fishR Vignette - Back-Calculation of Fish Length

Dr. Derek Ogle, Northland College

December 16, 2013

Francis (1990) defined back-calculation as,

"... a technique that uses a set of measurements made on a fish at one time to infer its length at an earlier time or times. Specifically, the dimensions of one or more marks in some hard part of the fish, together with its current body length, are used to estimate its length at the time of formation of each of the marks. ... The marks are generally annual rings associated with growth checks, ..."

Thus, back-calculation is the re-construction of the length of a fish at previous ages from measurements made on calcified structures.

This vignette requires functions in the FSA package maintained by the author which is loaded into R with

```
> library(FSA)
```

All analyses in this document use the West Bearskin Lake Smallmouth Bass (*Micropterus dolomieu*) data set used in Weisberg (1993). The West Bearskin Lake Smallmouth Bass data are in the SMBassWB data frame of the FSA package and are read into R with

> data(SMBassWB)

The analysis in this vignette will focus on only those fish captured in 1990 as found from the year-at-capture and age-at-capture with

```
> SMBassWByearclass <- SMBassWB$yearcap-SMBassWB$agecap
> wb90 <- Subset(SMBassWB,yearcap==1990)</pre>
```

This vignette begins with background information about back-calculation techniques in Section 1. Section 2 and Section 3 show how to organize the data for the back-calculation methods and how to use the techniques to back-calculate previous lengths for all fish. Section 4 is final thoughts about back-calculation methods.

1 Background

1.1 Growth Measurement Types

Two types of measurements can be made on calcified structures. First, a *radial* measurement is the total distance from the center of the structure (e.g., focus of scale or nucleus of otolith) to the anterior edge of an annulus. Second, an *incremental* measurement is the distance between two successive annuli. Each type of measurement is shown in Figure 1. Radial measurements are needed for back-calculating fish length. Thus, incremental measurements must be converted to radial measurements for back-calculation of fish length.

1.2 Terminology

Let L and S represent measurements of body length and scale (or other calcified structure) radius, respectively. Specifically, L_C and S_C represent the length and scale radius at the time of capture and L_i and S_i represent the length and scale radius at age i (Figure 1).



Figure 1. Depiction of (A) anterior, (AL) antero-lateral, and (L) lateral axes for making radial (S_i) or incremental (G_i) measurements on an age-4 ruffe (left) and an age-3 herring (right). The ruffe was captured in May, 1989 before new growth had begun; thus, the last increment shown is considered to be an annulus. The herring was caught in September, 1983 after new growth had begun; thus, the last increment shown is not a full year's growth and is labeled as "+" growth.

A back-calculation model is used to estimate L_i from known values of S_i , L_c , and S_c . The particular back-calculation model usually defines a family of lines for a sample of fish with one line that depends on L_c , and S_c used for a particular fish. The family of lines are discussed with each model further below.

The relationship between fish length and scale radius is a key component of back-calculation models. Depending on assumptions of the model, a function of the mean scale radius for fish of a given length (i.e., E(S|L)) or a function of the mean length for fish of a given scale radius (i.e., E(L|S)) will be used. These functions are not required to be linear, but often are, and in their linear form are represented as

$$E(S|L) = a + bL \tag{1}$$

$$E(L|S) = c + dS \tag{2}$$

where E(S|L) is estimated from the regression of S on L and E(L|S) is estimated from the regression of L on S.

1.3 Common Back-Calculation Models

1.3.1 Dahl-Lea Model

The first back-calculation model was jointly developed by Knut Dahl and Einar Lea and appeared in Lea (1910). This model is used rarely now but it provides a good starting point for considering the derivation and use of other back-calculation models.

The underlying concept of the Dahl-Lea model is that scale growth is in exact proportion to the growth in length of the fish. With this, the ratio of scale radius at age i to scale radious at capture is the same as the ratio of fish length at age i to fish length at capture. For example, if the scale was 50% of its captured size at age-1 then the length of the fish at age-1 should be 50% of its captured length. This concepts is shown with

$$\frac{S_i}{S_C} = \frac{L_i}{L_C}$$

Rearrangement of these ratios yields the Dahl-Lea back-calculation model

$$L_i = \frac{S_i}{S_C} L_C \tag{3}$$

The Dahl-Lea model describes a family of straight lines that pass through the *origin* and each observed (S_C, L_C) point. Visually (Figure 2), the estimated length at age *i* for a particular fish is found by locating S_i along the x-axis, moving vertically until the straight line for that fish is met, and then moving horizontally to the y-axis. The value located on the y-axis is the back-calculated length for the fish at that previous age.



Figure 2. Plot of length-at-capture versus scale radius for West Bearskin Lake Smallmouth Bass in 1990. All four methods of backcalculation are shown for fish 704 ($S_2 = 3.49804$, $L_C = 218$, and $S_C = 7.44389$; black point and line) with calculational steps shown in red and blue. Fish 701 is shown as the gray line for comparative purposes.

1.3.2 Fraser-Lee Model

Fraser (1916) was the first to describe but Lee (1920) was the first to formally put forth the back-calculation model derived from the concept that "the growth *increment* of the scale is, on the average \dots , a constant

proportion of the growth *increment* of the fish" (Francis 1990). In practice, Fraser and Lee modified the Dahl-Lea model by allowing that a fish is already of some length when the scale first forms (i.e., $L \neq 0$ when S = 0). Thus, if fish length is adjusted for the length of the fish when the scale forms (i.e., L = c when S = 0) then the Dahl-Lea ratios become

$$\frac{S_i}{S_C} = \frac{L_i - c}{L_C - c}$$

Rearrangement of these ratios yields the Fraser-Lee back-calculation model

$$L_i = \frac{S_i}{S_C} (L_C - c) + c \tag{4}$$

The constant c is generally taken from one of three sources. First, as was Lee's intention, c is the length of the fish at the time of scale formation. Second, which is what Lee actually did, c is the intercept of the length-scale relationship regression (e.g., from (2)). Third, c may come from published "standards" for a species (e.g., Carlander (1982)).

The Fraser-Lee model describes a family of lines with an intercept of c that pass through the (S_C, L_C) point (Francis (1990); Figure 2).

1.3.3 Scale Proportional Hypothesis

The scale proportional hypothesis (SPH) was named by Francis (1990) but was first recognized by Whitney and Carlander (1956) when they said

"If the scale was 10 per cent larger when the fish was caught than the average scale for that size of fish, [then] the scale would be 10 per cent larger than normal throughout the life."

If "average" and "normal" are considered to be expected values then this hypothesis can be written as

$$\frac{S_i}{E[S|L_i]} = \frac{S_C}{E[S|L_C]}$$

Assuming that the scale-length relationship is linear than the two expected values in these ratios are computed by plugging L_i and L_c , respectively, into the scale-length relationship (i.e., (1)) to produce

$$\frac{S_i}{a+bL_i} = \frac{S_C}{a+bL_C}$$

Solving for L_i yields the general SPH back-calculation model

$$L_i = \frac{S_i}{S_C} \left(L_C + \frac{a}{b} \right) - \frac{a}{b} \tag{5}$$

The linear SPH model produces a family of lines that all have an intercept of $-\frac{a}{b}$ and pass through each observed (S_C, L_C) point (Figure 2).

The SPH model is the same as the Fraser-Lee model except that the intercept from (2) is replaced with $-\frac{a}{b}$. Further note that the SPH model is the same as the Dahl-Lea model if a = 0.

1.3.4 Body Proportional Hypothesis

The body proportional hypothesis (BPH) was also named by Francis (1990) and was also first recognized by Whitney and Carlander (1956) when they said,

"If a fish at time of capture were 10 per cent smaller than the average fish with that size of scale, [then] the fish would be 10 per cent smaller than the expected length for the size of that scale throughout life."

This hypothesis can be written as

$$\frac{L_i}{E[L|S_i]} = \frac{L_C}{E[L|S_C]}$$

If the length-scale relationship is linear then the expected values can be found by plugging S_i and S_c (2) to get

$$\frac{L_i}{c+dS_i} = \frac{L_C}{c+dS_C}$$

Solving for L_i yields the general BPH back-calculation model

$$L_i = L_C \frac{c + dS_i}{c + dS_c} \tag{6}$$

The linear BPH model produces a family of lines that have an intercept of $\frac{cL_C}{c+dS_C}$ and pass through each observed (S_C, L_C) point (Figure 2). In contrast to the other back-calculation models, the BPH model uses lines with a different intercept for each fish.

The linear BPH model is the same as the Dahl-Lea model if c = 0.

2 Data Organization & Manipulation

All back-calculation models, with the exception of the Dahl-Lea method, requires the scale measurement data to be in two formats. The first format, called one-fish-per-line format¹, is used to compute the scale-length (1) and length-scale (2) relationships. The second format, called one-measurement-per-line format², is used to back-calculate previous fish length from scale measurements.

2.1 Convert from One-Fish-per-Line to One-Measurement-per-Line

Most files containing growth data are organized in a one-fish-per-line format. In this format, all information about a single fish, including all of the growth measurements, are stored on a single line in the data file. The wb90 data frame, a portion of which is shown below, is recorded in one-fish-per-line format.

> head(wb90)

¹The one-fish-per-line format is often referred to as "wide" format

²The one-measurement-per-line format is often referred to as "long" format

	specie	s l	ake	gear	yea	ircap	fish	agecap) 10	encap	anu1	anu2	anu3 a	anu4	anut	5 anu6	anu7
8	SM	В	WB	E		1990	482	1		75	1.511	NA	NA	NA	NA	A NA	NA
9	SM	В	WB	E		1990	768	1		75	1.580	NA	NA	NA	NA	A NA	NA
10	SM	В	WB	E		1990	428	1		71	1.487	NA	NA	NA	NA	A NA	NA
65	SM	В	WB	E		1990	478	2	2	116	1.606	2.736	NA	NA	NA	A NA	NA
66	SM	В	WB	E		1990	379	2	2	117	1.596	3.055	NA	NA	NA	A NA	NA
67	SM	В	WB	E		1990	477	2	2	111	1.321	2.449	NA	NA	NA	A NA	NA
	anu8 ai	nu9	anu	10 a	nu11	anu1	2 rad	lcap									
8	NA	NA		NA	NA	N N	A 1.	511									
9	NA	NA		NA	NA	N N	A 1.	580									
10	NA	NA		NA	NA	N N	A 1.	487									
65	NA	NA		NA	NA	N N	A 2.	736									
66	NA	NA		NA	NA	N N	АЗ.	055									
67	NA	NA		NA	NA	N N	A 2.	449									
> tail(wb90)																	
	speci	es	lake	gea	r ve	arcap	fish	ageca	D :	lencar	anu	1 anu2	anu	3 a.	nu4	anu5	anu6
425	SI	MB	WB	0	E	1990	443	}	7	283	3 1.560	2.930	4.08	2 5.1	243 6	6.500	7.834
426	SI	MB	WB		E	1990	699)	7	289	9 1.70	5 3.088	4.278	3 5.4	445 6	5.968 8	3.023
427	SI	MB	WB		E	1990	429)	7	279	1.65	5 3.345	4.404	1 4.9	951 5	5.732 7	7.002
428	SI	MB	WB		E	1990	697	7	7	294	1.63	2.973	4.15	7 5.0	015 6	6.354 8	3.454
442	SI	MB	WB		E	1990	388	3	9	300	1.08	5 2.035	3.22	74.0	634 5	5.534 6	5.532
443	SI	MB	WB		E	1990	389)	9	329	9 1.059	9 2.188	3.55	1 4.4	408 5	5.786	7.582
	anu7	a	nu8	an	u9 a	nu10	anu11	anu12	r:	adcap							
425	8.577		NA		NA	NA	NA	NA NA		8.577							
426	9.530		NA		NA	NA	NA	NA NA		9.530							
427	8.447		NA		NA	NA	NA	NA NA	. 8	8.447							
428	9.371		NA		NA	NA	NA	NA NA		9.371							
442	7.278	8.	081	9.3	81	NA	NA	NA NA		9.381							
443	8.321	9.	464	10.4	35	NA	NA	NA NA	1	0.435							

The partial data file shown above illustrates two important points about organizing data for performing back-calculations in R. First, each data file must contain enough columns to hold growth measurement for the first through the maximum observed age in the data set. In this example, that means there must be a column for the first through the twelfth radial measurements (i.e., the columns labeled with the prefix *anu*). For fish younger than the maximum age the columns corresponding to ages older than the observed age must be filled with NA (for "not available"). Second, the column names containing the growth measurements should have a common prefix (e.g., *anu*) followed by the number of the measurement. For example, the second measurement should be something like *anu2*. The common prefix will allow efficiency in the analyses and the numbers allow efficient labeling of the age corresponding to the growth measurement.

Radial scale measurements were recorded in this data file, as is required for the back-calculation techniques. If the file had contained incremental measurements then it would need to be converted to radial measurements with gConvert(). The gConvert() function requires the data frame with the incremental measurements as the first argument, the prefix (in quotes) for the columns containing the incremental measurements in the in.pre= argument, and the type of measure to *convert to* in the type= argument (the options are "inc" or "rad"). The result from gConvert() should be saved to an object that has a different name than the original data frame (so that the original data frame is preserved). For example, if it were needed, one would make this conversion with

> wb90A <- gConvert(wb90,in.pre="anu",type="rad")</pre>

The back-calculation techniques cannot be applied to data in the the one-fish-per-line format; rather the data must be "reshaped" to one-measurement-per-line format. Each row in the one-measurement-per-line format contains all information known about a particular growth measurement (rather than an individual fish). Thus, information about an individual fish is spread across several rows. Data that is in one-fish-per-line format can be converted to one-measurement-per-line format with gReshape(). The gReshape() function

requires the data frame with the *radial* measurements as the first argument and the prefix for the columns containing the radial measurements in the in.pre= argument³.

The gReshape() function also needs to know whether "plus growth" was recorded in the data file or not. "Plus growth" is growth on the margin of the scale that is not complete and does not represent a full year of growth (see the herring scale in Figure 1). If "plus growth" is present, then the last recorded "radius" will need to be ignored by gReshape(). If "plus growth" is recorded then the last.plus= argument should be set equal to the variable name that contains the age-at-capture ("agecap" in the West Bearskin Lake example). If "plus growth" is not recorded then the last.plus= argument should be set equal to NULL, which is the default. "Plus growth" was not recorded in the West Bearskin Lake data (as noted by the age-at-capture being equal to the number of recorded measurements).

The results of gReshape() should be saved to an object. The West Bearskin Lake data was reshaped with

```
> wb90r <- gReshape(wb90,in.pre="anu")</pre>
> head(wb90r)
  species lake gear yearcap fish agecap lencap radcap prvAge
                                                                  anu
                  Ε
                        1990
                              482
                                       1
                                              75 1.511
                                                              1 1.511
1
      SMB
            WB
2
      SMB
            WB
                   Е
                        1990
                              768
                                        1
                                              75
                                                  1.580
                                                              1 1.580
                   Е
                                                              1 1.487
3
      SMB
            WB
                        1990
                             428
                                        1
                                              71
                                                 1.487
4
      SMB
            WB
                   Е
                        1990
                              478
                                        2
                                             116 2.736
                                                              1 1.606
5
            WB
                   Е
                        1990
                              379
                                        2
      SMB
                                             117 3.055
                                                              1 1.596
6
      SMB
            WB
                   Е
                        1990
                                        2
                                             111
                                                  2.449
                              477
                                                              1 1.321
```

2.2 Convert from One-Measurement-per-Line to One-Fish-per-Line

As noted in the previous section, most data is originally stored in one-fish-per-line format. However, if the data is stored in one-measurement-per-line format then it will also be necessary to convert to one-fish-per-line format. This conversion is rare enough that a specific function for performing it has not been developed. Fortunately, the conversion can be obtained by careful use of reshape(). The reshape() function requires the one-measurement-per-line data frame as the first argument, the name of the variable that contains the "data" (i.e., radial measurements) in the v.names= argument, the age variable in the timevar= argument, all of the variables that are constant for each fish in the idvar= argument, and the direction="wide" argument to indicate conversion to what R calls "wide", but we are calling "one-fish-per-line", format. For example, the one-measurement-per-line wb90r data frame is converted to a one-fish-per-line data frame with

```
> str(wb90r)
                  # to see the variable names
'data.frame': 767 obs. of 10 variables:
$ species: Factor w/ 1 level "SMB": 1 1 1 1 1 1 1 1 1 ...
        : Factor w/ 1 level "WB": 1 1 1 1 1 1 1 1 1 ...
$ lake
$ gear
         : Factor w/ 1 level "E": 1 1 1 1 1 1 1 1 1 ...
: int 482 768 428 478 379 477 472 465 460 451 ...
$ fish
$ agecap : int
               1 1 1 2 2 2 2 2 2 2 ...
$ lencap : int
               75 75 71 116 117 111 112 114 115 113 ...
$ radcap : num
               1.51 1.58 1.49 2.74 3.06 ...
$ prvAge : int 1 1 1 1 1 1 1 1 1 ...
$ anu
         : num 1.51 1.58 1.49 1.61 1.6 ...
> wb90z <- reshape(wb90r,v.names="anu",timevar="prvAge",</pre>
                idvar=c("species","lake","gear","yearcap","fish","agecap","lencap","radcap"),
                direction="wide")
> head(wb90z)
```

³This prefix will be "rad" if the radial measurements data frame was created with gConvert()).

	species	lake	gear	yearcap	fish	agecap	lencap	radcap	anu.1	anu.2	anu.3	anu.4	anu.5
1	SMB	WB	E	1990	482	1	75	1.511	1.511	NA	NA	NA	NA
2	SMB	WB	E	1990	768	1	75	1.580	1.580	NA	NA	NA	NA
3	SMB	WB	Е	1990	428	1	71	1.487	1.487	NA	NA	NA	NA
4	SMB	WB	E	1990	478	2	116	2.736	1.606	2.736	NA	NA	NA
5	SMB	WB	E	1990	379	2	117	3.055	1.596	3.055	NA	NA	NA
6	SMB	WB	E	1990	477	2	111	2.449	1.321	2.449	NA	NA	NA
	anu.6 ar	nu.7 a	anu.8	anu.9									
1	NA	NA	NA	NA									
2	NA	NA	NA	NA									
3	NA	NA	NA	NA									
4	NA	NA	NA	NA									
5	NA	NA	NA	NA									
6	NA	NA	NA	NA									

The process illustrated in this subsection is NOT needed for the example in the vignette. It is used here simply to illustrate how to convert one-measurement-per-line data to one-fish-per-line data if needed.⁴

3 Performing the Back-Calculations

3.1 Acquiring Coefficients from Regression Models

The scale-length (1) and length-scale (2) relationships must be fit using the length-at-capture and scale radius-at-capture data in the *one-fish-per-line format* data frame (e.g., wb90). These relationships are fit with lm() and the coefficients are extracted from the saved lm() object with coef(). These models are fit and the coefficients are saved to ojects with

```
> lm.sl <- lm(radcap~lencap,data=wb90)
> ( a <- coef(lm.sl)[1] )
(Intercept)
        -1.304
> ( b <- coef(lm.sl)[2] )
lencap
0.03537
>
> lm.ls <- lm(lencap~radcap,data=wb90)
> ( c <- coef(lm.ls)[1] )
(Intercept)
        41.65
> ( d <- coef(lm.ls)[2] )
radcap
27.36</pre>
```

3.2 Applying Back-Calculation Models

The back-calculation models are applied to the one-measurement-per-line format data frame (i.e., wb90r) by writing the model as an equation in R. In this dataframe note that S_i is the anu variable, S_c is the radcap

 $^{^4}$ Note that anu10, anu11, and anu12 in the original wb90 data frame contained all NAs.

variable, and L_c is the *lencap* variable. Thus, estimated lengths at previous ages are computed with the Dahl-Lea model and appended to the wb90r data frame as the *dl.len* variable with

```
> wb90r$dl.len <- with(wb90r,(anu/radcap)*lencap)</pre>
> head(wb90r)
 species lake gear yearcap fish agecap lencap radcap prvAge
                                                          anu dl.len
                                 1 75 1.511 1 1.511
               Ε
                    1990 482
1
     SMB
          WB
                                                              75.00
                                        75 1.580
2
     SMB
          WB
                Ε
                     1990 768
                                  1
                                                      1 1.580
                                                              75.00
3
                Е
     SMB
          WB
                    1990 428
                                  1
                                        71 1.487
                                                      1 1.487
                                                              71.00
                                  2
              E
4
     SMB
          WB
                    1990 478
                                       116 2.736
                                                      1 1.606
                                                              68.07
5
     SMB
          WB E
                    1990 379
                                  2
                                       117 3.055
                                                      1 1.596
                                                              61.13
                                   2
6
     SMB
          WB
                E
                    1990 477
                                       111 2.449
                                                      1 1.321
                                                              59.85
```

Use of with() above simply saves the user from having to type wb90r\$ for each variable name in the equation. Estimated lengths at previous ages computed with the Fraser-Lee, SPH, and BPH models are appended to the wb90r data frame in unique variables with

```
> wb90r$fl.len <- with(wb90r,(anu/radcap)*(lencap-c)+c)</pre>
> wb90r$sph.len <- with(wb90r,(-a/b)+(lencap+a/b)*(anu/radcap))</pre>
> wb90r$bph.len <- with(wb90r,lencap*(c+d*anu)/(c+d*radcap))</pre>
> head(wb90r)
  species lake gear yearcap fish agecap lencap radcap prvAge
                                                                                               anu dl.len fl.len
        SMB
                WB E
                                  1990 482 1 75 1.511 1 1.511 75.00 75.00
1
                                                      1 75 1.580
1 71 1.487
                                                                                        1 1.580 75.00 75.00
        SMB
                WB E
                                 1990 768
2

        SMB
        WB
        E
        1990
        428
        1
        71
        1.487

        SMB
        WB
        E
        1990
        478
        2
        116
        2.736

        SMB
        WB
        E
        1990
        379
        2
        117
        3.055

        SMB
        WB
        E
        1990
        477
        2
        111
        2.449

3
                                                                                       1 1.487 71.00 71.00
4
                                                                                       1 1.606 68.07 85.28
                                                                                       1 1.596 61.13 81.02
5
6
                                                                                         1 1.321
                                                                                                      59.85 79.05
   sph.len bph.len
     75.00
1
                 75.00
2
     75.00
                 75.00
3
     71.00
                71.00
4
     83.31
                 85.21
5
     78.74
                 79.72
6
     76.84
                 79.46
```

3.3 Some Summary Calculations

The mean and standard deviation of lengths-at-age estimated with the SPH model, for example, with

```
> Summarize(sph.len~prvAge,data=wb90r,digits=2)
Warning: To continue, variable(s) on RHS of formula were converted to a factor.
             mean
                      sd
                          min
                                  Q1 median
                                               Q3 max percZero
 prvAge
         n
      1 181 75.27 6.53 59.8 70.9
1
                                      75.1 79.7 92
                                                             0
2
      2 178 112.17 10.49 89.9 105.0 112.0 120.0 148
                                                             0
      3 155 145.66 14.26 118.0 136.0
                                     146.0 155.0 194
                                                             0
3
4
         71 170.88 15.66 140.0 160.0
                                     171.0 179.0 218
                                                             0
      4
5
      5
         64 199.75 17.51 160.0 188.0 199.0 209.0 246
                                                             0
6
      6 64 234.83 23.38 191.0 220.0 234.0 245.0 298
                                                             0
                                                             0
7
      7
         50 268.56 25.31 214.0 254.0 267.0 283.0 346
8
      8
         2 282.67 27.07 264.0 273.0 283.0 292.0 302
                                                             0
9
      9
          2 314.50 20.51 300.0 307.0 314.0 322.0 329
                                                             0
```

A plot (Figure 3) of back-calculated length-at-age can be constructed with



Figure 3. Back-calculated length-at-age for West Bearskin Lake Smallmouth Bass.

The mean back-calculated length separated by the age-at-capture of the fish can be found with,

```
> sumTable(sph.len~agecap*prvAge,data=wb90r,digits=1)
Warning:
          RHS variable was converted to a factor.
Warning:
          RHS column variable was converted to a factor.
           2
                  3
                                                  8
     1
                        4
                              5
                                     6
                                           7
                                                        9
1 73.7
          NA
                NA
                       NA
                             NA
                                    NA
                                          NA
                                                 NA
                                                       NA
2 77.5 113.3
                NA
                       NA
                             NA
                                    NA
                                          NA
                                                 NA
                                                       NA
3 74.1 111.2 148.9
                       NA
                             NA
                                    NA
                                          NA
                                                 NA
                                                       NA
4 67.5 119.5 159.9 194.0
                              \mathbf{N}\mathbf{A}
                                    NA
                                          NA
                                                 NA
                                                       NA
6 76.0 104.8 133.6 168.1 198.0 229.7
                                          NA
                                                 NA
                                                       NA
7 77.7 115.1 142.1 168.6 200.4 236.3 269.1
                                                 NA
                                                       NA
9 66.9 96.0 131.8 163.6 195.5 234.6 255.4 282.7 314.5
```

Finally, a plot (Figure 4) of length-at-age separated by the age-at-capture of the fish can be constructed with

```
> library(lattice) #required for xyplot
> xyplot(sph.len~jitter(prvAge)|factor(agecap),data=wb90r,pch=".",
ylab="Back-Calculated Length (mm)",xlab="Age (jittered)")
```



Figure 4. Back-calculated length-at-age for West Bearskin Lake Smallmouth Bass separated by age-at-capture.

4 Further Thoughts

4.1 Choosing The Back-Calculation Method

Francis (1990) argued that either proportional model (i.e., SPH or BPH) should be used to back-calculate fish length. He further argued that if lengths are back-calculated using both proportional methods then the difference in results is a measure of the minimum uncertainty in back-calculated lengths. Pierce *et al.* (1996) back-calculated the lengths of two species using the Fraser-Lee and both proportional methods and compared these to observed lengths-at-age. Their results showed little or no difference among the three methods and they hypothesized that this result was due to the tight body-scale relationship observed for the two species studied.

4.2 Non-Linear Length-Scale Relationship

The discussion above began with general SPH and BPH models and then showed specific models if the length-scale or scale-length relationships were linear, which will not always be the case. The SPH and BPH methods are general to any form of relationship; albeit, with a bit more work. The scale-length or length-scale relationship should be examined carefully and the functional form that best represents the relationship should be chosen to develop an appropriate back-calculation model from the general SPH and BPH hypotheses.

For example, suppose that the scale-length relationship followed a power function of the form $E(S|L) = aL^b$. In this case, the ratios of the SPH hypothesis would be

$$\frac{S_i}{aL_i^b} = \frac{S_C}{aL_C^b}$$

This can still be solved for L_i to get a non-linear SPH model that was attributed to Monasytrsky by Bagenal and Tesch (1978))

$$L_i = L_c \left[\frac{S_i}{S_C}\right]^{\frac{1}{b}}$$

4.3 Lee's Phenomenon

"Rosa Lee's" or just "Lee's" phenomenon is the tendency for back-calculated lengths at a given age in the same cohort of fish to be smaller as the fish they are computed from get older. It must be remembered that Lee's phenomenon is detected by examining lengths at the *same* age and *same* cohort back-calculated from progressively older fish (i.e., captured in later years). Reverse Lee's phenomenon is the tendency for the same back-calculated lengths to be larger as progressively older fish are examined.

For example Van Oosten (1928) carefully examined the scales of lake herring from Saginaw Bay Lake Huron. In particular, he examined the back-calculated lengths, using the Dahl-Lea back-calculation model, of lake herring from the 1918 and 1919 year-classes captured at various ages. His results (Table 1), although not tested for statistical significance, illustrate "Lee's" phenomenon for back-calculated lengths at ages one to three for both year-classes (Table 1).

Ricker (1969) described four possible explanations of Lee's phenomenon. The first explanation was that an incorrect mathematical model was used in the back-calculation of previous lengths. This problem can be identified by examining the annular measurements on the scales at a given age on the scales of progressively older individuals from the same cohort (Newman and Weisberg 1987). If Lee's phenomenon is not evident in the scales but it is evident in the back-calculated lengths then a problem exists with the model used to make the back-calculations.

Table 1. Back-calculated lengths-at-age for Saginaw Bay lake herring from two different year-classes captured at several different ages.

	Back-Calculated Length-at-Age									
	Ι	II	III	IV	V	VI				
Age		19	918 Ye	ar-Cla	ss					
IV	116	179	214	232						
V	114	171	205	229	241					
VI	113	166	203	228	247	263				
Age		19	919 Ye	ar-Cla	ss					
III	127	192	224							
IV	122	183	217	236						
V	119	179	213	237	251					
VI	116	168	199	237	248	267				

The remaining explanations involve the appearance of faster-growing but younger-aged individuals of a cohort in "earlier" samples of the cohort and slower-growing, older-aged individuals in "later" samples. The first of these explanations is that a biased form of sampling is used. An example, would be the use of gill-nets to select samples where the faster-growing individuals will be captured by nets at younger ages and the slowergrowing individuals at older ages. The second and third of these explanations are length-selective natural and fishing mortality. If faster-growing individuals are selectively removed from the population earlier in the history of the cohort then the older individuals captured in later years will be relatively slow-growing individuals that will have relatively smaller back-calculated lengths at previous ages.

Duncan (1980) offered another explanation. Annular increments on calcified structures may contract as calcium is resorbed from hard parts for metabolic purposes. This contraction would make growth and, subsequently size, appear smaller than what was actual.

4.4 Use With Otoliths

The back-calculation methods described above can be applied to annular measurements on otoliths. However, slow-growing fish tend to have larger otoliths than fast-growing fish of the same size (CITATIONS). This leads to an underestimation of previous lengths and the appearance of Lee's phenomenon. To avoid this problem, Campana and Jones (1992) proposed a back-calculation model that is similar to the Fraser-Lee method but uses a biologically-determined intercept that is determined by the mean size of the fish and the otolith at the larval or juvenile stage. Specifically, biological intercept is measured in the smallest fish where all larger fish yield a directly proportional linear length-otolith relationship. Very small fish for which the proportional relationship between length and otolith size has not yet developed are not used in the development of the back-calculation model. This leads to a method that does not depend on any regression method and is insensitive to growth-related variations in the length-otolith relationship. The specific back-calculation model is,

$$L_i = L_c + \frac{(O^* - O_c)(L_c - L^*)}{O_c - O^i}$$

where O_c is the otolith radius at capture, O_i is the otolith radius to the ith annulus, and L^* and O^* are the fish length and otolith radius at the biological intercept.

Otoliths have also been shown to grow continuously during periods when fish size is not growing (CITATION). This phenomenon tends to make the length-otolith relationship non-linear, but linear within a given age. Thus, Morita and Matsuishi (2001) developed a back-calculation model that incorporated an *age-effect* into the relationship regression.

4.5 Validation

As with any model, the results from the use of the model should be compared to actual observations to validate that the model is producing the results for which it was intended. In back-calculation, this means comparing back-calculated lengths at previous ages to observed lengths when the fish was that age. The strictest validation is the comparison of the back-calculated lengths of an individual fish to its own previous lengths. However, this is a difficult process and, thus, the usual form of validation is to show that the average back-calculated lengths-at-age are the same as the the observed average lengths at those ages. However, as Francis (1990) correctly points out, this latter method will only detect gross errors with the back-calculation technique because a sample of one-year-old fish is not the same as a sample of one-year-old fish that will survive to some later age (see comments above about length-selective mortality). Despite these difficulties every attempt should be made by a researcher to validate the back-calculation model for the population of fish being studied.

References

- Bagenal, T. B. and F. W. Tesch. 1978. Methods for Assessment of Fish Production in Fresh Waters, chapter Age and growth, pp. 101–136. 3rd edition, Blackwell Scientific Publications, Oxford, U.K. 12
- Campana, S. E. and C. M. Jones. 1992. Analysis of otolith microstructure data. In D. K. Stevenson and S. E. Campana (editors), Otolith microstructure examination and analysis, *Canadian Special Publication* of Fisheries and Aquatic Sciences, volume 117, pp. 73–100. 13
- Carlander, K. D. 1982. Standard intercepts for calculating lengths from scale measurements for some centrarchid and percid fishes. Transactions of the American Fisheries Society 111:332–336. 4
- Duncan, K. W. 1980. On the back-calculation of fish lengths; modifications and extensions to the Fraser-Lee equation. Journal of Fish Biology 16:725–730. 13
- Francis, R. 1990. Back-calculation of fish length: a critical review. Journal of Fish Biology 36:883–902. 1, 4, 5, 12, 14
- Fraser, C. M. 1916. Growth of the spring salmon. Transactions of the Pacific Fisheries Society 1915:29–39. 3
- Lea, E. 1910. On the methods used in the herring-investigations. Publ. Circonst. Cons. perm. int. Explor. Mer 108(1):14–22. 2
- Lee, R. M. 1920. A review of the methods of age and growth determination in fishes by means of scales. Fisheries Investigations, London Series 2 4((2)):1–32. 3
- Morita, K. and T. Matsuishi. 2001. A new model of growth back-calculation incorporating age effect based on otoliths. Canadian Journal of Fisheries and Aquatic Sciences 58:1805–1811. 13
- Newman, R. M. and S. Weisberg. 1987. Age and Growth of Fish, chapter Among- and within-fish variation of scale growth increments in brown trout., pp. 159–166. Iowa State University Press. 12
- Pierce, C. L., J. B. Rasmussen, and W. Leggett. 1996. Back-calculation of fish length from scales: Empirical comparison of proportional methods. Transactions of the American Fisheries Society 125:889–198. 12
- Ricker, W. 1969. Effects of size-selective mortality and sampling bias on estimates of growth, mortality, production, and yield. Journal of the Fisheries Research Board of Canada 26:479–541. 12
- Van Oosten, J. 1928. Life history of the lake herring (*Leucichthys artedi* Le Sueur) of Lake Huron as revealed by its scales, with a critique of the scale method. Bulletin of the Bureau of Fisheries 44:267–428. 12
- Weisberg, S. 1993. Using hard-part increment data to estimate age and environmental effects. Canadian Journal of Fisheries and Aquatic Sciences 50(6):1229–1237. 1
- Whitney, R. R. and K. D. Carlander. 1956. Interpretation of body-scale regression for computing body length of fish. Journal of Wildlife Management 20:21–27. 4, 5

Reproducibility Information

Version Information

- Compiled Date: Mon Dec 16 2013
- Compiled Time: 9:30:01 PM
- Code Execution Time: 3.29 s

R Information

- **R Version:** R version 3.0.2 (2013-09-25)
- System: Windows, i386-w64-mingw32/i386 (32-bit)
- Base Packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other Packages: FSA_0.4.3, knitr_1.5.15, lattice_0.20-24
- Loaded-Only Packages: bitops_1.0-6, car_2.0-19, caTools_1.16, cluster_1.14.4, evaluate_0.5.1, for-matR_0.10, Formula_1.1-1, gdata_2.13.2, gplots_2.12.1, grid_3.0.2, gtools_3.1.1, highr_0.3, Hmisc_3.13-0, KernSmooth_2.23-10, MASS_7.3-29, multcomp_1.3-1, mvtnorm_0.9-9996, nlme_3.1-113, nnet_7.3-7, plotrix_3.5-2, quantreg_5.05, sandwich_2.3-0, sciplot_1.1-0, SparseM_1.03, splines_3.0.2, stringr_0.6.2, survival_2.37-4, tools_3.0.2, zoo_1.7-10
- **Required Packages:** FSA, lattice and their dependencies (car, gdata, gplots, graphics, grDevices, grid, Hmisc, knitr, multcomp, nlme, plotrix, quantreg, sciplot, stats, utils)