Draft Working Paper for predissemination peer review only.

Working Paper 4.4 (supplementary)
April 15, 2008

A method to apportion landings with unknown area, month and unclassified market categories among landings with similar region and fleet characteristics

Michael Palmer

A Working Paper in Support of GARM Reference Points Meeting Term of Reference 4

Introduction

Previous Northeast groundfish stock assessments (e.g., Mayo and Terceiro, 2005) apportioned stock landings with unknown statistical areas among landings with known areas by applying an "x-ratio" (SAS code developed by R. Mayo, pers. comm.). A single x-ratio was applied to all stock landings with known statistical areas without any stratification of the data set (Eq. 1).

$$x_{ratio} = \frac{(\sum l_{known} + \sum l_{unknown})}{\sum l_{known}} \quad (1)$$

where,

l_{known} = landings records with known statistical area

$l_{unknown}$ = landings records with unknown statistical areas.

Likewise, the current Northeast Fisheries Science Center's (NEFSC) Commercial Data Biostatistical Analysis Program software application (BioStat v 5.10[1]) contains a method to apportion unclassified market categories among the classified market categories. The current BioStat method is analogous to the x-ratio approach.

The original x-ratio and BioStat methods do not use what known information exists in those landings with unknown areas or unclassified market categories. For example, often when statistical area is unknown, there is information on the port of landing, time of year, gear, ton class of a vessel and market category. This accessory data can be used to improve the precision with which landings with unknown attributes (e.g., area, market category) are apportioned among those landings where the information is known. By apportioning landings with unknown attributes among landings that have similar characteristics, the precision of the x-ratio method can be improved. Such a process can also be used to apportion landings records where the month of landing is unknown; an infrequent occurrence in the commercial fisheries database, CFDBS. Currently neither the original x-ratio or Biostat methods accommodate landings where the month of landing is unknown.

This paper describes a method to use the known information to more precisely apportion landings with unknown statistical area, month and/or unclassified market category by applying a stratified x-ratio. The method has been built into a landings extraction script that can be run on any species, or stock, for the entire range of data in the commercial landings database, CFDBS (1964 – 2004, *note: CF_AA tables are used from 1994 onward). The details of this new method are described below and the performance of the extraction code was assessed relative to the original x-ratio method using both Gulf of Maine and Georges Bank haddock landings as examples.

[1] *NOAA Fisheries Toolbox Version 2.10, 2006. Commercial Data Biostatistical Analysis Program 5.10. [Internet address: http://nft.nefsc.noaa.gov].*

Methodology

The stratified apportionment code is written in SAS[2] and can be run on both PC and Unix platforms. Because the code serves the dual purpose of data retrieval and apportionment of unknown landings, a direct link to the NEFSC's commercial fisheries database (CFDBS) is required. There are several input variables that a user must specify:

1. Species (CFDBS NESPP3 code);
2. Stock (or substock) definitions (define which statistical areas are included in which stock area);
3. The range of years for which the code is to be run on (can be run from 1964 onward);
4. The market assignments (i.e., market categories groupings);
5. Region specifications (this is used to group ports into broader geographic regions where vessels tend to fish on similar stocks. **Note necessarily the same as stock area*);
6. Gear groupings (group the gears used to target the species into major groupings, i.e., otter trawl, longline, gillnet);
7. Temporal groupings (month, quarter, semiannual, etc.);
8. How many stratification parameters to use in the allocation scheme (range: 1 – 6);
9. Which stratification parameters to use (statistical area, market, vessel ton class, gear group, region group, temporal group);
10. The order of the stratification parameters (because parameters are dropped from the strata set after each iteration, the order in which strata are dropped/retained can be important); and,
11. The file directory where the user wants to export BioStat and summary data files.

The code automatically determines the appropriate CFDBS table set to use for the data extraction based on the input year(s) (1964 – 1981, WOLANDS; 1982 – 1993, WODETS; 1994 – 2003 CFDETS_AA). All years are re-formatted to a common format to facilitate further data processing. The code then apportions data with unknown areas, months and/or unclassified market categories using the following procedure (the outline assumes all six stratification procedures are being used):

1. All landings records where statistical area is known (KNOWN), and where it is unknown (UNK) are summed grouping landings by region group, gear group, vessel ton class, temporal group, and market category. An expansion ratio is calculated for each stratum as (KNOWN + UNK) : KNOWN. **Note: because the method is attempting to apportion landings with unknown areas in the first step, the statistical area can not be used as a stratification parameter in this step (the code automatically drops it from the list of stratification parameters and will use it again when apportioning landings with unknown months and/or unclassified market categories.*
2. Multiplies the individual landings by the respective stratum expansion factor and zeros out those landings associated with the UNK landings.

[2] SAS Institute, Version 9.1.3, Cary, NC

3. Occasionally there are instances where there are no KNOWN landings for a particular strata grouping. In these instances the UNK landings can not be apportioned and are carried forward to the next step.
4. The code attempts to apportion any remaining UNK landings left after steps 1-2 by relaxing the stratification criteria by removing the last stratification parameter (e.g. market category) such that the expansion ratio of (KNOWN + UNK) : KNOWN is calculated only for the strata grouping that remains (e.g., region group, gear group, and vessel ton class, and temporal group). Step 2 is then rerun using the new expansion factors (note: this step is only performed if the number of defined strata parameters > 2).
5. If any UNK landings remain after step 3, the stratification criteria are relaxed further such that the third to last stratification parameter is removed and steps 1-2 are rerun note: this step is only performed if the number of defined strata parameters > 3).
6. If any UNK landings remain after step 4, the stratification criteria are relaxed further such that the fourth from last parameter is removed and steps 1-2 are rerun (note: this step is only performed if the number of defined strata parameters > 4).
7. If any UNK landings remain after step 4, the stratification criteria are relaxed further such that the fourth from last parameter is removed and steps 1-2 are rerun (note: this step is only performed if the number of defined strata parameters > 5).
8. Steps 1-7 are re-run with month being the unknown variable.
9. Steps 1-7 are re-run with market category being the unknown variable (*optional*).

Results

The extraction code has been tested on both Gulf of Maine and Georges Bank haddock stocks for the years 1994 to 2006. The performance of the code was assessed against the landings values reported in the 2005 GARM analyses (included landings through 2004, Mayo and Terceiro, 2005). It is important to note that the extraction code uses the new CFDETS_AA (Wigley et al., 2007a) tables rather than the CFDETS tables which were used in the 2005 GARM. There are minor differences in these landings which are described in detail in Wigley et al. (2007b).

In general the percentage of landings with unknown statistical areas is relatively small (< 1.0 % Table 1). The amount of landings with unknown statistical areas did increase after 2004. This increase is likely a result of electronic dealer reporting implementation which began in May, 2004. For both the Georges Bank (Fig. 1) and Gulf of Maine haddock stocks (Fig. 2), the landings estimates calculated by the new apportionment method correspond well with those of GARM 2005.

Sensitivity to order of stratification parameters

The code was run on 2003 haddock using three different orders of six stratification parameters:

Statistical area, temporal group, region group, ton class, gear group, market category
Statistical area, temporal group, region group, market category, ton class, gear group

Statistical area, temporal group, gear group, ton class, market category, region group

The impact of the order on the apportionment of the unknown landings to stock and market category groupings is shown in Table 2. In general there was little impact of changing the order of the parameters (≤ 0.01 %). Because there was little impact on 2003 haddock landings does not necessarily imply that other years and species will not be impacted. These results are a product of the properties of the underlying data. The sensitivity of stock landings to parameter ordering should be evaluated for each species.

Summary

This new method provides a systematic and objective way to apportion species landings with unknown attributes among those landings where the information is known. For haddock, the new method results are nearly identical to the results obtained by the previous method. This may not hold for all species, particularly those species where a high percentage of landings have unknown attributes. Analysts should compare the results of the new method to previously employed methods and assess the sensitivity to the ordering of the stratification parameters.

It is important to note that this method is species-specific but not vessel-specific. It will not preserve vessel landings. It may be possible to modify the code to preserve vessel landings, but the existing code does not and therefore should not be used for the purpose of sector allocations or monitoring.

References

Mayo, RK, Terceiro, M, editors. 2005. Assessment of 19 Northeast groundfish stocks through 2004. 2005 Groundfish Assessment Review Meeting (2005 GARM), Northeast Fisheries Science Center, Woods Hole, Massachusetts, 15-19 August 2005. *U.S. Dep. Commer., Northeast Fish. Sci. Cent. Ref. Doc.* 05-13; 499 p.

Wigley, S., Hersey, P., Palmer, J.E. 2007a. A description of the allocation procedure applied to the 1994 to present commercial landings data. GARM data meeting, October, 2007.

Wigley, S., Legault, C., Brooks, E., Cadrin, S., Col, L., Hendrickson, L., Mayo, R., Nitschke, P., Palmer, M., Sosebee, K. and Terceiro, M. 2007b. Annual comparisons of the trip-based allocated and the single-species prorated cmmercial landings, biological samples and numbers of landed fish at age. GARM data meeting, October, 2007.

Tables

Table 1. Total Georges Bank and Gulf of Maine haddock stock landings with unknown statistical areas assigned by the apportionment method, 1994 to 2006.

Year	Stock	Total landings (mt)	Landings assigned from unknown apportionment process (mt)	Percent of total landings (%)
1994	Georges Bank	206.4	1.4	0.7
1995	Georges Bank	231.4	1.9	0.8
1996	Georges Bank	320.3	2.9	0.9
1997	Georges Bank	880.2	1.1	0.1
1998	Georges Bank	1913.9	4.7	0.2
1999	Georges Bank	2572.1	28.1	1.1
2000	Georges Bank	3202.8	3.2	0.1
2001	Georges Bank	4819.7	17.0	0.4
2002	Georges Bank	6531.8	47.7	0.7
2003	Georges Bank	5759.5	9.6	0.2
2004	Georges Bank	7375.3	135.6	1.8
2005	Georges Bank	6603.7	561.0	8.5
2006	Georges Bank	2642.8	88.2	3.3
1994	Gulf of Maine	120.1	0.6	0.5
1995	Gulf of Maine	173.0	1.0	0.6
1996	Gulf of Maine	246.6	0.5	0.2
1997	Gulf of Maine	588.6	0.1	0.0
1998	Gulf of Maine	885.2	2.3	0.3
1999	Gulf of Maine	542.5	3.8	0.7
2000	Gulf of Maine	737.9	1.6	0.2
2001	Gulf of Maine	929.2	2.0	0.2
2002	Gulf of Maine	976.9	1.5	0.2
2003	Gulf of Maine	1023.0	5.3	0.5
2004	Gulf of Maine	946.5	43.7	4.6
2005	Gulf of Maine	961.5	81.6	8.5
2006	Gulf of Maine	618.2	32.6	5.3

Table 2. Example results from three different runs of the apportionment procedure on 2003 haddock landings. In each run a different stratification order was used.

Year	Stock area	Market category	Unapportioned landings (kg)	Stratification: area, time, region, market category, ton class, gear		Stratification: area, time, region, ton class, gear, market category		Stratification: area, time, gear, ton class, market category, region	
				Apportioned landings (kg)	% increase	Apportioned landings (kg)	% increase	Apportioned landings (kg)	% increase
2003	GB	1470	3,139,921	3,142,810	0.09	3,142,824	0.09	3,142,831	0.09
2003	GB	1475	2,494,362	2,498,777	0.18	2,498,765	0.18	2,498,756	0.18
2003	GB	1479	116,686	116,713	0.02	116,712	0.02	116,718	0.03
2003	GOM	1470	501,994	504,183	0.44	504,194	0.44	504,181	0.44
2003	GOM	1475	505,153	507,891	0.54	507,880	0.54	507,888	0.54
2003	GOM	1479	11,417	11,614	1.73	11,615	1.73	11,614	1.73
2003	OTHER	1470	1,759	1,760	0.01	1,760	0.01	1,760	0.01
2003	OTHER	1475	2,075	2,075	0.01	2,075	0.01	2,075	0.01
2003	OTHER	1479	59	59	0.00	59	0.01	59	0.00
2003	UNK	1470	5,078						
2003	UNK	1475	7,154						
2003	UNK	1479	224						

Figures

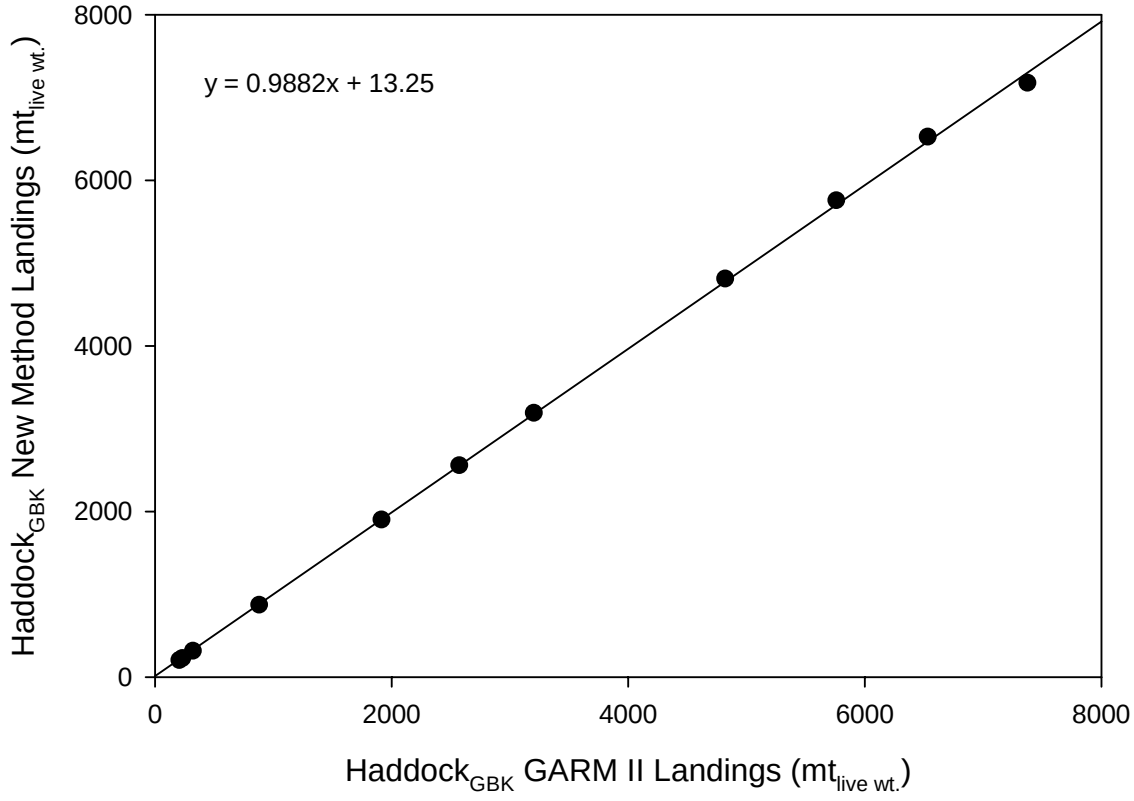

Fig.1. Comparison of Georges Bank haddock landings estimates derived from the new apportionment method relative to GARM II stock landings estimates where the x-ratio was applied, 1994 to 2004. Dashed line indicates 1:1 identity line.

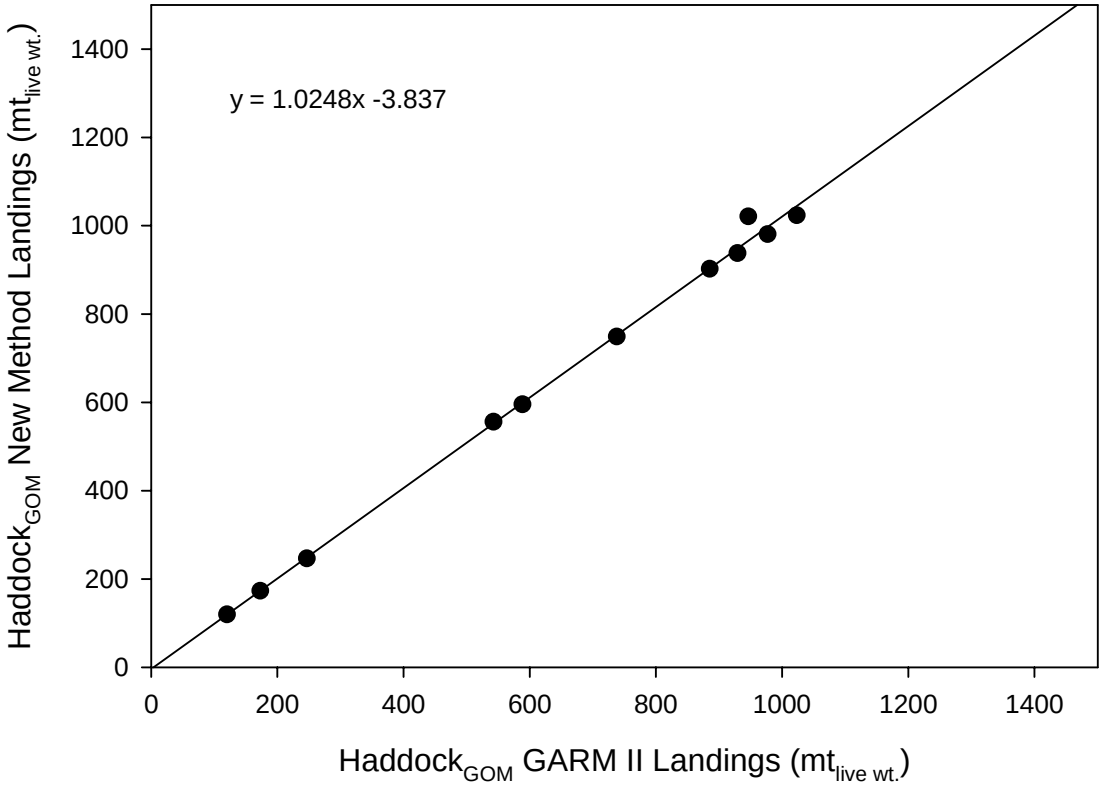

Fig. 2. Comparison of Gulf of Maine haddock landings estimates derived from the new apportionment method relative to GARM II stock landings estimates where the x-ration was applied, 1994 to 2004. Dashed line indicates 1:1 identity line.