The use of recreational CPUE statistics in groundfish stock assessments

July 14, 2004

Cynthia M. Jones, Ph.D.,
Center for Quantitative Fisheries Ecology
Old Dominion University
800 W 46th St
Norfolk, VA 23701
Phone 757.683.4497
FAX 757.683.5293
Email cjones@odu.edu

Executive Summary

Modern age-based stock assessment models rely on "tuning" indices to calibrate the model to relative changes in fish-population abundance. Tuning indices are typically fisheries-independent surveys, such as trawl surveys. In the U.S. West Coast groundfish fisheries (encompassing over 80 species) no such fishery-independent survey exists to characterize relative population abundance. One of few potential tuning indices available for these species is fishery-dependent recreational CPUE. Recreational CPUE is rarely used for tuning and its use in the West Coast groundfish stock assessment requires various validation and standardization procedures.

The major problems that have been identified in using recreational CPUE as a tuning index are: 1) lack of regional fine-scale recreational surveys with sufficient historical length to capture changes in fish-population abundance for all but the major species, 2) aggregation of CPUE measures over species and areas, and 3) uncertainty in CPUE time series due to the effects of regulatory restrictions, changes in species behaviors, changes in angler preferences over time, and change over time in the value of a unit of recreational fishing effort due to technological changes, among others. To overcome these problems, ongoing survey procedures need to be changed (for example, using better sampling protocols in state surveys; adding questions on area fished to existing surveys), and historic data corrected for factors that are distorting CPUE as a proxy for abundance.

The ability to standardize recreational CPUE time series is not universal and standardization is more difficult with historic data. These data weren't collected to be a proxy for abundance and it must be modified to serve this purpose. Often these historic data are sparse, some lack sufficient detail, and some were not collected following statistically-valid protocols that insure that the data are representative of the true recreational catch and effort. Even for those data sets where CPUE data are collected rigorously, there are data gaps and sparse data for important species. These characteristics of the available CPUE time series add to the uncertainly in the putative proportionality of CPUE to fish abundance.

Report to the CIE on recreational CPUE indices

A workshop was held in Santa Cruz, California, from June 29-30, 2004 to address methods of improving CPUE time series as tuning indices. During the meetings, scientists demonstrated that methods could be successfully applied in some species to correct for regulatory artifacts in the CPUE data. Similarly, scientists demonstrated that different datasets could be combined to fill in for sparse data, and methods to disaggregate data and mine disparate databases. Clearly, the quality of the recreational time series of CPUE can be improved to provide a better tuning index. Nonetheless, there are real limitations with these dataset because they are not detailed enough to be useful for all species.

Background

Traditionally, age-based stock assessments use data obtained from commercial fisheries to build catch-at-age matrices as input. These models do not rely on fisheries-dependent data alone, but are "tuned" with independent indices of fish-population abundances. Fishery-independent surveys provide among the most reliable tuning indices. However, fishery-independent surveys are not always available for the target species or species group or the fishery-independent survey is conducted at inappropriate times or locations. When fishery-independent indices are unavailable, then fishery-dependent data are used. In the case of West Coast groundfish, stock assessment scientists are exploring the use of recreational CPUE time series as an index of fish abundance to use in their models. Unlike commercial fisheries data, recreational data are more difficult to obtain and has a shorter historic record. In commercial fisheries, there are fewer fishermen with larger vessels which are docked at identified ports. Hence, they can be sampled more directly and easily than recreational anglers. Recreational angling is characterized by much greater participation, smaller vessels (with the exception of the for-hire or charter industry), and with much greater geographic assess to the water. The characteristics of the recreational fisheries have caused difficulties in sampling for catch and effort and data are sparser than for the commercial counterpart. Also, recreational fisheries data have not been routinely collected prior to 1979 for most of the U.S. In 1979, the NMFS launched its Marine Recreational Fisheries Statistics Survey (MRFSS), which now provides a longer than 20-year long time series of recreational catch and effort suitable to analysis of major species over broad regions. In an effort to provide tuning indices for declining West Coast groundfish stocks, stock assessment scientists are exploring this time series of CPUE and others that measure recreational CPUE in more restricted geographic locations. This report addresses the challenges encountered in using these data that were addressed in a workshop in Santa Cruz, California during July 29-30, 2004.

Summary of Work

I, *Review of background materials* - Prior to the workshop on June 29, 2004 I was sent a CD which contained the background materials for the workshop. I reviewed these documents, along with other referenced at the meeting, which included:

1. Letter from Ms. Karen Garrison (NRDC) to Mr. Rodney McInnis
2. Bocaccio Stock Assessment (2003)
3. Bocaccio STAR Panel Report (2003)
4. E.J. Dick Beyond 'Lognormal vs. Gamma': Discrimination Among Error Distributions for Generalized Linear Models - text and figures
5. Kevin Hill Hill, K. and N. Schneider. 1999. Historical Logbook Databases from California's Commercial Passenger Fishing Vessel (Passenger) Fishery 1936-1997. SIO Reference Series No. 99-19.
6. David Sampson Oregon's Ocean Recreational Boat Survey
7. Andi Stephens and Alec MacCall, A Multispecies Approach to Subsetting Logbook Data for Purposes of Estimating CPUE –text and figures

8. Wade Van Buskirk
(1) CPUE Workshop Letter
(2) RecFIN SSC Memorandum 5-27-04.pdf
(3) MRFSS Data User's Manual
9. Farron Wallace
(4) Washington OSP Manual
(5) Formats for WA Ocean Catch/Effort tables by Year

In additional, I also reviewed methodologies presented in:

1. Pollock, K.H., C.M. Jones, and T. Brown. 1994. Angler Survey Methods and Their Application in Fisheries Management. American Fisheries Society Special Publication 25. American Fisheries Society, Bethesda, Maryland.
2. Gutherie, D., J.M. Hoenig, M. Holliday, C.M. Jones, M.J. Mills, S.A. Moberly, K.H. Pollock and D.R. Talheim, Editors. 1991. Creel and angler surveys in fisheries management. American Fisheries Society Symposium 12, Bethesda, Maryland.
3. Hoenig, J.M., C.M. Jones, D. Wade, D.S. Robson, and K.R. Pollock. 1997. "Calculation of catch rates in roving creel surveys of Anglers". Biometrics 53:306-317.
4. Jones, C.M., D.S. Robson, H.D. Lakkis and J. Kressel. 1995. "Properties of Catch Rates Used in Analysis of Angler Surveys". Transactions of the American Fisheries Society 124(6):911-928.
5. Washington Department of Fish and Wildlife, User's Guide for Programs OCDAIL and STRATUM, June 21, 2004. Draft manuscript.
6. Guisan, A., T.C. Edwards, Jr and T. Hastie 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157: 89-100.

II. *Attendance at CPUE workshop*- I participated in the review of recreational CPUE analyses used in groundfish stock assessments at a workshop that was held in Santa Cruz, California, from June 29-30, 2004.

III. *Address three specific issues regarding the use of CPUE*

III.1. Are the basic data reliable and standardized? Standardization needs to address whether a unit of effort (say an angler day) is as effective today as it was 20 years ago?

a. RECFIN

The question of data reliability and standardization must first address which data sources can be used for CPUE, which data are collected, and the length of their time series of CPUE. In the workshop, participants identified three major (MRFSS, CPFV, and ORBS) and several minor sources. The major data sources are currently integrated under the leadership of the Recreational Fisheries Information Network (RECFIN). RECFIN started in 1995 but the basic data collection comes from the NMFS Marine Fisheries Recreational Fisheries Statistics Survey (MRFSS) that goes back to 1980. The goal for the integrated RECFIN database is to coordinate surveys between three states (California, Oregon, and Washington) and put survey information and data into a central, assessable electronic database.

MRFSS is the basis of RECFIN and RECFIN follows its methodology. RECFIN surveys obtain data on effort from a telephone survey and catch and effort from intercept surveys. The intercept surveys are at public access points and from roving surveys of the shoreline. Because effort is collected from coastal county residents, it does not include non-coastal anglers and a correction is obtained in the intercept survey. The RECFIN surveys sample 1% or less of total trips taken. Costs are about $12 per interview for the telephone survey and $36 for an interview onsite. For the onsite survey, the angler trip is the sampling unit, while for the telephone survey, the sampling unit is the angling household. The survey is stratified by time waves, region, and modes of fishing. Port and site are not strata. Allocation is optimized to putative effort. Recently RECFIN has concentrated more sampling effort on private boat trips than before. Previously they sampled manmade structures more heavily. The RECFIN surveys follow a well-reviewed protocol of data collection that rests on carefully developed statistical sampling designs. Data acquisition and entry follow strict quality-control guidelines. Thus these data are well standardized and are reliable. Problems occur when sampling is sparse and unusual events are encountered, as in with an angling household with unusually high fishing activity or an onsite-intercept interview where the angler has caught a great number of fish. These types of events result in unrealistic estimates of CPUE even though the data were acquired correctly.

With the transfer from MRFSS to RECFIN, and with survey sampling now done under the control of the states, some differences in procedures exist. In California data collection is similar to MRFSS. Sampling is bimonthly to monthly for two regions and six districts. Output formats and variables are similar to MRFSS. California conducts the onsite surveys while MRFSS continues to conduct the telephone survey. The data time series is not complete and there are gaps for years and during some seasons. In addition to the onsite surveys, California has a list frame of charter boats to call directly. Such a list frame, when kept current provides much greater efficiency and cost savings. However, because not all charter captains cooperate, California uses the onsite survey to make adjustments to expand for uncooperative charter trips. Such data are important because the charter catch and effort statistics have been poorly sampled in the past and can be an important component of the recreational fishery.

RECFIN has also been adding other survey data to their database to expand its usefulness. These include the Oregon Recreational Boat Survey (ORBS), The Washington OSP and the California Private Fishing Vessel Survey (CPFV). Although these surveys add much needed data, they each have limitations. The ORBS has data for effort and mean CPUE estimates only and ports and month data are summarized; in Washington’s OSP ports and weeks summarized; while in the CPFV only rockfish and lingcod trip types are sampled.

One problem that is recognized is the sparse data coverage in the historic survey. RECFIN intends to expand information from the on-site survey to include demographics, etc. They have added information on the location of the fishing trip (area of the sea) in the onsite survey. However, even with their new approaches, they are sampling only 1% of angler trips. When angler behavior is consistent, then this sampling may be sufficient.

Where regions, seasons, and target fish change angler behavior then this sampling intensity may prove too sparse to obtain stable estimates of CPUE.

A strength of the RECFIN program lies in the standardization of data, consistent data base management, and electronic access. Importantly for the calculation of CPUE, they record whether the fish are released dead or alive, but also try to assess if the fish thrown back alive will remain alive. For calculation of CPUE, sample variables are trip type (directed toward), catch species, catch location (since 1999), time, demographics, and avidity. When they calculate mean CPUE they employ weighting factors (effort strata), summarize across these strata, and also post stratify when necessary. Several of the data series have incomplete catch data, including missing weights, length or specific species identification (species are given only as “rockfish” for example). Potentially, these missing data could be improved with carefully considered imputation techniques.

One problem with the transfer from NMFS to state-controlled sampling is the lack of sufficiently long overlap, between MRFSS and the new surveys to estimate coefficients to use in standardizing the time series. It is hard to evaluate whether this is a major or minor issue.

b. California State Recreational Fishing Surveys

A major survey is conducted on the charter industry in California through the California Private Fishing Vessel Survey (CPFV). The commercial passenger fishing industry began by the turn of last century, with participation being cyclical. There are fewer vessels today than in the past but these vessels have larger capacity; this fleet is among the largest of its kind in the world. California instituted a mandatory logbook in 1936 that required operators to keep daily trip records. This survey was expanded in the 1990s to include more information on species, discards, etc. Compliance has historically been about 75-95%. The data available in these surveys are aggregated but assessment scientists can filter the database by depth strata, by season, and species groups to target specific species. With recent improvements to the data collection, scientists can now obtain trip-specific target species, block area, and species composition from 1994-2003.

Other issues that arise from these data are the issues of effort saturation, such as when more anglers don't mean more fish are captured and where fish can be locally depleted. When these problems occur, CPUE can be hyperstable and not linearly related to abundance of fish. Note here that bag limits will also have the effect of decoupling CPUE and fish abundance. As these change over the years, CPUE departs from its relation to abundance.

These data are useful for examining possible changes in effort by observing changes in the proportion of fishing trips that occur inshore versus offshore. We would anticipate fishing location to follow a pattern of moving further offshore as stocks have dwindled. This can also reflect a change in fishing power as vessels become more capable of fishing further offshore.

There are also caveats in using these data for standardized CPUE calculations. The CPFV data contains no trip-specific or target species information for many species of interest currently before the 1990s; fishing power has changed as effort has become more effective due to fish finders etc.; anglers' preferences have changed on decadal scales and as restrictions occur; only after the mid 1990s have discard data been collected; species lists have changed over time; rockfish catch information is poor in quality. Captains could only record effort in one block area even though they may have fished in many. Trip codes are given different values within the database, e.g. northern and southern California may have a different code value for the same species. Many of these are not insurmountable problems and just require care when extracting and interpreting these data.

California has an S-K project to recover some of its data still in trip-specific logbooks and not in electronic and assessable formats. Some species groups have been well collected historically on logbook forms, e.g. salmon, lingcod, but rockfish species have not consistently been documented. Now rockfish species are individually specified, e.g. boccacio. The logbooks list target species now and contain lots more detail recently e.g. fishing method, and bait types.

As part of their current analysis, California has reassessed historic records to project effort and catch in blocks of area offshore. This provides a long time series and shows that effort is highest in southern California and is growing in Mexico and Baja areas, with the greatest effort occurring in spring time, thus showing a strong seasonal component.

California augmented the CPFV survey with an Observer Program from 1987-1998. The purpose of this program was to improve information on rockfish and lingcod, and to specifically address whether landed catches were decreasing. This program provided location-specific data on catches of these species. In 1988 the coverage was expanded as the program tried to get more data from northern California. It covered 3-5% of trips. The survey sampler chose trips to accompany so that the coverage was not random. The program stressed good fish identification by training observers well. Location was recorded, but GPS was not available at this time. This program provides better quality data on CPUE even though participation by vessels was voluntary. Samplers recorded when and why they were refused. Even with its quality, these data are limited too. They sampled few multi-day trips. Over 2,000 trips have been observed over 12 years.

c. Oregon State Recreational Fishing Surveys

The major state-conducted survey of recreational angling in Oregon is the Oregon Recreational Boat Survey (ORBS). Initially it began as the Ocean Salmon Sampling Project, clearly with an emphasis on salmon fishing and fishing sites. Even though this was not its initial focus, the survey also obtained groundfish data. Data are taken on boat trips through on site surveys. The goal is to estimate catch as CPUE x effort, hence CPUE is available in this database. The value of the historic data is limited because until 1998, Oregon sampled from May to September or October with 75% of trip interviews taken for salmon species. Landings data for rockfish are recorded in aggregate without data on specific species. To estimate effort, they count boats and trailers at moorage slips, etc. At some ports, they count boats through choke points. They obtain data on catch by randomly interviewing returning boats at major sites. Hence they lack coverage at smaller sites. The value of these data to groundfish CPUE evaluations depend on how severe the differences in target species and catch rates are between high-use and low-use access sites. This is an undercoverage issue and there are no data to evaluate this issue. Oregon also obtains information on charter boat trip with a telephone survey, but this provides data on boat trips and not on angler trips.

There are fairly good data on lingcod. These data exist on magnetic tape that may be useful because lingcod are hard to mis-identify. However, reading the magnetic tape is not straightforward and these data are not yet assessable electronically.

The caveats on this dataset are that historic species composition is problematic and, because of the focus on salmon, there was incomplete sampling for species composition. There is also contamination from salmon and halibut trips where other species were caught when salmon were not, even though salmon were the target. There are also a lot of missing data for combinations of year x month x port x boat type. There is no historic data on discards. It was noted during the workshop that these data have been used under the assumption that fishing power hasn't changed; otherwise the CPUE isn't correctly interpreted.

In 1998 Oregon changed name of this survey to ORBS and focused more on groundfish. Now groundfish are identified to species. In the sampling survey, strata are ports (not all ports are sampled each time) and week. The survey is post stratified by boat type (charter vs. private), fishery type (estuary vs. ocean), and trip type (salmon, bottom, combo).

d. Washington State Recreational Fishing Surveys

Washington conducts two surveys of recreational fishing: one for coastal waters (Ocean Sampling Program) and one for Puget Sound. The surveys have found a change in angler preferences over the years. For example, in the beginning of their surveys, anglers used to discard black rockfish but have not done so since the 1980s as other fisheries declined. Today, black rockfish are 80% of the recreational catch. In 1984 groundfish survey questions were integrated into their Ocean Sampling Program (OSP) to capture catch statistics for this growing fishery.

The survey methodology for the OSP is based on on-site counts and interviews. Samplers measure effort from a vantage point where they can count boats entering and exiting the port. They correct these counts to eliminate nonangling trips by interviewing boats that return to port. The goal is to obtain exit counts of 20-30% of all boat traffic. Samplers interview boating parties to obtain information on their catch. These data are subsequently used to calculate CPUE. Most of the recreational fishing occurs in spring through fall, and little if any fishing occurs in winter. Hence, sampling occurs in spring and fall. Samplers monitor both major and some minor ports, and stratify weekday/weekend days, matching sampling effort to fishing effort. When they compare their compute catch estimates from this survey to MRFSS, the results don’t agree, but there are problems with the comparison, as the OSP has frame undercoverage and is a minimal estimator of catch.

The caveats in using these data are a lack of data availability historically on rockfish and that catch data are in summary form only before 1986. Additionally, trips targeting Halibut are grouped with bottom fish trips and these must be identified before calculating

CPUE. Another drawback of this database is that raw creel data is not available prior to 1990. Catch is recorded as landings only, and there is no information on discard. Hence, CPUE is a minimum estimate. Only since 2002 has sampling collected release data on the catch. Further, there is a history of bag limits for some species of rockfish that must be addressed in data analysis to avoid possible underestimate of CPUE.

The goal of the OSP is to provide in season quota management and to obtain more biological data on the catch. Salmon and halibut have in-season management already. The program wants to begin in-season management of bottom fish in 2005.

The second survey is the Puget Sound Survey. This is a phone survey from an angler list frame combined with on site surveys at recreational access points. Such combination surveys are very efficient if the list frame is up to date and covers all anglers. Most list frames are somewhat out of date and exempt certain classes of anglers, such as the youth or seniors. Beginning in 2003, state scientists started correcting for frame undercoverage in this survey. In estimating CPUE it is difficult to assess the impact of frame undercoverage without further information. CPUE will be incorrect if the undercovered groups differ substantially in what they target and how much they catch. For the Puget Sound Survey, they may be able to compare CPUE from their list frame with that calculated from the undercovered groups. However, we cannot assume that this information can be hindcast over historic data, but this may be a valid approach for the previous few years. One indicator that some corrections may be applied to immediately preceding years is that trip length has been stable over time, which may indicate similar targeting and fishing location offshore.

e. Summary

Among the datasets that we reviewed at the workshop, MRFSS has been standardized (methodologies of collection) most extensively, although the states question the accuracy of some of its estimates. It is assessable electronically and is standardized historically. Few of the other datasets are standardized historically. When compared with MRFSS, the estimates do not agree. The time span of overlap between surveys is too short to calculate meaningful correction factors. However, when used carefully, they provide much needed data on CPUE. Other types of standardization (consistency of the effectiveness of an angling trip, e.g. has technology changed the angler's ability to find fish) are not done extensively, consistently, or at all. This is an area of study that needs development.

III.2. Are the statistical methods (usually general linearized models) performed correctly and are they producing reliable indices with relevant confidence intervals?

a. General concerns

Special considerations must be addressed when developing reliable indices of CPUE as proxies for abundance when taken from fishery-dependent recreational surveys. Some of these considerations are: the level of data aggregation – this influences the availability of specific data on catch and the stability of the estimator; level of stratification – especially coherence between federal and state surveys; effect of regulations on relation between CPUE and abundance; potential changes in the value of an angler or boat trip as trips are made more efficient through acquisition of technology – can recreational effort be standardized; and changes in effort that occur as anglers fish further offshore, closer inshore, or change preferences. Not all of these topics were covered in the workshop in detail or at all.

The MRFSS obtains CPUE data from its onsite surveys. There is some under coverage at locked, private marinas, but on the whole sampling onsite is done according to rigorous statistical protocol. Sampling is stratified by mode, region, and wave. Estimates can be post stratified by area of fishing, catch type (landed, released), and species. The survey provides a weighted mean of bimonthly CPUE. Even though they try to weigh these means, it may not always be correctly weighted - sampling of different modes may not be proportional to effort. These data are somewhat aggregated, and the effect of this aggregation should be a minimization of the variability of CPUE of individual anglers upon which it is based. In California survey samplers now get information on location where they got their catch offshore, but this isn't available nationwide. Such data will help determine whether the “angler trip” occurs consistently in the same fishing areas or whether there are shifts offshore or inshore to indicate depletion.

The calculations of CPUE are typically made from “directed” trips. This accomplishes a degree of standardization as it uses data from anglers who have the correct gear and fish in areas where the species are normally found. The MRFSS survey records directed trips as those that sought the species as a primary or secondary target. However, there are a variety of definitions for "directed" trip in other databases and this complicates comparisons and auxiliary use of CPUE indices from other sources.

The federal and state surveys are limited by costs in how many samples they can take onsite. Onsite samples are relatively expensive compared to phone and mail surveys (Pollock et al. 1994). When scientists need disaggregated data at fine scale, as they do sometimes for CPUE, this results in empty “cells”. One way to handle the lack of data is through imputation. Hot-deck imputation (you fill an empty cell with the value of a similar cell) has been applied to some of the databases. Caution should be exercised when this is done to account for the lack of degrees of freedom caused by the imputation upon further statistical analysis or manipulation.

b. Application of Generalized Linear Models (GLM)

One of the assumptions of CPUE is that it is proportional to abundance, CPUE ~ qfN, where q is the catchability coefficient and f is fishing effort. But, we know that q is not constant over time. As fish populations have been depleted, their behavior can change and pockets of the remaining population may be found in less commonly fished habitat. Typically q is handled as a nuisance parameter within stock assessment models. Changes in q can be evaluated by comparing times when effort is constant to see if q shows evidence of change. We expect that f has also changed over time. This is especially true as technologies have developed, e.g. fish finders, GPS, etc. Also note that f is different when angler versus boat trip information is recorded. Each will give a different value for CPUE and different variance when there is more than one angler per boat. To my knowledge, procedures to standardize *f*, on the basis of fishing power similar to commercial effort, have not been done. Dr. Ralston suggested during the workshop that the CALCOFI database be used to estimate recruitment and in turn evaluate changes in fishing power based on increases in catch normalized to CALCOFI-derived recruitment estimates. This approach, or another, should be tried to calculate changes in recreational fishing power.

The use of GLM has increased over the past 30 years with the advent of greater computer speed and memory. Guisan et al (2002) provide an overview of this statistical method. The value of GLM is that it is not restricted to the use of the normal distribution, and offers a link function. GLM has 3 components: 1) a random component from exponential distribution family, 2) a linear predictor using continuous and categorical variables (same as normal linear model) and, 3) a link function (g) with linearity and ability to constrain the response variable. Often times the Gamma distribution is used for positive observations or the Binomial and Logit to model for proportions.

Model selection is complex when estimating CPUE from recreational fishing surveys. CPUE has lots of zeros, for most anglers catch few or no fish while very few anglers catch many fish. To model this behavior, one can use a mixed model and combine two GLM as in the Delta distribution. When the scientist chooses explanatory variables, he or she should base their choice on: parsimony, residuals plots, and hypothesis testing. The choice of “best” model can be made by using information –theoretic statistics such as AIC or BIC with penalty for additional parameters. The link function that is selected depends on the error distribution, plausibility of the range of fitted values, and the linear relation between linear predictor and covariates. Some extensions of GLM are quasi-likelihood GLMs, joint models (double GLM), and GAMS.

During the workshop we saw the application of GLM by Andi Stephens and Alec McCall and the most extensive evaluation used bocaccio as the model fish. In this recreational fishery there are party boats of about 10 anglers, and they fish in different offshore areas. Data was taken from the MRFSS onsite survey, and it was compared to the California state survey. Scientists used the logit function to model whether bocaccio were caught and a likelihood function based on the probability of catching bocaccio given that other species in a complex were caught. The California trip data worked best. During the modeling, the scientists noted a problem with targeted fishing. For example, if it is a good year for tuna, anglers will fish on this "better" species and it will look like no bocaccio were caught, when in fact they were not sought. The older data didn't have any information on what was sought and these older data have to be used. Therefore scientists had to develop a selection criterion that acts as a proxy for this information. Some of the other explanatory data are based on presence/absence and this dampens the problem with changes in *q*.

Another hurdle that was noted in the application of GLM to CPUE was the effect of bag limits. One concern was how data are recorded in the different surveys because data is needed on what is caught per boat versus per angler. The effect of bag limits may be lessened when anglers share catch within a boat to sidestep per angler bag limits. Beyond this, there are ad hoc treatments that include breaking CPUE time series into before/after regulation segments to understand better the effect of bag limits in decoupling CPUE from abundance. We were shown an example of how to use statistical properties of bag sizes to compute a conversion factor to correct the time series of CPUE for such regulatory effects. These are straightforward and relatively simple methods for correcting long time series for such regulatory effects.

We were presented with several different examples of how to correct for regulatory artifacts in the CPUE data. One example was for black rockfish. With this species one sees a “stacking up” effect of higher bag limits when the limit is in place. One approach at correcting this is to assess how severe is the regulatory effect by using the ratio of the frequency of bags of 8+ fish to the sum of frequency of all bags up to and including 7 fish caught. This ratio was preserved despite bag limit, hence there is little regulatory effect and CPUE was useful proxy despite the bag limit.

The second example was for the sheepshead fishery in southern California. Dr. McCall simulated the effect of bag limits and found that too-high bag limits save few fish. This modeling was done with the assumption that people will comply fully and so overestimates any effect.

A third example was Bocaccio in southern California. In this fishery, recreational catches have declined so dramatically that the historic CPUE is a meaningful proxy. Recently, the extremely low bag limits in this fishery have altered the value of CPUE as a proxy. There are, however, some indicators of population status. For example, the recreational fishery in the nearshore catches young fish and gives an indication of recruitment. Also, the bocaccio recreational fishery shows poor compliance to this extreme bag limit. When Dr. McCall applied the same ratio technique to bocaccio as he previously did to sheepshead, the truncation approach didn't work. The ratio of catches with the sum to a given bag limit is not a stable relationship as it was in sheepshead. Hence, although this approach to standardize CPUE works in some fisheries, it doesn’t work in all over the entire time series.

The final example of GLM application by Dr. MacCall for producing a standardized time series of CPUE was for the lingcod recreational fishery. Dr. McCall tried two ways of producing the CPUE time series from the MRFSS database. In one he used a GLM approach directly on the “A” and “B” types of catch and this caused problems with the response variable. He also used an indirect method which restricted the data to only retained fish (A) for the GLM and then used RECFIN to get the ratio of all catch (A+ B1+B2) over catch of “A”-type landings. The lingcod assessment rejected the use of these indexes because of too many unsolved problems. For southern California catch-type “A” gives misleading CPUE, especially in recent years.

Jason Cope presented an alternate approach to fit CPUE data in nearshore species, again using GLM approaches. These data were taken from the recreational-fishery dependent CPFV observer program. The reported trips landed juveniles to adult rockfish. Other available databases to measure juvenile rockfish recruitment are the TENERA midwater and benthic surveys, which sample juvenile to adult rockfish. The statistical methods were GLM using lognormal and gamma error structure. Using these databases, he developed stock assessment and fishery independent indices for Cabezon as his base-case model. Five survey indices were included for four species over multiple years. His approach was to use linear combinations of the indices in a series of nested models, fit with GLM also with different error structures. This approach results in a set of linear models parameterized with different combinations of the five datasets and these results have to be compared to evaluate the best fit. Dr. Cope used Akiake's information criteria (AIC) with a penalty for the number of variables used in the model (see Burnham and Anderson, 2002). He presented three different indices that were developed and evaluated the fit to observed CPUE visually. This approach would be more convincing with a statistically-based evaluation such as a run's test. Beyond this, the approach was difficult to evaluate because indices with different spatial origins were used and different levels of aggregation were compared.

III. 3. Is the resulting time series index really proportional to changes in stock abundance over time? It may not if the standardization is inadequate, or if changing spatial patterns in fishing effort relative to stock distribution or changing species targeting patterns cause a non-linear response. Clearly, it cannot be as standardized as the level of the trawl survey standardization. The big question is how to add this standardization uncertainty into the interpretation of the CPUE trend. If the recreational CPUE data are the only substantive data in the assessment model will follow the trend in those data, for better or worse.

To answer the question of whether the time series index is proportional to changes in stock abundance requires a complex assessment of the underlying assumptions of the relationship between CPUE and N (CPUE = *qf*N), the inherent "quality" of the recreational CPUE datasets, the regional coverage of the data surveys, and the stability of CPUE indices. Uncertainty is introduced at all of these levels. As mentioned in the previous section, to my knowledge, changes in fishing power and therefore *f* have not been evaluated.

One of the greatest concerns is the lack of stability of the CPUE indices because of the sparse data. The recreational surveys were not designed to produce indices of CPUE for use in stock assessment models. Their purpose has been to obtain, at least, minimal estimates of the recreational catch to characterize this fishery and for comparison with the commercial catch which has been historically much greater. In recreational fisheries, CPUE is calculated as the ratio of the means ($CPUE_1 = \frac{\sum c_i}{\sum f_i}$; where c_i is catch from an individual angler or vessel and f_i is an angler or vessel trip for example) when it is used for catch expansion from access point surveys (Jones et al. 1995, Hoenig et al. 1997). As an estimator of catch, this is the appropriate estimator to use for proportionality to species abundance. CPUE calculated as the mean of the ratios ($CPUE_2 = \sum_i \frac{c_i}{f_i}$) is a valid indicator of angler behavior (Pollock et al. 1994; Jones et al. 1995), and it may also induce more instability into the indices. These tradeoffs could be evaluated with simulation models, but have not been evaluated yet to my knowledge. Depending on the level of disaggregation, the actual CPUE may be the mean of the ratios and this should be documented. Another issue that has not been addressed well is the impact of aggregating data on the validity and stability of CPUE indices. This concern is not addressed by current research, even though different models use different levels of aggregation and some models combine different levels of data aggregation in the same model. This is a standardization issue that is important to evaluating the ability to use recreationally-derived CPUE indices. The use of aggregated and imputed data may be unavoidable in species that are rare historically, or have historically not been targeted and where data are sparse. This is an area where current standardization may be inadequate.

As species are depleted from more accessible locations, anglers venture further to catch them. Increased travel time was used to indicate declines of flatfish abundance caused by the North Sea commercial fisheries – even though CPUE was relatively constant when travel time was ignored. So too, increases in the distance that anglers go to fish recreationally may indicate declines beyond what is evident in the CPUE data. Information on distance traveled can be used as a variable in GLM and this was discussed at the workshop. Some of the databases have added questions on fishing locations to the surveys (RECFIN, CPFV), although these data are missing from the historic survey data. The ability to standardize CPUE time series for fishing location effects will be possible in the future as data accumulate. It will, however, be difficult to do this standardization with the historic data.

Regulatory restrictions, such as bag limits, may distort the proportionality between CPUE and abundance if release data are not obtained as part of a recreational fishing survey. However, this distortion can be corrected if not too severe. First, restrictions may be ignored and it is not uncommon to find anglers with illegal catch when conducting an angling survey. Beyond these obvious violations, anglers may share their catch when several are fishing in one vessel. Such sharing distorts individual angler CPUE but not the vessel or boat-trip CPUE. This is an example where slightly aggregated data may stabilize CPUE indices and provide more accurate estimates of CPUE, hence of fish abundance. Dr. McCall used a method of ratios of catch bag limits to evaluate the impact of regulations on bocaccio and sheepshead CPUE time series. This method standardized the sheepshead time series, but not the bocaccio series. Hence, although standardization techniques exist and can work well on the indices of some species, the techniques do not work with all species. Some species can’t be standardized this way when regulations affect the indices.

In summary, the ability to standardize CPUE time series is not universal and standardization is more difficult with historic data that was not collected rigorously. Even for those collected rigorously, there are data gaps and sparse data for important species. Both add to the uncertainly in the putative proportionality of CPUE to fish abundance.

Literature Cited

Burnham, K.P. and D.R. Anderson. 2002. Model selection and inference: a practical information-theoretic approach. 2nd ed. Springer-Verlag, New York.

Guisan, A., T.C. Edwards, Jr and T. Hastie 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157: 89-100.

Hoenig, J.M., C.M. Jones, D. Wade, D.S. Robson, and K.R. Pollock. 1997. "Calculation of catch rates in roving creel surveys of Anglers". Biometrics 53:306-317.

Jones, C.M., D.S. Robson, H.D. Lakkis and J. Kressel. 1995. "Properties of Catch Rates Used in Analysis of Angler Surveys". Transactions of the American Fisheries Society 124(6):911-928.

Pollock, K.H., J.M. Hoenig, C.M. Jones, D.S. Robson, and C.J. Greene. 1997. "Catch rate estimation of roving and access point surveys". N. Amer. J. Fish. Management 17:11-19.

Pollock, K.H., C.M. Jones, and T. Brown. 1994. Angler Survey Methods and Their Application in Fisheries Management. American Fisheries Society Special Publication 25. American Fisheries Society, Bethesda, Maryland.

Appendix 1: Statement of Work

Consulting Agreement Between The University of Miami and Dr. Cynthia Jones

May 4, 2004

General

The importance of recreational CPUE statistics in groundfish stock assessments has grown as Pacific Fishery Management Council (PFMC) sponsored assessments of bocaccio, cabezon, cowcod, black rockfish, lingcod, and yelloweye rockfish have evaluated time series of recreational CPUE as indicators of stock abundance. As more of the minor groundfish species
become assessed, CPUE statistics derived from catch rates in sport fisheries will play a major role in stock evaluation.

Due to concerns about the reliability of CPUE statistics derived from recreational data, and their importance to groundfish management, the Scientific and Statistical Committee (SSC) in cooperation with the Northwest and Southwest Fisheries Science Centers is sponsoring a workshop on Recreational CPUE Statistics to promote an exchange of data, information, ideas, and solutions. This two-day workshop, to be held in Santa Cruz, California, from June 29-30, 2004, will include a limited number of presentations as well as discussion among participants to promote and exchange ideas. The workshop will be divided into two phases: (1) sources of recreational data on the West Coast and (2) developing CPUE statistics to make inferences about stock abundance.

Workshop attendees shall complete the following tasks:

- Review methods previously used in West Coast Groundfish stock assessments;
- Review data methodologies used for recreational data analyses and summary in all states;
- Review of methods used in other regions where use of recreational data is significant;
- Develop recommendations for standard approaches to be used to include CPUE data in groundfish assessments.

The workshop will be open to the public and will be held at the NOAA Fisheries Santa Cruz Laboratory. Significant attendance by members of the SSC groundfish subcommittee, which will run the workshop, members of the west coast stock assessment community, and representatives from interested PFMC constituent groups is expected.

Specific

The consultant's duties shall not exceed a maximum total of 14 days: Several days prior to the meeting for document review; the two-day meeting; and several days following the meeting to complete the written report. The report is to be based on the consultant's findings, and no consensus report shall be accepted.

1) Become familiar with background materials.

2) Participate in the review of recreational CPUE analyses used in groundfish stock assessments at a workshop to be held in Santa Cruz, California, from June 29-30, 2004.

3) Address the following issues in a written report:

    a. Are the basic data reliable and standardized? Standardization needs to address whether a unit of effort (say an angler-day) is as effective today as it was 20 years ago.
    b. Are the statistical methods (usually generalized linear models) performed correctly and are they producing reliable indices with relevant confidence intervals?
    c. Is the resulting time series index really proportional to changes in stock abundance over time? It may not if the standardization is inadequate, or if changing spatial patterns in fishing effort relative to stock distribution or changing species targeting patterns cause a non-linear response. Clearly it cannot be as standardized as the level of trawl survey standardization. The big question is how to add this standardization uncertainty into the interpretation of the CPUE trend. If the recreational CPUE data are the only substantive data in the assessment model, the model will follow the trend in those data, for better or for worse.

4) No later than July 14, 2004, submit the written report[1] consisting of the aforementioned issues (see Specific Task 3), findings, analysis, and conclusions, addressed to the "University of Miami Independent System for Peer Review," and sent to Dr. David Die, via email to ddie@rsmas.miami.edu, and to Mr. Manoj Shivlani, via email to mshivlani@rsmas.miami.edu.

[1] The written report will undergo an internal CIE review before it is considered final. After completion, the CIE will create a PDF version of the written report that will be submitted to NMFS and the consultant.

ANNEX I: REPORT GENERATION AND PROCEDURAL ITEMS

1. The report should be prefaced with an executive summary of findings and/or recommendations.

2. The main body of the report should consist of a background, description of review activities, summary of findings, conclusions/recommendations, and references.

3. The report should also include as separate appendices the bibliography of all materials provided and a copy of the statement of work.

Appendix 2: Background material

Table of Contents
Background Materials CD for
The Recreational CPUE Statistics Workshop
June 29-30, 2004

General Background Materials

1. Workshop Agenda
2. Letter from Ms. Karen Garrison (NRDC) to Mr. Rodney McInnis
3. Bocaccio Stock Assessment (2003)
4. Bocaccio STAR Panel Report (2003)

Presenters Background Materials

1. E.J. Dick
   Beyond 'Lognormal vs. Gamma': Discrimination Among Error Distributions for Generalized Linear Models - text and figures

2. Kevin Hill
   Hill, K. and N. Schneider. 1999. Historical Logbook Databases from California's Commercial Passenger Fishing Vessel (Passenger) Fishery 1936-1997. SIO Reference Series No. 99-19.

3. David Sampson
   Oregon's Ocean Recreational Boat Survey

4. Andi Stephens and Alec MacCall
   A Multispecies Approach to Subsetting Logbook Data for Purposes of Estimating CPUE –text and figures

5. Wade Van Buskirk
   CPUE Workshop Letter
   RecFIN SSC Memorandum 5-27-04.pdf
   MRFSS Data User's Manual

6. Farron Wallace
   Washington OSP Manual
   Formats for WA Ocean Catch/Effort tables by Year