Transactions of the American Fisheries Society

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utaf20

Objectively Assigning Species and Ages to Salmonid Length Data from Dual-Frequency Identification Sonar

W. S. C. Gurney\textsuperscript{ab}, Louise O. Brennan\textsuperscript{c}, P. J. Bacon\textsuperscript{b}, K. F. Whelan\textsuperscript{d}, Martin O’Grady\textsuperscript{e}, Eileen Dillane\textsuperscript{f} & P. McGinnity\textsuperscript{f}

\textsuperscript{a} Department of Mathematics and Statistics, University of Strathclyde, Level 9, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, Scotland
\textsuperscript{b} Marine Scotland, Freshwater Laboratory, Pitlochry, PH16 5LB, Scotland
\textsuperscript{c} Earth and Ocean Sciences, School of Natural Sciences, National University of Ireland-Galway, University Road Galway, Ireland
\textsuperscript{d} School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland
\textsuperscript{e} Inland Fisheries Ireland, Swords Company, Swords Business Campus, Balheary Road, County Dublin, Ireland
\textsuperscript{f} School of Biological, Earth and Environmental Sciences, University College Cork, Distillery Fields Campus, North Mall, Cork, Ireland

Published online: 09 Apr 2014.

To cite this article: W. S. C. Gurney, Louise O. Brennan, P. J. Bacon, K. F. Whelan, Martin O’Grady, Eileen Dillane & P. McGinnity (2014) Objectively Assigning Species and Ages to Salmonid Length Data from Dual-Frequency Identification Sonar, Transactions of the American Fisheries Society, 143:3, 573-585

To link to this article: http://dx.doi.org/10.1080/00028487.2013.862185

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
ARTICLE

Objectively Assigning Species and Ages to Salmonid Length Data from Dual-Frequency Identification Sonar

W. S. C. Gurney
Department of Mathematics and Statistics, University of Strathclyde, Level 9, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, Scotland; and Marine Scotland, Freshwater Laboratory, Pitlochry, PH16 5LB, Scotland

Louise O. Brennan
Earth and Ocean Sciences, School of Natural Sciences, National University of Ireland–Galway, University Road, Galway, Ireland

P. J. Bacon*
Marine Scotland, Freshwater Laboratory, Pitlochry, PH16 5LB, Scotland

K. F. Whelan
School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland

Martin O’Grady
Inland Fisheries Ireland, Swords Company, Swords Business Campus, Ballyheary Road, County Dublin, Ireland

Eileen Dillane and P. McGinnity
School of Biological, Earth and Environmental Sciences, University College Cork, Distillery Fields Campus, North Mall, Cork, Ireland

Abstract
Fishery managers need robust ways of objectively estimating the quantitative composition of fish stocks, by species and age-class, from representative samples of populations. Dual-frequency identification sonar data were used to first visually identify fish to a broad taxon (Salmonidae). Subsequently, kernel-density estimations, based on calibrated size-at-age data for the possible component species, were used to assign sonar observations both to species (Atlantic Salmon Salmo salar or Brown Trout Salmo trutta) and age-classes within species. The calculations are illustrated for alternative sets of calibration data. To obtain close and relevant fits, the approach fundamentally relies on having accurate and fully representative subcomponent distributions. Firmer inferences can be made if the component data sets correspond closely to the target information in both time and space. Given carefully chosen suites of component data, robust population composition estimates with narrow confidence intervals were obtained. General principles are stated, which indicate when such methods might work well or poorly.

*Corresponding author: baconpj@marlab.ac.uk
Received September 2, 2013; accepted October 30, 2013
Modern fisheries management and stock assessment programs require accurate census data to parameterize predictive models and to determine the success of management initiatives regarding the achievement of conservation goals. While estimates of the population sizes of Atlantic Salmon Salmo salar, which are based on recreational and commercial fishery catches, provide large, data-rich samples upon which to base stock-size assessment, many of the fisheries upon which scientists depended for information to drive census calculations, particularly those from commercial fisheries, have been discontinued recently. Furthermore, much wild Atlantic Salmon production can occur in larger rivers, where it is difficult to separately count both salmon and Brown Trout Salmo trutta. In Ireland, for example, 15 of 140 recognized Atlantic Salmon rivers have the potential to produce 60% of the salmon (SSC 2008–2009). The engineering demands of building solid counting weirs (e.g., crump-weirs with associated resistivity counters) and the related costs are usually prohibitive. Video, photography, and viewing towers are often not viable options because of the color and turbidity of many temperate rivers (Brennan et al. 2008, 2009). Acoustic-imaging counters, based on sonar technology, offer novel and useful solutions to counting fish in such difficult environments (Fleischman et al. 2003; Maxwell and Gove 2007; Burwen et al. 2003, 2004, 2010).

In appropriate settings acoustic-imaging counters (hereafter acoustic images) are one way to get both taxonomic (from visual examination of images) and size (by measuring calibrated perpendicular images). This paper builds on a recent study into the utility of acoustic image counters, undertaken on the River Deel, a tributary of the River Moy in western Ireland (Brennan 2013). Key questions addressed were the numbers and sea-age composition of Atlantic Salmon in the river. The high quality acoustic images obtained at short ranges could be used to first visually distinguish, with high confidence, salmonids from other fish taxa (e.g., European Eel Anguilla anguilla, Eurasian Perch Perca fluviatilis, Northern Pike Esox lucius, Roach Rutulus rutilus. The only salmonids inhabiting the Deel are Atlantic Salmon and Brown Trout. Several life history forms of Brown Trout occur in the Deel, including “sea-trout” (the vernacular for individual Brown Trout that spend part of their life in the ocean) and possibly “ferox” trout (the vernacular for large, cannibalistic lake-dwelling Brown Trout). As hydroacoustic data are unable to distinguish these morphs, the term “trout” is used hereafter to refer to any fish determined to be Brown Trout. Even high-quality DIDSON images did not enable small adult Atlantic Salmon to be visually distinguished from large trout. Furthermore, the likely age-class compositions within the salmonid species were entirely unknown and not illuminated by simple inspection of the sonar images. However, size data measured from acoustic images (for both calibration and wild fish), could possibly be used to (1) infer the ratios of salmon to trout, and (2) similarly deduce the age-structures within each species.

Recent publications show that the average sizes of Atlantic Salmon in Scotland change systematically with seasonal date within the year (Bacon et al. 2009). More importantly, in the present context, the combination of seasonal-date and body-length of fresh-run Atlantic Salmon in Scotland gives a very reliable discrimination (about 95%) between 1 sea-winter (1SW) and multi-sea-winter (MSW) salmon (Bacon et al. 2011; see also http://www.mathsat.strath.ac.uk/outreach/salwrd/ for extended details).

However, while the approach of Bacon et al. (2011) suggested that the desired discrimination might well be possible on the Deel, it was apparent that their detailed approach was not ideal for the available Deel data. Their approach had not been calibrated with data on Atlantic Salmon from Ireland, and it relies on knowing the fish are definitely Atlantic Salmon. It also relies on data about fresh-run (including angled) fish, whereas the acoustic images of Deel fish moving in autumn could well relate to early-run (and smaller) salmon that had been in the river for some months.

A number of statistical techniques exist to reconstruct a target distribution (probability density function [PDF]) from two or more known potential component PDFs, which occur in unknown ratios in the target distribution (e.g., here, as the ratio of fish species). Using the target salmonid length PDF observed at the Deel (from instrument-setting-calibrated acoustic images), we applied kernel density estimation in a powerful and flexible framework to obtain best-estimates of both (1) the proportionate contributions by species, and (2) the proportionate age distributions within species. It provides a case study of how counts of total fish can potentially be separated into subunits that are more relevant to both population biology and to fish management.

**METHODS**

**Study area.**—The Moy catchment has 177 km of main river channel draining an area of approximately 2,000 km² (Figure 1). The system comprises two sixth-order subcatchment basins of approximately equal size. The eastern subcatchment consists only of the main river channel and tributaries, while the western subcatchment includes two large, interconnected lakes (loughs): a single main tributary, the Deel, which flows into Lough Conn (57 km²), and the Clydagh and Manulla rivers, which flow into Lough Cullin (8 km²). The Deel is approximately 37 km long and drains an area of around 229 km².

**Salmonid populations.**—The most recent (2006–2010) annual estimates of the run abundance of Atlantic Salmon from the Moy fishery are about 43,000 adults (range, 27,857–55,174), which includes the precoastal and freshwater fishery catches. Anglers at the Moy fishery (which includes the Ridge Pool and the waters downstream to the estuary) in Ballina, catch an average of 797 salmon (310–1,666) per year, during an open season between the February 1 and September 31. However, in recent years, the earliest salmon were not angled there until the third week of April. An estimated average of approximately 10,000 fish are angled on the river system as a whole, with peaks of large spring-run fish (2-sea-winter fish) captured in late April.
Brown Trout occur in the loughs and streams, and resident Brown Trout from the loughs migrate up rivers to breed. Both Brown Trout and salmon migrate into the Deel in autumn to spawn.

**Biometric data for Deel salmonids.**—Scale samples from 257 Atlantic Salmon were collected in 2008 from salmon pools at the tidal limit of the Moy in County Mayo, western Ireland (Figure 1) for age determination and genetic stock identification (GSI). Salmon were collected by rod anglers during the peak of the recreational rod fishing season (from April to September, but omitting February and March) and measured to 0.5 cm FL. Note that although all Deel salmon enter via the Moy estuary, not all Moy salmon ascend the Deel.

Because small numbers of salmon enter the Moy estuary in every month of the year, the achieved salmon sampling period was curtailed. This is likely to somewhat truncate the recorded size distributions in comparison to the true distributions, and differently so for MSW fish, with a peak run around April, compared with 1SW fish, which run from May onwards. Relatively few fish are thought to enter the estuary after October. Because the PDFs for salmon size by sea age were neither entirely random nor fully representative samples, this analysis also used an alternative description of the component salmon sea-age PDFs (see Scottish Salmon Data below).
Trout living in the Deel itself are mainly small parr, below the size range that can be accurately measured by dual-frequency identification sonar (DIDSON). However, larger Trout from the loughs migrate up the Deel to breed. Data on the likely sizes of large trout in the Deel system were accordingly approximated from gill-netting surveys conducted in Lough Conn in August (just prior to spawning) in 1994, 1998, 2001, and 2005. Trout captured in August were measured for fork length, and scale samples were read to determine ages.

Atlantic Salmon spawn in the Deel from late November through to the end of February, peak spawning occurring in the latter half of December (Brennan 2013). Trout spawning is generally believed to precede salmon spawning by a week or two.

Scottish Atlantic Salmon biometric data.—Given the truncated seasonal observations for which Irish size-at-sea-age data were available for salmon from the Moy, an alternative set of salmon size PDF data were also used. This comprised the 186,000-record subset of Scottish data described in Bacon et al. (2011; AllSA subset) and better represented early-run MSW fish. Measurement errors (repeatability for individual fish) of the Scottish data were about 0.5% (authors’ unpublished data).

DIDSON imaging system.—The DIDSON technology was installed in the Deel in October 2007 at a site approximately 1 km upstream of Lough Conn but below the available spawning areas. The equipment was operated continuously from November 13, 2007 until December 31, 2008. The river width at the selected site was 27 m; the counting zone was delimited by physical fish-barrier fences (Brennan et al. 2009) to a 10-m width.

Deel DIDSON operations.—A standard multi-beam DIDSON developed by Soundmetrics, Seattle, Washington (http://www.soundmetrics.com/), was used at the Deel site. The DIDSON was operated at 1.8 MHz (high frequency mode) with a beam width (two-way) of 0.3° horizontal by 14° vertical, with a total of 96 beams (Brennan 2013). The initial configuration was based on published recommendations (Cronkite et al. 2006; Maxwell and Gove 2007). Subsequent refinements were made to the software settings to optimize performance at the site to fine-tune the use of Convolved Samples over Threshold (CSOT) processing and motion detection (Brennan 2013).

Potential sites for the DIDSON’s location were assessed using bed-profile and substrate surveys, and the most suitable (best bed profile with least substrate reflective interference) was chosen following trial operations that tested the DIDSON at each potential site. Guide-fences were used to restrict fish (moving both upstream and downstream) to a distance of 2–12 m (usually <11 m) from the DIDSON. The extent and adequacy of the sonar beam were established using a tungsten sphere (~38.5 dB). The extent of the beam was mapped based on the location of the sphere within the beam (Brennan 2013).

The DIDSON’s motion detection function was used to detect and record potential fish signals, and subsequent fish size measurements for all fish >15 cm FL were obtained manually (as distinct from the manufacturer’s automatic option) for greater accuracy. The short ranges to fish targets (2–12 m) allowed salmonids (as opposed to other taxa and debris) to be identified visually with high confidence, based on near video-quality acoustic images of fish appearance and swimming behavior. A semi-automated procedure was adopted at the Deel by Brennan (2013), whereby the system’s motion detection automatically identified and saved image files (CSOT files) that included likely fish signals. These files were subsequently viewed by trained observers, assigned upstream or downstream movement categories, classified as salmonids or not salmonids, and manually measured (using the manufacturer’s Mark Fish software tool). Such manual measurements of fish length from DIDSON images have been shown to be more accurate than the software-automated values (Boswell et al. 2008).

Calibration of DIDSON fish-length recording.—An experiment was undertaken to calibrate the known lengths of fish to those recorded by the DIDSON. Fifty-two live fish were lightly hooked on rod and line and briefly passed through the DIDSON beam under typical flow conditions and instrument settings. Calibration fish of a variety of species (salmon, trout, perch, pike, and roach) and a wide range of field-measured fork lengths (31–110 cm) were measured in a variety of situations, including fish near the riverbed and close to the surface and at a variety of directions and distances from the sonar source (Brennan 2013). Lengths were recorded to the nearest 0.10 cm. Although such precision exceeds the accuracy of any single measurement, rounding it to a lesser accuracy would, by inflating the technique error, slightly impair estimation of the average, which is the target of the calibration exercise for population-level discrimination of species and age-classes.

Analysis of the calibration data showed that neither the distance of target fish from the sonar source nor the angle of the fish to the sonar affected the calibration accuracy, assessed as both single variable relationships and as an overall multiple regression (predicting true length from DIDSON length, range, and angle). This yielded a simple linear equation between DIDSON recorded lengths ($L_D$) and true fork lengths ($L_T$), correcting by just 2.07%:

$$L_D = 0.0 + 0.9797 \cdot L_T + \varepsilon_D,$$

where $\varepsilon_D$ represents the measurement error for the calibration data set. A regression forced through the origin was just as powerful, more logical, and not significantly different from a regression with a fitted intercept (slope = 0.9797, $P < 0.001$; $R^2 = 0.92$). The average correction to the DIDSON measurements (scaling by a factor of 0.9797) was smaller than the measurement error on single length-estimations, $\varepsilon_D$, which represented a coefficient of variation of 3%, and was, effectively, normally distributed. In subsequent analysis, both the 0.9797 correction factor and the single-measurement error coefficient (3%) were subsequently used to correct and fuzzify (i.e., broaden by adding the simulated measurement error) the physically...
measured salmonid fork length component distributions, so that they would properly match the overall target distribution, as derived from wild fish passing through the DIDSON beam (see below for details). This step could be much more critical, if in other situations the sets of \( L_D \) and \( L_T \) measurements matched less well.

Our analysis considers only data on fish moving upstream, whose orientation into the current provided a more dependable perpendicular view from which to reliably estimate their body lengths. The choice of upstream-moving fish also reduces (but does not eliminate) the potential for double-counting of any fish that may have been recorded passing in both directions, thereby improving the independence of the data points.

Statistical analysis.—Numerous workers (e.g., Aitcheson and Aitken 1976; Sanvincente-Añorve et al. 2003) have recognized that size distributions measured in the wild are generally composites of narrower distributions attributable to identifiable components (species, age-classes, etc.) of the system under observation. Kernel density estimates (Schartau et al. 2010) are frequently used to describe the overall distribution and then model the component distributions using standard distribution functions (such as the normal). In this work we followed Schartau et al. (2010) in using a kernel density estimate of the overall distribution, but then used ancillary component-specific data to also derive kernel density estimates for the component distributions. Numerical methods were then used to determine the set of relative sizes for the possible component populations, which would lead to the observed overall distribution.

A key aspect of such work is to ensure that the subcomponents would be structurally able to reproduce the target. In particular there should not be gaps in the component distributions where there are data in the target, and none of the components should completely overlap another. Ensuring that the available data adequately meet the assumptions of the method requires some careful initial data exploration, which for brevity and simplicity, we here describe under statistical methods (as it is needed to fine-tune the general approach to the constraints of the particular data) before the biological results deduced from the validated analysis procedures.

The relative (normalized) density functions at the heart of this paper are kernel density estimates computed using the R routine called “density” with 256 bins (classes of fish sizes) over the range of 0–120 cm. Default smoothing values were used, but for samples of less than 10 fish the smoothing bandwidth was adjusted. Equation (1) was used to adjust the component distributions (of \( L_T \), accurate fork length measurements). Each point in the original data were replaced by 100 simulated values with their mean equal to 0.9797 times the original value and a CV about that mean of 3%. This appropriately fuzzifies the original \( L_T \) fish measurements into the somewhat (in this example only slightly) wider range that would result via technique error from DIDSON measurements (\( L_D \)) of the same fish. Figures 2 and 3 below illustrate single subpopulation component distributions, showing, for each species and age-component, both true \( L_T \) and \( L_D \) fuzzified length distribution PDFs, and report both the number of points in the samples and the fuzzifying bandwidth for each distribution. With the measurement-error CV as low as 3%, as in this study, the two sets of distributions are actually nearly identical (Figures 2, 3).

Normalized density functions had units of \( cm^{-1} \) and were transformed into absolute density functions (fish numbers/centimeter length-class) by multiplying by the total number of individuals in the sample from which the relative density estimate was obtained.

To fit the absolute density function, \( D(l) \), estimated for the DIDSON data from the Deel, with a set of normalized basis functions, \( B_i(l), i = 1 \ldots N \), we asserted that

\[
D(l) = \sum_{i=1}^{N} N_i B_i(l)
\]

and then chose a set of subpopulations (\( N_i \)) so as to minimize the cumulative square error between the observed and predicted values of \( D(l) \). This minimization used a Nelder–Mead algorithm as implemented in the R routine “optim”, with parameter scaling set to the initial parameter set, relative tolerance equal to \( 10^{-8} \), and a limit of 5,000 iterations.

The relative density distributions for Lough Conn trout and Atlantic Salmon from the Moy, as well as the Scottish canonical salmon data set are contrasted in the center panel of Figure 4. No combination of the above initial subcomponent distributions formed a sound basis for a fitting exercise since many of them were almost wholly confounded. To obtain a reasonably well-conditioned basis for predicting the target distribution shown in the top panel of Figure 4, the combined length distributions for trout of three age groups—1 and 2 years, 3 and 4 years, and 5, 6, and 7 years—were formed. Because the samples from these freshwater scale-read age-groups of trout were of very different sizes, it was inappropriate to simply amalgamate the data and reestimate the relative density function. Instead we assumed that the survival system was in a statistically stationary state over years, with a year-to-year survival of \( S \). Thus, three length distributions of the relative contributions of amalgamated trout age-classes could be rewritten to be constrained to depend on \( S \) as follows:

\[
B_{12}(l) = b_1(l) + S b_2(l);
B_{34}(l) = b_3(l) + S b_4(l);
B_{567}(l) = b_5(l) + S b_6(l) + S^2 b_7(l).
\]

The assumed year-to-year survival was initially chosen, somewhat arbitrarily, as \( S = 0.5 \). In practice, for these Moy data, the results proved rather insensitive to the actual value of \( S \) used, (but note that with other data the magnitude of \( S \) could potentially be influential). The overall effect of such a general constraining mechanism, however, was central to subsequently obtaining clear biological discriminations.
FIGURE 2. Normalized length density distributions for the Irish (Lough Conn) Brown Trout data (ICTA) stratified by scale-read age. DIDSON observed length ($L_D$, black line) is related to directly observed length ($L_T$, red line) by $L_D = 0.9797L_T + \varepsilon_D$, where $\varepsilon_D$ is a normally distributed random variable with mean of zero and CV = 3%. The directly observed distribution is estimated from the raw data; DIDSON observed distributions are estimated from simulated data in which each directly observed point is represented by 100 simulated points. Both distributions are kernel density estimates obtained using R routine “density” with 256 bins over the range 0–120 cm and all other parameters set to their default values, except for the smoothing bandwidth in the case of the DIDSON simulated data, which is forced to the default value for the equivalent directly observed data set.

FIGURE 3. Normalized length density distributions for Atlantic Salmon data stratified by scale-read sea-ages. The upper row shows distributions from the canonical Scottish scale-aged data set (allSA of Bacon et al. 2011). The lower row shows distributions from the scale-aged data from the Irish River Moy (IMS). The DIDSON observed length ($L_D$, black line) is related to directly observed length ($L_T$, red line) by $L_D = 0.9797L_T + \varepsilon_D$, where $\varepsilon_D$ is a normally distributed random variable with mean zero and CV = 3%. For estimation details of $L_D$ see Figure 2.
FIGURE 4. Target DIDSON length distribution from the River Deel and its potential components, both transformed into DIDSON length units (see Figures 2, 3). Top panel shows the target data set, middle panel shows all its possible components, and bottom panel shows a rationalized subset of components used for the fitting (obtained by combining the raw age distributions for trout by assuming 50% mortality between year-classes). In the lower two panels, double width lines show Atlantic Salmon (continuous line = Scottish canonical; dashed line = River Moy) with ages of 1 = red, 2 = green, and 3 = blue) and single width lines are Lough Conn Brown Trout. In middle panel age 1 = continuous red line; age 2 = continuous green; age 3 = continuous blue; age 4 = dashed red; age 5 = dashed green; age 6 = dashed blue; age 7 = long dashed red. In the bottom panel, age 1 + age 2/2 = red, age 3 + age 4/2 = green, and age 5 + age 6/2 + age 7/4 = blue.

As might be expected after examining the amalgamation of the three combined trout age-classes with the distributions derived from the Moy salmon data (the lowest panel of Figure 4), the revised components did not provide a wholly satisfactory fit to the Deel density distribution. However, the Lough Conn trout combined age-class data plus the Scottish canonical length distributions for salmon of sea ages 1, 2, and 3 produced a fit that was sufficiently plausible for it to be pertinent to investigate and quantify how well the component age-class proportions could be identified. To answer this question a 50,000-element Markov chain–Monte Carlo (MCMC) simulation was constructed from the stationary distribution of the parameters by using the Metropolis–Hastings method implemented in R routine “metrop” (package mcmc). Completely uninformative priors were used, with the assumption that the 256 measurement-bands, which composed the estimate of the Deel absolute density function, were normally distributed. Hence, because both the SD of this assumed distribution and the unconditional distribution of the data cancel out of the Markov chain

FIGURE 5. The target DIDSON length distribution from the Deel fitted with subcomponents from Irish data only. The heavy black line shows the target, and the light black line the fitted envelope. Colored lines show the components as noted in Figure 4. Component amplitudes are $N_{1+2/2} = 1.22$, $N_{3+4/2} = 1.132$, $N_{5+6/2+7/4} = 1.645$, $N_{1+2} = 2.543$, $N_{3+2} = 2.507$, and $N_{3+3} = 1.06$. Cumulative square error = $1.6 \times 10^4$.

FIGURE 6. Target DIDSON length distribution from the Deel fitted with Irish trout subcomponents but Scottish salmon subcomponents of sea-ages 1, 2, and 3. The heavy black line shows the target and the light black line the fitted envelope. Colored lines show the component contributions, as in Figure 4. Component amplitudes are $N_{1+2/2} = 1.22$, $N_{3+4/2} = 1.132$, $N_{5+6/2+7/4} = 1.645$, $N_{1+2} = 2.543$, $N_{3+2} = 2.507$, and $N_{3+3} = 1.06$. Cumulative square error = $1.6 \times 10^4$. 
rule, it was possible to simply use the negative sum of the square errors (i.e., the negative of the objective function minimized in the previous fitting process) as our unnormalized log-likelihood.

To obtain satisfactory estimates of the stationary distribution required the rejection ratio for the Markov chain to be in the range 0.1–0.3, which was achieved by adjusting the SD of the proposal distribution. In the case of the four components with finite populations (N values), we set the proposal SD to be 0.1% of our initial guess at the mean, while for the two populations shown by the fitting process to be close to zero, we set the proposal SD to 1% of an initial crude estimate of the mean (which was set to 1). Bivariate parameter distributions were described using credibility contours calculated by making a 2D kernel density estimate for the appropriate pair of parameters in the MCMC sequence (using the R routine “kde2d”) and constructing contours on this estimated distribution.\(^1\)

**RESULTS**

**Component Size-at-Age Distributions of Lough Conn Trout**

The length distributions (PDF histograms) for the different sub-component age-classes of Lough Conn Brown Trout are shown in Figure 2. The sample sizes of freshwater (FW) ages 1, 6 and 7 trout were very low (n = 2, 8, 3, respectively), whereas n ranged from 30 to 80 for the other single-year age-classes. The corrections resulting from DIDSON calibration adjustment ranged from 30 to 80 for the other single-year age-classes.

**Component Size-at-Age Distributions for Atlantic Salmon**

The two importantly different potential sets of Atlantic salmon subcomponent size-at-sea-age data were available: the Scottish data and the Moy estuary (Irish) data. Their PDFs (Figure 3) are illustrated on the same horizontal scale as for Lough Conn trout (Figure 2). Note that the bulk of the Irish 1SW salmon were smaller than (to the left of) the modal value of the Scottish 1SW salmon (Figure 3; Figure 4 lower), and accordingly their mean lengths (54.4 and 61.6 cm, respectively) were very different. Similarly, a fair proportion of Irish 2SW salmon were between 50 and 60 cm, whereas almost no Scottish 2SW salmon were less than 60 cm. Furthermore, the Irish data show virtually no 2SW salmon above 85 cm, whereas appreciable numbers of Scottish 2SW salmon are bigger than this. Thus, although the mean lengths of the Irish and Scottish 2SW salmon were more similar (71.8 and 75.7 cm) than were the means of their 1SW counterparts, their two component PDFs were still very different.

**Deel Acoustic Image Target PDF and Potential Subcomponents**

The overall salmonid-size PDF, measured using the DIDSON, represents the target distribution and was recreated to a close approximation (from the suitably weighted proportions of the various potential subcomponents). The DIDSON target distribution is shown as the top panel of Figure 4; the lower two panels show the (similarly scaled) candidate subcomponents. The middle pane shows the full set of seven age-groups of Irish Lough Conn trout plus salmon segregated into both two sea age-groups (Irish data) and three sea age-groups (Scottish data). The lower pane shows the resultant three-component Lough Conn trout distributions, obtained following the survival-weighted recombination, that were used in the final fitting to predict the target distribution. The differences between the PDFs of Irish and Scottish salmon, for both 1SW and 2SW age-groups is emphasized in the lower pane of Figure 4 (note the Irish salmon data had no 3SW fish, but the Scottish did). The unhelpful broad overlaps between the original annual trout age-classes are clearly evident in the middle pane of Figure 4, while the much clearer separation between the combined trout age-classes is emphasized in the lower panel. Note particularly that the small sample of Irish trout aged 7 (middle panel) overlapped very considerably with Irish 1SW salmon, whereas the overlaps between Irish 1SW salmon and the oldest combined trout age-class is much less (lower panel). This clearly indicates why the combined age-classes gave improved discrimination.

**Fits Using Irish Trout and Irish Salmon Data**

The best fit result (using nonlinear optimization) based on Irish data for both the salmon and the trout subcomponents fitted rather poorly (Figure 5). The broad black line (the acoustic image target curve), shows appreciable deviations, both excesses and deficits, from the best estimate based on appropriately estimated proportions of the subcomponents (thin black line); this fit has a high cumulative square error of 1.5 × 10^5. The relative magnitudes that each subcomponent contributed to the best estimate (thin black line) are shown by the colored lines (see legend for details).

The deviations between the observed and best-fit lines are particularly instructive. The sequence of deviations between the observed target and prediction based on Irish Moy data are (peak to trough): 22 to 32, 32 to 48, 48 to 58, 58 to 72, 72 to 80 cm. The deep trough (low deviation) between the 58 and 72 cm is particularly marked and problematic for these Irish-only calibration data because it occurs in a size range where the only plausible fish subcomponents to fill it are (Irish) 1SW or 2SW salmon. This indicates that the present calibration sample of Moy estuary salmon may not be completely representative of the salmon run in the Deel River.

**Fits Using Irish Trout and Scottish Salmon Data**

Simple visual inspection of the corresponding initial best-fit results based on Irish trout and Scottish (instead of Irish) salmon components show (Figure 6) that the new best estimate (thin black line) fitted much better to the acoustic target PDF (thick black line) with very much smaller peak- and trough-deviations. In particular, the former big deviations in the

\(^1\)Both the univariate and bivariate parameter estimates are illustrated below (see Figure 7).
size-range 58–72 cm almost disappeared because the Scottish salmon calibration data include some 1SW and 2SW salmon of sizes not recorded in the Moy calibration sample; in addition, all other deviations are much reduced. The improved fit of the new estimate is confirmed by the greatly reduced cumulative square error (down by an order of magnitude from $1.5 \times 10^5$ to only $1.6 \times 10^4$). Figure 7 illustrates the resulting narrow credibility bounds around these population (parameter) estimates.

**Parameter Distributions**

The Scottish calibration data achieved greatly improved overall fits (achieved by nonlinear optimization). Moreover, the Scottish calibration data implied a very different ratio of 1SW to 2SW salmon in the Deel (101% instead of 55%, see Table 1). Interestingly, despite the very narrow limits (SDs) of the MCMC fits, the overall nonlinear optimization estimates were within the narrow MCMC credibility envelopes (see the Scottish data, 1SW and 2SW contributions; the estimated 3SW contribution is clearly too tiny to be accurate).

**DISCUSSION**

Fishery managers frequently need to make decisions based on sparse or incomplete data; the ideal information desired for scientific rigor is rarely available either widely or locally. Our results for the Irish Deel catchment show that, even when ideal data are missing, judicious choice of how the available data are used (e.g., our combination of trout age-classes) as well as the use of information from further afield (the Scottish data) can provide valuable insights that build into an informative

![Figure 7](image_url)

**FIGURE 7.** Markov chain–Monte Carlo (MCMC) parameter estimates for the fit to the Deel DIDSON data using Irish trout information (with some age-classes combined as before) but Scottish sea-ages 1, 2, and 3 sea-winter salmon. The MCMC sequence assumed an uninformative (uniform) prior and had a burn-in of 1,000 steps followed by 50,000 elements. The top two rows show histograms for the final contributions associated with each of the six components by species and age grouping. The bottom row shows 95%, 50%, and 25% credibility contours for a selection of bivariate marginal distributions. The mean ± 1 SD values for the parameters were $N_{t_1+2/2} = 0.14 \pm 0.23$, $N_{t_3+4/2} = 1.131 \pm 2.30$, $N_{t_5+6/2+7/4} = 1.644 \pm 2.50$, $N_{s_1} = 2.545 \pm 2.32$, $N_{s_2} = 2.505 \pm 2.43$, and $N_{s_3} = 0.1431 \pm 0.11$. 
TABLE 1. Estimated (from DIDSON observations) salmonid composition of the River Deel by species and age, as predicted from different component data sets and fitting methods. The cumulative square error (CSE) for the two nonlinear optimization fits are given. Credibility ranges of the population compositions (expressed as SDs) are shown for Markov chain–Monte Carlo fits. Brown Trout data were always from Lough Conn, Ireland; ratios of 1SW/2SW (where SW = sea winter) salmon and the two older combined trout age-groups (3–4)/(5–7) are given to show how these vary depending on which (Irish or Scottish) set of component salmon distributions were used. Predictions of <0.5 fish are shown as zero; na = not available.

<table>
<thead>
<tr>
<th>Salmon data source</th>
<th>Salmon sea-age composition</th>
<th>Ratio (%)</th>
<th>Trout freshwater ages (combined)</th>
<th>Ratio (%)</th>
<th>Whole model CSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1SW</td>
<td>2SW</td>
<td>3SW</td>
<td>1SW/2SW</td>
<td>1–2</td>
</tr>
<tr>
<td>Irish River Moy means</td>
<td>1,824</td>
<td>3,337</td>
<td>na</td>
<td>54.7</td>
<td>0</td>
</tr>
<tr>
<td>Scottish means</td>
<td>2,543</td>
<td>2,507</td>
<td>1</td>
<td>101.4</td>
<td>1</td>
</tr>
<tr>
<td>Scottish SD</td>
<td>2.545</td>
<td>2.505</td>
<td>0.14</td>
<td>101.6</td>
<td>0.14</td>
</tr>
<tr>
<td>Markov chain Monte Carlo</td>
<td>2.32</td>
<td>2.44</td>
<td>0.10</td>
<td>0.23</td>
<td>2.30</td>
</tr>
</tbody>
</table>

The Moy Case Study

Our results show that the proposed kernel density approach worked well and was able to provide close fits to the Deel DIDSON data, given suitable subcomponent distributions. An advantage of the procedures here described is the ability to the MCMC process to produce confidence intervals around the estimates. However, to obtain close fits, the approach fundamentally relies on having accurate and fully representative subcomponent distributions that do not largely overlap each other. Potentially confounding information was first filtered by taxon (salmonids or nonsalmonid), as allowed by the high-quality images available from the Deel (Brennan, 2013). Secondly, the subcomponents were carefully designed to have minimal confusion and maximum separation by size. The importance of this was well illustrated by the need to decompose the salmon data into sea age-classes and the trout data into carefully chosen combinations of freshwater age-classes. That extra flexibility allowed the kernel density estimation to obtain a better fit to the aged trout size data by effectively altering the age composition between the observed Lough Conn components and the trout using the Deel but retaining an ability to discriminate between the trout subcomponents. In addition, in this example, the limited sample sizes of some Lough Conn trout components required the use of combined age-classes, a procedure that can be usefully and realistically simplified by assuming that the age ratios should conform to a plausible survival pattern. In detail it should be noted that while sea-trout and lake Ferox trout might also use the Deel for spawning in small numbers, these were not represented in the Lough Conn trout sample.

Local data on the size-distributions and abundances of these potential life history components would be needed to investigate this further and to obtain more precise estimates of the likely detailed composition of the Deel.

The overall findings for salmon were broadly similar. The Irish 1SW salmon data (River Moy) included small fish not recorded in Scottish data and no large fish (>65 cm; Figures 3, 4). Furthermore, Irish 2SW salmon showed a very similar pattern that included smaller fish not recorded in Scotland (55–60 cm), no large fish (>85 cm), and also no 3SW fish (Figures 3, 4). A potential cause of the discrepancy could be if the small angled fish were a size-biased subset (i.e., more likely to take a lure) and the lack of larger salmon was due to the autumn-truncated seasonal rod sample. These size discrepancies raise interesting questions about salmon biology and fisheries between Ireland and Scotland, which more extensive data would help answer. In this regard a recent study (McGinnity, personal communication), using genetic stock identification to recognize biologically significant units of diversity within the Moy, shows that 1SW salmon returning in equivalent periods to the Deel tributary are on average 2 cm longer and 200 g heavier than fish returning to the western Moy. Thus the size distribution of the Deel salmon would indeed therefore be more typically reflected by the size distribution of the Scottish salmon, as described here.
Better fits were again obtained when the salmon component was subdivided into sea age-classes. It is entirely plausible that the ratio of 1SW to MSW salmon spawning in the Deel is not identical to that entering the Moy estuary. Furthermore, the sample of salmon sizes obtained from the Moy estuary was truncated in the spring (too few small fish) for MSW fish and in autumn for both 1SW and MSW salmon (too few large fish).

In this regard it is instructive that the Scottish data, comprising wider seasonal samples and three rather than two sea age-classes, allowed a better fit to the Deel DIDSON data. This somewhat strengthens the view that the poorer fit obtained with the Irish salmon components data were partly due to seasonal truncation because most of the Irish samples were collected between May and September, which could include overlooking any 3SW fish.

Technically it is clear that having more appropriate (suitable and locally relevant) subcomponents will allow the estimation procedure to achieve better overall fits. Our results indicate the sorts of improvements that can be obtained. However, unless the exercise is undertaken in a situation where the answer is also independently known, there remains a risk that over-enthusiastic complexity in the choice of subcomponents could lead to good, but spurious, fits. The keys to improved understanding would be more comprehensive and appropriate local data and fitting guided by good biological knowledge (McGinnity, persessional communications) and checking to ensure subcomponents are not themselves confounded.

The overall results of our enumeration of salmon and trout and their age-classes conformed to the broad expectations of Moy fishery management personnel. Our objective approach and the MCMC results define both the parameters (the relative contributions of the component species and age distributions) and their confidence intervals surprisingly well—in, to within 1% of the population-level composition ratios, when nontrivial numbers per class were observed (Figure 7; Table 1). However, those MCMC results are posterior distributions, and thus assume that both the component PDF data and the model structure are appropriate. The model structure and the component data are probably the weakest links because (1) the trout survivals, presently incorporated as a guess into the model structure, are not well known, and (2) the component distributions for the salmon are not ideal, being either time-truncated (Irish data) or data of Scottish origin. Given genuine trout mortality data and more appropriate salmon PDF estimates, the MCMC approach could be modified to reflect sensitivity to these aspects; unfortunately, with current data, such a sensitivity analysis would do little to increase biological understanding.

Accordingly, to use the method for serious monitoring purposes at the Moy, where an overall accuracy of about 5–10% is desired, then further field-work should clearly be done to allow calibration with more detailed and representative component data from the Moy system rather than rely on plausible inferences from eastern Scotland.

**Technical Implications**

Many authors have used hydroacoustic approaches to try and identify fish species and to monitor their populations (see references in Fleischman and Burwen 2003). However, Burwen et al. (2010; page 1306) note that “obtaining size related information from DIDSON images remains a largely unexplored area”. Our Deel case study shows that the use of size information, not only between species but also between age-classes within species, can, in favorable situations, greatly assist the accurate fitting of overall target size-distributions by subcomponents of species and age by size composition. The fitting approach used in our study is conceptually rather similar to the mixture models approach advocated by Fleischman and Burwen (2003), although we use fish lengths directly, rather than the more obscure echo-length standard deviation metric, which they reported.

A combination of careful site choice for the DIDSON on the Deel (Brennan, 2013), together with short distances to target fish (<12 m, which Burwen et al. 2010 state allows very accurate measurements) and evidently favorable fish behavior and orientations, lead to a low (3%) DIDSON single-fish measurement error, usefully less than the 5.76 cm RMSE (about 6.4% to 8.3% CV error, depending on fish-size) reported by Burwen et al. (2010) at ranges up to 21 m, and appreciably less than that in Fleischman and Burwen (2003). Our case study thus strengthens the suggestions of Fleischman and Burwen (2003) that such approaches, including the use of aged subcomponents, have merit, especially when accurate data can be obtained.

In discussing the accuracy and precision of such hydroacoustic measurements, Burwen et al. (2010) noted that taking multiple measures per fish echo record could reduce errors. Although such replicate measurements would clearly reduce the error (bias) for any single fish, unless one’s aim was to estimate lengths of particular fish, it is less clear that such replicate measurements would allow a better description of the size PDF for a population of fish than would the same number of single-estimates from a correspondingly larger sample of individual fish.

Fleischman and Burwen (2003) suggest that more sophisticated estimation procedures, such as combining sequential same-season size data from subsamples of netted fish, along with matching sequential hydroacoustic measurements, might allow improved identification and counting of fish by species across seasons within years. While this could be so in particular situations, our Moy case study, where a better fit was obtained from multiannual Scottish data than from recent local Irish (Moy) data, indicates a contrary risk. Unless the component samples utilized in fitting are fully representative, poor fits may result, in which case, unless the within-season subsamples were large and fully representative, a sequential approach might contribute more noise (error) than useful information. Although such sequential approaches could have merits, they probably need cautious evaluation in every new situation.
Fishery Management Implications

European stocks of Atlantic Salmon have declined recently, and their levels are still a cause of serious concern in both Ireland and Scotland (SSC 2008–2009; SALSEA-Merge 2011). With the closure of net fisheries and limited resources for monitoring wild salmon populations, the need for cost-effective and objective ways to monitor salmon stocks remains high (Brennan 2013). The monetary values of 1SW and MSW salmon to rod fisheries are very different, and the two sea-age components are reported as separate stocks by the International Council for the Exploration of the Sea (e.g., ICES 2009). Large river systems can be challenging in relation to stock assessment and catchment management, but the use of a DIDSON on the Deel provided objective observations of fish movements via a system that was easy to install and operate. This permitted the acquisition of real-time data and high-quality fish-length measurements that were not attainable using other counter technology (Brennan, 2013).

The SALWRD approach of Bacon et al. (2011) has shown that the sea-age of Scottish salmon can be assigned to 1SW or MSW with high reliability (95%) if length or weight and return date (even rod-capture date) are known. But this alone will probably not allow adequate stock monitoring in future. Both the present net and rod seasons are appreciably shorter than the salmon-run periods, and whether rod captures are fully representative of the entire run (i.e., not size biased) is unknown. The use of DIDSON data could supplement information on numbers and sizes outside the rod-fishing season; however, acoustic image data alone would not distinguish Atlantic Salmon from migratory trout (Brennan, 2013). But the combination of SALWRD style approach and kernel density estimation from hydroacoustic data, supplemented by appropriate small-scale random samples of the fish to check on the size by species and age distributions, would be much more powerful. Although assessments based on detailed locally sampled species and age subcomponent distributions are clearly more relevant and reliable, in the situations where such local data are completely lacking, initial approximations based on information from more distant sources could still be useful.

In fisheries management generally, kernel density estimates have the potential to disentangle species and age-class compositions in many situations, provided that subcomponents are not seriously confounded. The basic power of the method can be enhanced if some potential component species can be eliminated from high quality visual images (here the nonsalmonids), thus decreasing the separation needed by size alone. Discrimination on size will clearly be more feasible from high-precision sonar results, and more relevant when these are interpreted from subcomponents that are local in time and space. But note that many subcomponents can also prevent good discrimination, and in such situations forming fewer composite classes, based on likely survival schedules (here the trout age-classes), may yield better results.

ACKNOWLEDGMENTS

We thank Tiernan Henry (National University of Ireland, Galway) for helpful discussions and Nigel Bond (Marine Institute) for technical assistance during DIDSON operations. Declan Cooke kindly collected the salmon scales from the Moy. Funding for L.B.’s Ph.D. was provided by the Marine Institute and the Marine Research, Technology, Development, and Innovation (RTDI) Measure, Productive Sector Operational Programme, National Development Plan 2000–2006 (Grant-aid Agreement PhD/05/001). Equipment for the study (DIDSON dual frequency identification sonar) was purchased with the support of the Marine Institute and the Marine RTDI Measure, Productive Sector Operational Programme, National Development Plan 2000–2006, cofinanced under the European Regional Development Fund.

PMcG. was part supported by the Beaufort Marine Research award in Fish Population Genetics funded by the Irish Government under the Sea Change Programme.

REFERENCES


