Compare to Daisuke's Results

Derek H. Ogle

August 8, 2017

Initial Comparison (to Daisuke's Compiled Results)

Daisuke's results used 444 WBIC_YEARs. Our results, which included years after Daisuke's, had 683 WBIC_YEARs, but this was reduced to 655 WBIC_YEARs when "invalid" estimates were excluded. Our results shared 411 WBIC_YEARs with Daisuke's results. Of the 33 WBIC_YEARs in Daisuke's results that were not in our results, 3 were not in our file of PEs, 0 were not in our original file of raw data from the FMDB, 17 were not in our prepped file of raw data from the FMDB, and 13 were excluded because we considered the results to be invalid. A review of the data prepper log suggests that all of the WBIC_YEARs removed during data prepping occurred when the data were restricted to fish captured with fyke nets in the spring. Table 1 shows the reasons why we excluded the 13 WBIC_YEARs in Daisuke's results that were not in our results.

Table 1: Frequency of WBIC_YEARs in Daisuke's analysis that were excluded from our results by the reason we considered the results to be invalid.

reason	Freq
n<30	6
numAges<5	1
Age gaps issue	6

WBIC_YEARs in Daisuke's results that were not in our results are listed below.

```
WBIC_YEARs in Daisuke's results for which we do not have a PE.

[1] "1543300_1993" "2350500_2012" "2914800_1999"

WBIC_YEARs in Daisuke's results that were excluded in our prepping the FMDB file.

 [1] "378400_1999"  "716800_2006"  "968800_2006"  "995200_2005"
 [5] "1469100_2003" "1579900_2007" "1591100_2006" "1631900_2009"
 [9] "1631900_2012" "2106800_2012" "2271600_2011" "2316100_2003"
[13] "2316100_2011" "2393200_2012" "2435700_2002" "2865000_1997"
[17] "2949200_2005"

WBIC_YEARs in Daisuke's results that we considered invalid.

 [1] "995000_2012"  "1544700_1990" "1597600_1997" "1612200_2007"
 [5] "2098000_2010" "2382300_2010" "2641000_2011" "2654500_1991"
 [9] "2654500_1997" "2661100_2012" "2678100_2004" "2678100_2010"
[13] "2734200_1993"
```

Computation Results

Comparisons of our results to Daisuke's results for B, P, and P/B are in Figure 1, Figure 2, and Figure 3, respectively. Correlations between our and Daisuke's estimates are not as strong as I would have expected given that they originated from the same data (however, see the next section).

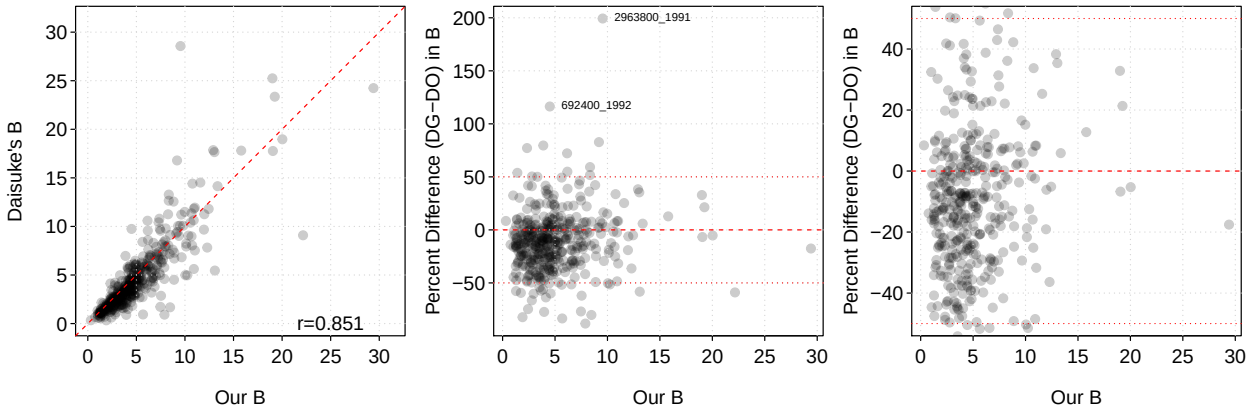

Figure 1: Daisuke's B (LEFT) and the percentage difference in B (MIDDLE and RIGHT) versus our B. Values that are more than 100% different are marked on the middle plot.

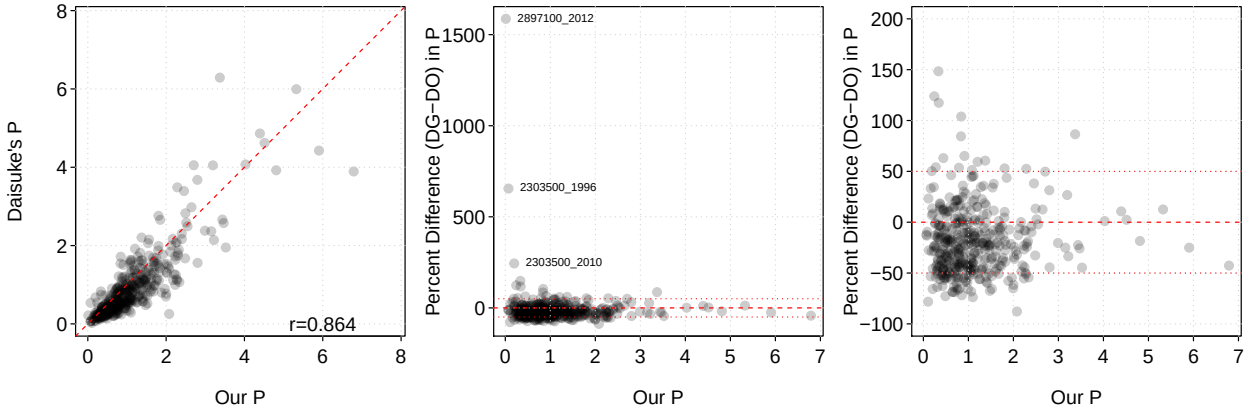

Figure 2: Daisuke's P (LEFT) and the percentage difference in P (MIDDLE and RIGHT) versus our P. Values that are more than 200% different are marked on the middle plot.

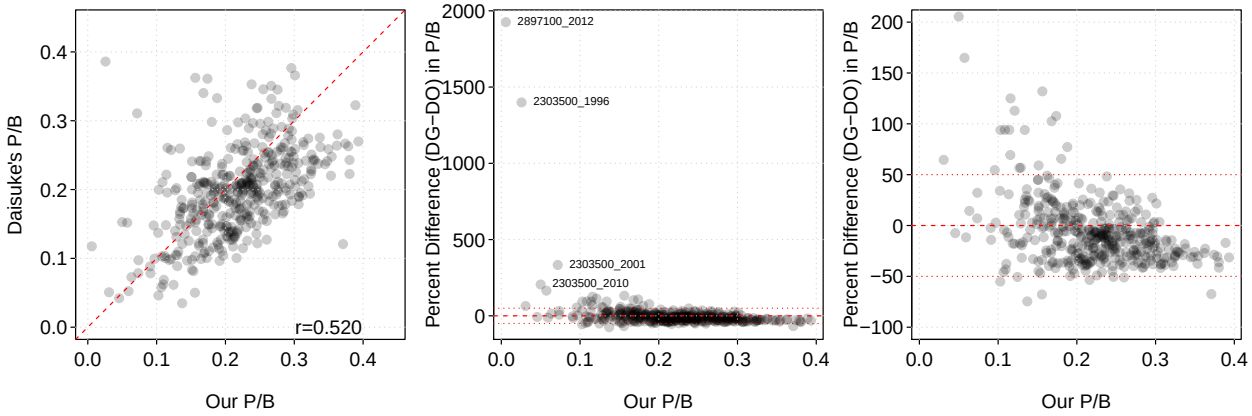

Figure 3: Daisuke's P (LEFT) and the percentage difference in P/B (MIDDLE and RIGHT) versus our P/B. Values that are more than 200% different are marked on the middle plot.

WBIC_YEARs for which our and Daisuke's B estimates were very different:

	wbic_year	B.DO	B.DG	pdiffB
24	692400_1992	4.503441	9.744144	116.3710
408	2963800_1991	9.543585	28.567052	199.3325

WBIC_YEARs for which our and Daisuke's P estimates were very different:

	wbic_year	P.DO	P.DG	pdiffP
224	2303500_1996	0.07128913	0.5382757	655.0600
226	2303500_2010	0.19712871	0.6779036	243.8888
380	2897100_2012	0.01283656	0.2165350	1586.8608

WBIC_YEARs for which our and Daisuke's P/B estimates were very different:

	wbic_year	PB.DO	PB.DG	pdiffPB
224	2303500_1996	0.025744563	0.3859967	1399.3329
225	2303500_2001	0.071683675	0.3107017	333.4344
226	2303500_2010	0.049975732	0.1526625	205.4732
380	2897100_2012	0.005800889	0.1175159	1925.8256

"Raw" Data Comparison

PEs

The PEs were the same for 411 WBIC_YEARs in common between our and Daisuke's analysis. **Thus, it does not appear that differences in PEs would explain the differences between our's and Daisuke's estimates of P and B.**

Lake Size

Lake size was exactly the same for 88.9% and within 1% of each other for 98.1% of the 216 WBICs in common between our and Daisuke's results (Figure 4). Our lake sizes for the four large differences in Figure 4 match what is available in the online lake app. None of the four lage differences in lake sizes matched with any of the problematic P, B, or P/B estimates identified above. **Thus, it does not appear that these differences in lake size data will explain (to any great extent) the differences between our and Daisuke's estimates of P and B.**

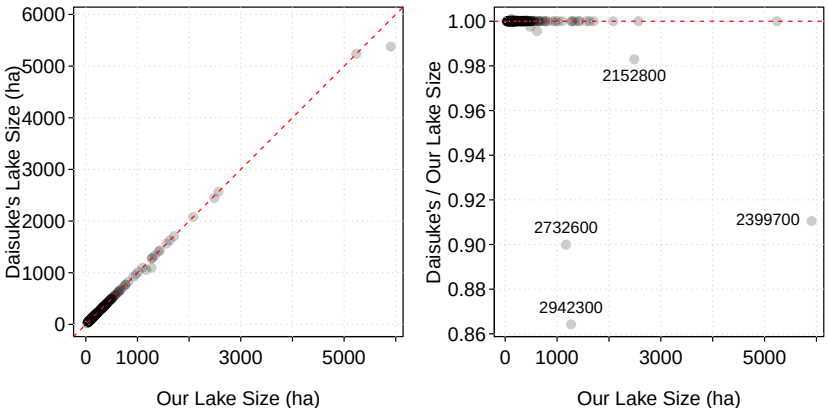

Figure 4: Daisuke's lake size (LEFT) and the ratio between Daisuke's and our lake sizes (RIGHT) versus lake size in our analysis. WBICs for which the difference was more than 1% are marked.

Weight-Length Relationships

We do not have access to Daisuke's weight-length relationship results. However, we computed the weight-length relationship for 350 WBIC_YEARs available (and matched what we had available) in what appears to be his weight-length data (were restricted to only WBIC_YEAR regressions that we considered valid). Of these, 99.7% were within 1% of each other for $log(a)$, 99.7% were within 1% of each other for b, 99.1% were within 1% of each other for n, and 99.7% were within 1% of each other for r^2 (Figure 5). Note that one of the most different regression results was from 2013, which was not used in Daisuke's analysis. None of the most different regression results matched with any of the problematic P, B, or P/B estimates identified above. **Thus, it does not appear that the actual weight-length relationships explain the difference between our and Daisuke's P and B estimates** (this assumes that Daisuke's weight-length relationships are the same as what we calculated here). However, **we could have used the weight-length relationships differently than he did.** For example, we may have used a different decision ladder to decide to use a weight-length regression other than the one for the specific WBIC_YEAR than Daisuke did (and we had regressions by lake class rather than region as Daisuke did). Or, Daisuke may have used the weight-length relationship to compute mean weight from mean length, rather than predicting weight for individual fish and then computing mean weight-at-age as we did.

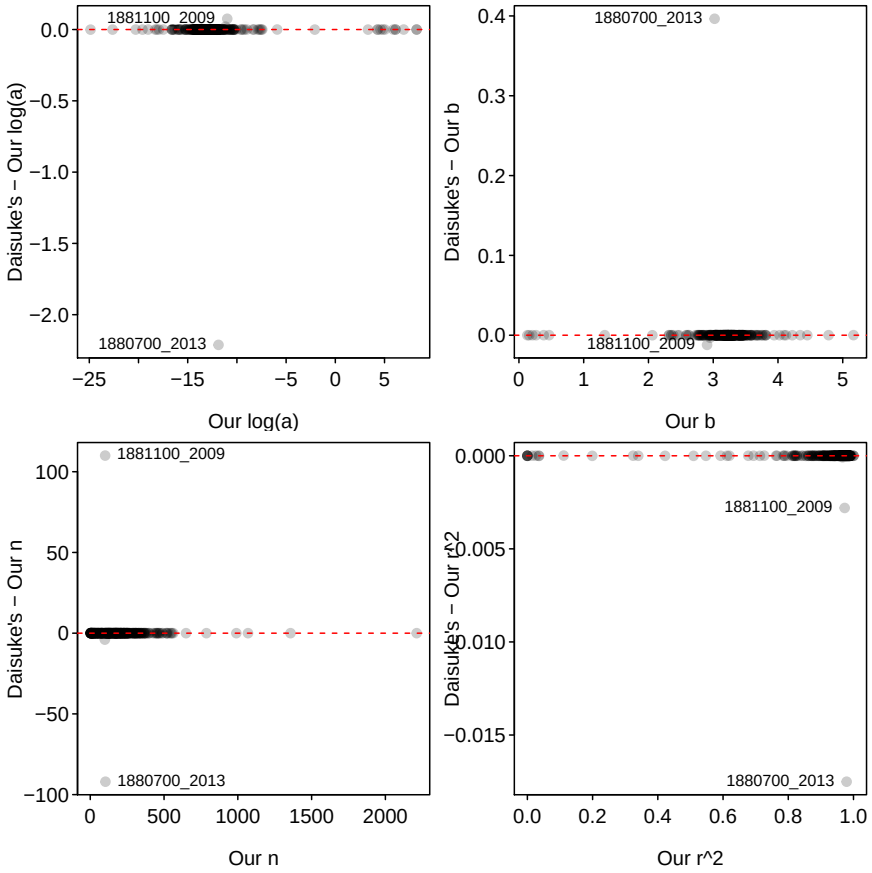

Figure 5: Differenes in Daisuke's and our weight-length regression results versus our weight-length regression results for four summary statistics.

Age-Length Data

We did not have access to the specific age-length keys that Daisuke used. However, we could compute the sample size, mean length, and mean age for fish in the 345 WBIC_YEARs available (and matched what we had available) in what appears to be his length-age data. The most striking issue here is related to a series of WBIC_YEARs where our sample size was exactly or very nearly exactly twice as large as Daisuke's sample size. The WBIC_YEARs corresponding to these sample size issues do not seem to correspond to any great extent to mean lengths or mean ages that are very different between our and Daisuke's results. With this observation and an examination of our original data it appears that some of the records for these WBIC_YEARs are duplicated in our original data file. Andrew thinks that this is related to a past database migration issue and is investigating this further.

There appear to be a handful of other discrepancies in these data. However, I won't pursue these further until we address the possible duplication issue.

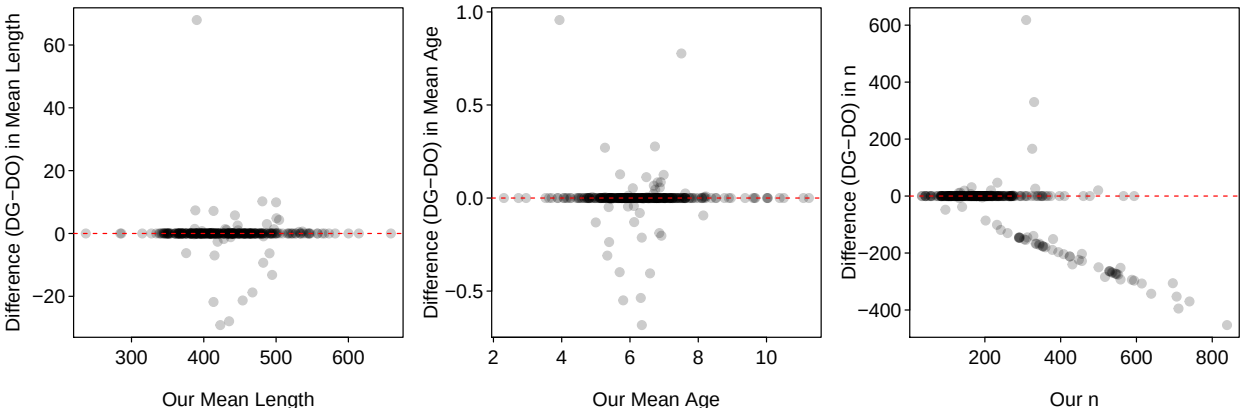

Figure 6: Differenes in summaries of Daisuke's and our length-age data versus our summaries for three summary statistics.

Length Data in FMDB

There are MANY differences in the lengths data in our respective versions of "raw data from the FMDB" (this is AFTER our data had been prepped). Note that seven of the WBIC_YEARs had a difference in sample size of more than 5000 individuals! It is possible that I have misinterpreted which of Daisuke's files were his "lengths" (I used "WIwalleye_len_pe.csv").

It is possible that some of the WBIC_YEARs for which our data set are larger is due to data that have been entered into the FMDB since Daisuke's data run in 2012. That is some of the very large differences shown in the figures below have dates in the "FDV_ENT_DATE" field as post-2012. However, these WBIC_YEARs don't have any entries with dates before 2012 (so, not sure why they are in Daisuke's file).

Assuming that I have chosen the correct file for Daisuke's lengths, this is likely the primary driver for the wide differences in our results.

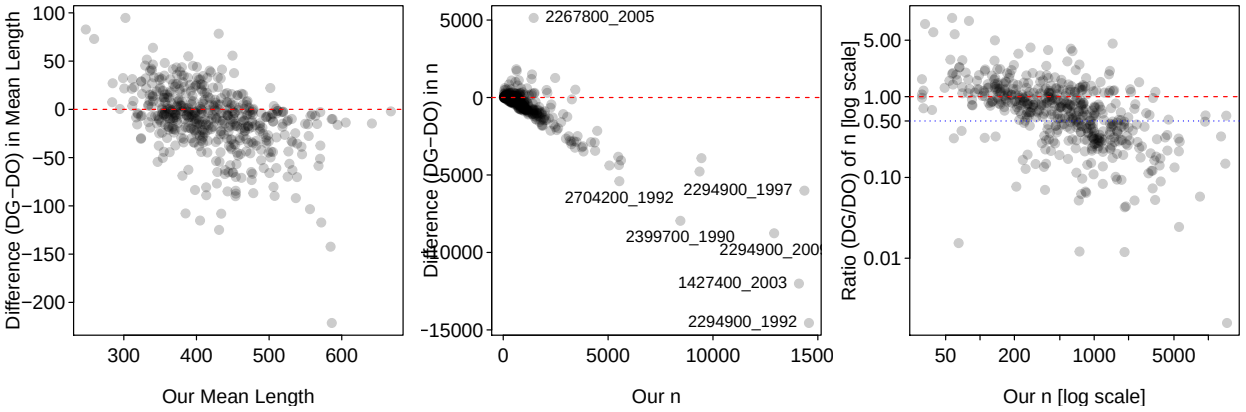

Figure 7: Differences in summaries of Daisuke's and our length data versus our summaries for three summary statistics.

Some examples (haphazardly picked and then filtered in Excel to compare to our results):

- "1001300_1999" has 93 records in our original raw FMDB file with a total of 130 fish (all data entry appeared to be in 2002 ... assumes that the "FDV_ENT_DATE" field is the date of data entry). When this is reduced to fyke nets in the spring there is a total of 70 fish. This matches our results (see below).
- "1018500_2012" has 370 records in our original raw FMDB file with a total of 1708 fish (all data entry appeared to be in April and October of 2012). When this is reduced to fyke nets in teh spring there is a total of 1227 fish. There were 11 fish under 7 in that we would have removed because the smallest 3-year-old was 7 in. These results match our results (see below).
- "1596300_1997" has 728 records in our original raw FMDB file with a total of 1320 fish (685 records for 1245 fish were entered in 2006, 43 records with 75 fish were added in 2014). When this is reduced to fyke nets in the spring there is a total of 801 fish. These results match our results (see below).
- "2295200_1992" has 791 records in our original raw FMDB file with a total of 4083 fish (all records appeared to be entered in 2014). When this is reduced to fyke nets in the spring there is a total of 1191 fish. These results match our results (see below).

```
       wbic_year n.DO mnlen.DO n.DG mnlen.DG mnlen.diff n.diff   n.ratio
1    1001300_1999   70 532.0937   92 469.1270  -62.96676     22 1.3142857
10   1018500_2012 1227 259.1338  359 332.1103   72.97648   -868 0.2925835
100  1596300_1997  801 451.2036  670 415.3469  -35.85670   -131 0.8364544
200  2295200_1992 1191 353.1922  163 365.3393   12.14704  -1028 0.1368598
```

Is it possible that Daisuke's results were not "expanded" by fish counts to include length measurements (or some approxiation) for when multiple fish were binned??? [*Note how his sample sizes are close to the original record numbers.*]

Biggest Questions for Daisuke

1. What "length data" did you use?
    a. Which file is that?
    b. Were multiple fish in binned lengths expanded to individual lengths? How?
    c. What criteria for inclusion/exclusion did you use?
2. How did you apply the weight-length regressions to ultimately get mean weights at age?
    a. If you converted mean lengths at age to mean weights at age, how did you get your mean lengths (related to ALK question below ... by assigning ages to individual fish or using the ALK to convert mean lengths in the aged sample to the larger sample)?
3. How did you apply the age-length keys to get ages?
    a. Did you assign ages to individual fish or use the ALK to expand the proportions in ages in the aged sample to proportions in the ages in the larger sample (the lengths)?
4. What criteria did you use to identify which weight-length relationship to use (i.e., WBIC_YEAR, WBIC, region, overall)?
5. What criteria did you use to identify which ALK to use (i.e., WBIC_YEAR, WBIC, region, overall)?
6. What criteria did you use to reject a final P and B calculation as being invalid (small n of lengths, small number of ages, gaps in ages, etc.)?