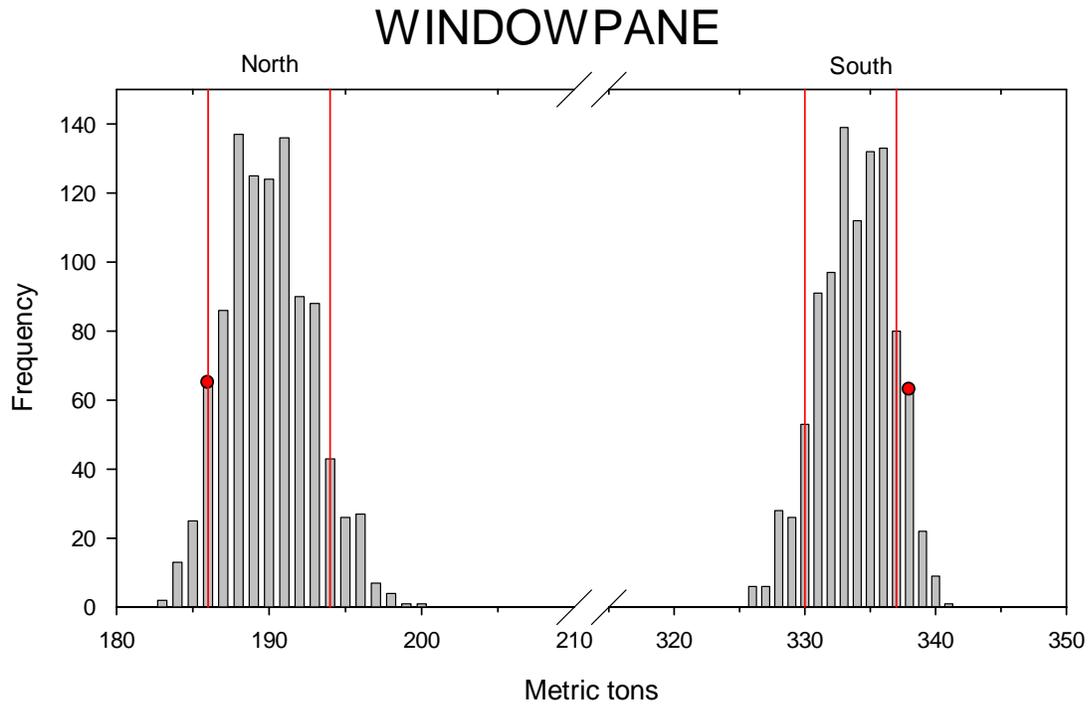
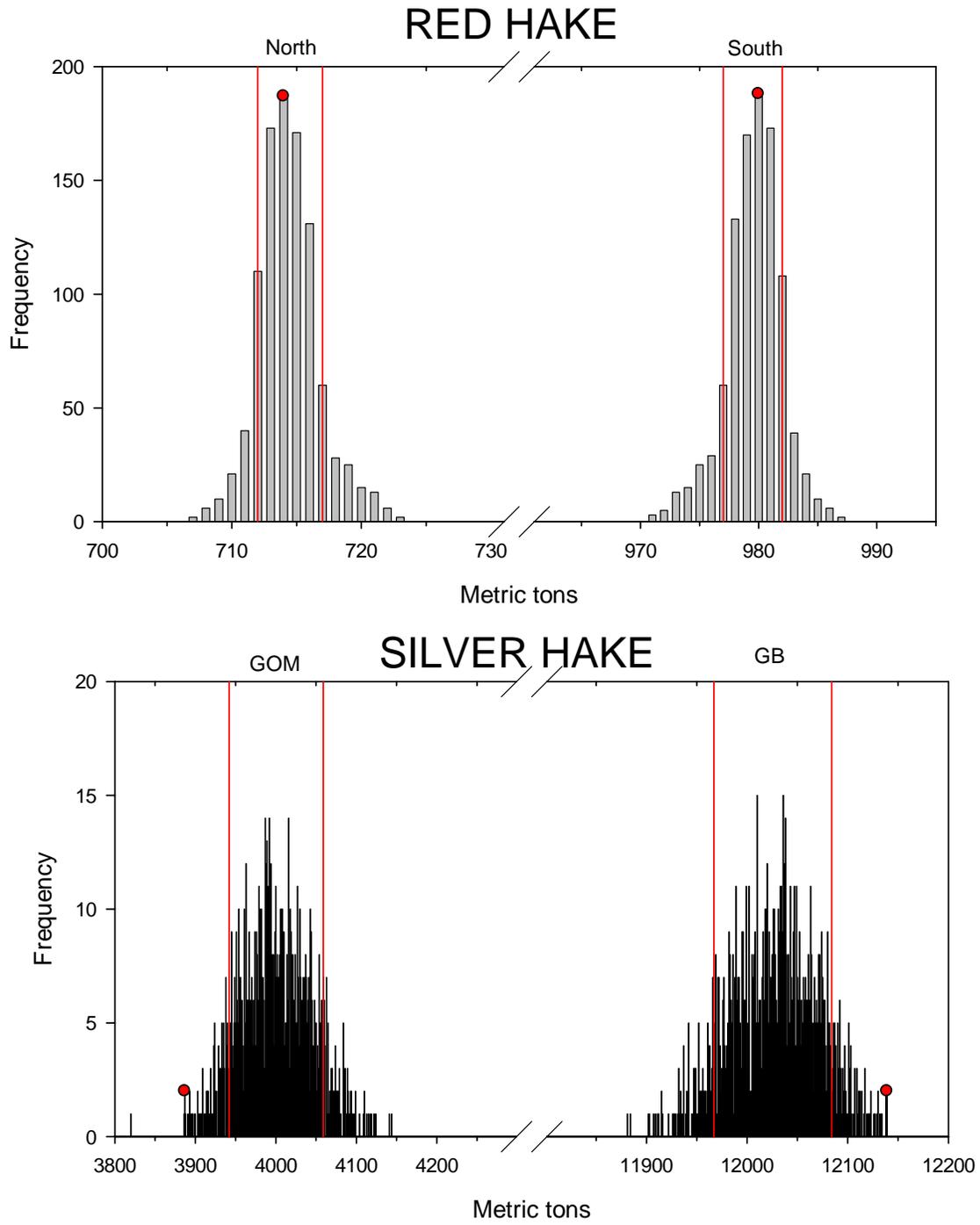


Figure 6. continued



Appendix Table 1.

Summary of header record scenarios in the allocation when the Dealer trip has a **SINGLE** transaction.

Dealer Header Record	VTR			Newly created Dealer Header Record
	Level A		Level B, C, D	
	Split trip	non-split trip	non-split trip	
one trip with one transaction p-m-d-did1-docn1 See Example 1 below	sub1 ntrip = 0.9 sub2 ntrip = 0.1			p-m-d-did1-docn1-sub1 ntrip = 0.9 p-m-d-did1-docn1-sub2 ntrip = 0.1 nsubtrips = 2 sum of ntrips = 1 (based on effort) ntrip partition based on VTR effort
		sub1 ntrip = 1		p-m-d-did1-docn1-sub1 ntrip = 1 nsubtrip = 1
			sub1 ntrip = 1	p-m-d-did1-docn1-sub1 ntrip = 1 nsubtrip = 1
	---	---	----	NO MATCH (for effort <i>ELevel</i> = X) p-m-d-did1-docn1-sub1 ntrip = null nsubtrip = null
consolidated trip d = '00'	---	---	sub1	p-m-d-did1-docn1-sub1 ntrips = number of docns, nsubtrip = null, subtrip = null

p = permit; m=month; d= day; did= dealer tripid; doc= docn; sub= subtrip.

Note: must use d|tripid|docn|subtrip to uniquely define records

Appendix Table 2.

Summary of header record (wort=1) scenarios in the allocation when the Dealer trip has **MULTIPLE** transactions.

Dealer Header Record	VTR			Newly created Dealer Header Record
	Level A		Level B, C, D	
	Split trip	non-split trip	non-split trip	
one trip with multiple transactions p-m-d-did1-docn1 p-m-d-did1-docn2 See Example 2 below	sub1 ntrip = 0.12 sub2 ntrip = 0.88			p-m-d-did1-docn1-sub1 ntrip = 0.06 p-m-d-did1-docn2-sub1 ntrip = 0.06 p-m-d-did1-docn1-sub2 ntrip = 0.44 p-m-d-did1-docn2-sub2 ntrip = 0.44 nsubtrips = 2 sub1: ntrip sums to 0.12, ntrips partitioned evenly among subtrips sub2: ntrip sums to 0.88 ntrip partitioned evenly among subtrips ntrips partitioned among sub1 and sub2 is effort based da and df is partitioned evenly within subtrip
		sub1 ntrip = 1		p-m-d-did1-docn1-sub1 ntrip = 0.5 p-m-d-did1-docn2-sub1 ntrip = 0.5 nsubtrip = 1 sum of ntrip = 1 ntrip partitioned evenly among docn-subtrips (i.e. not effort based) da and df is partitioned evenly within subtrip
			sub1 ntrip = 1	p-m-d-did1-docn1-sub1 ntrip = 0.5 p-m-d-did1-docn2-sub1 ntrip = 0.5 nsubtrip = 1 sum of ntrip = 1 ntrip partitioned evenly among docn-subtrips (i.e. not effort based) da and df is partitioned evenly within subtrip
	---	---	---	NO MATCH (for effort <i>ELevel</i> = X): p-m-d-did1-docn1-sub1 ntrip =null p-m-d-did1-docn2-sub1 ntrip =null nsubtrip = null sum of ntrip = null ntrip partitioned evenly among docn-subtrips (i.e. not effort based) da and df is partitioned evenly within subtrip
consolidated trip d= '00'	---	---	sub1	p-m-d-did1-docn1-sub1 ntrip = number of docn p-m-d-did1-docn2-sub1 ntrip = number of docn p-m-d-did1-docn3-sub1 ntrip = number of docn nsubtrip = null, subtrip = null

p = permit; m=month; d= day; did= dealer tripid (dltrpid) ; docn = document number; sub= subtrip.

Note: must use dltrpid|docn|subtrip to uniquely define records

Example 1: two transactions matching a split VTR

BEFORE:

DLRTRPID	DOCN	SUBTRIP	NSUBTRIP	NEM	NEG	MESH	NTRIPS	DA	DF	A	E
09700142	700151	1	2	621	100	3.2	.12	.12	.13	A	A
09700142	700142	1	2	621	100	3.2	.12	.12	.13	A	A
09700142	700151	2	2	621	181	0	.88	.88	1	A	A
09700142	700142	2	2	621	181	0	.88	.88	1	A	A

AFTER:

DLRTRPID	DOCN	SUBTRIP	NSUBTRIP	NEM	NEG	MESH	NTRIPS	DA	DF	A	E	
09700142	700151	1	2	621	100	3.2	.06	.06	.065	A	A	
09700142	700142	1	2	621	100	3.2	.06	.06	.065	A	A	
09700142	700151	2	2	621	181	0	.44	.44	.50	A	A	
09700142	700142	2	2	621	181	0	.44	.44	.50	A	A	
Totals:							1.0	1.0	1.13			

Example 2: two transactions matching a non-split VTR

BEFORE:

SOLE_SQL> select permit, month, day, dlrtrpid, docn, subtrip, nsubtrip, nemarea, negear, mesh, ntrips, da, df, alevel, elevel, vtrserno from cfnew.cfraw1994t_match_bak where dlrtrpid = '09240165'

DLRTRPID	DOCN	SUBTRIP	NSUBTRIP	NEM	NEG	MESH	NTRIPS	DA	DF	A	E
09240165	240166	1	1	514	050	6	1	1	.13	A	A
09240165	240165	1	1	514	050	6	1	1	.13	A	A

AFTER:

DLRTRPID	DOCN	SUBTRIP	NSUBTRIP	NEM	NEG	MESH	NTRIPS	DA	DF	A	E	
09240165	240166	1	1	514	050	6	0.5	0.5	.065	A	A	
09240165	240165	1	1	514	050	6	0.5	0.5	.065	A	A	
Totals:							1.0	1.0	0.13			

Technical note:

Multiple Transactions

Because effort is divided among the transactions for a trip, the effort for a trip with multiple transactions can not come from different sources (for example: *ELevel* = c and *Elevel* = D).

Thus, the matching of effort for these types of trips will be driven by the docn with the most catch (the docn used in the dlrtrpid). The alevel, elevel, nemarea, cumprob, prob, ntrips, da, etc. from this docn will be used on all the other transactions for the trip.

Appendix Figure 1. Flow chart of Oracle tables created and used during the allocation.

