Estimates of population status of humpback chub (*Gila cypha*) in the Grand Canyon based on a length-based mark-recapture assessment model.

Steven J. D. Martell

UBC Fisheries Centre,

2202 Main Mall,

Vancouver, BC

V6T 1Z4,

CANADA

2011-09-27

Abstract

A length-structured population model that incorporates mark-recapture data was developed to estimate abundance of humpback chub (*Gila cypha*) in the Grand Canyon between 1989 and 2011. The model was fit to observations on catch-at-length and capture-recapture data from sampling programs that employed a variety of gear types including electrofishing, tramel netting, and hoop nets. Previous assessment models were age-based and have been shown to produce biased estimates of recent recruitment estimates. The age-structured model was also limited to data on fish that were greater than 150mm total length; the minimum size required for tagging. The length-structured model developed here makes no assumptions about age of individual fish and is also fit to catch-at-length data for fish that are too small to tag. Model outputs are based the number of fish greater than a specified size, or in a fixed size interval, there are no age-based results.

Contents

1 Introduction

In this report, I develop a length-structured mark-recapture model (hereafter, LSMR) where the accounting system for population numbers is based purely on length. The model is a statistical catch-at-length model, where the initial length distribution and recruitment each year are treated as latent variables to be estimated by fitting the model to a series of catch-at-length observations take over a period of time. A separable function is developed to estimate the year and size effect in observed catch-at-length data. The statistical nature of the model is very similar to that of **?**, but is based on catch-at-length rather than catch-age data. The model is also fit to length-based capture-recapture data, where the growth and survival of tagged animals is updated at each time step and the predicted ratio of marked and unmarked fish-at-length is used to estimate the length-based recapture rates.

2 Methods

There are two major methodological components in this length-based model: (1) the development of an individual based model (IBM) for simulating a capture recapture program, and (2) a statistical catch-at-length mark-recapture model to estimate the number of individual fish in each length-class in each sampling year. A detailed description of the IBM simulation model is provided in the appendix; in short, this simulation model generates a matrix of the number of fish captured-at-length, a matrix of the number of newly marked fish released-at-length, and a matrix of marked fish recaptured-at-length. The remained of this section is a detailed analytical description of the statistical catch-at-length model used to estimate the abundance-at-length of humpback chub in the Grand Canyon.

The following is a description of the analytical model for the length-structured mark-recapture model (hereafter, LSMR) used in this assessment. I present the analytical model in the form of a table where the order in which model equations are presented also represent the order in which the calculations proceed in the computer code. Equations presented in each table are referenced, for example, as (T1.1), where the T1 refers to Table 1, and the .1 refers to the first equation in that table. The LSMR model was implemented in AD Model

Builder (**?**), and the template code is available in the appendix of this document as well as a Git code repository (`https://code.google.com/p/lsmr-project/`). The description of the Length-Structured Mark-Recapture (LSMR) model is broken down into: input data, estimated parameters, dynamics of numbers-at-length, capture probability, and negative log-likelihoods and prior densities.

The following notation is used to define the dimensions of various variables. Vector quantities are designated with an an arrow ($\vec{x}$) or with a single subscript, and matrix is denoted by boldface uppercase letters ($\mathbf{X}$) and where two subscripts are shown denotes the element specific calculation. Higher dimensional arrays are indicated by normal upper case letters with 3 or more subscripts.

2.1 Input data

The model dimensions consists of time intervals (year indexed by i) and length intervals (index by j, T1.1). Capture-recapture data for the humpback chub have been collected on an annual basis since May 1, 1989, and the latest capture record in this analysis is February 27, 2012. The principle input data for LSMR consists of model dimensions (e.g., years, length intervals), catch-at-length $\mathbf{C}_{ij}$ for each year, the number of new marks released at length $\mathbf{M}_{ij}$, and the number of recaptured marks at length $\mathbf{R}_{ij}$.

2.2 Estimated parameters & parametric functions

An array of estimated parameters is denoted by Θ and consists of $8 + (I-1) + J$ unknowns where I is the total number of years and J is the total number of length intervals. The average initial numbers-at-length is defined as $\dot{N}$, and the average number of new recruits at each time step is denoted by $\bar{R}$. To initialize the number of individuals in each length interval Λ in the initial year $i = 1$, the average initial number is multiplied by a length specific lognormal deviate η_j (T1.11) with the added constraint $\sum_j \eta_j = 0$ to ensure that $\dot{N}$ is identifiable. Similarly, the average number of newly recruiting fish in each time step is based on deviates at each time-step from the mean recruitment (T1.12), where again the constraint $\sum_i \nu_i = 0$ ensures that $\bar{R}$ is identifiable.

Natural mortality is a function of length (T1.6), where the natural mortality at the asymptotic length M_∞ is estimated from the data. Selectivity is also assumed to be a parametric function of length (T1.7) where l_x and g_x represent length-at-50% vulnerability and the standard deviation of the logistic function, respectively. Note that in (T1.6) the estimated natural mortality rate is confounded with asymptotic length l_∞ which is also an estimated parameter along with the von Bertalanffy growth coefficient k. The growth parameters are used to calculate a vector of growth increment $\vec{\Delta}$ assuming von Bertalanffy growth (T1.8). An additional parameter β is used to characterize the variability in annual growth increments for individual fish.

2.3 Growth transition

The asymptotic length l_∞ is defined as the average asymptotic length for a population of fish. It is assumed in (T1.8) that individuals greater than l_∞ continue to grow at a much reduced rate k; this is accomplished by exponentiating the growth increment equation $((l_\infty - x_j)(1 - exp(-k\tau)))$, adding 1.0, and taking the natural logarithm ensuring that (T1.8) remains positive for all positive values of l_∞, k, and Λ. The probability of transitioning from one length interval x_j to length intervals greater than x_j is based on a gamma density function (T1.9), where the mean is denoted by Δ_j and a variance equal to $\Delta_j \beta$. Each row of $\mathbf{P}_{j,j'}$ is normalized to sum to 1.0, and $\mathbf{P}_{j,j'} = 1.0$ when $j = j' = n$, where n is the number of length intervals (i.e., individuals in the last length interval represent a plus group).

2.4 Dynamics of numbers-at-length

2.5 Length-based capture probability

2.6 Negative log-likelihoods and prior densities

2.7 Length-structured mark-recapture model (LSMR)

Table 1: Data, parameters, and analytical procedures for the length-based mark-recapture model.

INDEXES, DATA & CONSTANTS	
index for time, index for length interval	i, j (T1.1)
time step	τ (T1.2)
set of midpoints of length intervals	$\Lambda = \{x_1, \ldots, x_J\}$ (T1.3)
catch, new marks, recaptures	$\mathbf{C}_{i,j}, \mathbf{M}_{i,j}, \mathbf{R}_{i,j}$ (T1.4)
PARAMETERS & DERIVED VARIABLES	
estimated parameters	$\Theta = \{\dot{N}, \bar{R}, M_\infty, l_\infty, k, \beta, l_x, g_x, \vec{\eta}, \vec{\nu}\}$ (T1.5)
mortality-at-length	$\vec{m} = \dfrac{M_\infty l_\infty}{\Lambda}$ (T1.6)
selectivity-at-length	$\vec{s} = \dfrac{1}{1 + \exp(-(\Lambda - l_x)/g_x)}$ (T1.7)
growth increment	$\vec{\Delta} = \ln(\exp[(l_\infty - \Lambda)(1 - \exp(-k\tau))] + 1)$ (T1.8)
length transition probability	$\mathbf{P}_{j,j'} = \displaystyle\int_{x_j - x^*}^{x_j + x^*} \dfrac{x_{j'}^{(\Delta_j/\beta - 1)} e^{x_{j'}/\beta}}{\beta^{\Delta_j/\beta} \Gamma(\Delta_j/\beta)} dx,$
	$\displaystyle\sum_{j'=1}^{J} P_{j,j'} = 1$ (T1.9)
total numbers, marked numbers	$\mathbf{N}, \mathbf{T}$ (T1.10)
INITIAL STATES ($i = 1$, $j = 1$)	
initial numbers-at-length	$\mathbf{N}_{i,j} = \dot{N} \exp(\eta_j), \quad \text{where} \displaystyle\sum_j \eta_j = 0$ (T1.11)
new recruits	$\mathbf{N}_{i,j} = \bar{R} \exp(\nu_i), \quad \text{where} \displaystyle\sum_i \nu_i = 0$ (T1.12)
DYNAMIC STATES ($i > 1$)	
capture probability	f_i (T1.13)

Figure 1: Growth increments by tag year.
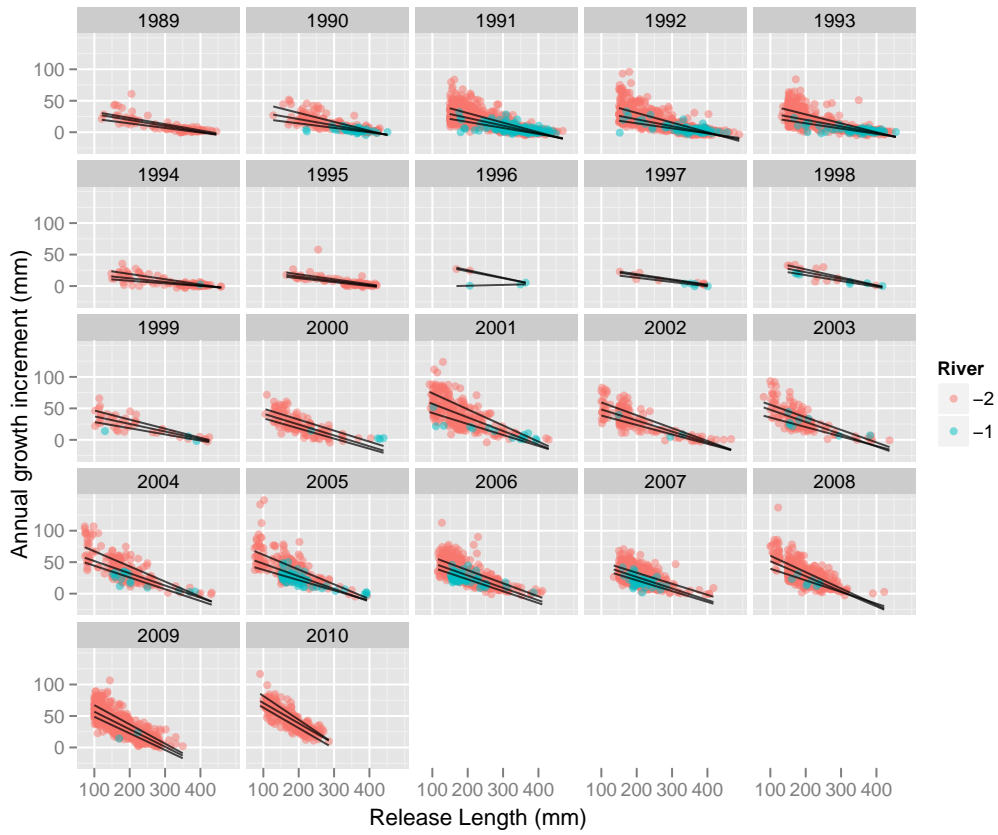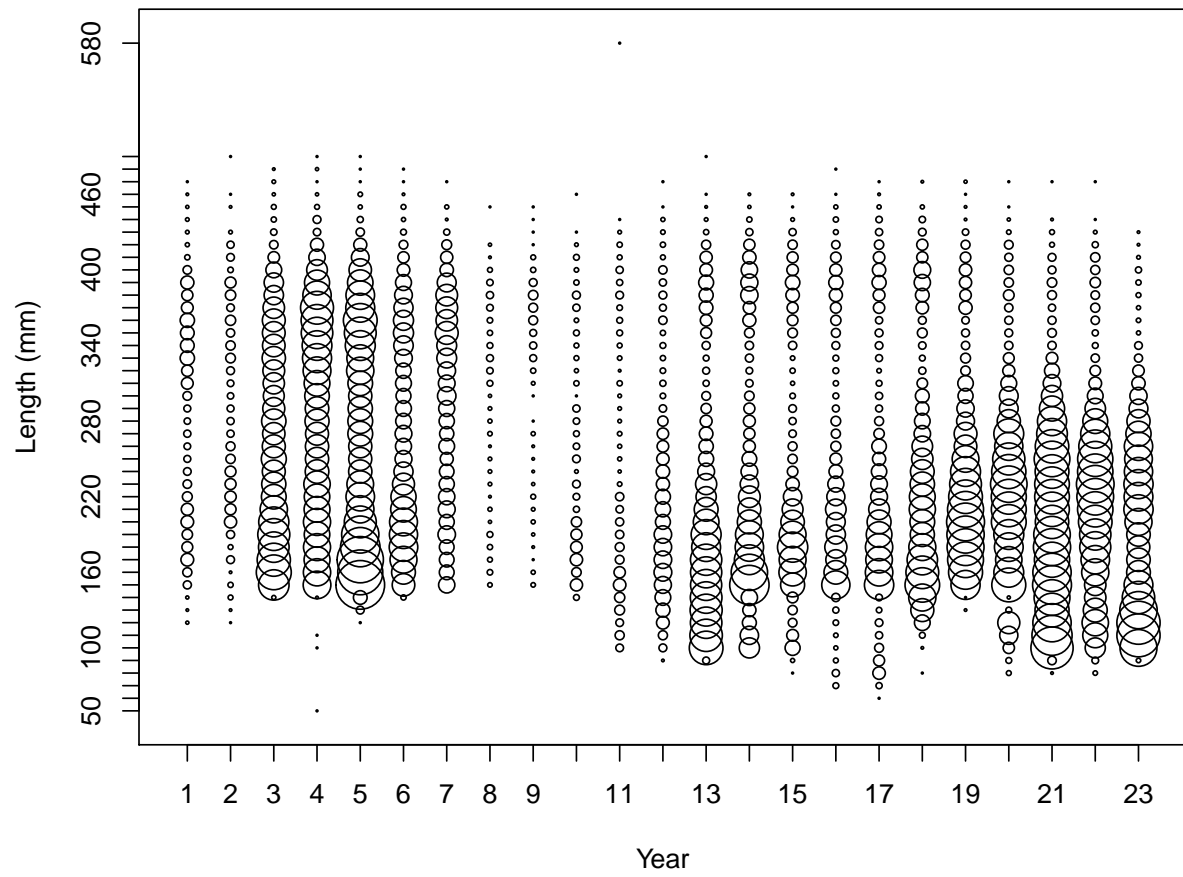
Figure 2: Length frequency by year for all gear types. Area of circle is proportional to abundance of measured fish.
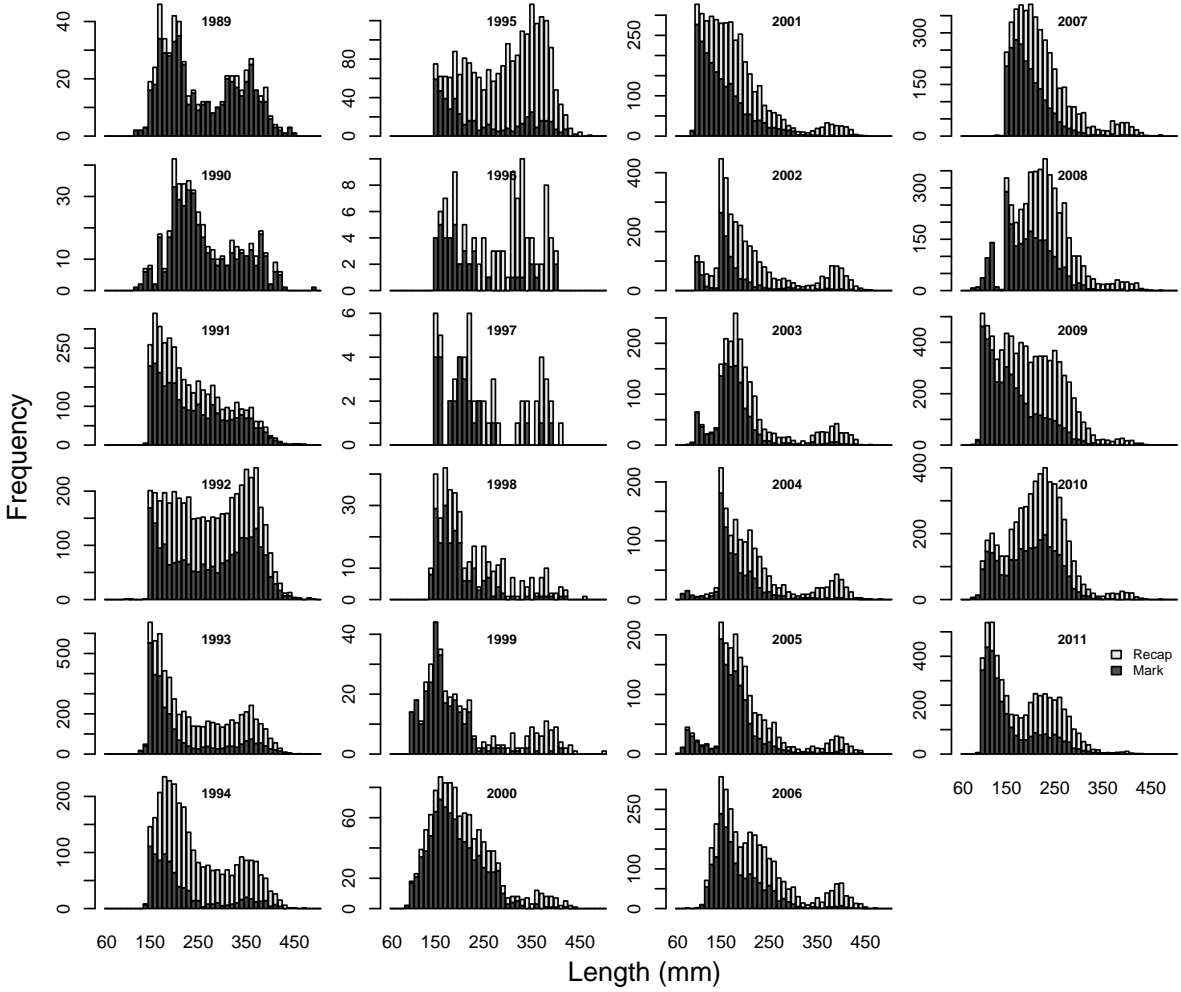
Figure 3: Annual capture and recapture history of HBC using hoop nets (all sizes & bait) in the LCR and COR reaches from 1989 to 2011.
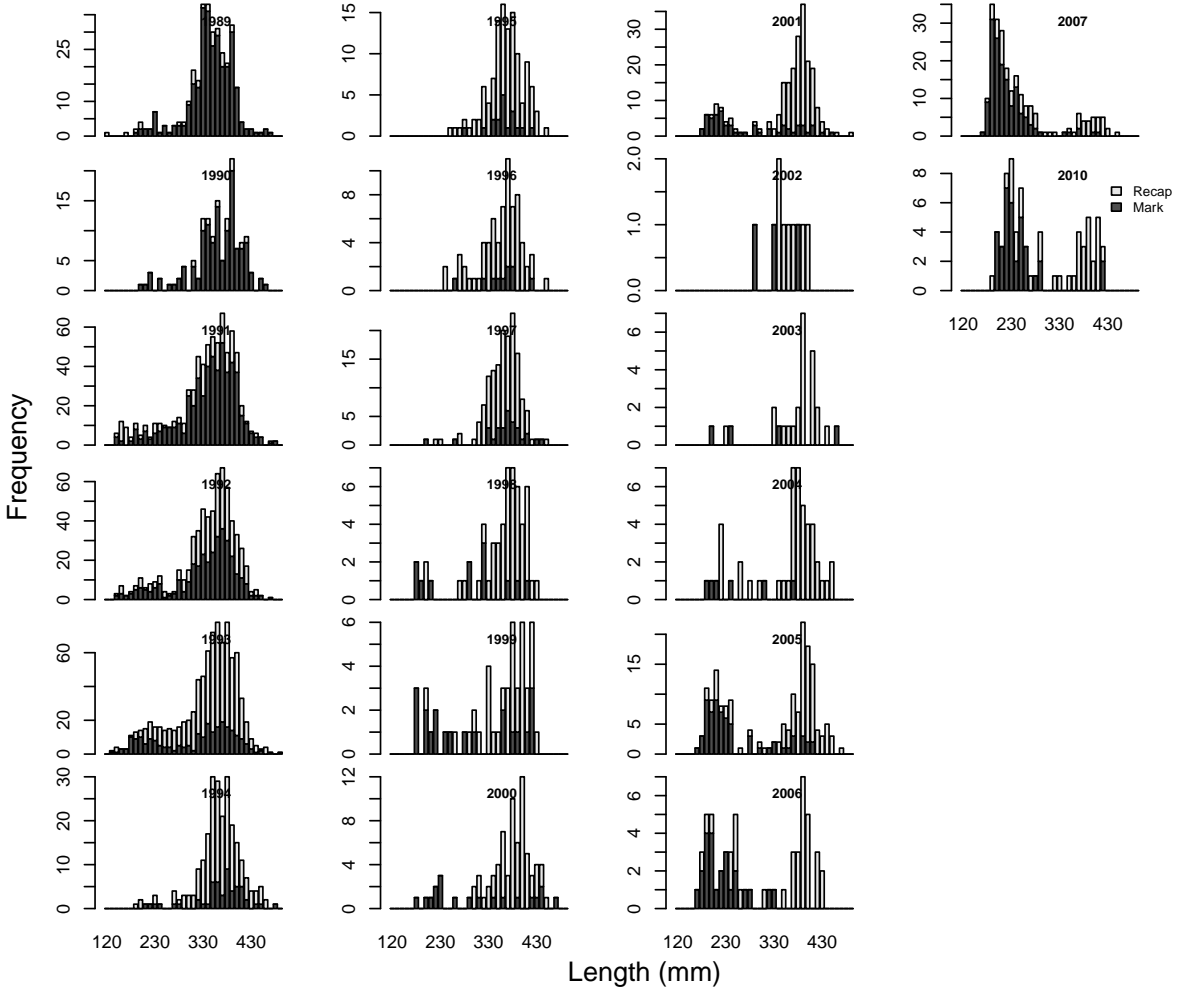
8

Figure 4: Annual capture and recapture history of HBC using tramel nets of all sizes in the LCR and COR reaches from 1989 to 2011.

9

Table 2: Number of fish measured by year and month, sampled by all gears in all reaches of both the LRC and COR.

						MONTH							
YEAR	1	2	3	4	5	6	7	8	9	10	11	12	(all)
1989	0	0	0	0	887	0	0	0	0	0	0	0	887
1990	0	0	0	408	125	0	0	0	0	43	42	0	618
1991	79	3	135	9	285	394	1624	1054	536	291	200	176	4786
1992	168	340	781	1293	493	1181	369	145	144	290	267	0	5471
1993	117	153	1132	759	1359	822	953	1093	270	228	250	239	7375
1994	154	201	296	658	707	410	198	154	152	129	128	55	3242
1995	226	231	383	915	476	83	0	0	2	0	0	0	2316
1996	0	2	20	96	47	6	0	0	27	0	0	0	198
1997	0	0	7	42	114	18	0	0	32	0	0	0	213
1998	0	0	1	234	47	36	39	65	5	18	0	0	445
1999	20	0	0	210	52	18	0	0	62	56	53	0	471
2000	20	0	0	418	21	271	2	44	20	333	151	13	1293
2001	0	0	1	37	491	1347	9	163	85	1180	929	0	4242
2002	0	4	0	980	1066	0	20	0	789	502	0	0	3361
2003	16	16	48	599	434	0	55	13	293	694	21	0	2189
2004	18	25	51	760	353	23	27	15	166	542	32	0	2012
2005	23	12	44	515	207	285	93	19	799	476	0	0	2473
2006	24	35	160	1098	921	679	179	56	270	317	0	0	3739
2007	0	0	2	937	1545	900	10	0	805	569	0	0	4768
2008	0	5	3	1265	1735	528	1020	195	718	928	0	0	6397
2009	0	0	6	1225	4180	2267	551	168	895	1323	443	245	11303
2010	0	0	2	0	1466	2882	523	14	1043	708	0	0	6638
2011	0	0	0	0	2796	2537	127	128	544	1042	63	49	7286
2012	28	61	0	0	0	0	0	0	0	0	0	0	89
(all)	893	1088	3072	12458	19807	14687	5799	3326	7657	9669	2579	777	81812

Table 3: Number of fish captured by gear type listed in the GCMRC database for each year.

					YEAR					
YEAR	ANGL	DIP	ELEC	GILL	HOOP	PA	SEINE	TRAP	(all)	NA
1989	6	0	0	321	560	0	0	0	887	0
1990	2	0	3	140	473	0	0	0	618	0
1991	4	0	44	711	3978	0	11	1	4786	37
1992	3	0	68	636	4740	0	24	0	5471	0
1993	2	0	80	876	6313	0	102	1	7375	1
1994	1	0	1	239	2997	0	2	1	3242	1
1995	1	0	7	118	2190	0	0	0	2316	0
1996	0	0	4	72	113	0	9	0	198	0
1997	0	0	2	153	58	0	0	0	213	0
1998	0	0	5	58	377	0	4	1	445	0
1999	0	0	17	54	392	0	1	7	471	0
2000	0	0	6	81	1150	0	55	1	1293	0
2001	5	0	1	233	4003	0	0	0	4242	0
2002	0	0	5	10	3346	0	0	0	3361	0
2003	0	1	102	27	2058	0	1	0	2189	0
2004	2	0	108	49	1621	226	0	0	2012	6
2005	0	0	228	174	1919	152	0	0	2473	0
2006	0	0	138	58	3442	100	1	0	3739	0
2007	0	0	9	225	4352	181	1	0	4768	0
2008	3	0	8	0	4838	1527	21	0	6397	0
2009	0	0	6	0	7706	3589	0	2	11303	0
2010	3	0	0	71	5294	1269	0	0	6638	1
2011	0	0	0	0	5376	1910	0	0	7286	0
2012	0	0	0	0	0	89	0	0	89	0
(all)	32	1	842	4306	67296	9043	232	14	81812	46

Table 4: Number of fish captured by length by all gear types listed in the GCMRC database for each year.

t		YEAR																					
	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
50	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	11	1	0	0	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	45	0	0	7	2	6	0
90	0	0	0	0	0	0	0	0	0	0	0	2	14	0	5	7	35	2	0	10	21	13	6
100	0	0	0	1	0	0	0	0	0	0	17	18	327	118	65	5	23	10	0	37	514	117	393
110	0	0	0	1	0	0	0	0	0	0	24	24	304	97	40	6	15	71	0	95	465	180	538
120	3	1	0	0	18	0	0	0	0	0	17	47	286	56	22	9	17	153	0	140	424	201	539
130	2	2	0	0	54	8	0	0	0	0	28	57	297	50	25	11	9	213	2	10	334	165	403
140	3	7	5	2	684	147	0	0	0	10	30	71	289	76	34	20	13	345	1	3	371	132	304
150	19	8	268	221	585	163	76	6	6	43	45	83	280	447	166	229	248	310	245	333	436	132	239
160	24	2	353	215	618	209	63	9	5	26	37	91	281	382	216	162	218	267	333	257	424	213	162
170	47	18	318	194	450	235	62	7	1	44	22	89	283	259	208	116	204	210	371	201	359	235	159
180	34	7	270	211	400	229	62	8	2	39	23	88	240	233	264	141	215	168	394	238	399	278	148
190	31	19	296	188	295	224	88	10	3	35	20	88	259	222	208	106	190	176	415	259	385	283	159
200	46	43	262	213	216	189	65	3	5	30	19	64	190	167	159	102	165	198	374	349	322	325	211
210	42	35	215	193	235	182	82	5	4	11	16	67	164	151	125	122	124	192	328	352	345	362	247
220	28	37	176	188	205	139	76	2	7	13	20	65	118	135	97	94	81	171	291	354	347	390	239
230	21	35	168	201	156	105	66	4	3	17	6	50	129	97	51	74	79	162	258	385	346	409	246
240	17	34	152	161	153	82	61	4	2	4	3	52	81	80	32	54	68	131	206	338	329	361	231
250	14	25	180	156	154	74	49	4	2	17	5	44	65	62	25	40	50	122	147	299	368	358	221
260	13	23	153	156	177	81	71	2	2	12	5	39	59	41	23	29	61	78	150	229	324	313	232
270	15	15	144	152	177	70	58	6	5	10	6	38	53	36	16	24	31	70	92	257	271	244	183
280	12	15	171	167	170	72	67	5	1	12	4	30	41	41	15	27	23	51	86	155	245	178	154
290	14	14	140	160	155	64	74	4	0	15	4	16	30	35	17	14	14	55	65	101	183	121	102
300	22	12	126	174	175	72	98	3	1	1	4	9	22	26	9	10	15	32	69	101	129	71	85
310	40	13	131	191	215	68	80	10	4	8	8	11	11	16	5	8	10	15	26	72	89	46	51
320	37	18	137	223	207	89	90	12	8	5	2	9	14	13	8	10	6	10	28	42	65	32	38
330	59	27	157	244	259	109	113	14	14	7	6	8	10	19	6	10	14	16	20	34	39	20	19
340	54	25	147	255	292	116	113	10	15	7	7	10	19	20	20	18	17	28	18	20	20	10	17
350	54	20	149	287	329	116	151	8	14	10	10	8	34	27	27	23	18	18	16	20	22	11	5
360	59	30	151	292	242	105	133	9	23	6	10	20	39	49	33	30	15	18	50	19	19	20	6
370	40	17	133	313	236	90	137	13	23	15	9	14	53	44	35	34	37	45	38	20	15	24	5
380	36	31	108	231	187	73	135	15	26	18	17	18	57	86	52	49	40	54	43	33	23	21	7
390	50	35	108	181	145	46	103	12	18	8	11	15	64	84	60	58	57	76	44	25	26	31	8
400	21	9	72	123	90	37	52	7	8	6	15	20	46	77	38	43	51	81	32	19	18	24	11
410	8	17	38	77	49	22	43	2	7	10	7	7	43	51	35	32	38	36	23	22	18	25	4
420	5	15	21	49	20	10	28	3	1	4	8	8	24	36	25	20	25	36	12	6	7	11	2
430	4	4	12	16	11	5	11	0	1	1	4	6	8	15	15	7	15	17	2	3	7	4	2
440	4	0	8	18	5	7	2	0	1	0	1	5	4	8	4	8	12	12	2	1	2	1	0
450	3	2	7	7	6	2	5	1	0	0	0	1	2	3	1	5	5	5	1	1	0	0	0
460	2	1	3	3	1	1	1	0	0	1	0	0	1	2	2	1	2	0	3	0	1	0	0
470	1	0	4	1	1	1	1	0	0	0	0	1	0	0	0	0	1	2	0	1	0	1	0
480	0	1	2	3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
490	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
580	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Table 5: Number of new marks released by year and size interval.

LI	YEAR																			
	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2010
120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
140	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150	0	0	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
160	0	0	2	3	3	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
170	0	0	0	2	3	0	0	0	0	0	0	0	0	0	0	1	3	2	1	0
180	0	0	2	3	10	0	0	0	0	2	3	1	2	0	0	1	9	4	9	0
190	1	1	8	5	9	0	0	0	1	1	0	0	6	0	0	1	7	4	31	0
200	2	1	3	6	10	1	0	0	0	0	2	1	5	0	1	1	9	1	26	4
210	2	3	7	5	6	1	0	0	0	1	1	1	6	0	0	0	7	2	19	3
220	2	7	3	4	9	1	0	0	0	0	2	2	7	0	0	0	6	3	15	7
230	7	6	6	6	8	1	0	0	0	0	0	3	3	0	1	1	5	1	8	6
240	1	2	7	8	5	1	0	0	0	0	1	0	3	0	0	0	0	2	13	2
250	3	0	9	1	4	0	0	1	0	0	1	1	1	0	0	0	0	0	6	5
260	1	1	9	2	4	1	0	0	1	0	0	0	1	0	0	0	3	1	5	3
270	3	1	9	3	2	1	0	0	0	0	1	0	1	0	0	0	0	1	3	1
280	3	2	11	10	5	0	1	0	0	2	0	1	0	0	0	0	3	0	2	2
290	3	4	6	4	4	0	1	0	0	0	1	0	3	1	0	0	0	0	1	0
300	9	0	25	9	5	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0
310	15	4	20	18	2	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
320	14	2	34	17	12	2	1	1	3	3	0	0	2	0	0	0	1	1	0	0
330	37	10	25	23	10	1	0	0	1	0	0	0	0	0	0	0	2	0	1	0
340	36	11	40	19	18	1	2	1	3	0	0	1	1	0	0	0	0	0	1	0
350	26	8	45	24	13	6	2	1	3	0	0	0	3	0	0	0	1	0	0	0
360	29	14	38	32	16	6	5	1	6	1	2	1	2	0	0	1	3	0	2	0
370	20	5	52	36	19	3	1	2	4	0	0	0	1	0	0	0	0	0	0	0
380	20	10	37	30	16	9	3	2	3	1	1	0	3	1	0	0	3	0	1	0
390	30	20	42	22	14	4	1	0	1	1	1	1	3	0	0	0	2	0	1	0
400	14	7	37	13	11	5	1	0	2	0	1	0	1	0	0	0	2	0	0	0
410	4	7	15	11	9	5	0	1	0	1	1	0	3	0	0	0	0	0	0	0
420	2	8	11	8	6	2	1	1	1	0	3	1	0	0	0	0	0	0	0	2
430	2	3	6	2	3	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0
440	1	0	4	2	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0
450	2	2	4	2	2	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
460	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
470	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
480	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
490	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 6: Number of new marks released by year and size interval.

LI	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10	11	1	0	0	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	0	12	0	5	15	41	0	0	7	2	6	0
90	0	0	0	1	0	0	0	0	0	0	0	2	277	97	64	7	30	2	0	10	16	13	5
100	0	0	0	0	0	0	0	0	0	0	14	17	235	53	36	4	21	9	0	37	462	92	343
110	2	1	0	0	0	0	0	0	0	0	18	21	206	14	21	6	13	54	0	95	412	146	437
120	2	2	0	0	0	0	0	0	0	0	10	34	182	9	23	2	15	111	2	140	370	142	423
130	3	6	1	1	15	7	0	0	0	0	21	38	159	8	31	6	9	130	1	10	246	118	311
140	16	7	1	1	40	111	0	4	4	8	24	48	142	264	136	13	11	239	203	1	244	75	215
150	18	2	204	169	553	97	59	5	4	29	44	64	123	185	160	181	193	205	257	289	304	73	165
160	34	17	211	141	395	86	47	4	4	18	33	72	130	115	154	123	150	168	280	194	275	121	109
170	29	6	187	95	389	97	39	4	2	30	17	67	99	76	156	79	133	114	267	130	222	120	76
180	28	17	152	102	232	84	28	5	2	18	16	63	82	40	123	77	139	84	218	137	193	154	58
190	33	33	161	64	200	64	39	2	4	22	18	59	54	38	72	45	115	76	195	151	162	148	58
200	35	29	160	68	124	39	23	3	3	18	14	46	54	28	56	41	91	87	154	173	110	156	72
210	25	27	117	70	69	37	12	2	2	6	10	44	38	12	30	48	51	77	137	154	120	158	87
220	11	32	97	73	56	32	16	3	1	6	14	40	39	18	21	36	30	65	104	147	116	181	77
230	15	31	90	65	39	14	16	0	2	10	5	32	33	10	8	20	26	46	81	116	106	197	82
240	9	21	89	52	28	14	6	0	0	3	1	35	21	7	8	12	17	58	66	99	101	160	67
250	11	17	105	51	26	3	9	1	0	6	2	27	21	5	4	13	20	45	41	64	95	148	80
260	12	12	78	65	34	8	12	0	1	7	1	24	21	6	1	8	13	19	28	68	77	136	68
270	8	10	69	55	38	8	7	1	0	1	2	24	18	7	4	7	9	25	22	42	66	102	52
280	10	8	104	61	29	10	5	0	0	4	2	25	16	11	4	6	7	17	17	16	50	71	49
290	11	10	82	49	26	5	6	0	0	2	2	10	14	8	2	1	6	11	11	22	23	33	24
300	20	8	64	67	28	6	8	1	0	0	0	3	3	4	1	2	3	5	9	16	22	23	21
310	19	12	63	72	38	8	5	1	0	1	1	4	3	4	0	2	2	3	2	5	17	12	14
320	17	10	66	83	39	10	10	1	0	0	0	5	2	3	0	3	1	3	4	5	4	10	11
330	14	12	70	87	34	16	13	1	1	0	0	2	2	1	6	3	3	1	3	5	8	3	6
340	19	11	78	114	49	20	20	0	0	1	2	0	2	5	1	3	3	4	4	2	3	4	6
350	25	13	69	113	66	17	25	2	0	0	1	0	2	3	2	2	1	3	1	2	2	1	0
360	16	8	69	115	74	12	9	1	0	1	1	3	1	3	3	1	3	1	6	3	3	2	1
370	12	18	44	131	53	13	16	0	1	1	0	0	0	7	6	3	2	4	4	2	0	2	1
380	12	11	44	97	55	14	16	0	0	1	1	1	1	5	6	1	2	4	1	2	0	0	1
390	6	2	33	82	40	4	15	0	1	1	2	1	0	3	2	3	4	5	3	4	2	0	0
400	3	6	19	42	25	7	7	2	0	0	0	1	1	3	1	1	7	7	3	0	0	1	1
410	2	5	12	29	21	4	4	0	0	1	2	0	1	4	2	1	0	3	3	2	0	0	0
420	1	1	6	17	15	7	7	0	0	1	0	1	1	0	1	2	1	2	2	2	0	0	1
430	3	0	2	7	2	0	0	0	0	0	0	1	3	1	2	1	1	1	1	1	1	0	1
440	1	0	1	7	2	1	0	0	0	0	0	0	1	0	0	1	2	1	0	0	0	1	0
450	3	0	3	2	0	1	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0
460	1	0	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
470	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
480	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
490	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
580	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

A Work plan

The following outlines a work plan for the assessment of abundance for humpback chub. The objective is to develop a much more flexible length-based model that can be used to explore alternative hypotheses about natural mortality rates, cumulative effects of release mortality for intensive sampling periods, and to potentially integrate other sources of environmental variations such as the effects of turbidity on capture probability and recruitment variation.

Analytical approach

I will develop a statistical catch-at-length model in the AD Model Builder software and additions R-scripts for manipulating data and summarizing model results. Input data for the model will consist of a matrix of the total catch-at-length in 5 to 10 mm size intervals for all years, a matrix of the number of marks released by size and year, a matrix of the number of marks recaptured by size and year, and a three dimensional array of the number of marks recaptures by size for each tag-cohort released (optional).

Estimated parameter will include: natural mortality rate, growth parameters, a vector of the initial numbers in each length interval, and a vector of age-0 recruits each year. Propagation of the numbers-at-length to the next time step will be based on a size-transition matrix, which is a function of the growth parameters and variation in growth. Observation models will include a probability of capturing an animal of a given length each year, the probabilities of capturing a marked and unmarked animal of a given length, and optionally the probability of recapturing a specific tag-cohort of a given length. These predicted observations will be compared with the empirical data using a negative binomial likelihood function. The negative binomial model is more suitable here because it can account for over-dispersed data and accommodate zero observations in cases where there is sparse information.

Detailed work plan

Major components of the project

The following is list of milestones for this project. Each of these items will be expanded upon in the section on project details.

1. Data acquisition and processing.

2. Development of an operating model for simulating data with known parameter values.

3. Development of a length-based assessment model to be fitted to data on capture and recapture information by length interval.

4. Simulation testing; exploring the precision and bias of the assessment model in jointly estimating recruitment, size-specific capture probability, and growth and natural mortality using simulated data sets.

5. Application of the length-based model to the HBC data.

6. Quantifying uncertainty in model parameters and estimates of recruits using Markov Chain Monte Carlo methods to integrate the joint posterior distribution.

7. Report & presentation to the Technical Working Group.

Project details

Data acuisition & processing The necessary data required to conduct the analysis will require information from the following fields in the GCMRC database: FISHNO, TRIP_ID, DATES, RM, RIV, TL, TAGNO. The following SQL statement was used in a previous study to extract the necessary information. Note that the following code has been modified to obtain all fish lengths, not just those greater than 150 mm.

```
--****
--All
--****

CREATE OR REPLACE VIEW FISH.V_ASMR_2009_ALL
(
    FISHNO,
    TRIP_ID,
    DATES,
    RM,
    RIV,
    TL,
    TAGNO
)
AS
SELECT "CAPTURE_ID" fishno, "TRIP_ID" trip_id, "START_DATE" dates,"START_RM" rm,"RI
      FROM FISH.CAPTURE_HISTORY_20091211_0832
    WHERE SPECIES_CODE = 'HBC'
    AND (
       (RIVER_CODE = 'COR' AND START_RM >= 57 AND START_RM <= 68.5)
       OR RIVER_CODE = 'LCR'
       )
    AND START_DATE >= TO_DATE('04/01/1989', 'MM/DD/YYYY')
    AND TOTAL_LENGTH >= 00
```

An R-script will be developed for post processing of the data to assign the length capture and recapture information into discrete length intervals for assembling input data into the assessment model.

Operating model An individual based model will be developed to generate simulated data sets with known natural mortality rates, recruitment vectors, growth rates and capture probabilities. The pseudo code for the operating model is as follows:

1. Specify a vector of absolute recruitment from 1950-2011.

2. For each individual recruit in each year apply the following procedure:

   (a) boolean trail for survival, if the animal survives then go to (b) else, individual died and restart at step 2.

   (b) boolean trail for capture:

      i. Captured: obtain length of fish, if greater than 150mm then tag and release fish, go.

      ii. Recaptured: obtain length of fish, goto step (a).

      iii. Not captured: goto step (a).

3. Store information about individual capture history and length into simulated database.

The above algorithm is intended to generate the exact data that is currently collected in the HBC monitoring program. Specific details about factors that affect capture probability and survival would be incorporated into the boolean probabilities defined in 2a and 2b.

Length-based assessment model A statistical length-based assessment model is similar in nature to a statistical catch-age model in that numbers-at-age, or in this case, numbers-at-length are propagated forward in time. The previous Age-Structured Mark-Recapture Model (ASMR) was dependent on catch-age information; age data for HBC were inferred from an analytical age-length key (i.e., based on inferences about growth, not empirical age-length data). Estimates of uncertainty were overly precise due to the pre-processing of the data to be used with ASMR. The length-based model makes no such inferences about the age of individual fish and is based strictly on the length observation data.

In a length based model, individuals in a given length interval are propagated in time by redistributing these individuals into new length bins based on a length transition probability (Fig 5). The length transition probability is a function of growth and the time interval between sampling events. The graphical example in Fig 5 does not account for mortality over time, its only meant to show the transition of individuals from one length bin to subsequent length bins.

Simulation testing The purpose of simulation testing is to (i) demonstrate that the model is able to estimate the model parameters given perfect information and satisfying all of the model assumptions, (ii) examine precision and bias in parameter estimates when faced with observation, process, and structural errors, and lastly (iii) to examine the
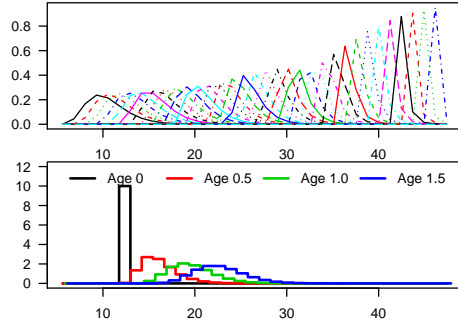
Figure 5: A graphical example of a length-based transition. Starting with 10 age-0 individuals, these 10 would then be distributed to the age-0.5 distribution distribution. The age-0.5 distribution then transitions to the age-1.0 distribution and so on.

estimability, bias, and precision of model parameters when underlying structural assumptions are not met. Simulation testing will be conducted to examine the ability to jointly estimate recruitment, capture probability, growth and natural mortality in a reliable and unbiased manner.

Application to the HBC data The length-based assessment model will then be applied to the length-based capture and recapture data for the humpback chub monitoring program. Model outputs will include estimates of recruits (and associated uncertainty), estimated model parameters (and uncertainty). It is anticipated that more reliable estimates of recruits will be available with the length based model because the model is not limited to length information that is greater than 150mm, as is the case in the ASMR model.

Quantifying uncertainty Estimates of uncertainty will jointly consider uncertainty in the mark-recapture data, growth and mortality. To do so the joint posterior distribution of the data will be constructed numerically using a Markov Chain Monte Carlo procedure (using the Metropolis Hastings algorithm implemented in AD Model Builder). Uncertainty in model parameters as well as outputs will be cast in the form of marginal posterior densities.

Report and presentation This is original research and the methods outlined for a length-based model that incorporates growth increment data from a mark recapture program has not been previously published to my knowledge. Ideally these results will be disseminated in the primary literature, but will also be presented to the Technical Working Group and be available as a technical report (e.g., USGS Open File Report). Also source code, scripts, and documentation will be hosted on an open source repository with version control. The intention here is to create a repository for continued development of the software and to document the historical changes over time.

B Individual based model for simulating the dynamics and sampling of humpback chub in the Grand Canyon

B.1 Introduction

The following is a detailed description of the simulation model that was used for simulating the population dynamics and data collection programs for humpback chub in the Grand canyon. We first describe in detail the life-history trajectory of an individual fish: the survival probability, growth, and capture history and provide the documented code to implement this process. I then describe the data structures that resemble a databased of individual capture histories for both tagged and untagged fish. The individual based model (IBM) is then used to estimate the fate and capture histories of a known number of recruits starting life at a 40mm length interval.

The simulation model was constructed using R (?). The algorithm populates three matrixes that contain the total number of fish captured in year t at length interval x, the total number of newly marked fish, and a matrix of the total number of recaptured individuals. At each time step the fish is alive, information on length and age is stored along with information on capture history, the tag number if the fish was large enough to tag.

B.2 R-code

```
## ---------------------------------------------------------------- ##
## Simulation  model  for  humpback  chub                           ##
## Author:  Steven  Martell                                         ##
## DATE:  September  26,   2011                                     ##
## Contact:  martell.steve@gmail.com                                ##
##                                                                  ##
## FUNCTIONS:                                                       ##
##   - .ibm => given N recruits returns list with the capture       ##
##             history of each recruit. It also fills the           ##
##             global variables CM and R for the assessment         ##
##             model.                                               ##
##   - .runIBM Loops over syr:nyr and applies the .ibm function     ##
##             to each of Rt recruits.                              ##
##                                                                  ##
##                                                                  ##
## ---------------------------------------------------------------- ##


## -- Required  libraries  -------------------------------------- ##
require( PBSmodelling )
```

```r
require( Hmisc )
require( ggplot2 )


## -- Simulation controls ---------------------------------- ##

#   Model Dimensions
syr        <- 1950       #initial year
nyr        <- 2011       #final year
dt         <- 3/12       #time step
tmax       <- 50/dt      #max longevity index (50 yrs)
dyr        <- seq(syr, nyr, by=dt)
xbin       <- seq(3, 55, by=1)

#   Growth (units in cm)
lmin          <- 4
lmin.sig      <- 0.4
m.linf        <- 0.08
linf          <- 40    #Asymptotic length (cm)
k             <- 0.18 #Growth coefficient

#   Selectivity
lh            <- 10    #length @ 50% selectivity
gamma         <- 1.5   #std in selectivity

sample.yrs  <- seq(1989, 2011, by=dt)
fishing     <- dyr %in% sample.yrs
min.size    <- 15.0 #minimum size for tagging (15 cm)



#   Parameter vector for the IBM model
THETA <- list( lmin=lmin,
               lmin.sig=lmin.sig,
               m.linf=m.linf,
               lh=lh,
               gamma=gamma )

set.seed(991)
iclr <- rev(topo.colors(length(syr:nyr),0.5))
Rt<-floor(rlnorm(length(syr:nyr), log(300), 0.9))
h1 <- 0.04
```

```r
h2 <- 0.20
t  <- min(sample.yrs)
T  <- max(sample.yrs)
pr <- 1
Et <- h1+(h2-h1)*cos(0.5+(sample.yrs-t)/(T-t)*pr*pi)^2
#Et<-runif(length(sample.yrs), 0.05, 0.2)  #Effort


#  Global objects for storing data
fish.id     <- 0     #Unique id for individual fish
tag.id      <- 0     #Unique tag no. for individual
#  Total catch, marked, recap, unmarked by year at length interval
C   <- matrix(0,nrow=length(sample.yrs), ncol=length(xbin))
M   <- matrix(0,nrow=length(sample.yrs), ncol=length(xbin))
R   <- matrix(0,nrow=length(sample.yrs), ncol=length(xbin))

## -------------------------------------------------------------- ##




## -------------------------------------------------------------- ##
## PROCEDURES                                                     ##
## -------------------------------------------------------------- ##


.ibm <- function( Nt, jyr, THETA )
{
    #ARGUMENTS:
        #Nt    = Number of recruits from cohort
        #jyr   = brood year (year the cohort was age-0)
        #THETA = a list of simulation parameters for the ibm.

    #RETURNS:
        #df    = a list w records of capture date, len,  tag.no

    with(THETA, {
        #jyr <- syr+(ii-1)
        ## indexs i = individual; j=time

        ## Calculate length & mortality for time j.
        lj <- linf*(1.-exp(-k*dt*1:tmax))
```

```r
mj <- m.linf*linf/lj


## Draw initial recruit size (gamma).
scale   <- (lmin.sig)^2/lmin
shape   <- lmin/scale
li <- rgamma(Nt, shape, 1/scale)



## Growth trajectory
growth <- function ( li )
{
    linf.i <- rnorm(1, linf, 0.1*linf)
    li+(linf.i-li)*(1-exp(-k*1:tmax*dt))
}
L <- sapply(li, growth)

## Survival trajectory
## S is a list of lengths of individuals at
## the time they were alive.
survival <- function ( lj )
{
    sj <- exp(-m.linf*linf/lj*dt)
    pj <- rbinom(tmax, 1, sj)
    nj <- which.min(pj)
    return(lj[1:nj])
}
S <- apply(L, 2, survival)



## Capture history
pcap <- function ( lj )
{
    #Arg: lj is the length at times j
    #Algorithm:
    #-1. determine index for sampling time periods (fyr)
    #-2. determine capture probability at sampling times
    #-3. if captured fish > 150 mm then assign a tag no.
    if(!is.null(lj))
    {
        # 1. Index for sampling time periods (fyr)
        n   <- length(lj)
        d1  <- which(dyr==jyr)  #min index of dyr for jyr
```

```r
d2    <- min(d1+n-1, length(dyr))
jj    <- d1:d2
iyr <- dyr[jj]              #years  alive
fyr <- fishing[jj]         #years  fished  fyr==1


## Did not survive into sampling years
if(length(iyr[fyr])==0) return(NULL)

# -2. Selectivity at time of sampling
len <- (lj[1:length(jj)])[fyr]
iyr <- iyr[fyr]
# TODO Add annual effort to sj capture probability
eyr <- findInterval(iyr, sample.yrs)
#print(cbind(sample.yrs[eyr],Et[eyr]))
sj  <- Et[eyr]*plogis(len, lh, gamma)

cj  <- rbinom(length(sj), 1, sj)
rid <- which(cj==1)  #row index for capture
if(length(rid)==0) return(NULL) #never captured

tmp <- c(iyr[rid], len[rid],
    tag.no=rep(NA,length=length(rid)))
tmp <- matrix(tmp, ncol=3)


# -3. Assign tag.id if greater than 150 mm
if(max(tmp[,2])>=min.size)
{
    tag.id <<- tag.id + 1
}
tmp[len[rid]>=min.size, 3] <- tag.id

# -4. Populate global variables with catch history
tt <- findInterval(tmp[,1],sample.yrs)
xx <- findInterval(tmp[,2],xbin)
for(i in 1:length(tt))
    C[tt[i], xx[i]] <<- C[tt[i], xx[i]] + 1

#print(cbind(tmp,sample.yrs[tt],xbin[xx],tmp[,3]))

# - Newly marked fish into matrix M
```

```
                    im    <- !is.na(tmp[,3])
                    if(sum(im)>0) #fish was large enough to tag
                    {
                        it <- min(which(im))
                        iy <- findInterval(tmp[it,1],sample.yrs)
                        ix <- findInterval(tmp[it,2],xbin)

                        #print(cbind(tmp, sample.yrs[iy],xbin[ix]))
                        M[iy, ix] <<- M[iy, ix] + 1

                        ir <- which(im)[-1]
                        ry <- findInterval(tmp[ir,1],sample.yrs)
                        rx <- findInterval(tmp[ir,2],xbin)
                        for(i in 1:length(ry))
                            R[ry[i], rx[i]] <<- R[ry[i], rx[i]] + 1
                    }

                    return(tmp)
                }
            }
            P <- sapply(S, pcap)
            return(P)
        })
}

.runIBM <- function()
{
    j <- 1
    A <- NULL
    for(i in syr:nyr)
    {
        A <- c(A, .ibm(Rt[j], i, THETA))
        j <- j+1
    }
    return(A)
}

.writeLSMRdata <- function()
{
        # This function writes the data file for LSMR model
        fn <- "simLSMR.dat"
        C  <- cbind(floor(sample.yrs),sample.yrs%%1*4+1,C)
```

```
            M  <- cbind(floor(sample.yrs),sample.yrs%%1*4+1,M)
            R  <- cbind(floor(sample.yrs),sample.yrs%%1*4+1,R)

            write("#Simulated_data_from_HBCsim.R", fn)
            write("#Model_Dimensions", fn, append=T)
            write("#Years,_timestep", fn, append=T)
            write(c(range(sample.yrs), dt), fn, append=T)
            write("#Array_dimensions(C,_M,_R)", fn, append=T)
            write(dim(C), fn, append=T)
            write("#Number_of_length_intervals", fn, append=T)
            nbin=length(xbin)+1
            write(nbin, fn, append=T)
            write("#Length_intervals_(cm)", fn, append=T)
            write(c(xbin,xbin[nbin-1]+diff(xbin[(nbin-2):(nbin-1)])), fn, append=

            write("#Catch_at_length_by_period", fn, append=T)
            write("#Year,_Period,_Count_...", fn, append=T)
            write.table(C, fn, append=T, col.names=F, row.names=F)
            write("#Number_Marked_at_length_by_period", fn, append=T)
            write("#Year,_Period,_Count_...", fn, append=T)
            write.table(M, fn, append=T, col.names=F, row.names=F)
            write("#Number_Recaptured_at_length_by_period", fn, append=T)
            write("#Year,_Period,_Count_...", fn, append=T)
            write.table(R, fn, append=T, col.names=F, row.names=F)

            write("#End_of_file", fn, append=T)
            write(999, fn, append=T)

            file.copy(fn, "../ADMB/srcLSMR/simLSMR.dat", overwrite=TRUE)
}

## ----------------------------------------------------------------- ##
## MAIN                                                              ##
## ----------------------------------------------------------------- ##
#A  <- .ibm(1000, 1980, THETA)
A <- .runIBM()
.writeLSMRdata()


## ----------------------------------------------------------------- ##

par(mfcol=c(6, 4), las=1, mar=c(2, 2, 1, 1), oma=c(3, 3, 1, 1))
for(i in seq(1,length(sample.yrs)-1/dt,by=1/dt))
```

```
{
        ut <- colSums(C[i:(i+3),]-M[i:(i+3),]-R[i:(i+3),])
        mt <- colSums(M[i:(i+3),])
        rt <- colSums(R[i:(i+3),])
        tmp <- rbind(ut, mt, rt)

    barplot(tmp,names.arg=xbin,xlab=""
    ,ylab="", main=sample.yrs[i], col=1:3)
}
mtext(c("Length (cm)","Frequency"),
    c(1, 2), outer=T, las=0, line=1)
```