Estimates of reproductive activity in red snapper by size, season, and time of day with non-linear models

Nikolai Klibansky

SEDAR41-DW49

Submitted: 20 July 2015
Addendum added: 28 August 2015

****Addendum is added to reflect changes made during the Data Workshop. Final information is found in the addendum.****

Please cite this document as:

Klibansky, N. 2015. Estimates of reproductive activity in red snapper by size, season, and time of day with nonlinear models. SEDAR41-DW49. SEDAR, North Charleston, SC. 22 pp.

Estimates of reproductive activity in red snapper by size, season, and time of day with nonlinear models

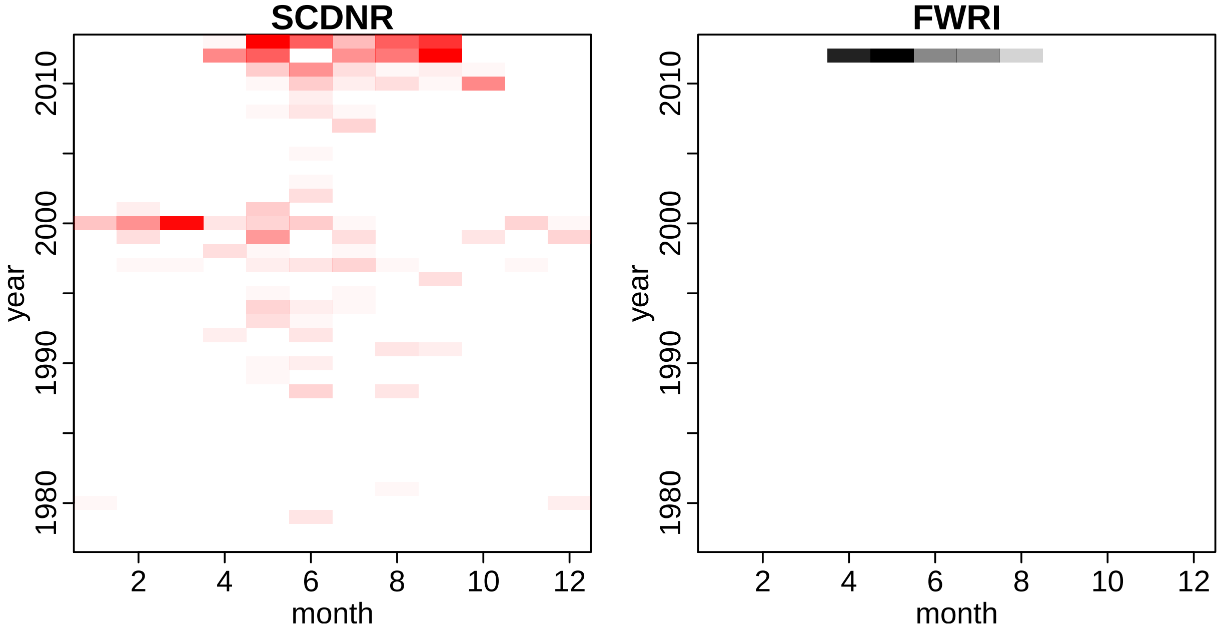
Nikolai Klibansky

Purpose

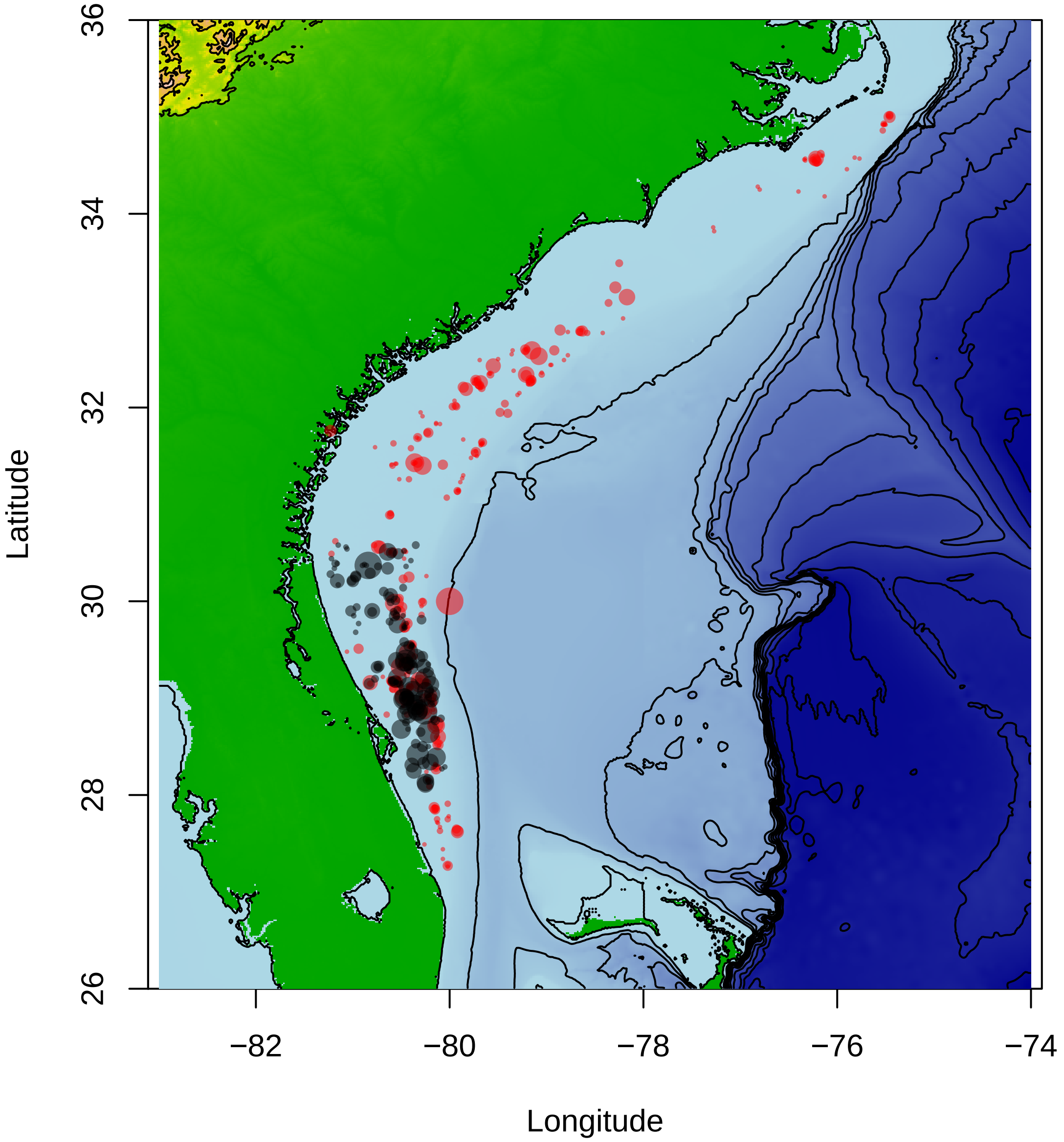
This working paper documents analyses of histology data from South Carolina Department of Natural Resources and the Florida Fish and Wildlife Research Institute (FWRI) for red snapper (*Lutjanus campechanus*) in the US South Atlantic. The purpose of the analyses is to compare maturity-at-age versus maturity-at-size regressions, and to apply a nonlinear model to binary reproductive data to estimate the date of peak spawning, batch number-at-length, and to estimate spawning indicator durations.

Temporal and spatial distribution of data

As noted above, the data set analyzed here derives from two sources. Both are considered large, valuable, sources of reproductive data for red snapper (FWRI 730 records; SCDNR 3014 records), but they differ greatly in the temporal and spatial distribution of the data. While SCDNR does have a lot of records from recent years, their data are otherwise spread out over four decades. By contrast the FWRI data are from a single year (i.e. 2012), when SCDNR also collected many samples.



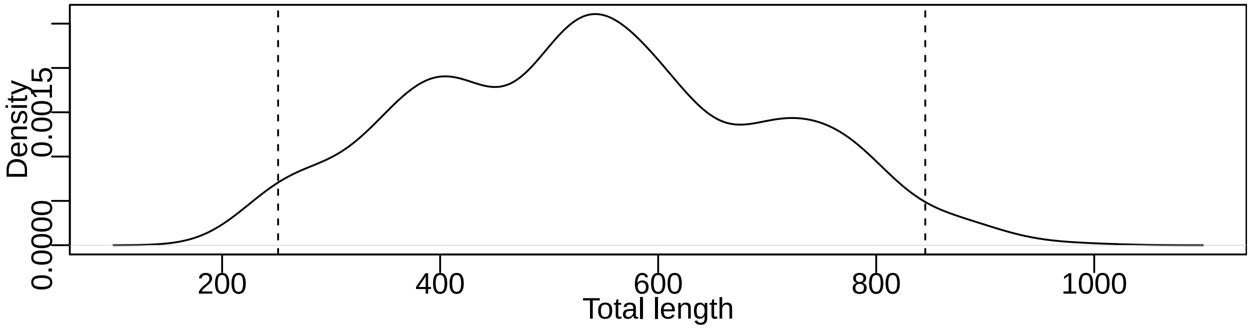
In addition the spatial range of the data is limited to Florida. This is evident in the map below where data collected by SCDNR is plotted as red points, with point size scaled to the maximum number of records from a single location; FWRI data are plotted as black points, with point size scaled similarly but independently.



Given that a general goal of the reproductive data analyses is to identify average patterns and parameter values over time and space, including the FWRI data will tend to add undo weight to fish caught in Florida in 2012. There are also some methodological differences in the way that gonad analyses were conducted, which complicate the interpretation of certain analyses. In an earlier stage of this analysis, I analyzed all the data together. When I removed the Florida data, the results of some of the analyses changed noticeably. Though I reiterate the value of the FWRI data for other analyses, for the above reasons the analyses below were conducted on the SCDNR data exclusively.

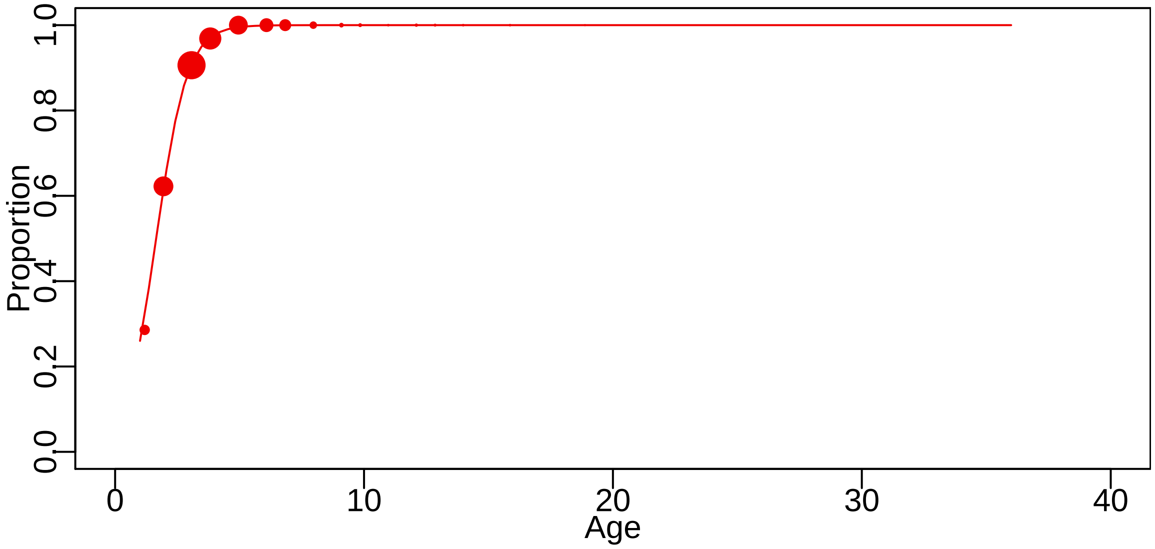
Reproduction at size

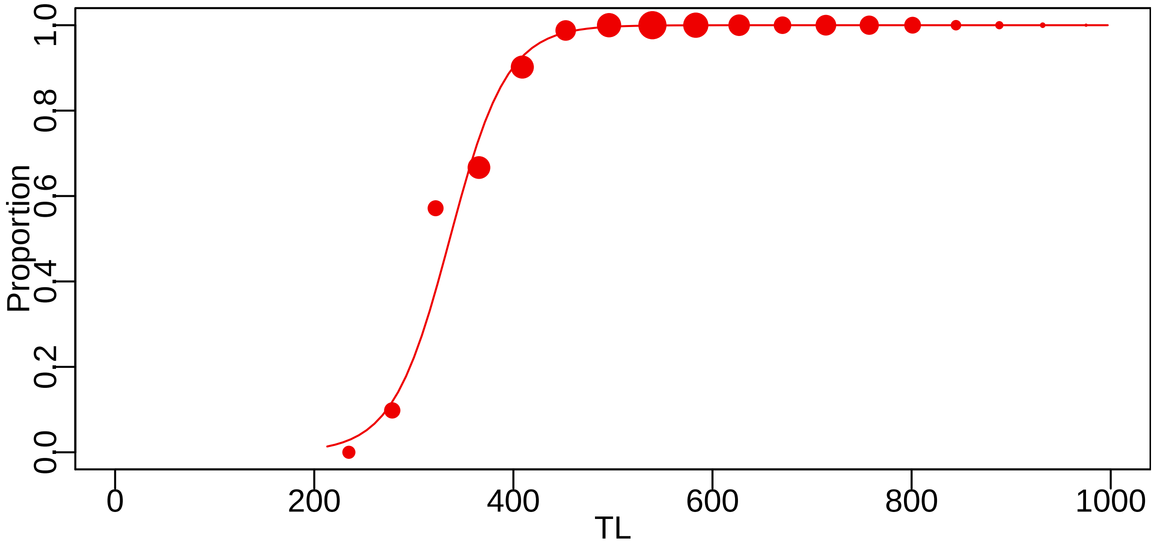
Although the assessment model is age-based and ultimately requires a vector of reproductive value-at-age, it is probably better to conduct statistical analysis of reproductive data relative to body size, then convert size-based to age-based vectors at the end. This is likely true for two reasons: 1. biologically, reproductive investment is actually determined more by the size of a fish than by its age, 2. the ages of red snapper included in the data set are very skewed while the size data are not. This second point is evident in the density plots below, where dashed vertical lines mark the upper and lower limits bounding 95% of the data. For those reasons my analyses were conducted with respect to length instead of age.



Revisiting maturity at size

To illustrate that length is a better predictor of the reproductive importance of a fish, I compare models where either age or length was used as a single predictor of maturity.



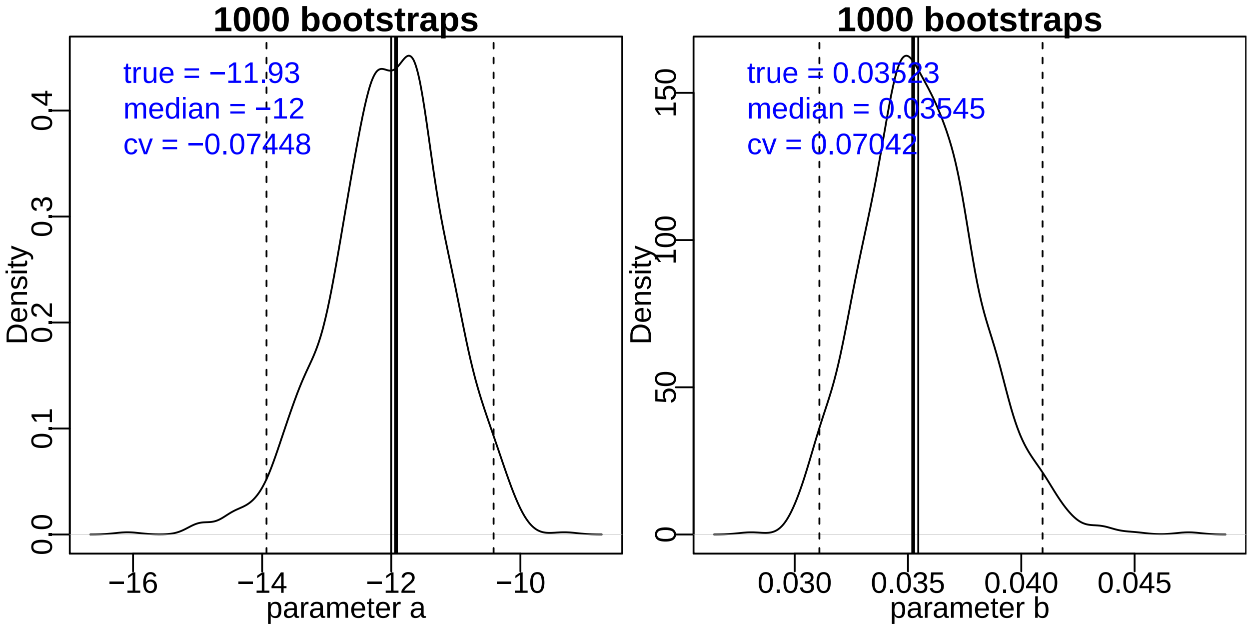


It may not be apparent from the plots that length is such a better predictor, but the Akaike's Information Criterion (AIC) value for age is 592.15 compared with 303.72 for total length. A lower AIC represents a better fitting model. As a rule of thumb, a difference in AIC values, or ΔAIC of >10 is a considered a very strong indication that two models are different (Bolker 2008). Considering a related statistic, Menard's r^2 or pseudo-r^2, describing the proportion of variation in data explained by a model, values for the age and length-based fits are 0.33 and 0.66, respectively (Quinn and Keough 2003).

Considering these differences, and looking back at the plotted functions, note the large proportion of observations evident along the upper asymptote in the TL-based function versus the age-based function. Another way of looking at this is that, considering TL one can predict that above a certain size cutoff, most females are mature; if one tries to identify a similar cutoff based on age, one is able to predict maturity status of far fewer females.

The equations for maturity as a function of TL is: $P(Mature) = \frac{1}{1+e^{-(a+bTL)}}$ where a = -11.7 and b = 0.03481.

Bootstrap results for the two parameters of the fitted length-based logistic function show minimal bias in estimated parameters and narrow confidence intervals.



Seasonal spawning activity

In standard calculations, spawning season duration is calculated as the number of days between the first and last evidence of a spawning indicator. The date of peak spawning may be estimated simply by observing the month of the year wher the proportion of fish found in a spawning condition appears to be highest. In standard calculations of batch number, one multiplies the duration of the spawning period by spawning rate in units of spawns per day, also known as spawning fraction (note that spawning frequency is simply the reciprocal of spawning fraction; Murua et al. 2003).

In this analysis I fit a plateau-shaped nonlinear model to binary spawning indicator data by Julian date, to yield more rigorous estimates of these quantities.

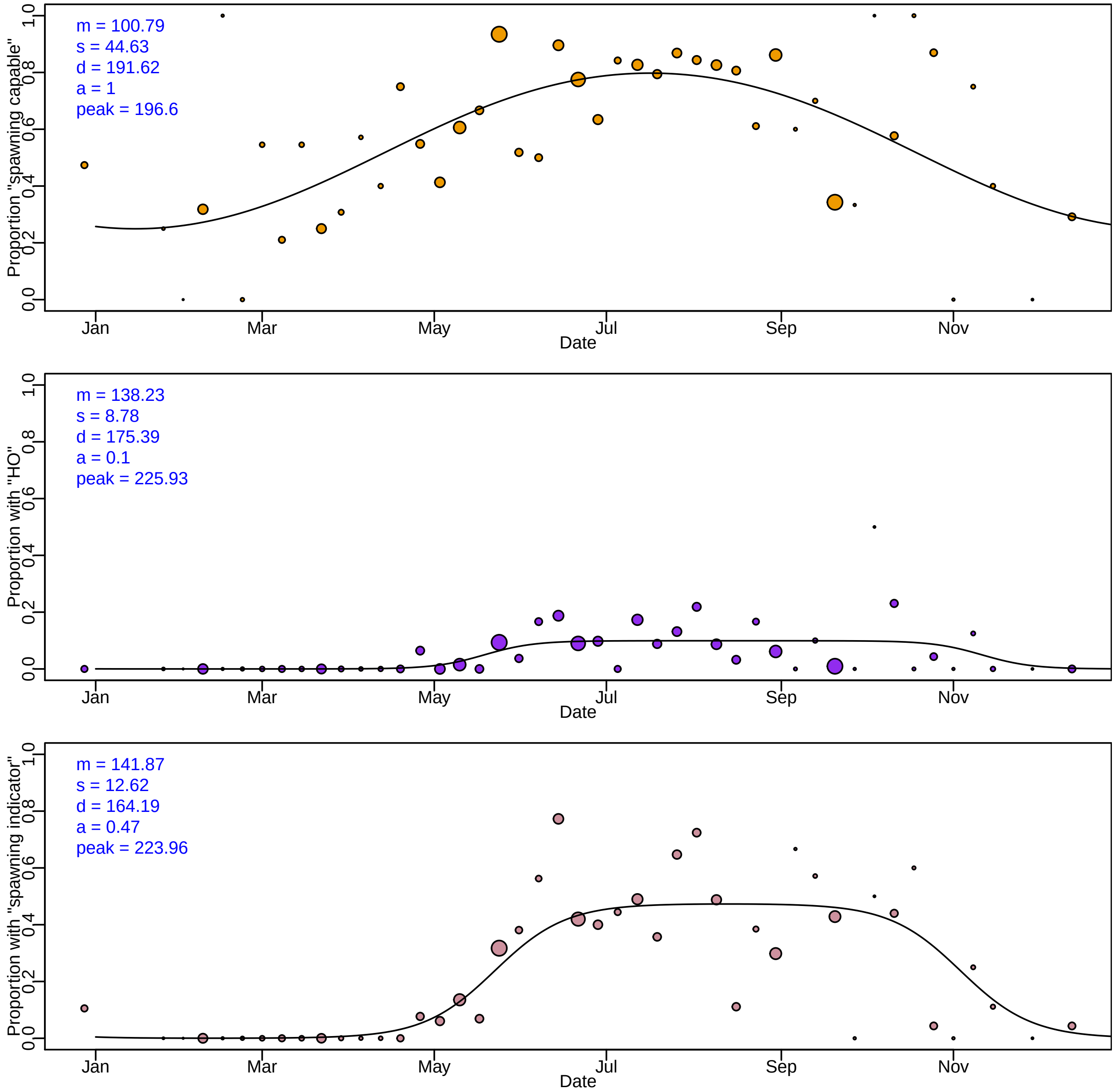
The form of the model fitted to these data is:

$$y = \frac{a}{1+e^{(\frac{-(x-(m-0.5))}{s})}} - \frac{a}{1+e^{(\frac{-(x-((m-0.5)+(d+1)))}{s})}}$$

Fitting this plateau model allows us to estimate the parameters with biological meaning:

- m is the average Julian date when a binary trait appeared
- s is the variation around m, representing asynchrony among females
- d is the average duration a binary trait is present once it appears
- a is the asymptote of each of the two logistic functions in the model, but not always the height of the resulting function
- peak is the Julian date at the peak in the model (e.g. date of peak spawning), calculated from estimates of m and d

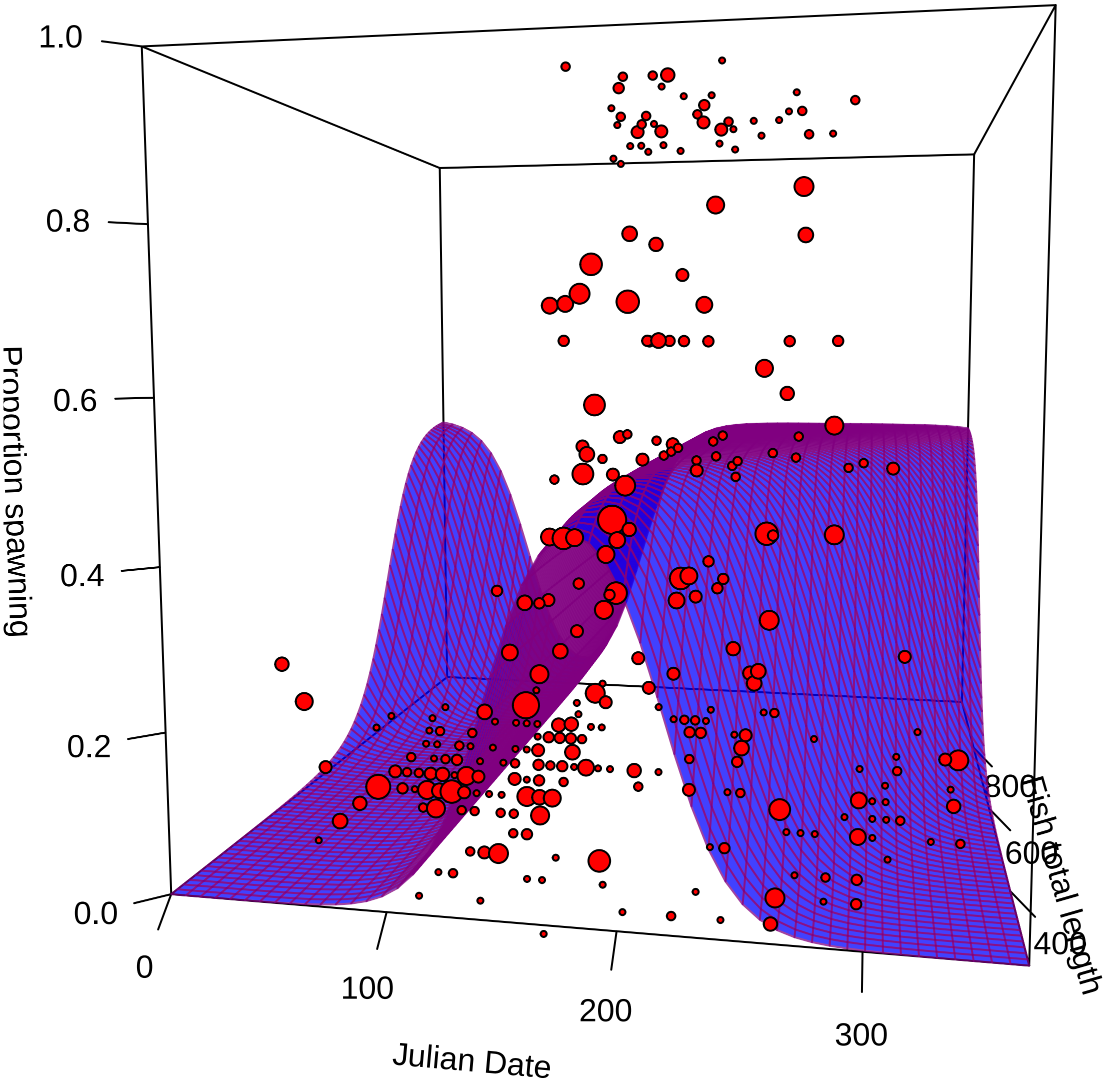
The figure below shows the plateau model fit to three different binary reproductive variables.



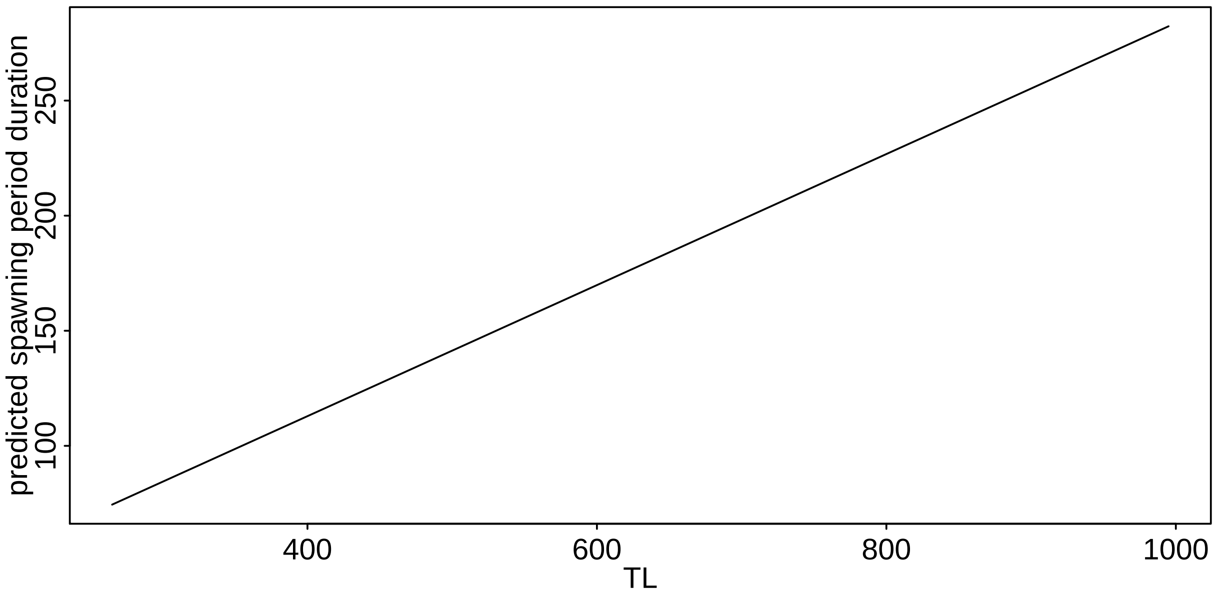
The plateau model was extended to simultaneously estimate d as a function of total length. This was done by replacing d in the plateau model with a linear function:

$$d = d_0 + d_1$$

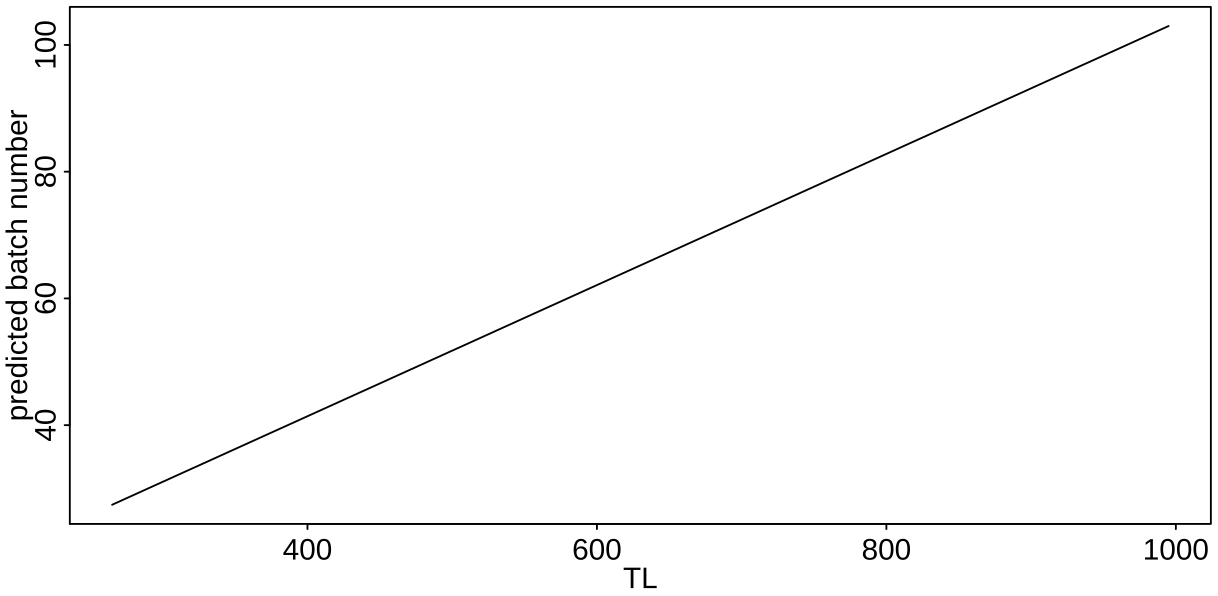
The fit of the model including size-dependent d, can be plotted as a surface:



Using AIC to compare the basic plateau model with the length-based plateau model, I observed a ΔAIC of 28.7. Recall that a ΔAIC >10 is considered a strong improvement in fit. Thus, the length-based model is preferable. Predicted spawning period duration as a function of total length is plotted below.



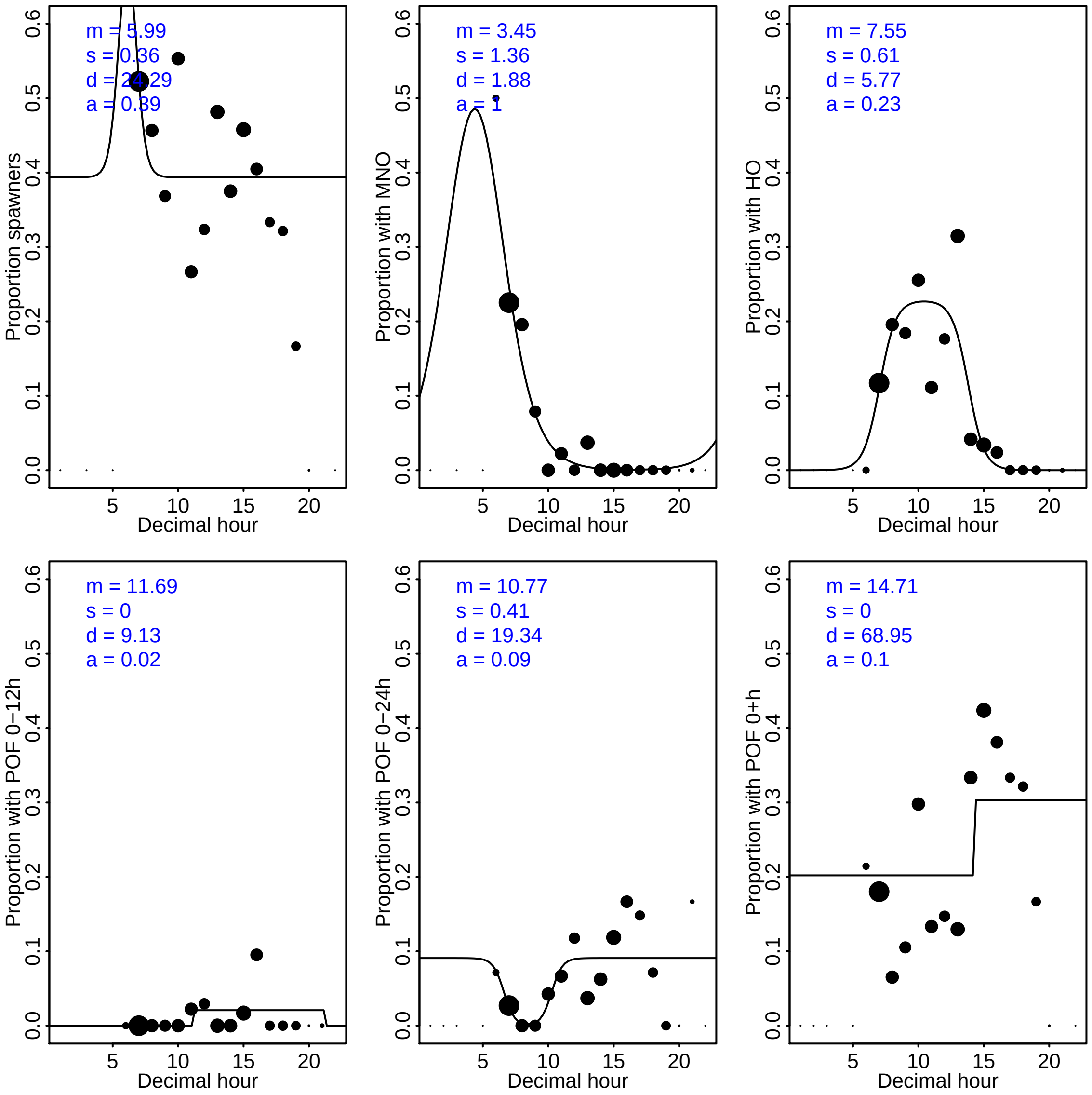
In standard calculations of batch number, one multiplies the duration of the spawning period by spawning rate in units of spawns per day, also known as spawning fraction (note that spawning frequency is simply the reciprocal of spawning fraction; Murua et al. 2003). One can alternatively calculate the spawning fraction during smaller time units within the spawning season (i.e. months or days), calculate batch number within each of these time units and then sum them to calculate the annual batch number (Hunter and Leong, 1981). Extending these methods, we can calculate length-based batch number in red snapper by summing estimated spawning fraction on each day of the year, by total length. Essentially this method takes the integral of spawning fraction from day zero to day 365, for each length. Integrating under the fitted surface by TL yields the following relationship between batch number and TL. Note that, in this plot, estimated batch number was multiplied by a correction factor of 0.71 based on previous estimates of spawning indicator duration of 34 hours.



Diel spawning activity

I also investigate the durations of several potential spawning indicators by fitting the plateau model to spawning indicator presence data, over a 24 hour period. Fit this model to the "proportion of spawners", but also created several binary variables from the histological maturity data categories, based on presence/absence of migratory nucleus oocytes, hydrated oocyte, post-ovulatory follicles (POF) 0-12 hours old, POF 0-24 hours old, and POF 0+ hours old.

In this analysis I fit the plateau model to each binary variable to estimate all parameter values. For variables where previous estimates of structure duration (d), were available, I also fixed d and fit the remaining three parameters. By comparing the fits of these two versions of the model, I was able to assess the estimate of d using this method. Fits of these models are below.



Solid lines in each panel indicate predicted proportions fitting all parameters. Black points represent observed proportions by hour, with the size of each point scaled to the number of observations. The fitted parameter values are printed in each panel in blue.

It has been proposed that the duration of HO in red snapper is 10 hours, and durations of POF were given as 12 or 24 hours. The duration of the composite spawning indicator associated with "Spawners" is estimated at 34 hours (i.e. the sum of HO and 24 hours POF). In this analysis, the estimate of *d* for Spawners was 24 hours, but *d* was not significant when compared with the model with *d* fixed at 34 hours. Fitting all parameters to the HO presence data, estimated average HO duration at 5.77 hours, which was significant compared with the reduced model with *d* fixed at 10 hours. The analogous comparison for POF1 (aka POF 0-12 hours) was not significant, but the fit for POF2 (aka POF0-24 hours) estimated d at 19.34 offered an improved fit over the reduced model. These results suggest that the duration of a spawning indicator combining HO and POF, actually has a duration of 25.12 hours rather than 34 hours. The results of these tests are presented in the table below.

Table 1: Model comparisons for testing parameters describing diel patterns in ovarian structure presence.

y.name	par.R	par.F	delta.AIC	G2	G2.p	Men.r2
Spawner	m,s,a	m,s,d,a	1.072	3.072	0.080	0.004
Spawner	a	m,s,a	4.425	8.425	0.015	0.010
HO	m,s,a	m,s,d,a	8.019	10.019	0.002	0.024
POF1	m,s,a	m,s,d,a	-2.043	-0.043	1	-0.001
POF1	a	m,s,a	4.996	8.996	0.011	0.117
POF2	m,s,a	m,s,d,a	14.996	16.996	0.00004	0.058

Literature cited

Bolker, B. M. (2008). Ecological Models and Data in R. Princeton, NJ: Princeton University Press.

Hunter, J. R. & Leong, R. (1981). The spawning energetics of female northern anchovy, Engraulis mordax. Fishery Bulletin 79, 215-230.

Murua, H., Kraus, G., Saborido-Rey, F., Witthames, P. R., Thorsen, A. & Junquera, S. (2003). Procedures to estimate fecundity of marine fish species in relation to their reproductive strategy. Journal of Northwest Atlantic Fisheries Science 33, 35-54.

Quinn, G. P. & Keough, M. J. (2003). Experimental design and data analysis for biologists. Cambridge, UK: Cambridge University Press.

SEDAR41-DW49: Estimating annual fecundity of Red Snapper by size using with nonlinear models

Nikolai Klibansky

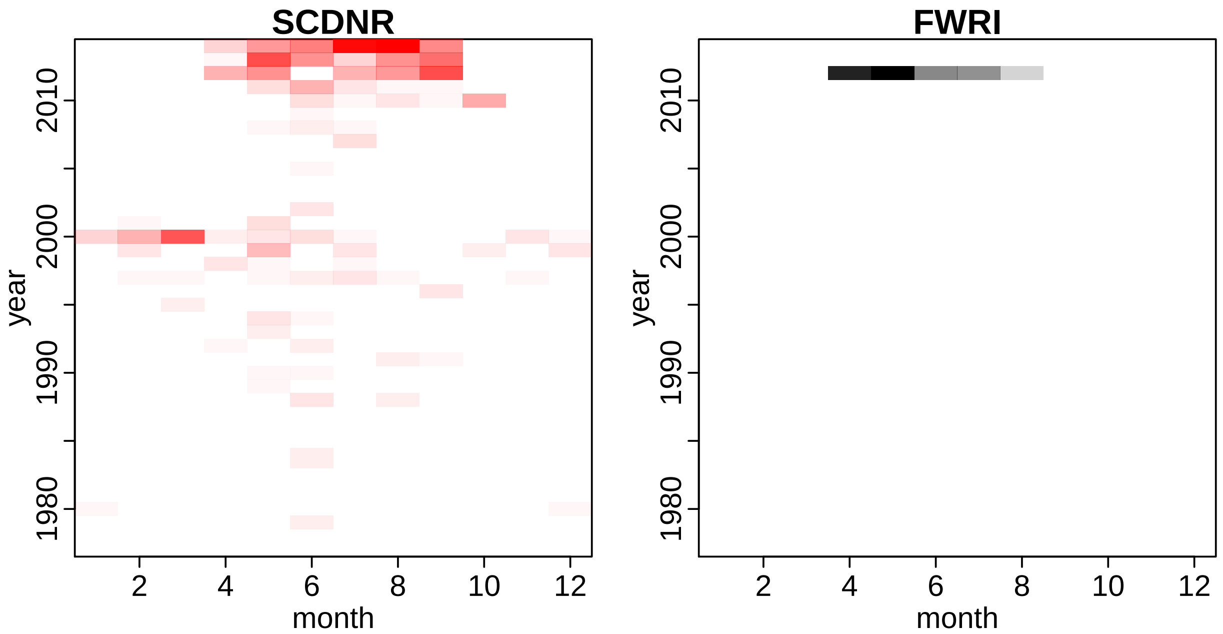
Purpose

Since Red Snapper (*Lutjanus campechanus*) exhibit indeterminate fecundity, the number of eggs produced per female per year (i.e. annual fecundity) is calculated as the product of batch fecundity (described above) and batch number (the number of batches produced per female per year). This working paper documents statistical approaches used to estimate batch fecundity and batch number by size, to estimate age from size, and ultimately yields estimates annual fecundity by age for Red Snapper in the US South Atlantic. Data used in the analyses were produced by the South Carolina Department of Natural Resources (SCDNR) and the Florida Fish and Wildlife Research Institute (FWRI).

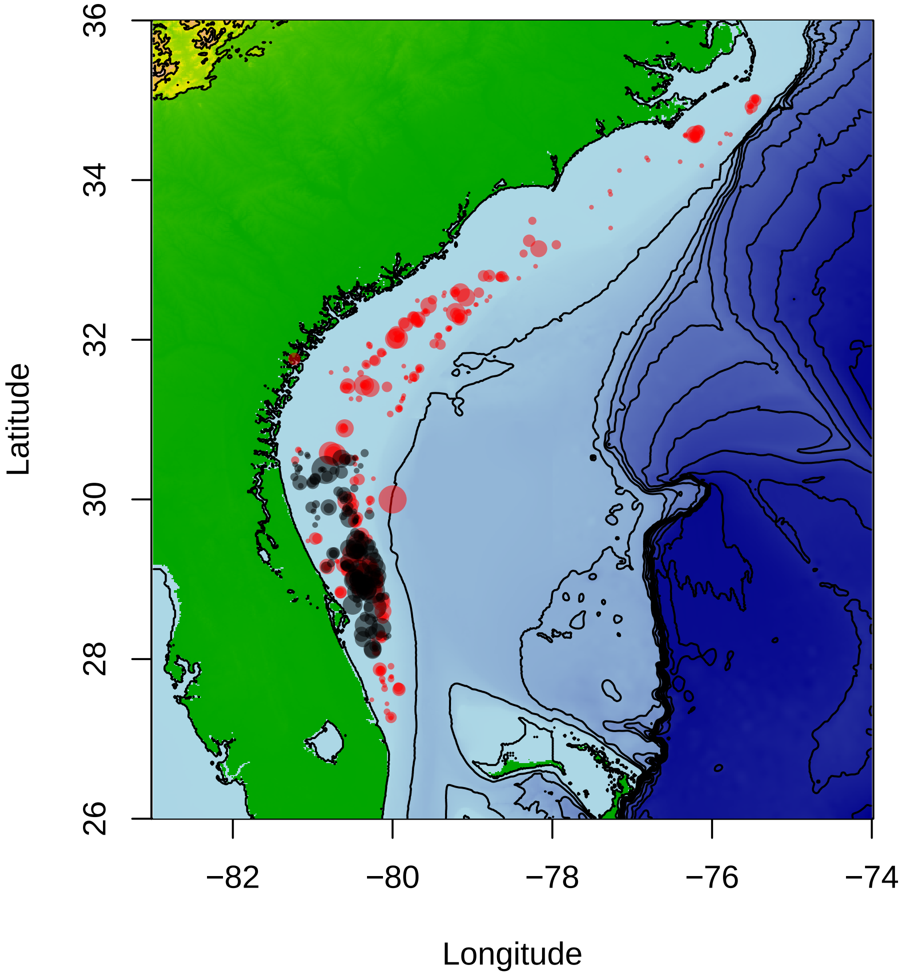
Temporal and spatial distribution of data

Data from the two sources named above were aggregated to produced two data sets. A hydrated oocyte (HO) count data set was used for batch fecundity estimation while a spawning status (i.e. presence/absence of spawning indicators) data set was used to estimate batch number. As is often the case, females in the maturity phase used to produce valid HO counts were relatively scarce and sample sizes for estimating batch fecundity were relatively small for both sources (FWRI = 44, SCDNR = 25). Previous analyses also showed that estimates from both sources were fairly similar. For these reasons, the combined HO count data sets were used in this analysis.

The spawning status data set was much larger for both sources (FWRI 730 records; SCDNR 3844 records), but the sources differed greatly in the temporal and spatial distribution of the data. The FWRI spawning status data were from a single year (i.e. 2012), while the SCDNR data were spread out over four decades. The SCDNR data does contain substantially more samples in recent years, including 2012.



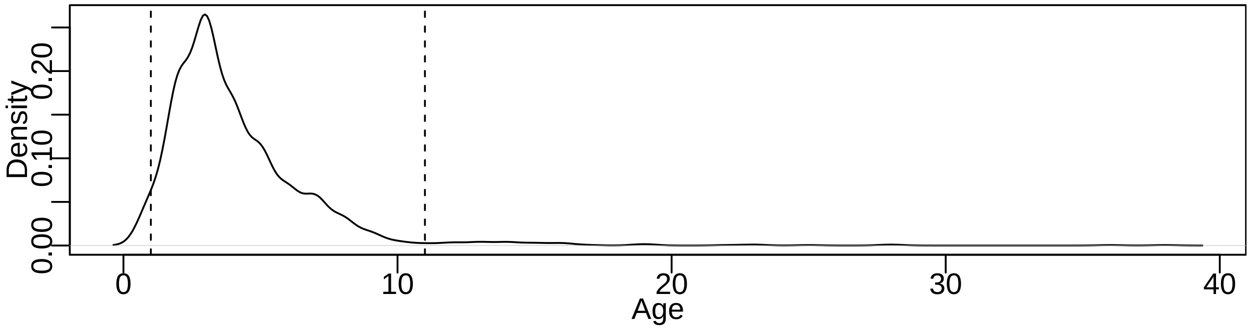
In addition the spatial range of the FWRI data was limited to Florida. This is evident in the map below where data collected by SCDNR is plotted as red points, with point size scaled to the maximum number of records from a single location; FWRI data are plotted as black points, with point size scaled similarly but independently.

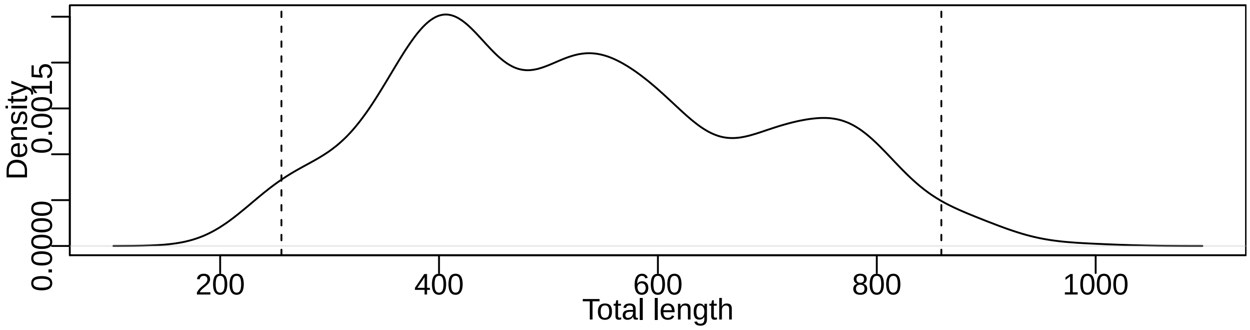


Given that a general goal of the reproductive data analyses is to identify average patterns and parameter values over time and space, I judged that including the FWRI spawning status data in estimating batch number would tend to add undo weight to fish caught in Florida in 2012. Though I reiterate the value of the FWRI data for other analyses, for the above reasons, estimation of batch number by size was based solely on the SCDNR spawning status data.

Use of size versus age

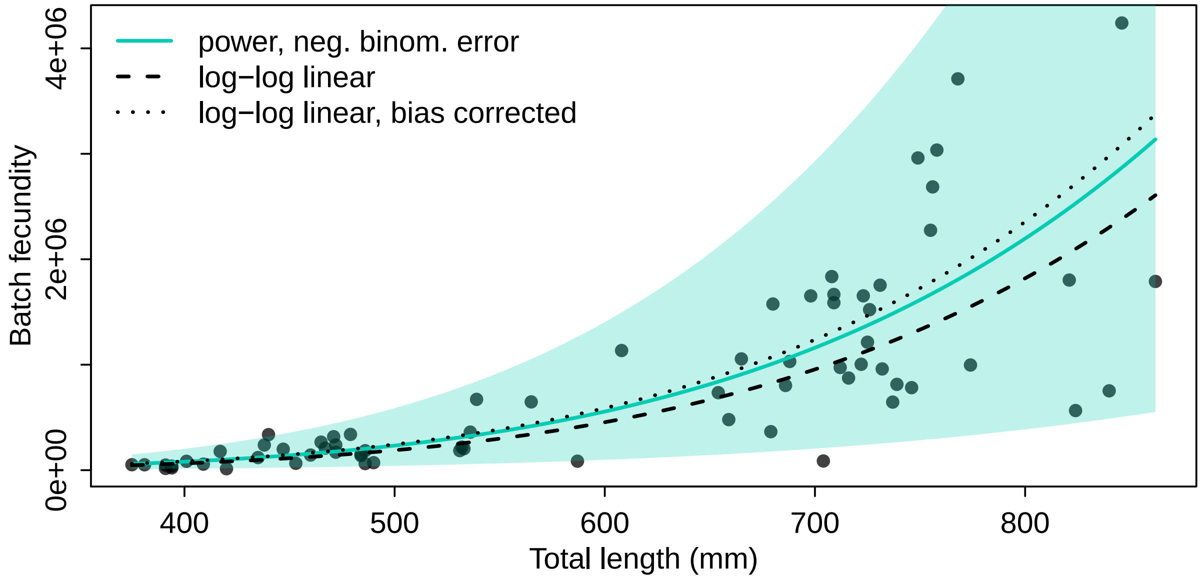
Although the assessment model is age-based and ultimately requires a vector of reproductive value-at-age, I chose to evaluate body size (i.e. total length) as a predictor of batch fecundity and batch number for two reasons: 1. biologically, reproductive investment is determined more by the size of a fish than by its age, 2. the ages of Red Snapper included in the data set are very skewed while the size data are not. This second point is evident in the density plots below, where dashed vertical lines mark the upper and lower limits bounding 95% of the data.





Batch fecundity

Fish fecundity tends to exhibit a power relationship (i.e. a*L^b) with fish length (Wootton, 1979). The parameters of this relationship are usually estimated by logarithmic transformation of both length and fecundity, and fitting a linear model to the transformed variables. Though this is a convenient way to estimate these parameters, it is also known to be biased (Sprugel 1983; Miller 1984) because the model is essentially fit to median instead of mean values. Bias corrections can be applied which eliminate much of the bias (Miller 1984). Though a better method is to simply fit the power function directly, using optimization methods. In this analysis, I used a optimization approach in R (function `optim()`) to model batch fecundity as a power function of total length with negative binomial error, following suggestions by Bolker (2008). The negative binomial distribution is often appropriate when error around the mean response increases with the predictor, which evident in many fecundity data sets. This may occur in part because oocyte density, and subsequently any estimation errors, are multiplied by gonad weight which also increases with length. Thus larger errors for longer fish may be due in part in an increase of observation error.



The equation for the fitted power model is:

$f_b = 3.012e - 08TL^{4.775}$

where f_b is batch fecundity and TL is total length. The overdispersion parameter (k) of the negative binomial distribution was estimated as $k = 2.617$.

Batch number and peak spawning date

Batch number is typically estimated by multiplying estimates of spawning period duration and spawning rate, known as spawning fraction (Hunter and Macewicz 1985; Murua et al. 2003). This method is analogous to calculating a definite integral (i.e. calculus); essentially, determining the area under a function over an interval. In this case, the function is a horizontal line with a y-intercept equal to the overall spawning fraction, and the interval is usually defined by the first and last dates that spawning indicators were observed.

In this analysis, I fit a more complex, four-parameter, plateau-shaped function to these same data, (i.e. presence/absence of spawning indicators by Julian date), to yield more rigorous estimates of these quantities.

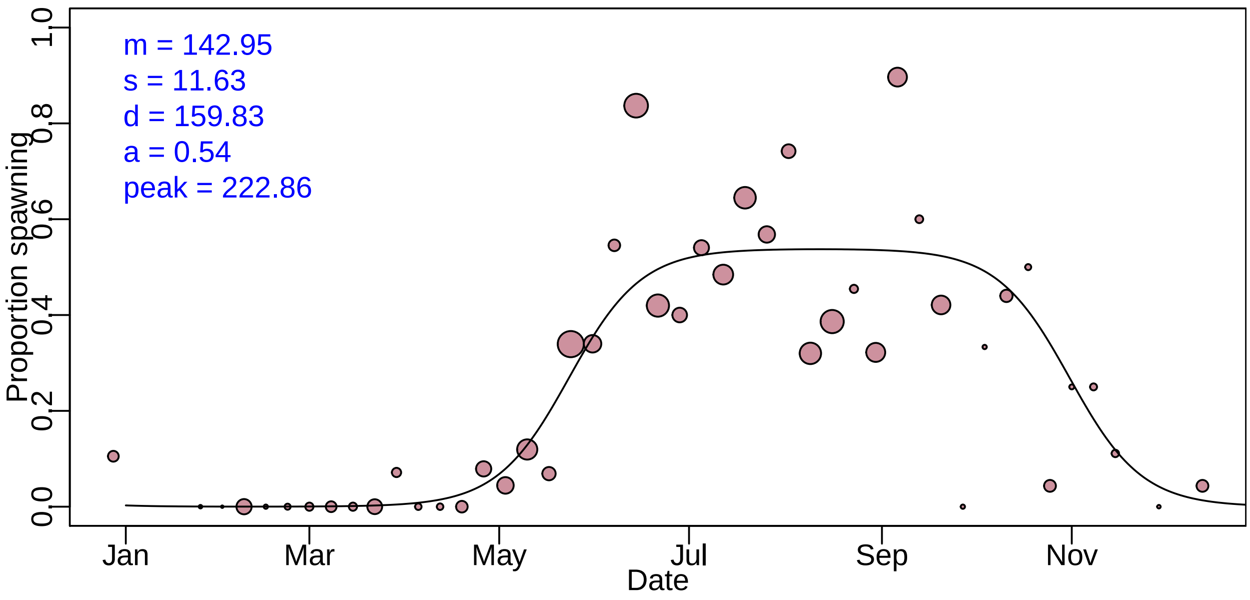
The form of the model is:

$P = \frac{a}{1+e^{(\frac{-(DATE-(m-0.5))}{s})}} - \frac{a}{1+e^{(\frac{-(DATE-((m-0.5)+(d+1)))}{s})}}$

where m is the average Julian date when a binary trait appeared; s is the variation around m, representing asynchrony in the start of spawning among females; d is the average duration for which a binary trait is present (i.e. average spawning period duration); and a is the asymptote of each of the two logistic functions in the model, but not always the height of the resulting function, and will sometimes approximate the standard calculation of spawning fraction. An estimate of peak spawning ($peak$) can be calculated as $m + d/2$, indicating the point the Julian day by which 50% of spawns have occurred in the population.

ADDENDUM

The figure below shows the plateau model fit to spawning status by Julian day, for data aggregated over all years. Estimates of the four model parameters and *peak* are printed on the figure.



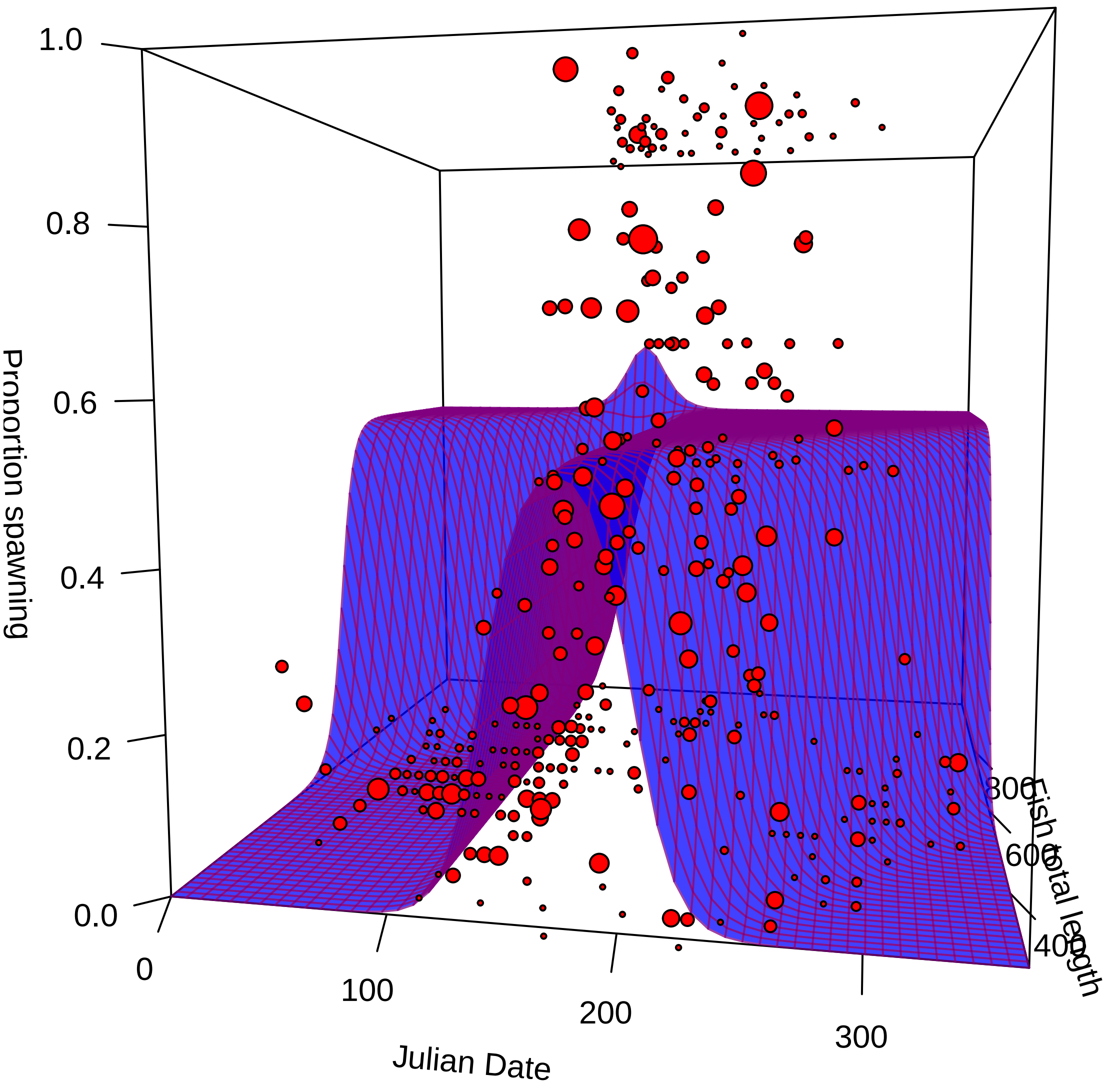
Since the curve is symmetrical and asymptotically approaches the x-axis early and late in the year, the area under this curve can be calculated over the entire year, yielding a size-independent estimate of batch number. Since spawning fraction has been shown to increase with size or age in multiple species (Fitzhugh et al. 2012), the plateau model was extended to simultaneously estimate d as a function of total length. Based, in part, on work by Porch et al (2015), the plateau model was extended by replacing d in the plateau model with a linear function:

$$d = d_0 + d_1 TL$$

This yields a five parameter model:

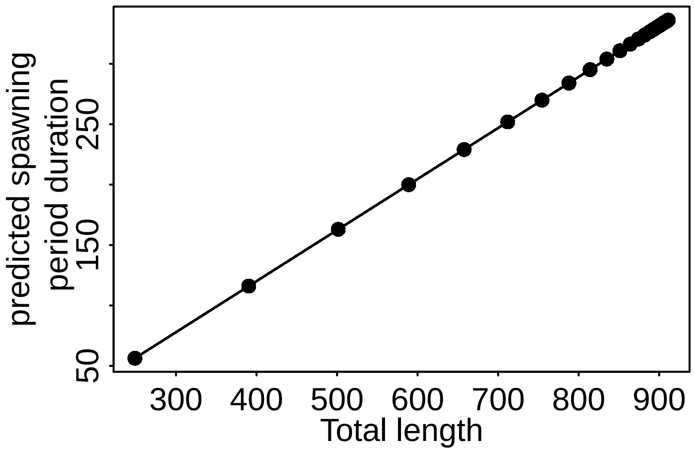
$$P = \frac{a}{1+e^{(\frac{-(DATE-(m-0.5))}{s})}} - \frac{a}{1+e^{(\frac{-(DATE-((m-0.5)+((d_0+d_1 TL)+1)))}{s})}}$$

The fit of the model including size-dependent d, can be plotted as a surface:

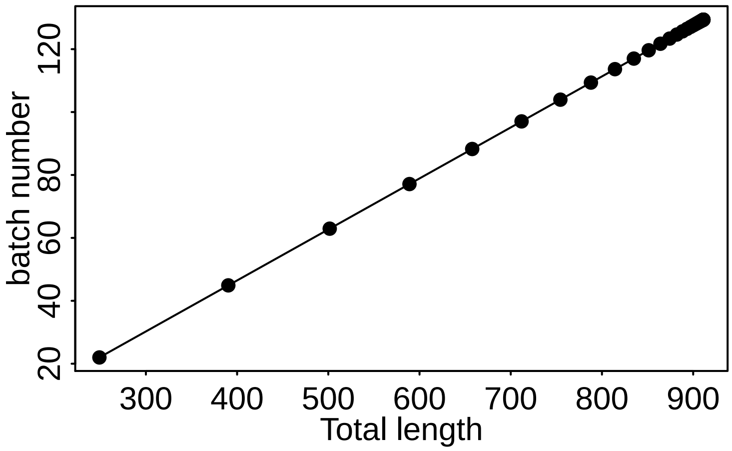


Using AIC to compare the basic plateau model with the length-based plateau model, I observed a ΔAIC of 42.5. As a rule of thumb, a difference in AIC values, or ΔAIC of $>$10 is considered a very strong indication that two models are different (Bolker 2008). Thus, the length-based model is preferable. The values of the optimal paramter estimates are $m =$ 143.51, $s =$ 8.57, $d_0 =$ -48.93, $d_1 =$ 0.42, and , $a =$ 0.54.

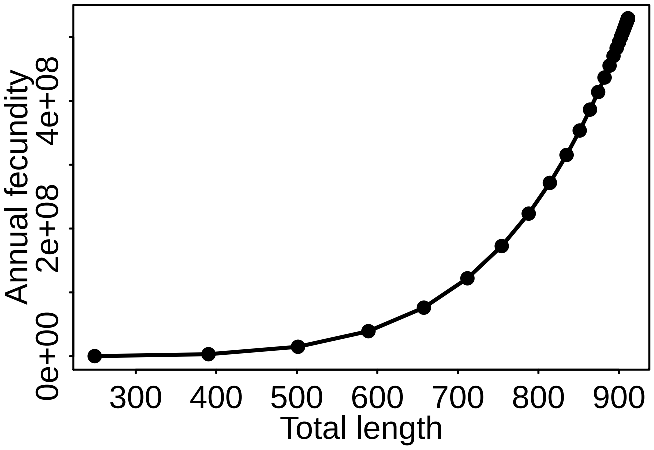
Predicted spawning period duration as a function of total length is plotted below.



Integrating under the fitted surface by TL yields the following relationship between batch number and TL. Note that, in this plot, estimated batch number was multiplied by a correction factor of 0.71 based on previous estimates of spawning indicator duration of 34 hours (Jackson et al. 2006).



Multiplying predicted batch fecundity by predicted batch number yields the following relationship between annual fecundity and TL.



Reproduction by age

Using the Von-Bertalanffy growth equation with parameters provided to me by Jennifer Potts at SEDAR41-DW (L_∞ = 911.36, k = 0.24, t_0 = -0.33), I calculated TL at ages 1 to 38, and then calculated batch fecundity and batch number at age, using my fitted equations, based on the computed TL values. These vectors are provided in Table 1.

Table 1: Predicted values by age and total length (TL), of batch fecundity (fb) with lower (fb95Lo), and upper (fb95Up) 95 percent confidence limits; batch number (nb), annual fecundity (fa), and the sample sizes used to estimate batch fecundity (fbN) and batch number (nbN).

Age	TL	fb	fb95Lo	fb95Up	nb	fa	fbN	nbN
1	249	8,348	1,469	21,081	22	183,656	0	35
2	390	71,391	12,568	180,433	45	3,212,595	7	241
3	502	236,216	41,586	596,476	63	14,881,608	22	403
4	589	508,885	89,592	1,284,998	77	39,184,145	7	251
5	658	862,372	151,826	2,177,599	88	75,888,736	6	183
6	712	1,257,887	221,459	3,176,323	97	122,015,039	12	108
7	754	1,659,875	292,233	4,191,392	104	172,627,000	9	94
8	788	2,042,348	359,570	5,157,182	109	222,615,932	1	53
9	814	2,389,568	420,701	6,033,957	114	272,410,752	0	26
10	835	2,694,182	474,329	6,803,144	117	315,219,294	0	8
11	851	2,954,714	520,199	7,461,021	120	354,565,680	1	4
12	864	3,173,326	558,688	8,013,041	122	387,145,772	0	6
13	874	3,354,115	590,518	8,469,555	123	412,556,145	1	7
14	882	3,501,970	616,549	8,842,910	125	437,746,250	0	7
15	888	3,621,859	637,656	9,145,644	126	456,354,234	1	4
16	893	3,718,428	654,657	9,389,493	126	468,521,928	0	5
17	897	3,795,813	668,279	9,584,899	127	482,068,251	0	1
18	900	3,857,576	679,155	9,740,858	128	493,769,728	0	0
19	903	3,906,717	687,808	9,864,944	128	500,059,776	0	3
20	904	3,945,719	694,671	9,963,429	128	505,052,032	0	0
21	906	3,976,615	700,113	10,041,445	128	509,006,720	0	0
22	907	4,001,052	704,413	10,103,153	129	516,135,708	0	1
23	908	4,020,359	707,814	10,151,905	129	518,626,311	0	1
24	909	4,035,598	710,497	10,190,385	129	520,592,142	0	0
25	909	4,047,617	712,613	10,220,735	129	522,142,593	0	1
26	910	4,057,092	714,281	10,244,660	129	523,364,868	0	0
27	910	4,064,557	715,593	10,263,510	129	524,327,853	0	0
28	910	4,070,437	716,630	10,278,358	129	525,086,373	0	2
29	911	4,075,067	717,446	10,290,050	129	525,683,643	0	0
30	911	4,078,712	718,087	10,299,254	129	526,153,848	0	0
31	911	4,081,582	718,591	10,306,499	129	526,524,078	0	0
32	911	4,083,840	718,993	10,312,201	129	526,815,360	0	0
33	911	4,085,617	719,302	10,316,688	129	527,044,593	0	0
34	911	4,087,015	719,550	10,320,219	129	527,224,935	0	0
35	911	4,088,115	719,746	10,322,997	129	527,366,835	0	0
36	911	4,088,981	719,894	10,325,183	129	527,478,549	0	1
37	911	4,089,662	720,014	10,326,902	129	527,566,398	0	0
38	911	4,090,197	720,107	10,328,255	129	527,635,413	0	1

Literature cited

Bolker, B. M. (2008). Ecological Models and Data in R. Princeton, NJ: Princeton University Press.

Fitzhugh, G. R., Shertzer, K. W., Kellison, G. T. & Wyanski, D. M. (2012). Review of size- and age-dependence in batch spawning: implications for stock assessment of fish species exhibiting indeterminate fecundity. Fishery Bulletin 110, 413-425.

Hunter, J. R. & Leong, R. (1981). The spawning energetics of female northern anchovy, Engraulis mordax. Fishery Bulletin 79, 215-230.

Hunter, J. R. & Macewicz, B. J. (1985). Measurement of spawning frequency in multiple spawning fishes. NOAA Technical Report NMFS 36, 79-88.

Jackson, M. W., Nieland, D. L. & Cowan, J. H. (2006). Diel spawning periodicity of red snapper Lutjanus campechanus in the northern Gulf of Mexico. Journal of Fish Biology 68, 695-706.

Miller, D. M. (1984). Reducing transformation bias in curve fitting. The American Statistician 38, 124-126.

Murua, H., Kraus, G., Saborido-Rey, F., Witthames, P. R., Thorsen, A. & Junquera, S. (2003). Procedures to estimate fecundity of marine fish species in relation to their reproductive strategy. Journal of Northwest Atlantic Fisheries Science 33, 35-54.

Porch, C., Fitzhugh, G., Lang, E., Lyon, H. & Linton, B. (2015). Estimating the Dependence of Spawning Frequency on Size and Age in Gulf of Mexico Red Snapper. Marine and Coastal Fisheries 7, 233-245.

Quinn, G. P. & Keough, M. J. (2003). Experimental design and data analysis for biologists. Cambridge, UK: Cambridge University Press.

Sprugel, D. (1983). Correcting for bias in log-transformed allometric equations. Ecology 64, 209-210.

Wootton, R. J. (1979). Energy costs of egg production and environmental determinants of fecundity in teleost fishes. Symposium of the Zoological Society of London 44, 133-159.