MRIP Calibration Workshop White Paper

Pacific RecFIN Comparisons – Searching for a Calibration

Summary of the Pacific RecFIN Statistics Subcommittee Report:
"Comparison of Effort Estimates from Pacific Coast
Marine Recreational Fisheries Surveys: 2003-2005"

by Han-Lin Lai and Dave Van Voorhees

Pacific RecFIN Comparisons – Searching for a Calibration

Summary of the Pacific RecFIN Statistics Subcommittee Report:
"Comparison of Effort Estimates from Pacific Coast
Marine Recreational Fisheries Surveys: 2003-2005"

by Han-Lin Lai and Dave Van Voorhees

The Pacific Recreational Fisheries Information Network (RecFIN) Program implemented both new and modified survey designs in 2003 due to changing management needs and perceived problems with the effort estimates produced by the Marine Recreational Fisheries Statistics Survey (MRFSS). In 2004, the RecFIN Technical Committee tasked its Statistical Subcommittee to look into developing a "calibration" of the new sampling methods against the MRFSS that could be used to revise the MRFSS time series of effort and catch estimates. This paper is a condensed summary of the report produced by the Subcommittee entitled "Comparison of Effort Estimates from Pacific Coast Marine Recreational Fisheries Surveys: 2003-2005".

Background

NMFS had conducted the MRFSS in 1981-1989 with the help of contractors and the Pacific States Marine Fisheries Commission (PSMFC). NMFS stopped conducting the MRFSS on the Pacific coast in 1990 due to insufficient funds, but started it again in 1993 as a key component of the new Pacific RecFIN Program established as a cooperative effort involving NMFS, PSMFC, Washington Department of Fish and Wildlife, Oregon Department of Fish and Wildlife, and California Department Fish and Game. Specialized surveys already being conducted by the states, such as the Oregon Ocean Recreational Boat Survey (ORBS) and the Washington Ocean Sampling Program (OSP) also comprised components of the new Program, and RecFIN did not conduct the MRFSS in some of the time periods and fishing modes covered by those surveys to avoid unnecessary overlaps. To address concerns that the MRFSS was producing effort and catch estimates that appeared to be consistently higher than what the state surveys were producing, RecFIN decided in 1998 to conduct the MRFSS alongside of the state surveys in Oregon and Washington for four months in each of two years to allow side-by-side comparisons and evaluation of any differences.

In 2002, the Pacific RecFIN Executive Steering Committee decided to stop conducting the MRFSS and directed the states to develop new survey designs and/or modifications of their existing survey designs that would replace the MRFSS starting in 2003. The major concern was that the MRFSS effort and catch estimates appeared to be consistently higher than those produced by the state surveys. NMFS continued to conduct the random-digit dialing Coastal Household Telephone Survey of the MRFSS for a couple of years to allow side-by-side comparisons with new telephone surveys developed by the states that were based on sampling from lists of their licensed anglers.

Objectives

The objectives of the Pacific RecFIN Statistics Subcommitte report were fourfold:

1. Quantify the differences between effort estimates derived from the old and new marine recreational fisheries surveys during the period 2003-2005.
2. Identify survey features that might have contributed to any observed differences.
3. Evaluate a possible calibration of "old" and "new" survey estimates that could be applied to produce a more accurate time series of recreational fishing effort statistics.
4. Identify and discuss potential improvements to the new state surveys.

Comparisons of Survey Estimates

The report compared fishing effort estimates produced by the MRFSS and state survey designs in Washington and Oregon. The Subcommittee planned to include comparisons of MRFSS estimates with those produced by the new California Recreational Fishery Survey Program (CRFS), but California Fish and Game was not ready to share those estimates before the report was completed. The survey designs compared included the following:

1. The Marine Recreational Fishery Statistics Survey provided effort estimates from data collected in two surveys:
    a. The Coastal Household Telephone Survey (CHTS) provided estimates of effort by coastal zone residents.
    b. The Access Point Angler Intercept Survey provided estimates of ratios needed to adjust CHTS estimates to account for effort by non-residents of the coastal zone.
2. The Washington Ocean Sampling Program (OSP) provided effort estimates based on on-site boat entrance counts.
3. The Washington Puget Sound Sampling Program (PSSP) provided effort estimates from data collected in a new Angler License Directory Telephone Survey (ALDTS). This survey provided estimates of effort by holders of Washington fishing licenses.
4. The Oregon Ocean Boat Survey (ORBS) provided effort estimates based on on-site boat exit counts.
5. The Oregon Shore and Estuary-Boat Survey (SEBS) effort estimates from data collected in a new Angler License Directory Telephone Survey (ALDTS). This survey provided estimates of effort by holders of Oregon fishing licenses.

The Subcommittee compared domain estimates produced by the different surveys in 2003-2005 by geographic area, fishing mode, and year. The specific comparisons for different geographic areas were as follows:

1. Washington Puget Sound:
    - Total annual fishing effort estimated by the PSSP and the MRFSS.
    - Mean angler fishing effort estimated by the ALDTS and CHTS.
2. Washington Ocean:
    - Total annual fishing effort estimated by the OSP and MRFSS.
3. Oregon Inland:

- Total annual fishing effort estimated by the SEBS and the MRFSS.
- Mean angler fishing effort estimated by the ALDTS and CHTS.

4. Oregon Ocean:
    - Total annual fishing effort estimated by the ORBS and MRFSS.

Results of Comparisons

The comparisons showed relatively consistent differences in annual effort estimates for private boat and shore fishing, but very inconsistent differences in annual estimates for charter boat fishing. In general, the MRFSS estimates of annual shore fishing effort tended to be higher than the PSSP and SEBS estimates. The MRFSS effort estimates for private boat fishing in the ocean also tended to be higher than the estimates produced by the OSP and ORBS. However, the MRFSS effort estimates for charter boat fishing showed no consistent directional difference in any of the paired comparisons.

The comparisons of CHTS and ALDTS estimates of the mean number of fishing days per angler showed no consistent differences, indicating that the differences in annual effort estimates observed in the MRFSS vs. PSSP and MRFSS vs. SEBS comparisons appeared to be mostly attributable to differences in the methods used for (1) expanding sampled data to estimate population totals and (2) adjusting telephone survey estimates to account for effort by anglers not covered in the telephone survey sample frames.

Consideration of a Calibration

This third objective of the report seemed very reasonable at first glance. The Subcommittee was asked to assess whether it would be possible "to calibrate the old and new survey estimates in order to maintain a continuous time series of recreational fishing effort". However, for various reasons explained below, the Subcommittee decided that a meaningful and reliable calibration is very problematic and essentially impossible for any of the comparisons of the MRFSS with any of the state surveys.

For almost all purposes of stock assessment and fishery management, calibration between different survey estimates is misleading and is not necessarily the end goal. In any given application, such as a stock assessment, the need is to make good or optimal use of all available data from the different surveys: past data from "old" surveys, recent data from "new" surveys, and both kinds of data from "continuing" surveys. So the stated concern, "calibration" of effort estimates, may be viewed more appropriately as a concern over best methods for "combination" of different estimates. As detailed below, various useful combination methods exist and should be considered.

The report discussed three main topics related to the search for a calibration:

1. Calibration concepts,
2. Calibration of MRFSS and state surveys, and
3. Combination of estimates.

The discussion of the last topic covered a variety of methods. No one method was recommended, or elaborated in working examples. The reason was simply that fishery applications which call for suitable combination of survey estimates can vary greatly in the available data and the purposes to be served. For similar reasons, the Subcommittee's main conclusion was the following caveat. "The choice of combination method in any application - for instance in a stock assessment - will best be made not through default to a would-be recipe but through collaboration on the analysis team between both subject-matter experts and survey statisticians."

Calibration Concepts

In applied science, a method M1 of measurement (available but perhaps not standard) is "calibrated" to another method M0 (regarded as a standard or reference), within an acceptable accuracy criterion. Calibration requires two stages as follows:

1. finding a calibration function f(M1) and
2. using f(M1) to convert the M1 measurements to the M0 standard values.

The calibration function may be linear, such as calibrating Fahrenheit temperature to Celsius temperature, or nonlinear, such as calibrating fish weight or age to length. Depending on the data and the model specification, the calibration functions may not even be monotonic. Furthermore, M0 need not necessarily represent the "truth" any more than does M1 because both M0 and M1 measurements are subject to variations described by mean square errors.

"Scientific" calibration is concerned with experimentally validating the use of a proposed calibration function. For this purpose, the National Institute of Standards and Technology (NIST) set up validation criteria which call for "traceability": complete information about every step in a process chain of comparisons, with accompanying uncertainty statements (http://ts.nist.gov/MeasurementServices/Calibrations/policy.cfm.).

"Statistical" calibration is concerned with theoretic methodology for defining suitable calibration functions and uncertainties between given measurement methods. Frequently, this field is referred to as inverse regression (Krutchkoff 1967) or inverse prediction (Neter et al. 85, p.172). Two articles, Hunter and Lamboy (1981) with discussions and Osborne (1991), merit note for their overviews of statistical calibration. In most familiar cases, one method is deemed a standard or reference, and so "absolute" calibrations (Williams 1969) are made to it.

In other cases, no one method can be deemed as a standard and calibrations among all methods are "relative" or "comparative". Examples of comparative calibration are given by Barnett (1969), Williams (1969), Theobald and Mallinson (1978), and Kimura (1992). Note that Barnett's approach is different from the others due to the assumption on the choice of reference method. This choice is on technical grounds when there is a single controlled method and a set of experimental alternatives (Osborne 1991). When there are no technical grounds to make the choice of reference, Theobald and Mallinson (1978) point out that there is symmetry about the calibration problem which is missing in Barnett's approach. Kimura (1992) points out that comparative calibration is best understandable via measurement error models (Fuller 1987, Ch. 4.1).

Calibration of MRFSS with State Surveys

For the following reasons, the Subcommittee did not recommend calibration – that is, attempting to construct or use one or more calibration functions based on comparing estimates from MRFSS and state-run surveys.

1. All the surveys were subject to both sampling errors and non-sampling errors. None of the surveys would meet the NIST standard for serving as the "reference" or "gold standard" measurement. In particular, none of the surveys were set up as controlled experiments with a set of experimental alternatives, and there is no way to perform "ground truthing".
2. There are no replicate specimens within each survey domain (i.e., mode and wave) for use in a NIST-type calibration experiment that would establish an unbroken chain of comparisons and provide a statement of uncertainty within and between surveys.
3. There was vast diversity in comparisons among survey domains. A simple calibration that aggregates all domains would introduce untraceable bias. Meta-analysis would require further auxiliary data (now lacking) so that diversity could be interpreted unequivocally.
4. The surveys had very limited temporal overlap. Any calibration function between the estimates produced by the surveys would in large part be a very problematic extrapolation to times where one or both surveys offered no data.
5. There was obvious variability both within and between surveys. The variability was mode-and wave-dependent. In addition, particularly for the new state surveys, a given survey may have nominally been the "same" survey from one year to the next while its methods actually changed to some degree. Consistent calibration within, as well as between, surveys is therefore doubtful.
6. The form and features of appropriate calibration functions are unknown, and it would be very difficult to determine whether such functions should be linear, nonlinear, univariate, or multivariate.

Combined Survey Estimators

Estimators from several different surveys may be combined to define an improved or "compromise" estimator. Many different approaches exist for this, but the underlying basic concept is to find the compromise as a weighted sum of the individual survey estimates.

Calibration Estimators: There is a family of calibration estimators in survey sampling (Zieschang 1990, Deville and Särndal 1992, Théberge 1999). For a single survey, the calibration estimator serves as a calibration on known population totals of auxiliary variables to modify the sampling design weights (inverses of inclusion probabilities; these weights appear in the Horvitz-Thompson estimator) subject to the calibration equation (or benchmark constraint), such that , for all sampled units. The calibration weights are used for estimation in survey sampling to achieve internal consistency. This weighting control procedure is conducted through a generalized regression (GREG, Valliant et al. 2000). Renssen and Nieuwenbroek (1997) and Wu (2004) extend calibration estimators to multiple surveys on the same target population when the same auxiliary data are jointly collected in all surveys but population totals of auxiliary variables are unknown. The weight adjustment aids obtaining consistent estimates. However, there is a drawback. Due to the enlarged number of independent variables, GREG can produce negative weights: an undesirable feature for many users.

Linear Compromise Estimators: The early work by Graybill and Deal (1959) proposed to use the linear (actually convex) combination estimator which is unbiased with minimum variance. The pairings of MRFSS with the OSP and MRFSS with the ORBS seem to be good candidates for this approach. See Särndal, Swensson and Wretman (1992, p.370-372) for the solution for the optimal analytical solution. With three or more surveys, combined estimators become more difficult but a numerical solution may be found, for instance by convex programming using selected constraints.

Dual-Frame Estimators: A sample survey of retail stores (M. A. Bershad, U.S. Bureau of Census, in Hansen et al. 1953, Ch. 12) is the one of the earliest uses of two-frame surveys. Hartley (1974) developed a general theory for the dual-frame method. We can briefly describe the method here. Suppose two surveys use different sampling frames (Frames A and B), neither with complete coverage of the target population, but with some degree of overlap as shown in the following diagram. The domain labeled D1 contains N1 members that are listed in Frame A only; Domain D2 contains N2 matched members that are in both frames; and Domain D3 contains the N3 members that are listed in Frame B only. Each of the three domains has domain total, denoted respectively by $Y1$, $Y2$, and $Y3$. The population total is $Y = Y1 + Y2 + Y3$. Estimates of $Y1$ and $Y3$, and of these estimates' variances, are straightforward from the samples taken from Frame A and B respectively. There remains to get a compromise estimate of $Y2$, using the two estimates that can be derived from the sampling based on Frame A and from that based on Frame B. The two estimated variances are used to obtain the weights needed for this compromise estimate.

Multiple Frame Estimators: Henning et al. (1978) extended the dual-frame method to multiple surveys. On their own, readers can readily extend the above discussion. Combination of MRFSS and all California survey estimators would be one natural application of the multiple-frame method. These surveys do use different kinds of frames which do not merely overlap. For example, MRFSS uses a frame of households while ALD uses a frame of angler licenses. However, the multiple-frame method allows combining estimators if certain linkages are established between frames.

Measurement-Error Models (also known as error-in-variables models): Measurement error models are a very general approach toward use of multiple sources of information. Schnute (1994) and Schnute et al. (1990) introduced this approach into fishery stock-assessment models. Error-in-variables models are a common subset of the larger class of measurement error models, as discussed in Fuller (1987). For the same reasons that calibration is problematic with these data, measurement error models would be difficult to implement. A sound measurement error model requires an understanding of the ways in which measurement errors are incorporated in the model, as well as sound estimates of the consequent measurement errors.

Multiple Indices: Within much modeling, a powerful yet routine practice is the use of several estimates (indices) of the same quantity but from different sources. In particular, a typical fishery stock-assessment model uses several indices of each stock's abundance. As this use illustrates (for abundance rather than effort) the aim of analysis need not require prior construction of a compromise estimate. In fact, use of the separate estimates, rather than of just a compromise estimate, allows the model to use more information. Moreover, the model itself, through its population dynamics or other components, may well be able to reconcile (at least in part) the differences among the separate estimates.

Unresolved Problems and Major Uncertainties

It is important to point out that the limitations of surveys and their analyses can contribute to major uncertainties. These limitations are common in almost all other surveys, fishery or non-fishery. Provided that we understand these limitations, the comparisons of effort estimates among surveys provide us a rare opportunity to try and study the merits and deficiencies of alternate survey designs. The Subcommittee concluded that all surveys are subject to improvement for the following reasons.

1. We don't know which survey is closer to the truth.
2. The time period over which the data series overlap is relatively short.
3. The data collection programs have not been static.
4. Differences are not systematic, and both environmental and regulatory changes may play a major role.
5. All methods could be improved to better address the following issues:
    a. Non-sampling errors
    b. Appropriate weighting of data in estimation and imputation
    c. Problems with implementation of protocols for sampling and data processing
    d. Most are multi-purpose surveys which may not be optimized for all estimates
    e. Zero, or null, point or variance estimates resulting from small sample size

Which survey is closer to the truth?

All surveys are dependent on probability sampling design. The ideal condition for any surveys is to use a perfect sampling frame, to draw sufficient number of samples according to a specified design, to correctly measure the sampled units and record measurements, and to use the correct statistical method corresponding to the sampling design. With limited amount of resources, a small sample size from a probability sampling design would increase the possibility to deviate from this ideal survey. The deviation would yield uncertainties (inaccuracy and imprecision), and surveys can produce very unimpressive results when based on very small samples.

Keep in mind that the particular sample selected by a surveyor is one of many possible samples from a targeted population. Different samples as well as different survey designs will give different results. A calculated confidence interval incorporates uncertainty due to sampling error because it is based on one particular sample and a normality assumption on the parameter estimate. Moreover, we never know when or whether the central limit theorem will take over and we are not able to repeat sampling a large number of times. The problem is the intractable and dynamic nature of recreational fishery populations over spatial, temporal, and demographic scales.

Was overlap period long enough?

The comparisons are based on very short time period over which survey data series overlap (Tables 1 and 4). Therefore, it is important to keep in mind that the results of comparison should not be extrapolated. This is one of the main reasons that we do not recommend one approach nor do we offer a case study/example for calibration because projecting a calibration factor back in time where data series do not overlap would likely yield biased results.

Were data collection programs static?

The sampling units were different among MRFSS CHTS (households), state ALDS (anglers) and state vessel sampling surveys (boat-trips, e.g., WA OSP, and ORBS). The demographical profiles of sampling frame across these surveys were also different. The durations for recall of trips differed across MRFSS and state surveys. The two estimation procedures for MRFSS CHTS exemplify how effort estimators can differ from the enumeration of recalled trips. In order to make valid comparisons, adjustments on data and analytical methods are necessary. However, such adjustments have yet to be evaluated.

Were differences systematic?

As pointed out in the Results section, the differences exist year to year, wave to wave, and across modes (Figures 2 to 31). There are also environmental and regulatory changes over time, which have yet to be investigated.

Were any methods free of bias?

Non-sampling errors: This report discusses a variety of nonsampling errors that affect the interpretation of discrepancies of the estimates among surveys and estimation methods. Nonsampling errors are categorized into three types: coverage errors, nonresponse errors, and measurement and implementation errors. It is therefore not surprising that estimates from different marine recreational sample surveys differ.

Estimation and imputation: Some of the current methods did not appropriately account for the sampling design (selection probability of primary sampling units and subunits) and population domains. Also, imputation for data with nonresponse errors was not consistently used across surveys. MRFSS uses hot-deck imputation, however, the use or lack thereof in the other surveys is unknown. The effects on point and interval estimates have not been evaluated.

Implementation: There were sufficient differences among survey protocols, and it was difficult to determine which survey was most free of bias. The multiple frame approach and combination of multiple survey estimates provide a possible venue for this solution. However, all surveys should be implemented with certain common design and protocols for data collection, such as sampler training, data QA/QC, database management, and socio-psychological factors (face-to-face vs. phone).

Multi-purpose surveys: MRFSS and state surveys have been planned for multiple purposes. For examples, the majority of anglers and charters in Puget Sound targets on salmon rather than marine fishes, and access-point intercept surveys provide information on non-fishing households (or unlicensed anglers) and catch rates over spatial-temporal and mode specifications. Accuracy and precision are frequently compromised with the requirement of cost-efficiency planning for multi-purpose surveys. Although stratification and post-stratification and domain estimations have been applied to the surveys, efficiencies of existing sampling designs and estimation methods have yet to be evaluated. Such evaluation would be highly desirable before - and as input to - survey redesign efforts.

Zero estimates and/or variances: One challenge in the comparison studies was the production of zero (or null) values for certain point estimates or variance estimates. In many cases, small sample size (n) was a problem, especially when the samples were divided into geographic domains. Effects of small sample size on estimation can be seen when; (i) n=1, which is not sufficient to make statistical estimation, and (ii) n<5, which is less than the requirements for applying small sample theory where the chance that all measurements of the sampling units are identical with zero variance is high. Furthermore, if the variation among measurements is extremely small, computer software and hardware limitations can cause rounding errors that yield small negative variances. The Subcommittee recommended future research be directed at methods for imputation and/or synthetic small-area estimation.

Strategies for Improvement

During the course of this study, the Subcommittee identified the following strategies for improvement of the Pacific RecFIN surveys:

1. Develop more complete and efficient sampling frames through implementation of angler and vessel registries and consider possible use of multiple frames to reduce errors due to undercoverage.
2. Reduce the potential for nonresponse errors by redesigning questionnaires, conducting better outreach, and collecting relevant variables to aid appropriate imputation or weighting of data.
3. Reduce measurement and implementation errors by providing or upgrading training for samplers and data processing staff, automating data quality control checks, and improving database management.
4. Develop improved estimation methods that are more in accord with the sampling designs, evaluate possible improvements in imputation methods, examine possible use of domain and small area estimation methodologies.
5. Investigate possible methods for combining multiple survey estimates using expert collaboration to obtain and apply optimum weights.
6. Review issues of survey costs versus sample size to make more efficient use of available funds while maximizing statistical precision.
7. Develop more complete documentation of all survey sampling and estimation methods.

This document was the result of a rare opportunity to perform side-by-side comparisons from multiple surveys. More such comparisons should be performed in future, since this approach provides a way of assessing the strengths and weaknesses of similar recreational fishing surveys and survey methodologies.

The RecFIN Statistical Subcommittee shares the visions of the Pacific Fishery Management Council's Scientific and Statistical Committee, Groundfish Management Team, and the National Research Council of the National Academy of Sciences to provide better recreational data for use in fisheries management and stock assessments. We hope this project will bring RecFIN administrators and analysts, along with their constituents, closer together for future development of the RecFIN data system.

Acknowledgements

The authors of the Pacific RecFIN Report summarized here were Rob Andrews (NMFS-HQ), Jennifer Cahalan (WDFW), David Cassell (PSMFC), Meisha Key (CDFG), Han-Lin Lai (NMFS HQ), Todd Lee (NMFS-NWFSC), Rod Pedersen (PSMFC), Keith Sakuma (NMFS-SWFSC), Nick Tolimieri (NMFS-NWFSC) , Wade Van Buskirk (PSMFC), Dave Van Voorhees, NMFS-HQ), Joe Weinstein (CDFG), and Bryan Wright (ODFW).

References

Barnett, V.D. 1969. Simultaneous pairwise linear structural relationships. Biometrics 25: 129-142.

Cochran, W.G. 1977. Sampling techniques, 3rd ed. John Wiley & Sons, New York. 428 pp.

Deville, J., and Särndal, C. 1992. Calibration estimators in survey sampling. Journal of the American Statistics Association 87: 376-382.

Fuller, W.A. 1987. Measurement error models. John Wiley & Sons. New York. 440 pp.

Graybill, F.A., and Deal, R.B. 1959. Combining unbiased estimators. Biometrics 15: 543-550.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. 1953. Sampling survey methods and theory. Volume I. John Wiley & Sons, New York. 638 pp.

Hartley, H.O. 1974. Multiple frame methodology and selected applications. Sankhya Ser. C 36: 99-118.

Henning, C.R., Woltman, H.F., and Isaki, C.T. 1978. An application of multi-frame methodology and measurement error research for the 1976 registration and voting survey. Proceedings of Social Statistics Section, American Statistics Association 1978: 542-547.

Hunter, W.G., and Lamboy, W.F. 1981. A Bayesian analysis of the linear calibration problems, with discussions and response. Technometrics 23: 323-350.

Kimura, D.K. 1992. Functional comparative calibration using an EM algorithm. Biometrics 48: 1263-1271.

Krutchkoff, R.G. 1967. Classical and inverse regression methods of calibration. Technometrics 9: 425-439.

Lohr, S.L. 1999. Sampling: design and analysis. Duxbury Press. 493 pp.

National Research Council (NRC). 2006. Review of Recreational Fisheries Survey Methods. http://newton.nap.edu/catalog/11616.html (accessed November 21, 2006).

Neter, J., Wasserman, W., and Kunter, M.H. 1985. Applied linear statistical models, regression, analysis of variance, and experimental design, 2nd Ed. Irwin, Homewood, IL. 1127pp.
Osborne, C. 1991. Statistical calibration: a review. International Statistical Review. 59: 309-336.
Renssen, R.H., and Nieuwenbroek, N.J. 1997. Aligning estimates for common variables in two or more sample surveys. Journal of the American Statistics Association 92: 368-374.
Särndal C., Swenson, B., and Wretman, J. 1992. Model assisted survey sampling. Springer. 694 pp.
Schnute, J.T. 1994. A general framework for developing sequential fisheries models. Canadian Journal of Fisheries and Aquatic Science 51: 1676-1688.
Schnute, J.T., Mulligan, T.J., and Juhn, B.R. 1990. An error-in-variables models with an application to salmon hatchery data. Canadian Journal of Fisheries and Aquatic Science 47: 1453-1467.
Théberge, A. 1999. Extensions of calibration estimators in survey sampling. Journal of the American Statistics Association 94: 635-644.
Theobald, C.M., and Mallinson, J.R. 1978. Comparative calibration, linear structural relationships and congeneric measurements. Biometrics 34: 39-45.
Valliant, R., Dorfman, A.H., and Royall, R.M. 2000. Finite population sampling and inference, a prediction approach. John Wiley & Sons. New York. 504 pp.
Williams, E.J. 1969. Regression methods in calibration problems. Bulletin of the International Statistics Institute 43: 17-28.
Wu, C. 2004. Combining information from multiple surveys through the empirical likelihood method. Canadian Journal of Statistics 32: 1-12.
Zieschang, K.D. 1990. Sample weighting methods and estimation of totals in the consumer expenditure survey. Journal of the American Statistics Association 85: 986-1001.