Review of methods used to estimate trawl discards in the West Coast groundfish fishery, Seattle, 15-17 July 2008

External, independent peer review report by

John Cotter,
Cefas, Lowestoft, UK
4 August 2008

Contact: john.cotter@cefas.co.uk
Tel. +44 1502 524323
Centre for Environment, Fisheries and Aquaculture Science
Lowestoft, NR33 0HT
United Kingdom

Contents

0. Executive summary

The goal for this review of the West Coast groundfish observer program (WCGOP) was to improve the methods used for estimating and communicating quantities of fish discarded by commercial fishing vessels. My role as an independent reviewer was to comment on strengths and weaknesses of the current approaches, to identify uncertainties in estimated and projected quantities of discards being used to assess uptake of optimal yields for each fishery, and to recommend alternative approaches for the WCGOP, if appropriate.

Most striking to me was the contrast in approaches between WCGOP and European observer programs familiar to me. WCGOP estimates weight of discards as the priority, whereas European programs estimate length frequencies and other biological variables. Weights are estimated in European programs, if required, using the allometric growth equation based on numbers-at-length. Secondly, WCGOP focusses on quantities of fish discarded, taking quantities retained from official archives (PacFIN) of quantities landed, whereas many European programs use the observers to estimate quantities of fish retained haul-by-haul, as well as quantities discarded. There are some good reasons for these contrasts, e.g. the Magnuson-Stevens Act which relates to optimal yields (OY) in terms of weights, and the generally large sizes of fish catches on the West Coast that can make the sampling of retained fish difficult. Nevertheless, the focus on weights and the need to draw data on retained quantities from outside the observer program are curtailing the biological information available from WCGOP, are resulting in significant possibilities for errors in data processing, and are causing lengthy delays before finalised data can be produced.

I have concluded that there are several sources of uncertainty in the estimates of discards produced by WCGOP. The estimation model uses a detailed geographic and seasonal post-sampling stratification of observed hauls but this does not coincide with the way that vessels and trips were sampled. Results for hauls within a locality are therefore likely to be spatially autocorrelated because many would have been observed on the same vessels, trips, and during the same 2-monthly trip-limit periods. These effects are likely to be causing bias and increased variance among strata, and a lack of independence within strata, though I cannot estimate the magnitudes of the effects. The dependence of the fleet-level estimation method on ‘discard ratios’ (weights discarded/weights retained) is also likely to be adding to, and complicating error variance because weights retained, whether of one target species or all, are in the denominator and can be very small. Heavy reliance by WCGOP on the accuracy of fish tickets seems risky because there appear to be good reasons for fishers to under-declare the weights they land, even if this is illegal. An observer program is better targeted in my view when it can be applied to assess the accuracy of landings declarations, among other tasks. For this to be possible, retained quantities must be estimated independently by the observers. Another uncertainty in current estimation methods arises from the process of matching observer records to fish tickets and logbooks in PacFIN. There are many practical reasons why this is error-prone and subjective. It is also one of the main reasons for weeks of delay in the production of WCGOP discard estimates.

My general impressions of WCGOP in practice were of competence, diligence, and conscientiousness. I agree with the distribution of observers along the coast in

relation to fishing activities. I strongly recommend that they should be estimating quantities retained as well as discarded for each haul, and that they should give priority to length measurements, not weightings. Both changes would make the WCGOP so much more informative about fishing and biological processes, e.g. selectivities, recruitment and growth, that they are well worth considering even if it means accepting large sample-to-catch expansion (a.k.a. 'raising') factors, and restricting the species of interest and, perhaps, other observer duties. If, nevertheless, weight continues to be the primary variable estimated onboard, I would like to see special efforts made to obtain estimates of observer precision which, at present, is another uncertainty in the program. I have suggested a way of achieving this. Quality control of data collected for WCGOP requires continuing high vigilance because of the geographic distances between observers. Observers should report to different debriefers (if not done) so as to maintain consistent standards coast-wide. Tests of observers' species identification abilities are crucial to WCGOP and should require that the core of priority species be identified virtually perfectly. The current standard of '80% of all species' does not seem specific enough.

I also considered the way in which vessels and trips are selected for observations. Randomisation is restricted by requiring that all vessels be observed in a sampling cycle of 8-10 months, and that each selected vessel be observed for a full 2-month trip-limit period. These restrictions are likely to be causing some bias and, probably more importantly, sampling inefficiency. I have suggested how the existing sampling scheme could be fitted into a cluster or multi-stage scheme from Sampling Theory to avoid these problems, and to avoid the need for stratification of hauls or to refer to PacFIN, meaning that results could be produced much more quickly, easily, and would, in theory, be unbiased. These estimators, obtainable from textbooks, could be tried out with existing data now. However, I also think that the existing sampling scheme should be revised at an early opportunity. For this, I have recommended that better sampling efficiency and interpretability could be obtained by randomly selecting and observing individual trips on different vessels using a season-based sampling cycle. Standard sampling formulae are 'design-based' estimators that make minimal assumptions about the data, only about the sampling design. Modelling could possibly improve the precision of estimation as well as the understanding of the discarding process. I have tentatively suggested a preliminary linear mixed model (that will certainly need adjusting and development) for fitting to the rich data resources now available from WCGOP.

Lastly, I have suggested that NWFSC consider whether more effort should be put into communicating WCGOP results to the fishing industry, e.g. post-observation private data reports, public annual reports on the fisheries, a web-site, etc..

1. Introduction

I was told that the general purpose of the review was to improve methods for estimating and communicating quantities discarded from the observer programme. The focus was on the west coast bottom trawl groundfish ("limited entry trawl", LE) fisheries because they were the most extensively observed west coast fisheries and the methods used for observing them were similar to those used for other west coast fisheries. Specific terms of reference given to me as a Statement of Work (SoW) were:

1. Review background materials and become familiar with the West Coast groundfish bycatch model and methods to estimate total mortality;
2. Actively participate in the review panel, July 15-17, 2008, Seattle;
3. Comment on the primary sources of uncertainty in the estimates or projections;
4. Comment on the strengths and weaknesses of current approaches;
5. Recommend alternative model configurations or approaches, as appropriate;
6. Complete independent peer review report after the completion of the STAR panel meeting in accordance with the terms of reference (ToR).

Members of the review panel held at NMFS, Northwest Fisheries Science Center (NWFSC), Seattle were:

Chris Harvey (Chair)
Jim Hastie
Janell Majewski
Marlene Bellman
Eliza Heery
Jennifer Cahalan (independent reviewer)
John Cotter (independent reviewer)
Jon Cusick

In addition, Liz Clarke and other NWFSC scientists attended the panel from time to time.

My activities for the review included study of all documents submitted to me beforehand (as listed in the SoW), participation in the review panel during which I raised and discussed most of the findings reported here, and preparation of this report with the help of printed versions of NWFSC presentations plus a list of data quality checks for which I asked. I am grateful to all review staff for the clarity of their submitted material and presentations, and their willingness to discuss all aspects of their work.

Annex 1 of the SoW describes the format and contents required of this report which I have attempted to follow. In section 2, I display what I learned about the west coast fisheries from the review and documents. The text also serves as introductory background for this report. Section 3 summarises my understanding of the main features of the West Coast groundfish observer program (WCGOP). Section 4 presents my findings on WCGOP and reasoning supporting them. The numbering of paragraphs in section 4 is intended to match that used in section 3 for ease of cross-

referencing between the current situation and my findings related to them. Section 5 presents my main conclusions and recommendations with cross-references to other sections where fuller reasoning is given.

2. Background

NMFS documents submitted to me before the review and presentations at the review explained numerous facts which are summarised here as background as well as to display the extent of my understanding.

2.1 The fisheries

More than 90 species are found in commercial catches, 60 of which are classed as rockfish. Pacific hake, fished mainly in summer, provide the largest volume of fish and the largest total revenues. Sablefish are next in financial value, followed by rockfish and flatfish. Aside from bottom trawls, the types of gear used in west coast fisheries included midwater trawls, fish traps, and long lines. However, the balance of vessel types in the fleet frequently changes in response to changing regulations, fishing opportunities, and markets. Fleet effort is capped by limiting the numbers of fishing licenses and the lengths of the vessels. For example, there are now approximately 270 trawl permits, of which 9 or 10 are for factory trawlers. Trawlers tend to catch between 1000 and 10000 lbs in a single haul and to make trips of 3 to 5 days.

Fish are discarded in west coast fisheries because

- they have low market value;
- they are over trip limits for the species;
- they belong to species prohibited for landing, e.g. Pacific halibut, salmon, Dungeness crabs.

There is a tendency for fishers to discard over-fished and re-building species (see below) to avoid inadvertent infringements of trip limits.

In addition to commercial fisheries, there are significant recreational fisheries for some species, notably rockfish.

2.2 Management

Management measures for the fishery stem from stock assessments conducted annually on 28 species. These are used to find an ‘optimum yield’ (OY) in terms of total catch weight for each species for the forthcoming fishing year. This is to conform with the Magnuson-Stevens Fishery Conservation and Management Act which stipulates prevention of over-fishing while maintaining optimum yield from each fishery. In order to monitor uptake of the OYs, observer data on weights discarded have to be added to official data for weights landed so as to estimate weights caught. Scientists’ current views are that seven species of rockfish, ling cod, and Pacific hake are in the category “over-fished”. Other species thought to be recovering from over-fishing are categorised as “re-building” stocks. Most other stocks are being maintained at levels above, or well above OY. However, total revenues from the fisheries have declined somewhat in recent years.

Management of the west coast fisheries by the Pacific Fisheries Management Council (PFMC) has the following mandates:

- to prevent over-fishing;
- to rebuild over-fished stocks; and
- to attain optimum yields (OY) for each target species.

OYs have gradually reduced since the 1990s, particularly for re-building species. The effect has been an 80% reduction of landings.

One important management measure is "trip limits", i.e. catch quota set for around 20 individual species or groups of species per vessel. Fish caught that are over the trip limit for the species must be discarded. In earlier years, trip limits were set for individual trips. Subsequently, many of the limits were set for 2-monthly periods, the intentions being (a) to reduce the need to discard on every trip and (b) to encourage fishing at different localities or with different strategies as the lowest quota for each of the most vulnerable species became exhausted in turn. However, it remains legal for vessels that have reached their landing limits for one species to continue fishing for another while discarding the first. Other benefits of the 2-monthly system include (c) a spreading of catch throughout the year allowing processing industries on shore to work efficiently, and (d) discouragement of increases in engine power since higher catch rates merely mean that vessels have to be tied up earlier in each 2-month period.

Management measures used for the west coast fisheries also include:

- A rockfish conservation area (RCA) set parallel with the coast between two depth-bands; trawling is only permitted shorewards or seawards of the RCA. A cowcod conservation area also exists as a box off California.
- Constraints on trawl design.
- Permit buybacks to reduce the size of the fishing fleet.
- Inspections at sea or in port to enforce regulations.

3. The west coast groundfish observer program (WCGOP)

The WCGOP is well described in the pre-review documents and by presentations at the review. Here I summarise my understanding of aspects that are relevant to a scientific assessment of the program.

3.1 Goals

The WCGOP goals include collection of quantitative, species composition, and biological information on the bycatch and discards of west coast groundfish species. Fishing variables such as effort are also recorded. The information is processed by NWFSC and used by fisheries managers, stock assessors, and other scientists. The primary quantitative variable of interest is weight of each species discarded because of the relevance for assessing uptake of the OY. Collection of biological information was given little emphasis in the review; I noted that it is 10th priority among observers' duties set out in the WCGOP observer manual.

3.2 Distribution of observer sampling

Approximately 40 observers, depending on season, are based in fishing towns spread fairly evenly along the west coast and, so far as practically possible, observer time

spent at sea is matched to fishing effort. Most fishing vessels are also based consistently in these towns but interchange of a minority of vessels between ports dependent on fishing opportunities and market prices, etc., does occur. Sampling is stratified in time so as to match the 2-monthly trip limits, and by gear. In 2006, eight distinct fishery types were treated as sampling strata.

3.3 Vessel selection

All vessels licensed to fish in the limited-entry trawl fishery are observed during a "sampling cycle". Typically, it must be 8 to 10 months long in order to allow all vessels to be observed. The order of sampling the vessels within each port group is decided by a random draw. The first drawn are observed for all trips made during the first 2-monthly trip-limit period within the sampling cycle, and so on, such that, for an 8-month cycle, approximately 20% of the vessels will be observed in each of four 2-monthly periods. One reason for observing all trips by a selected vessel within a 2-monthly trip limit period is that vessels change their discarding practices as the end is approached, as modelled by Helser et al. (ICES CM 2002/V05). Another reason is that it is harder for a skipper to fish atypically while under observation for this long period; i.e. the "observer effect" is thought to be diminished. There are exceptions to the overall sampling plan; for example, observation is delayed if the vessel is temporarily outside the fishery or no observer is available, and is abandoned if the vessel is unsafe. Vessels in other west coast fisheries are selected for observation in similar ways depending on length of fishing season, and WCGOP priorities.

3.4 Observers

Observers have the right to be present on fishing vessels but do not serve to enforce fishery regulations. They work for a contractor to NMFS, are graduates in biology, and are trained intensively in species identification and sampling methods (among other essential subjects) for a 2-3 week period prior to their first solo sailing. They must be able to identify 80% or more of species correctly. Refresher training is provided annually. Observers are assigned to one of 12 port groups for the period of their contract, with the number in each group depending on fishing activities there.

3.5 Observers' duties at sea

Observer duties on board a fishing vessel are covered by a comprehensive manual which lists the priorities for work on each haul. A large part of the job consists of (a) estimation of the weight of the total catch by one of various methods depending on circumstances, (b) separation and weighing as individual species of over-fished and rebuilding species from the discarded portion of the catch, the latter as determined by the crew, and (c) separation and weighing of other fish as individual species or as groups of species called 'categories'. Categories are likely to be sampled to determine species composition. This does not usually cause statistical problems because most or all categorised species (c) are common, but there may sometimes be problems in separating look-alike species such as small thornyheads or mixed flatfishes. Weights of the various species groups and categories discarded are usually determined from the volume occupied within bins or baskets on deck together with estimates of density which varies depending on species composition and packing of the fish within the containers. Weights of fish retained in the different categories, as well as start and end times and locations for each haul, are usually recorded from the trawler logbook filled in by the skipper from information hailed by the crew on deck (see below). Information on fishing variables, and by-catch such as marine mammals, seabirds, etc. is routinely recorded. Biological information (length, age, maturity,

sex) is collected from any tagged fish found but has a low priority for species in the bulk of the catch. Data are first recorded on waterproof paper and arethen entered into a computer.

3.6 Debriefers

Debriefers are NMFS employees experienced in observer work at sea. There are five debriefers, and their job is to review each observer's sampling methods, calculations, and data at the end of every month. The data are returned to the observer who corrects any errors on the data sheets and in the database. Every two months, the observer is interviewed and evaluated, and final checks and logging of data carried out using standardised database procedures. I was shown a 7-page printout of standard data entry checks, and the comprehensive and detailed manual for debriefers. Completion of the quality verification process, referred to as "phase 1", takes about 2 months. Prior to further processing of observer data, a final, automated quality check is run on the entire data set available for a chosen period. A debriefer reviews all errors flagged. This is the beginning of "phase 2". Debriefers may reject observer data that are of poor quality. Extreme and unlikely results are accepted if they can be supported. Debriefers are themselves assessed every 2 months.

3.7 Processing of observer data using fish tickets

The rest of phase 2 of the data processing involves the matching of observed trips with landing records for each trip so as to estimate total catch weights (landings + discards), and to relate results for the observed part of the fleet to those for any part which could not be observed. Landings records come from fish-buyers' receipts, referred to as "fish tickets" and archived in the PacFIN database after receipt from the state agencies and harmonisation of field codes across states. There is generally a 2-month lag in this process before 90% of fish tickets are archived and, by April, about 99% of tickets are present for the previous calendar year. The linking of fish tickets to some observer trips is complicated and labour intensive for NWFSC because:

- Trips may have more than one fish ticket and all have to be found.
- Observers may not be able to record all of these tickets for each trip.
- Conversely, fish tickets may be cancelled and replaced by others.
- Species may be recorded differently by observers, skippers, and fish buyers.
- There are possibilities for double counting of landed weights in the fish ticket system.
- The port of return observed may not be that where the fish are sold.
- Procedures and formats vary in the three Pacific states.
- Species groupings on fish tickets are usually coarser than those used by observers.

Linking of observed trips to fish tickets is partly automated but associated investigations can take 2 weeks or more.

Phase 3 of data processing firstly involves estimation of haul-level weights of species or species groups in the observer records for cases when data records represent samples of catches. The haul-level results are then summed to the trip level to match with fish tickets. The analyst must assign consistent species names, catch categories, port and port grouping codes to the observer and fish ticket data sets before they can be merged. The analyst is assisted in this by biological sampling of the landings by port staff of the three states to determine species composition exactly. The samples

are from individual trips undertaken during a chosen period of variable length, usually months, but the estimated average composition for the period and port group is a composite result that is applied to the aggregated landings occurring within that same stratum. Fish ticket data from PacFIN having the original species composition data, much of which is given as groups of species, are referred to as "line item" data; those having the groups broken down into individual species are called "distributed" data. Resolving discrepancies and verifying matches between observed trips and fish tickets may involve extensive investigations and some subjective judgements for each individual case.

Having completed the matching and merging process, the analyst next adjusts the weight of retained catch reported by the observer so that it is the same as the total weight reported on fish tickets for the trip. This also involves adjusting all of the observed haul weights discarded by corresponding factors for individual species. The same is true for the weights retained as taken from the trawl logbooks (see below). Estimation of discarding rates for the observed trips by species or species groups is then relatively straightforward using the fish ticket weights landed and the observed weights discarded. "Coverage" of the observer program, defined as weight observed over weight landed, is also estimated for each port group and fishery in order to see how well the observer program followed the fishery. The absence of a *fishery* field in the PacFIN data does not help completion of this task. Phase 3 takes approximately four weeks to complete.

3.8 Trawl logbooks

Completion of trawl logbooks is mandatory on all trawlers. The same logbook form is used across all states but there are differences in processing by states including a different rate of feed into PacFIN. Generally, the logbook data for a calendar year are only about 90% complete by the following July, and a minority of trips are never reported in logbook records on PacFIN due to incomplete submissions. Weights of fish categories in the logbooks are recorded by the skipper as hailed by the crew processing the fish on deck. Since these are only estimates they are subsequently adjusted in PacFIN so as to align with quantities recorded on the fish tickets for the trip. Adjustment involves no loss of information because both the original and the adjusted weights are stored in PacFIN. The logbooks also provide the start and end times and locations of each tow, as used by observers. GIS analysis of logbook data for 2006 has indicated that the locations of LE trawl tows observed in WCGOP matched well with the pattern of fishing by the whole LE trawl fleet.

3.9 Forward projection of total fishing mortality

A trawl by-catch projection model, developed in 2002, is in use by the PFMC to predict catches and by-catches by each LE trawler during the current year's entire 2-monthly trip limit periods (that are set at the beginning of the year). The purpose of the model is to allow pre-season planning of the next year's trip limits, as well as in-season revisions of them, or of trawl closure areas, if necessary in the light of recent catch rates and compositions, so as to avoid exceedence of the OYs for any species. The model achieves this by calculating the depth distributions of vessels from logbook data, by-catch rates by area and depth zone from observer data, and overall landings from fish ticket data. The data are updated at least annually but observer data are 9 to 21 months old before they go into the model, so some may be nearly 3 years old at the time of the next annual update. The projection model was reviewed

by PFMC's Scientific and Statistical Committee in 2003 and 2004 and was not examined in detail in the present review.

3.10 Current estimation of total fishing mortality in the Limited Entry groundfish bottom trawl fishery

Fleet-wide discard estimates in the LE trawl fishery for past time periods are derived from observer data, fish tickets, and trawl logbooks. In order to restrict estimation to only this fishery, several specific filters are used to remove inappropriate fish ticket and logbook records obtained from PacFIN; the same filters are applied to observer data from NMFS. All data at the haul level are grouped by area (N or S of 40 deg 10 min), depth (6 depth ranges from 0 to >300 fathoms but excluding the RCA), and season (winter and summer). These groupings create a post-sampling stratification. Neighbouring strata not having many hauls may be combined to give reasonable sample sizes.

'Discard ratios' are then estimated for all strata; the ratios are total discarded weight/ total retained weight for individual target species, and total discarded weight of one species/ total retained weight for all target species for individual by-catch species. The different denominator is needed for by-catch species because many show zero retained weights. In these ratios, retained weights are being used as a proxy for trawl effort for the purpose of expanding observed trips to the entire LE trawl fleet.

Next, weights discarded by the entire fleet are calculated as the products of the discard ratios and the total weights retained by the fleet as summed from logbooks, using individual or total target species weights for target and by-catch species respectively. Because logbooks are not available for 100% of trawl trips, these estimates are further expanded by the ratio of fish ticket weights/ logbook weights, again using individual or total target species as appropriate. This assumes that the fish ticket weights provide accurate measures of landings.

3.11 Assessments of alternative re-stratification and expansion schemes

NWFSC is aware that its current method of expanding observer estimates to fleet-wide estimates of discarding is only one of several possible methods. Therefore, three alternative re-stratification schemes have been investigated. The existing stratification of hauls with 2 areas and 6 depths was labelled S1. S2 was 2 areas, 3 depths; S3 was 3 areas, 3 depths; and S4 was 6 areas and 3 depths. The thinking behind these alternatives was that the numbers and sizes of strata affect variance and the accuracy with which tows are assigned to the correct stratum. Extensive results cannot be summarized in a few words but most were within +/- 25% of the original discard estimates with S1. Also, larger numbers of strata sometimes led to more inestimable discard ratios due to the lack of retained fish in the small strata.

Two alternative methods for expanding to fleet level were also investigated. In the original method, use of retained weights as effort metrics in the discard ratios was motivated by the structure of the projection model (see above), and because retained weights given on fish tickets are expected to be accurate for both legal and commercial reasons, whereas trawling effort is a self-reported, non-legally binding quantity in logbooks, some of which are unavailable. Nevertheless, two alternative effort metrics were tested: tow-duration, and number of tows. Interestingly, relationships between by-catch weights discarded and tow duration or weights of target species retained were weak at the haul level, but somewhat stronger at the

vessel x 2-monthly period level. The many results of using different expansion factors cannot be summarised briefly and accurately. Suffice to say that many discard estimates were little affected by the different expansion factors, but with a few exceptions reaching nearly +/- 100% of the original estimates.

3.12 Bootstrap studies of estimation uncertainties

NWFSC investigated the uncertainty associated with fleet-wide estimates of total mortality using bootstrap re-sampling at the trip and haul level within strata. The four alternative re-stratifications of hauls and the alternative expansion methods for deriving fleet-wide estimates assessed in the study were the same as described in 3.11 above. Summarising, it was found that bootstrap estimates tended to be normally distributed and centered around the point estimates for each stratification and expansion scheme, but with several exceptions that may have been attributable to the random re-combination of hauls in the double-sampling procedure. CVs tended to be smallest under the existing, S1 stratification scheme.

4. Findings

The following paragraphs are numbered so as to correspond to paragraphs in section 3 that describe aspects of the current observer program. This is intended for easy cross referencing.

4.1 Findings on WCGOP goals

WCGOP goals contrast with the goals of European observer programs in two significant respects discussed in the next two subsections.

4.1.1 Estimation of weights discarded

WCGOP focusses on weights discarded because of the importance of OY under the Magnuson-Stevens Act, but under the European Data Collection Regulation (EC 1639/2001, see Chapter III, section H1(e); http://www.leglatext.ee/text/en/T60672.htm), European fishery agencies are required to estimate length and age distributions of discards of significant commercial species, but not weights, even though management of European fisheries is primarily by TACs expressed in weight terms as on the US west coast. In practice, European observers sample length distributions for most of the fish species caught, including by-catch species, from almost every haul. Estimated relative volumes (not weights) of sample and catch are only used for expanding the sample to the catch level.

At Cefas, our reasons for not estimating weights at sea were (a) motion-compensated balances cannot be carried to or used on most national fishing vessels, (b) other methods of estimating weights of fish were generally considered difficult to apply with reasonable accuracy, (c) weights are less informative biologically than length and age distributions, e.g. they give no indication of whether fish are big or small, or of year classes or growth, and (d) ICES fish stock assessments require numbers-at-age, not weights. When required, weights can be estimated easily from length distributions using the allometric growth equation ($W = \sum_L N_L a L^b$) , but the same is not true in reverse. It is necessary to estimate a and b for each species by sampling from hauls scattered over regional and seasonal strata, then plotting $\log W$ versus $\log L$ for each stratum.

4.1.2 Estimation of quantities retained

Another contrast in goals is that WCGOP observers do not routinely sample and estimate retained fish caught on the vessels, whereas many European programs (certainly the Cefas program) do for each haul (though this is not a requirement of the EC regulation). One advantage of estimating retained quantities on the fishing vessel is that the proportion of the catch discarded for each haul is then available together with all the associated fishing variables, thereby gaining maximum information for understanding and modelling the discarding process. Under the WCGOP, discarding can only be modelled as a function of trip variables representing time-averaged haul variables, a situation that offers less explanatory power.

A second advantage of estimating retained quantities is that there is then no compulsion to refer to other databases to find retained quantities and, thence, percentage discarding rates. I and others have argued elsewhere (see Fish and Fisheries (2004), vol 5, pp. 235-254) that data sets contributed to stock assessments should be kept mutually independent if each is to add genuinely new information. The reasons are (a) to avoid reduplication of common errors and biases, (b) to prevent occurrence of spurious relationships across data sets, and (c) to prevent over-fitting of models due to over-optimistic estimation of degrees of freedom. Observer data that include both discarding and retention can be completely independent of official archives of logbook and landings data, but WCGOP estimates of just discarding require linkage of all three data sets which, aside from reasons (a) to (c) just referred to, is a highly complicated process that directs much scientific effort to the making of routine judgements about the matching and adjustment of records. It also forces long delays in the production of observer data while PacFIN is brought up to date by three states. I conclude that the linking process, whilst obviously carried out with great care at NWFSC, is potentially error-prone, and almost unverifiable externally because of the time and specialised knowledge required. I am also concerned that the complexity of the linking and adjustment process may make fisheries data collection processes seem like ‘black boxes’ to fishers and managers leading, possibly, to distrust of the scientific work.

4.2 Findings on distribution of sampling

I agree that distribution of observer effort among towns along the west coast in relation to fishing effort in each region is a sensible, practical way to organise the program. It was re-assuring to be shown GIS plots that demonstrated that the geographic distributions of hauls observed in 2006 closely matched those of hauls fished as taken from logbooks.

The word ‘stratification’ was used much during the review but mostly in relation to the grouping of observed hauls for estimation purposes, not in relation to the original sampling process. It can be asked, therefore, whether stratified sampling utilising different sampling rates per unit of effort in the different regions would be more efficient than the existing scheme. Bearing in mind the highly multivariate (i.e. multispecies) nature of observer sampling, I doubt this because sampling could only be optimised for one species at the expense of inefficiency for others. Nevertheless, a linear modelling study of existing WCGOP results might reveal consistent differences of discarding with latitude within each season and gear type. This might be helpful for optimising sampling allocations to different port-linked strata but it

could also be of wider interest for management of the fishery generally. Suggestions for a linear modelling approach are given in section 4.10.

4.3 Findings on vessel selection

The WCGOP vessel selection procedure represents a practical approach to sampling, bearing in mind the possible trip-limit and observer effects. However, the protocol of observing all trips by each selected vessel during a 2-month period represents a constraint on randomisation implying that bias may be occurring and that sampling precision would probably be over-estimated using standard sampling formulae. The reason for this is that a sequence of trips made by one vessel is likely to find a more biased mean discarding rate, and less variance than the same number of trips made by randomly selected vessels from the entire fleet. In WCGOP, the 'sampling unit' is the (vessel, 2-month period) duple, whereas under simple random sampling in which individual trips are observed on randomly selected vessels, the sampling unit is the (vessel, trip) duple or 'trips' for short. If, on average, *n* trips are made by a vessel during 2-months, the effective sample size is theoretically increased by *n* times when trips are sampling units and the same number is observed on different vessels; the variance of the mean discarded quantities should be reduced correspondingly by dividing by *n*. This could represent a substantial gain in sampling precision at little or no extra cost, though practical constraints on sampling may prevent full realisation of the theoretical gain.

Whether or not the observer effect is actually diminished by a 2-month sampling period as opined at the review is likely to remain a matter of subjective judgement because observer effects are extremely difficult to estimate satisfactorily. My view is that it is best simply to try to diminish observer effects by maintaining good relationships between observers and fishers, and by avoiding all enforcement roles for observers. Concerning trip-limit effects, i.e. increases in discarding at the end of each 2-month trip-limit period, application of a model such as that by Helser et al. is needed for the precision of estimation to benefit from knowledge of the effect. Estimates then become dependent on the assumptions of the model, particularly on whether it is truly generally applicable. I understand that such modelling is not being routinely applied in this way and I tend to doubt whether, if it were, it would produce the multiplicative gains in precision that can be expected from switching to a trip-based sampling unit.

There remains therefore a two-way trade-off between the unknown observer effect which is reasonably expected to be reduced by 2-monthly sampling on each vessel, and the greater effective sample size that would be obtained by putting observers on newly selected vessels for every trip. My preference is for independent random sampling of trips because of its clear theoretical benefits to precision. It could also turn out to be more practical than a 2-monthly observation period per vessel since observers are not tied to a vessel that may turn out to be only sporadically active during the period. Also fishers do not have the possible inconvenience of an observer on board for a whole 2 month period.

I have a second reservation about the current vessel selection procedure. Although randomised, all vessels are selected during a sampling cycle so it is the sequencing, not the sample that is randomised. By selecting *all* vessels for one sampling cycle, the results in one 2-month period become dependent on those of previous periods

because previously observed vessels will not be observed again during the cycle. By contrast, a random sample of vessels could consist of any combination of vessels, either all different or with the possibility of repeated selection of some vessels depending on whether sampling was *with*, or *without* replacement, respectively.

My recommended strategy is to pick trips *without* replacement from a population of trips made by all vessels during a preferred sampling period, e.g. a 2-monthly trip-limit period or a 6-monthly season. The following procedure achieves this: (1) randomly select vessels *with* replacement before the period begins, then (2) observe them in the sequence of drawing so far as practically possible. With this scheme, any vessel can occur more than once, but each trip made in the period can only occur once in the sample. By the theorems of expectation, an unbiased estimate of the average discard rate is the average of the observed trips, and the total discard rate is the average multiplied by the total trips made by the fleet - if that number is known. Alternatively, the number of trips made by only the observed vessels during the sampling period might be an easier expansion factor to find. Then each observed vessel's total discarding for the sampling period can be estimated first, and simply expanded by the number of vessels in the fleet to get the fleet-level estimate. The two estimates would differ somewhat because of the different information put into them. Some simple examples of expansion of observer data to fleet level are discussed in Fish and Fisheries (2007), vol 8, pp. 123- 152.

4.4 Findings on observers

Having a large team of observers spread over a long coastline raises concerns about quality control of the results they bring back. The current system of initial and refresher training, de-briefing and regular meetings between NWFSC staff and observers provides a high level of quality control but there remains a question in my mind of whether it can and should be higher. Much depends on turn-over rates for contracting companies and for the observers employed by them since experience counts for a lot in the observer job. There is also a need for good relations with NWFSC staff, and empathy with the program's objectives, both of which can be assisted by long association. Though I learned of no problems in these areas in the WCGOP, it is possible that data quality could be controlled better with observers as NWSFC employees. At Cefas, most observers are employees, and all new observers are accompanied to sea on their first trip by an experienced observer who signs them off on file as competent and safe before they are permitted to sail solo. Later, other accompanied trips are made occasionally to check standards. Being our employees, they can always be reached should any data quality problems surface.

Species identification is particularly important for the WCGOP program with so many species being of relevance to management. Being able to identify '80% of species', the expected level of proficiency, is not quite the same, or as relevant, as being able to identify all priority species confidently. I did not examine this area of the program carefully but suggest that, if it does not currently exist, a priority list of species should be drawn up that observers should know well-nigh perfectly. Assigning discarded amounts to the wrong species is worse than not sampling at all and could go unnoticed for a long time without occasional accompanied quality-control trips.

4.5 Findings on observers' duties at sea

I have already raised a question about the need for estimating weights of fish at sea. Given that this practice must continue, the observer manual appears satisfactorily to

describe the several options for attempting the task or, at the least, for preparing for practical training. I agree with the approved approach on deck of isolating over-fished and re-building stocks for weighing purposes since they would otherwise be very poorly estimated if they formed a small proportion of the overall catch, as would often be the case.

We did not discuss at the review how big samples of 'categories' of mixed species should be to adequately represent species composition, nor do I claim to have practical experience of this aspect. I mention it as being worthy of attention if it has not already been considered. It is a multinomial sampling problem with, I suggest, sample size determined by the need for a minimal number of the rarest priority species to be present in the sample.

I am not comfortable with the idea of estimating the density of fish in containers on deck in order to estimate weight from volume; I suspect that this would be very difficult to accomplish with an accuracy that improves on that obtained with a single, standard density factor for all fish or, perhaps, each species. Also, I note that NWSFC has no estimates of the precision of observers' estimates of weights. Precision could be estimated using a set of trips, each of which is specially organised to have two observers on board who then estimate weights of different categories independently and without conferring until the end of the trip. The average log ratios of weight estimates by the pair of observers on board on each trip would, if near zero (=log 1), add to the credibility of WCGOP of, if distant from zero, might lend support for changing observer duties from estimation of weights directly to estimation of numbers-at-length with weights estimated subsequently by the allometric growth equation. Of course, additional, similar sea-going work could be organised to estimate the precision of weights by the latter, indirect method. Variance arises in it through sub-sampling of the length distributions and their expansion to haul level.

4.6 Findings on debriefers

The debriefing process is bound to depend on the skill and alertness of the debriefers. The only options I can imagine for improving the process (if they are not already in place) are (a) two-person data checking with one reading out loud from the original paper sheets, the other verifying the computer records, and (b) regular geographic rotations of debriefers to different port groups along the west coast so that all locales are maintained at the same standards.

4.7 Findings on the use of fish tickets

I have already commented on the complexities and delays of matching observer records to fish tickets, see section 4.1.2. The task could be obviated by having observers return quantities retained as well as discarded for each haul. This would require observers to design their catch sampling procedures afresh for almost every new vessel they sample, and it may also involve large volumetric expansion factors to convert from the sample of discarded or retained components to the whole catch. Neither task is impractical for a well-trained team. Sampling of both catch components has been achieved by observer teams elsewhere and the advantages of observer data that are timely, directly informative about factors affecting discarding rates, and independent of PacFIN could well outweigh the difficulties of re-organising catch sampling procedures and re-training observers. Observer estimates of quantities retained will vary from fish-ticket quantities landed because of the sampling process in WCGOP. If it is still required that the two estimates match exactly, it would be

better to manipulate the two final estimated totals in a research report with all methods clearly documented rather than continue with the existing undocumented (as I understand) record-by-record adjustments that depend so highly on the knowledge and diligence of the data processor.

4.8 Findings on trawl logbooks

Use of trawl logbooks by WCGOP raises objections similar to those I have raised about use of fish tickets, except that reference to logbook data on PacFIN adds significantly to the problems of incompleteness, matching, and delay. Ideally, the observer program would not derive weights of fish retained from trawl logbooks via PacFIN because of the severe difficulties of doing so in some cases. I also have doubts about the value of using quantities of retained fish hailed to the skipper by the crew for recording in the logbook, and their subsequent adjustment to align with fish ticket weights. My reasons are the cross-linkages between data sets, the delays involved, the intricacy of the process of adjustment, and the assumed supremacy of fish ticket weights. It is fair to state that fish tickets are legal, commercial documents, but this does not guarantee that they will be reliable, e.g. when an honest fish ticket would reveal infringement of a trip limit.

On the other hand, logbooks are useful for obtaining positions and times if they can be read directly in the wheelhouse during the observed trip. I understand that progress is being made by NWFSC with equipping observers with GPS devices that will allow them to find haul positions independently of the ship's logbook and this seems to me to be a worthwhile effort particularly if the data recorded in the wheelhouse are not directly available or reliable on every trip. Our experience at Cefas is that regular way-points should be recorded for trawl tows as well as start and end positions because many trawlers tow along circular paths. We are usually permitted to read wheelhouse navigation equipment.

4.9 Findings on forward projection of total fishing mortality

I comment on one aspect. The age of the most recent observer data supplied to the projection model is a matter of possible concern; it is a question of how much the discarded quantities for each managed species vary from year to year and season to season. If the variation is large, a 1- to 3-year lag in the supply of data could be a significant impediment to effective management of the fisheries. However, for this to be true, the variation must be real, not sampling error, so it is important to know what the sampling error is. The bootstrap variances presented during the review appear helpful for judging this aspect. If the time-dependent variance is thought to be significantly greater than the observer sampling variance, then there is a clear need to try to improve the timeliness of observer data used in the projection model. If, contrarily, sampling error is dominant, there probably is not.

I have already noted that observer data could be produced more quickly if linkages with PacFIN were broken. However, if that were done, a further check would then need to be run on the sampling variance of observer data which might have changed as a result of new estimation procedures.

4.10 Findings on current estimation of total fishing mortality

The grouping of haul-level data into strata by geographic area, season, and depth for the purposes of the estimation model is inconsistent with the nature of the WCGOP randomised sampling scheme, suggesting that bias and sampling inefficiency could be

reducing the benefits of the program. I think there are several specific problems. At the end of this section, I suggest an alternative, design-based method of estimation, plus a linear mixed model for haul-level data that would be instructive and may also provide the required fleet-level estimates.

The clustering of samples around vessels, trips, and 2-monthly periods is not being recognised in the estimation model. Instead, hauls are being re-grouped without recognition of their sampling origins. As a result, mean levels of discarding in a stratum of hauls classified by an area, season, and depth range are not just affected by those factors but also by the possibly much larger effects of the small numbers of vessels and trips that visited that stratum during each 2-monthly period. Since a vessel is likely to fish in a restricted locality on trips during a 2-monthly period, since some vessels are much better at catching fish than others, and since some trips are much more productive than others, a high degree of spatial auto-correlation can be expected in the haul-level data, implying that variability among strata is likely to be higher than if hauls were the independently, randomly chosen sampling units - something that would require a helicopter to achieve!

The use of discard ratios in the estimation model for the purpose of expanding observed discarded weights to the fleet-within-stratum level is likely to be imparting an unusual and difficult error structure to the estimates. This is because the denominator in the ratio is the quantity retained, either of one target species or of several. Ratios are inestimable when the denominator is zero as sometimes occurs but, more seriously, the ratios are likely to be highly variable, multiplicatively so, when the denominator is small, as it may be in several strata. Conversely, the variance of the ratio is driven by the variance of the quantity discarded when the denominator happens to be large. Thus the variance is itself likely to have a very large variance depending on the relationship between numerator and denominator which may vary with species and circumstance. For example, catching of large numbers of young fish would cause a high discard ratio and large variance of the estimated weight discarded. These deductions imply that use of discard ratios is imparting large and possibly erratic components of variance to the fleet-level estimates of discarding. Large sample sizes may be reducing these effects to acceptable levels (through the central limit theorem) but do not alter the low efficiency of discard ratios in the estimation process. Subsequent expansions of estimates using logbooks and fish tickets appear sensible to me given that this method of estimation is in operation.

Having in mind these reservations about the current estimation model, I have thought further about design-based estimation methods which we touched upon during the review. The WCGOP sampling scheme appears to fit within the cluster sampling mould of Sampling Theory. Using standard terms, the (vessel, 2-month period) duples are primary sampling units (p.s.u.); trips within p.s.u. are secondary sampling units (s.s.u.); and hauls within s.s.u. are tertiary sampling units (t.s.u.). The population from which the p.s.u. are drawn randomly is the set of all (vessel, 2-month period) duples within a sampling cycle. Standard cluster sampling estimators are designed for this sort of sampling structure given that all trips and all hauls are sampled for each p.s.u.. If not, multi-stage sampling estimators may be needed. Total discard.lbs per average vessel in a 2-monthly trip-limit period is estimated first, and this value expanded to a fleet-level estimate using the total number of (vessel, 2-

month period) duples in the sampling cycle. Design-based estimation appears to me to be well worth exploring for estimating total discarding (or catch) mortality because:

- it is consistent with the sampling scheme and therefore should theoretically be unbiased;
- it makes minimal assumptions, in particular, it is model-free; and
- it is quick and easy to calculate.

However, design-based formulae are not necessarily the most precise way of estimating total mortality. Application of more information, e.g. hours of trawling, numbers of trips, if available, using ratio estimators might produce better estimates (though knowing this could be difficult).

Linear modelling of haul data is another possible approach to estimation and is also likely to assist with understanding of the discarding process. Linear models would have a more regular error structure than arises from the current discard-ratio (i.e. non-linear) estimation model. Models usually evolve in structure as one tries to fit them and, since I have no experience of this with WCGOP data, my suggestions have to be considered as tentative. With this health warning, a linear mixed model for starting the modelling process might be (omitting regression coefficients and error terms):

log(*discard.lbs.per.haul*) = (*vessel, 2-month.period*) + *trip.nested* + *area* + *depth* + *season* + *year*

trip.nested would be a random effect nested within (*vessel, 2-month.period*) which, along with all other explanatory factors, would be fixed effects. [Random effects can only be nested within fixed effects because the two must be statistically independent when fitting with the usual methods of linear mixed models.] The benefit of such a modelling exercise would be that effects due to *area*, *depth*, and *season* are estimated with allowance for the clustering variables, i.e. (*vessel, 2-month.period*), and *trip*. The estimated effects could then be used to find mean(log(*discard.lbs.per.haul*) for each stratum in the S1 (or other) schemes. An expansion factor chosen from whatever information is generally available is still needed to estimate fleet-level discarding. Mixed-effects models can be fitted in R using lme(); see Pinheiro, J.C. and Bates, D.M. (2000) 'Mixed-effects models in S and S-plus', Springer, New York.

4.11 Findings on alternative re-stratification and expansion schemes

As already noted, the haul-level data uncorrected for mean vessel, trip, and 2-monthly-period effects are likely to display significant spatial autocorrelation. Re-stratification of the haul-level data without differential weighting of different strata appears to amount to little more than a different viewpoint on the same, spatially related information. If I am right, this would explain the generally small effects of different re-stratification schemes. The variation that was observed may be explicable just in terms of the adjusted quantities retained in different strata leading to large variability in the discard ratios (see 4.10), and the reported uncertainties associated with assigning hauls to strata for trips that crossed stratum boundaries. My view comes from making a clear distinction between pre- and post-sampling stratification; the first is a subject of generalised Sampling Theory, and the second is done for modelling or administrative purposes and is always conditional on the particular sample of data that is analysed.

By contrast, the tests of expansion methods did inject new information into the calculations, namely the tow-durations and numbers of tows. I am not really surprised that relationships between tow duration and weights discarded were very weak at the haul level; there is typically such a high degree of variability in fishing operations that weak effects can easily be masked. I reason, nevertheless, that quantities caught and discarded must bear some relationship to tow duration when averaged over many vessels for a significant period because, if they did not, most fishers would take the easy option of very short tows. Also, the chances of encountering fish must surely increase as tow duration increases. It follows that tow duration and numbers of tows – the former carrying the most information about fishing effort – *are* worthy candidates for expanding observer data. Furthermore, since they then occur in the numerator line of calculations, not in the denominator, they are less likely to impart erratic error structures to the fleet-level estimates of discarding. The main question in my mind is whether it is worth taking fleet effort information from PacFIN – with all the associated problems as already discussed. I tend to think not. Instead, I prefer the relatively straightforward, easily explicable, and quickly implemented solution of estimation with design-based statistics which expand to fleet level according to numbers of vessels, or numbers of trips if that data is readily available without resorting to PacFIN. Another, long-shot possibility would be to have port-based observers and biological samplers conduct a separate survey of fishing effort by interview, observations of presence in port, or by sampling wheelhouse logs, if this fits well with their other responsibilities and is acceptable. The information might provide useful effort-based expansion factors in a more timely way than is available from PacFIN. Other basic information on the fishery might also be obtainable in this way, possibly with benefits to relations with the industry from the additional presences on the quayside.

4.12 Findings on bootstrap studies

I find that the bootstrap studies provide useful preliminary estimates of the reliability of discard estimates but that they are to an unknown extent affected by the same weaknesses that I found in the estimation studies (3.11, 4.11), namely that vessel fishing power, trip, and 2-monthly period have not been allowed for. The resulting spatial autocorrelation of haul-level data may therefore account for much of the variation of different stratification and expansion schemes found with the bootstrap method. Indeed, it is quite possible that the bootstrap CVs are actually larger than would be found using design-based estimation, or modelling and correction for the spatially correlated effects.

4.13 Communications

In my opinion, communications from scientists to the fishing industry concerning an observer program and its results are very important both for smoothing the work of the observers and for gaining co-operation from the industry with management. I feel that communications from WCGOP could be improved since I understand that there are very few at present, aside from heavyweight technical data reports. I acknowledge that there are constraints on communication, namely the need for confidentiality of information, e.g. about individual, and small groups of fishers, but these constraints need not prevent all communications.

1. One communication we consider very important in England is the confidential trip report sent as a courtesy to the vessel skipper or owner after sampling by an

observer. I understand that WCGOP could only send such a report to the permit holde,r but he or she probably would not object to distributing it to skipper and crew. Our trip report is prepared after data have been entered and checked on the computer. It contains quantities retained and discarded of the main species illustrated with pie charts, possibly also a plot of the fishing course and a break-down of catches by haul. The report must be produced quickly and automatically using a simple program.

2. Another possible communication is some kind of annual report. This is more labour intensive. Cefas managed one for 2004, see www.cefas.co.uk/publications/techrep/tech134.pdf, but it has not been followed up unfortunately, possibly because it was too complicated. With care and an investment of time, it could be designed to be simple, semi-automatic to produce, and informative, not just for the fishing industry but also for fisheries managers and scientists. It could be web-based to keep costs down.

A third type of report we used at Cefas was circulated confidentially among scientists. After each trip, the observer writes down any non-numerical information learned in a fixed format with headings: (a) fishing and gear information, (b) biological information, (c) views of fishers, and (d) safety. Such reports can be quite informative to specialists and help to prevent surprises when fishers have an opportunity to air their views in public meetings. The reports need to be indexed or searchable electronically by scientists.

5. Conclusions and recommendations

Observer sampling of fishing trips to estimate discarding is a difficult task at the best of times and, even though many observer programs exist around the world, there seems still to be no clear consensus on the right way to do the job. Part of the problem is the variety of fisheries observed; a method that works in one may be unsuited to another. Another part is the adaptation of ideas from Sampling Theory to a fishery context; the text books are written in a general way and seldom mention fishing fleets, their unpredictable activities, movements from one group to another, the serendipity of catches, and numerous other factors that blur the application of theory. A third part of the problem comes from the natural search for precision; small CVs can be obtained by restricting which vessels and trips are observed, e.g. by looking for “typical” examples, but, in that case, precision for a given sample size is likely to be obtained at the cost of extra bias, an error that does not show itself in the CV. Yet another aspect concerns the uses to which the observer data are put. Usually they are applied to management and assessment of the fishery together with other information such as landings statistics, biological sampling of landings, fishers’ logbooks, tagging data, and so on. These data sets are often treated as separate sources of information, able to enhance or modify conclusions drawn from other data sets when, in fact, the final numbers being used by the management or assessment committee are the product of intricate prior cross-linkages between data sets, a process that leads to a tendency for each of the sets to impart nearly the same message, perhaps without the users realising it.

Readers may be interested to know something about my underlying opinions as background to the conclusions and recommendations of this report. In my opinion:

- Observer programs should conform as closely as can be organised in practice with the ideas of Sampling Theory because it represents a consolidation of generalised sampling logic built up over many years. Estimators may be design- or model-based, or a blend of both.
- Observers should record as much information as is safely and reliably obtainable, wtihout excessive tiredness, under most conditions at sea.
- Discarding is sometimes a short-term issue, e.g. when related to large, young year classes in an area. Observer data should therefore be available for use quickly, e.g. within three months, preferably within one. Speed is usually preferable to detailed refinements of precision.
- The best models of the relationships among variables are fitted to multivariate observations from the same sampling unit. Eg. measurements of discarding and retention and possible explanatory variables are, ideally, made for the same hauls. Conversely, when discarding is measured on hauls but retention can only be estimated for, say, the entire trip, information on the relationship between discarding and retention is significantly diminished. Consequently, 'poor' estimates of retention per haul, if available, are likely to be more informative than 'good' estimates for the trip.
- Communication between scientists and the fishing industry may seem like a one-way process but it is still very important, particularly when catches decline for some reason and times are hard for fishers. Observer programs can achieve a lot towards this goal.
- Databases should be self-contained, independent repositories of objectively verifiable data only. That way, they can be looked to for fresh information uncontaminated by relationships with other databases.

The following paragraphs summarise conclusions and recommendations taken from section 4.

5.1 Primary sources of uncertainty in the estimates or projections

5.1.1 Spatial relationships among hauls

The estimation model uses estimation of discard ratios from hauls grouped by geographic, depth-related, and seasonal strata. I am concerned that hauls are not independent sampling units. The hauls in a locality will tend to have been made by a small number of vessels, and a restricted set of fishing trips. A spatial relationship among hauls can therefore be expected as a result of the fishing powers of particular vessels and the successes of particular trips. This may be as, or more important than spatial variability arising from region, depth, and season *per se*. The consequences of this mismatch between the estimation model and the sampling scheme are difficult to gauge but, I suspect, it adds variability within and between years, the latter representing an annual bias. Reliability of apparent trends in discarding are therefore likely to be diminished. See 4.10.

5.1.2 Variance from discard ratios

The use of retained quantities of fish based on PacFIN data in the denominators of discard ratios means that many estimates of fleet-level discarding are derived with a factor that is potentially highly variable, since retained weights range from zero to very large. This translates directly into varying uncertainties for the discard estimates. Unfortunately, it may not be possible to see which estimates are affected badly, and

which not, because it is the retained weights in the individual strata that are relevant, not the overall fleet-level of retention shown in data tables with fleet-level discarding. See 4.10. There is also a step-change in variance between target species that have retained quantities of the same species in the denominator, and bycatch species that have retained quantities of all target species in the denominator. This complicates the task of users of the data who have to try to understand the reliability of discard estimates for a range of species.

5.1.3 Fish tickets

Fishers and collaborating merchants have high incentives to under-declare landings of species controlled by trip limits. This is especially the case given that trip limits may be adjusted downwards in mid year for some species if it appears that the OY is being approached too closely. Inaccurate landings data have long been acknowledged in Europe (Report 1985/86 by ICES statistician. ICES CM 1986/D:19) and I was surprised that fish tickets seem to be treated as a kind of 'gold standard' on the US west coast, valid for adjusting observer and logbook data. A check on this could be obtained from the observer program if it can be re-organised to sample and record quantities retained as well as discarded at sea. Not only would this give better estimates of discarding rates (% of catch), but it may also be instructive to stock assessors, especially if standard errors were available.

5.1.4 Mis-matching of observer records with fish tickets and logbooks in PacFIN

Fleet-level estimates of discarded and retained fish made with the current estimation model involve retrieval of adjusted data from PacFIN, then matching of records with observer records. Both adjusting and matching can be laborious and somewhat subjective. Errors must surely be made. Unfortunately, options aside from "assume everything is correct" are hard to find and implement so the rates and consequences of errors are hard to estimate. See 4.1.2.

5.1.5 Variability of discard weights estimated at sea.

I understand that there are no estimates of the precision of the catch and discard weights estimated by observers at sea. Is it +/-5%, +/-20%, or +/-75%? A method I have suggested may be feasible. See 4.5.

5.1.6 Time variability of discard data used in the projection model

The projection model is used to plan and adjust trip limits. Observer data are an important input to this process but, because of processing delays, they tend to be a year or more old when they are used. The importance of this point depends on how much real, time-dependent variability of discarding is occurring in the time series. If there is a lot, then the projection model is likely to be giving poor results because the discard data are old. On the other hand, if most of the variability in the time series is attributable to sampling error, there is less of a problem – aside from the fact that discarding is only estimated within wide limits. The reported bootstrap studies of estimation variance provide some assistance with this matter. See 4.9.

5.2 Strengths and weaknesses of current approaches

The strengths of the current approach to observing and estimating discards include:

- The fact that it works and produces required data; it is also conforms with the Magnuson-Stevens Act.

- It conforms with the design of the projection model. This means that differences between projections and observations are less likely to be modelling artifacts; however, I doubt whether it removes all risk of spurious differences.
- The distribution of observers along the west coast in relation to expected local fleet sizes assures that results are not biased for the coastal fleet by exclusion of distant or inaccessible vessels. See 4.2.
- The 2-month observation period for every selected vessel is a reasonable approach for diminishing possible effects on normal fishing of having an observer on board. However, I am not convinced that this is giving more benefit for sampling efficiency than would be gained by selecting individual trips on different vessels. See 4.3.
- NWFSC is making considerable efforts to ensure that data returned by observers are of high quality. Possibly, these procedures could be improved still further.

The weaknesses of the current approach, aside from the uncertainties listed in section 5.1, include:

- The importance attached to direct estimation of weights of fish by WCGOP: Weights tell nothing about age structures of fish stocks, recruitment, size, growth, or gear selectivity effects related to size, gender or reproductive maturity. As an example, consider table 6 in the Data Report for 2006; tabulated discard rates and percentages by weight give no indication of whether a few large individuals were discarded – perhaps due to trawl damage – or whether a year class just recruited to the fishery has been virtually wiped out. Most readily collectable biological variables are collected routinely for priority commercial species in European observer programs, though admittedly we have fewer species to consider and lighter catches in general. Weights can be estimated indirectly from numbers-at-length and may in fact be more accurate than weights estimated without weighing machines on fishing vessels. See 4.1.1 and 4.5.
- The lack of observer estimates of quantities of fish retained: One consequence is that landings data must be taken from PacFIN causing potential for additional errors and biases, linkage between the observer database and PacFIN to the detriment of independence; diversion of extensive scientific time to deal with these computing tasks, substantial delays in the production of discard results, and impediments to understanding of haul-by-haul relationships between discarding and retention. See 4.7. Another consequence arises if PacFIN fish tickets are sometimes under-declared, as seems likely (see 5.1.3): estimated discarding rates would then be too high.
- The reliance on logbooks from PacFIN for expansion of observer estimates of discarding by geographic stratum to fleet-level estimates in the discard estimation model: My objections are much the same as for estimation of retained quantities from PacFIN, above, though the delays are worse for logbooks than fish tickets. Other expansion factors based on fishing effort, e.g. number of vessels, number of trips, if available without using PacFIN, would be much easier and quicker to use than logbooks. See 4.8.
- The constraints on randomisation: Vessels are selected so that all are observed in a random sequence during a sampling cycle. This is not the same as a random sample of vessels taken in a season of interest, suggesting that bias and serial dependence are resulting from the WCGOP procedure, particularly when results are presented for time periods that do not match with the sampling cycle. The 2-

month observation period on each vessel has supporting practical arguments but represents another constraint on random sampling suggesting further bias is a possibility. Reduced sampling efficiency also follows. See 4.3.

5.3 Alternative model configurations or approaches

If vessel selection and observation under WCGOP is to continue in much the same way, I recommend that the cluster or multi-stage sampling approaches offered by Sampling Theory be applied, and without use of PacFIN or geographic post-sampling stratification of hauls. See 4.10 for detailed suggestions for making fleet-level estimates. Quite possibly, the sampling formulae could be easily applied to existing observer data so that the performance of such a system could be assessed before any changes were considered. If so, I recommend such a study at an early opportunity.

The primary sampling unit under the current approach is the vessel, 2-month period,which unfortunately confounds vessel and date effects. The 2-monthly observation period reduces sampling efficiency. Also, the length of the sampling cycle is arbitrary, depending on the number of vessels, not the factors that affect fishing. I therefore recommend that the sampling scheme be changed to use trips as sampling units, and to use seasonal periods for the sampling cycle. See 4.10. Precision depends on the number of trips observed and one would hope for at least 30 or so in each season and stratum; hence the estimable seasonal detail would depend on the observer resources and any geographic stratification of whole trips thought necessary. My preference would be to continue with 6-monthly 'seasons' and a minimum of geographic strata. Sampling strata improve detail but reduce precision; one cannot have it both ways.

I recommend that WCGOP switches to observing quantities retained as well as discarded even if precision is low for each haul. Precision of the mean result will be improved by large numbers of observations, as should be available under WCGOP. See 4.1.2

I also recommend switching observers' duties from estimating weights to estimating numbers-at-length and biological variables if this is politically acceptable and practically feasible for WCGOP. See 4.1.1. It may be necessary to restrict observers' attention to a manageable number of listed priority species so that 'zero catch' does not become confusable with 'no information'.

The WCGOP program now has a large resource of data for the purposes of modelling and understanding the fisheries, possibly also to improve on estimation with purely design-based estimators. Tentative suggestions for linear modelling are put forward in section 4.10.

I recommend that WCGOP aims to produce more, simply-formatted informative material for the fishing industry, especially those who have co-operated with the observer program. The material might be printed, or it might be on a web-site – particularly if the delays in finalising estimates can be reduced. See 4.13.

6. Appendix 1: Statement of Work

Statement of Work for John Cotter (CEFAS)

External Independent Peer Review by the Center for Independent Experts

Review of the West Coast Groundfish Bycatch Model and

Methods Used to Estimate Total Mortality

Project Background:

The Northwest Fisheries Science Center (NWFSC) has produced reports documenting total fishing mortality for groundfish for the years 2004-2006, which are available on the NWFSC's website (Hastie 2006, Hastie and Bellman 2006, 2007). A central element of these analyses is the estimation of discard occurring in the non-hake groundfish bottom trawl fishery. This estimation process incorporates data from at-sea observations, logbooks, and fish tickets. In addition to the estimation of discards for prior years, the NWFSC has developed a model used in the projection of trawl catch (landings and discards) in upcoming fisheries. The methodology used in the catch-bycatch model was initially reviewed by the Pacific Fishery Management Council's Scientific and Statistical Committee (SSC) in early 2003. We have determined that a review of major science products every five years is a best practice and the catch / bycatch model should be appropriately reviewed at this time.

Overview of CIE Peer Review Process:

The Office of Science and Technology implements measures to strengthen the National Marine Fisheries Service's (NMFS) Science Quality Assurance Program (SQAP) to ensure the best available high quality science for fisheries management. For this reason, the NMFS Office of Science and Technology coordinates and manages a contract for obtaining external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of stock assessments and various scientific research projects. The primary objective of the CIE peer review is to provide an impartial review, evaluation, and recommendations in accordance to the Statement of Work (SoW), including the Terms of Reference (ToR) herein, to ensure the best available science is utilized for the National Marine Fisheries Service management decisions.

The NMFS Office of Science and Technology serves as the liaison with the NMFS Project Contact to establish the SoW which includes the expertise requirements, ToR, statement of tasks for the CIE reviewers, and description of deliverable milestones with dates. The CIE, comprised of a Coordination Team and Steering Committee, reviews the SoW to ensure it meets the CIE standards and selects the most qualified CIE reviewers according to the expertise requirements in the SoW. The CIE selection process also requires that CIE reviewers can conduct an impartial and unbiased peer

review without the influence from government managers, the fishing industry, or any other interest group resulting in conflict of interest concerns. Each CIE reviewer is required by the CIE selection process to complete a Lack of Conflict of Interest Statement ensuring no advocacy or funding concerns exist that may adversely affect the perception of impartiality of the CIE peer review. The CIE reviewers conduct the peer review, often participating as a member in a panel review or as a desk review, in accordance with the ToR producing a CIE independent peer review report as a deliverable. The Office of Science and Technology serves as the COTR for the CIE contract with the responsibilities to review and approve the deliverables for compliance with the SoW and ToR. When the deliverables are approved by the COTR, the Office of Science and Technology has the responsibility for the distribution of the CIE reports to the Project Contact. Further details on the CIE Peer Review Process are provided at http://www.rsmas.miami.edu/groups/cie/

Requirements for CIE Reviewers:

One CIE reviewer shall participate in the Review Panel for the West Coast groundfish bycatch model and estimation of total mortality. The CIE reviewers shall have expertise in the collection and analysis of bycatch information.

The CIE reviewer's duties shall not exceed a maximum total of 7 days for pre-review preparations involving review of documents, participation during the 3-day panel review meeting, and completion of the CIE independent peer review report.

The CIE reviewers shall have the requested expertise necessary to complete an impartial peer review and produce the deliverables in accordance with the SoW and ToR herein.

Statement of Tasks for CIE Reviewers:

The CIE reviewers shall conduct necessary preparations prior to the peer review, conduct the peer review, and complete the deliverables in accordance with the ToR and milestone dates as specified in the Schedule section.

Prior to the Peer Review: The CIE shall provide the CIE reviewers' contact information (name, affiliation, address, email, and phone), including information needed for foreign travel clearance when required, to the Office of Science and Technology COTR no later than the date as specified in the SoW. The Project Contact is responsible for the completion and submission of the Foreign National Clearance forms (typically 30 days before the peer review), and must send the pre-review documents to the CIE reviewers as indicated in the SoW.

Foreign National Clearance: The CIE shall provide the necessary information (e.g., name, birth date, passport, travel dates, country of origin) for each CIE reviewer to the COTR who will forward this information to the Project Contact. The Project Contact is responsible for the completion and submission of required Foreign National Clearance forms with sufficient lead-time (30 days) in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations at the Deemed Exports NAO link http://deemedexports.noaa.gov/sponsor.html

Pre-review Documents: Approximately two weeks before the peer review, the Project Contact will send the CIE reviewers the necessary documents for the peer review, including supplementary documents for background information. The CIE reviewers shall read the pre-review documents in preparation for the peer review.

Documents to be provided to the CIE reviewers prior to the STAR Panel meeting include:

- Estimated 2006 Discard and Total Catch of Selected Groundfish Species
- Estimated 2005 Discard and Total Catch of Selected Groundfish Species
- Revised Estimated 2004 Discard and Total Catch of Selected Groundfish Species, September 2006
- Additional supporting documents as available.
- An electronic copy of the data, the parameters, and the model used for the estimation of total mortality (if requested by reviewers).

Additional background documents may also be provided.

Peer Review during Panel Review Meeting: The CIE reviewer shall participate and conduct independent peer review during a panel review and the dates and location of the meeting is specified in the Schedule of Milestones Deliverable, and attached tentative Agenda (Annex 2). The Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The CIE Program Manager can contact the Project Contact to confirm the facility arrangements.

Independent CIE Peer Review Reports:

The primary deliverable of the SoW is the CIE reviewer shall complete and submit an independent CIE peer review report in accordance with the ToR, and the CIE report shall be formatted as specified in Annex 1.

The CIE report is to be based on the CIE reviewer's findings in accordance with the ToR, and no consensus report shall be accepted. The CIE reviewer shall comment on whether the best available science was utilized in the bycatch model, and shall not comment on management decisions.

Terms of Reference:

1) Review background materials and become familiar with the West Coast groundfish bycatch model and methods to estimate total mortality;
2) Actively participate in the review Panel to be held in Seattle, Washington from July 15-17th, 2008; Participants are strongly encouraged to voice all comments during the review panel meeting so the analysts can address and respond to comments during the Panel meeting;
3) Comment on the primary sources of uncertainty in the estimates or projections;
4) Comment on the strengths and weaknesses of current approaches;

5) Recommend alternative model configurations or approaches, as appropriate;
6) Complete a final CIE independent peer review report after the completion of the STAR Panel meeting in accordance with the ToR and the Schedule of Milestones and Deliverables.

Schedule of Milestones and Deliverables:

13 June 2008	CIE shall provide the COTR with the CIE reviewer contact information, which will then be sent to the Project Contact
1 July	The Project Contact will send the CIE Reviewers the pre-review documents
15-17 July	Each reviewer shall participate and conduct an independent peer review during the panel review meeting
7 August	CIE shall submit draft CIE independent peer review reports to the COTRs
21 August	CIE will submit final CIE independent peer review reports to the COTRs
29 August	The COTRs will distribute the final CIE reports to the Project Contact

Acceptance of Deliverables:

Each CIE reviewer shall complete and submit an independent CIE peer review report in accordance with the ToR, which shall be formatted as specified in Annex 2, via email to Manoj Shivlani (shivlanim@bellsouth.net) and Dr. David Die (ddie@rsmas.miami.edu). Upon review and acceptance of the CIE reports by the CIE Coordination and Steering Committees, CIE shall send via e-mail the CIE reports to the COTRs (William Michaels William.Michaels@noaa.gov and Stephen K. Brown Stephen.K.Brown@noaa.gov) at the NMFS Office of Science and Technology by the date in the Schedule of Milestones and Deliverables. The COTRs will review the CIE reports to ensure compliance with the SoW and ToR herein, and have the responsibility of approval and acceptance of the deliverables. Upon notification of acceptance, CIE shall send via e-mail the final CIE report in *.PDF format to the COTRs. The COTRs at the Office of Science and Technology have the responsibility for the distribution of the final CIE reports to the Project Contacts.

Request for Changes:

Requests for changes shall be submitted to the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the Contractor within 10 working days after receipt of all required information of the decision on substitutions. The contract will be modified to reflect approved changes. The Terms of Reference (ToR) and list of pre-review documents herein may be updated without contract modification as long as the role and ability of the CIE

reviewers to complete the SoW deliverable in accordance with the ToR are not adversely impacted.

Key Personnel:

Contracting Officer's Technical Representative (COTR):

William Michaels
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Stephen K. Brown
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Stephen.K.Brown@noaa.gov Phone: 301-713-2363 ext 133

Contractor Contacts:

Manoj Shivlani, CIE Primary Coordinator
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Project Contact:

Stacey Miller
NMFS, Northwest Fisheries Science Center
2725 Montlake Blvd. East, Bldg ANNEX B, Seattle, WA 98112
Stacey.Miller@noaa.gov Phone: 206-437-5670

Elizabeth Clarke
NMFS, Northwest Fisheries Science Center
2725 Montlake Blvd. East, Bldg ANNEX B, Seattle, WA 98112
Elizabeth.Clarke@noaa.gov Phone: 206-860-3381

ANNEX 1:

Format and Contents of CIE Independent Reports

1. The report should be prefaced with an Executive Summary with concise summary of goals for the peer review, findings, conclusions, and recommendations.

2. The main body of the report should consist of an Introduction with
    a. Background
    b. Terms of Reference
    c. Panel Membership
    d. Description of Review Activities

3. Summary of Findings in accordance to the Term of Reference, including answers to each question in the Statement of Work

4. Conclusions and Recommendations in accordance to the Term of Reference

5. Appendix for the Bibliography of Materials used prior and during the peer review.

6. Appendix for the Statement of Work

7. Appendix for the final panel review meeting agenda.

8. Appendix for other pertinent information for the CIE peer review.

Please refer to the following website for additional information on report generation:
http://www.rsmas.miami.edu/groups/cimas/Report_Standard_Format.html

ANNEX 2:

***DRAFT* AGENDA**

Review of the West Coast Groundfish Bycatch Model and

Methods Used to Estimate Total Mortality

July 15-17, 2008
NOAA Fisheries NWFSC
2725 Montlake Blvd. E
Seattle, WA 98112

Tuesday, July 15, 2008

8:30 am	Welcome and Introductions
9:00 am	Overview Presentations: West Coast Groundfish Observer Program (WCGOP) Data: WCGOP, logbooks, fish tickets West Coast Groundfish Management
12:00 pm	Lunch
1:00 pm	Presentation*:* Bycatch Projection Model and Total Mortality Estimation
4:00 pm	Discussion
5:00 pm	Adjourn For the Day

Wednesday, July 16, 2008

8:30 am	Discussion Continued
10:00 am	Reviewers Develop List of Questions
12:00 pm	Lunch
1:00 pm	Reconvene to Discuss List of Questions
5:00 pm	Adjourn For the Day

Thursday, July 17, 2008

8:30 am	Discussion Continued
12:00 pm	Lunch
1:00 pm	Conclusions and Wrap Up
3:00 pm	Review Adjourned