Tanner Crab Appendix C
22_02: Revised Input NMFS Survey Sample Sizes from Boostrapping

William T. Stockhausen (AFSC, NMFS)

11 May, 2022

Contents

Introduction

The Tanner crab assessment model currently uses the multinomial probability distribution to describe the likelihood associated with size composition data from the EBS shelf survey (Stockhausen, 2021). The precision or, in contrast, the uncertainty, associated with a sampled multinomial distribution is characterized by the sample size: the higher the sample size, the more precisely an observed size composition reflects the true composition. When the population being observed is sampled in a truly random fashion, so each individual in the population has an equal chance of being sampled, the sample size is equal to the number of samples taken (i.e., individuals measured). However, because crab (and groundfish) in the EBS are patchily distributed in groups of similar size, the assumption of random sampling is violated in the EBS shelf survey and the number of crab measured in the survey overestimates the precision of the resulting size compositions. In trawl surveys like the EBS shelf survey, in which a population is sampled in a two-step process by taking a number of hauls distributed across the range of the population, with size (and other biological characteristics such as sex, maturity state, and shell condition) measurements taken on a number of individuals caught in each haul, the "true" sample size that reflects the precision/uncertainty of the observed size composition typically varies between the total number of animals measured and the number of hauls in which animals were captured and measured (hauls in which no animals were captured and measured are important in determining the extent and size of the population being sampled, but contribute nothing to understanding the distribution of sizes within the population). For Tanner crab, then, the "true" sample sizes for annual size compositions from the EBS shelf survey can range

between 10's-100's (hauls at which crab were found and measured) and 1,000's (numbers of crab measured), depending on sex, maturity state, and shell condition.

Because the Tanner crab is an integrated assessment model that seeks to balance model fits to data across a variety of data types and sources (e.g., indexes of abundance and biomass, survey and fishery size compositions, growth and maturity data) in order to estimate underlying parameters governing population dynamics, reflecting fishery prosecution, and addressing management issues such as sustainable yield, it is important to accurately characterize the precision/uncertainty associated with the those data. The relative weights assigned to different data sources in the parameter estimation algorithm can influence the values of the estimated parameters, particularly when different data sources encompass conflicting information. The Tanner crab model uses a maximum likelihood approach to parameter estimation with input sample sizes for survey size compositions fixed at 200 per sex in all years, where the value of 200 reflects a practical choice for the assumed precision of the size compositions that allows the estimation algorithm to converge. In past work to improve on using a fixed sample size when fitting the model, several iterative re-weighting schemes were tried but did not produce reasonable results. The Dirichlet-Multinomial distribution (Thorson et al. 2016), as a substitute for the multinomial, allows for the estimation of the "effective" sample size as a parameter with the input sample size as an upper bound. Application of this distribution in previous models resulted in the "effective" sample size being estimated as the same as the input sample size, suggesting the default input sample size of 200 may not be a bad practical choice.

To investigate this issue more thoroughly, a "bootstrapping" technique (Stewart and Hamel, 2014) was applied to EBS survey data to estimate annual input sample sizes for survey size compositions for male Tanner crab, as well as immature and mature females. The resulting bootstrapped sample sizes were then used as input sample sizes to re-fit the 2021 assessment model (referred to here as 22.02) and compare changes in results with the original (21.22a).

Bootstrapping

Two-stage bootstrapping as described in Stewart and Hamel (2014; hauls within survey strata, individuals within hauls) was applied to Tanner crab EBS survey data on an annual basis for the entire survey dataset (1975-2021) to derive effective sample sizes for male, immature female, and mature female size compositions (Tables 1, 2; Figures 1-4). The average number of males measured in the survey was 6292, while the average number of stations at which crab were found was 215. The arithmetic mean effective N was 394, while the average harmonic mean effective N was 266. The average (across all years) relative reduction from the measured number of males to the harmonic mean effective N ($N_{measured}/N_{harmonic}$) was 96. For immature females, the average number measured in the survey was 2300, while the average number of stations at which crab were found was 215. The arithmetic mean effective N was 145 and the average harmonic mean effective N was 266, while the average relative reduction from the measured number of immature females to the harmonic mean effective N was 96. The average number of mature females measured in the survey was 1698, while the average number of stations at which crab were found was 126, the arithmetic mean effective N was 234, and the average harmonic mean effective N was 129. The average relative reduction from the measured number of mature females to the harmonic mean effective N was 92.

Stewart and Hamel (2014) examined several estimators for bootstrapped effective N's and found the harmonic mean to have the least bias ($< 1\%$) and subsequently recommended its use. The average harmonic mean effective N's over the survey time series (82 for males, 395 for females) agreed surprisingly well with the default value of 200 used to fit sex-specific survey size compositions in the

assessment model. However, the harmonic effective mean varies appreciably (on the order of 30%) over the time series (as do the arithmetic average effective N, the number of non-zero hauls, and the numbers measured), so using annual values of the harmonic mean effective N as input sample sizes for survey size composition data in the assessment model may better reflect the associated variability in precision (or uncertainty, conversely) of the data compared with using a fixed value.

References

Stewart, I.J. and O.S. Hamel. 2014. Bootstrapping of sample sizes for length- or age-composition data used in stock assessments. Can. J. Fish. Aquat. Sci. 71: 581–588 (2014) dx.doi.org/10.1139/cjfas-2013-0289

Stockhausen, W. 2021. 2021 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. In: Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands: 2021 Final Crab SAFE. North Pacific Fishery Management Council. Anchorage, AK. https://meetings.npfmc.org/CommentReview/DownloadFile?p=acbf6b6c-18ba-4b1a-abb5-84a87cfbaea3.pdf&fileName=3%20Eastern%20Bering%20Sea%20Tanner%20Crab%20SAFE.pdf

Thorson, J.T., K.F. Johnson, R.D. Methot, and I.G. Taylor. 2016. Model-based estimates of effective sample size in stock assessment model using the Dirichlet-multinomial distribution. Fisheries Research. http://dx.doi.org/10.1016/j.fishres.2016.06.005.

List of Tables

List of Figures

[1] TRUE

Tables

Table 1: Size composition sampling information for Tanner crab in the NMFS EBS shelf survey, 1975-1999. measured: number of crab measured; non-0 hauls: number of hauls with measured crab; avg(N): arithmetic mean effective sample size; har(N): harmonic mean effective sample size

	female								male			
	immature				mature				all			
year	measured	non-0 hauls	avg(N)	har(N)	measured	non-0 hauls	avg(N)	har(N)	measured	non-0 hauls	avg(N)	har(N)
1975	1,047	73	72	33	2,567	95	275	150	7,287	127	253	161
1976	1,097	88	62	31	1,615	96	174	113	4,734	132	262	167
1977	776	69	48	24	1,921	83	438	187	4,234	117	154	95
1978	1,949	88	89	38	1,945	103	220	120	5,227	158	361	238
1979	429	43	73	34	597	51	75	30	1,829	110	133	77
1980	1,491	103	118	66	2,041	108	193	65	7,530	175	517	314
1981	579	71	89	55	2,525	122	115	55	6,988	182	732	480
1982	823	87	109	70	2,841	129	157	63	5,204	202	610	437
1983	2,113	102	113	53	2,355	115	264	132	4,648	187	225	137
1984	1,879	135	147	80	1,815	107	173	77	3,854	184	328	196
1985	847	141	125	82	829	91	90	46	1,900	188	288	225
1986	1,588	162	110	67	522	107	145	95	3,137	228	216	143
1987	4,230	189	165	90	837	129	180	107	6,463	229	334	199
1988	3,735	206	230	144	2,283	169	272	148	8,312	253	340	216
1989	3,271	204	118	76	2,123	170	279	148	9,245	243	241	155
1990	3,114	198	159	88	3,013	178	403	253	9,598	253	503	357
1991	2,259	163	115	64	3,851	174	362	174	9,946	241	443	304
1992	1,494	107	146	96	3,025	167	388	222	6,929	231	483	353
1993	869	99	112	69	1,882	155	384	233	5,593	230	665	494
1994	921	97	112	59	1,441	120	327	198	3,832	213	495	398
1995	834	115	90	61	1,197	116	235	139	2,789	191	320	232
1996	883	115	101	66	1,072	125	197	124	2,705	190	264	184
1997	1,329	116	201	102	672	111	246	169	2,207	195	251	160
1998	1,710	146	195	115	504	96	195	124	3,052	195	344	244
1999	2,628	138	185	108	765	105	223	134	3,933	186	189	109

Table 2: Size composition sampling information for Tanner crab in the NMFS EBS shelf survey, 2000-2021. measured: number of crab measured; non-0 hauls: number of hauls with measured crab; avg(N): arithmetic mean effective sample size; har(N): harmonic mean effective sample size

	female								male			
	immature				mature				all			
year	measured	non-0 hauls	avg(N)	har(N)	measured	non-0 hauls	avg(N)	har(N)	measured	non-0 hauls	avg(N)	har(N)
2000	2,249	142	198	115	587	89	195	122	4,117	206	347	229
2001	3,678	164	159	88	1,008	109	226	136	5,482	227	231	151
2002	3,585	155	138	68	850	105	129	73	5,459	213	233	141
2003	2,834	153	110	63	1,675	128	244	97	7,003	214	309	192
2004	3,922	175	222	118	1,083	124	143	70	7,468	257	370	262
2005	3,352	201	135	83	1,562	129	113	70	7,529	267	264	194
2006	4,364	211	172	104	2,659	180	261	154	12,035	271	471	279
2007	2,430	186	148	95	2,707	185	221	123	9,586	275	328	196
2008	1,747	153	112	74	2,363	167	269	160	7,389	253	722	536
2009	2,408	171	207	116	1,680	140	248	135	5,977	241	561	395
2010	3,180	186	165	103	1,186	126	190	113	6,624	240	400	280
2011	5,044	193	185	118	1,176	137	286	177	9,151	223	348	235
2012	3,611	195	203	114	1,662	144	167	96	8,386	230	388	254
2013	2,917	163	192	105	2,419	157	258	154	9,611	214	474	289
2014	2,211	165	153	88	2,066	148	295	165	10,861	235	722	550
2015	1,455	118	179	115	1,808	115	212	111	7,413	251	827	648
2016	1,373	110	143	81	1,618	100	240	114	7,073	266	636	453
2017	2,033	131	185	62	1,338	118	221	142	6,206	251	451	257
2018	4,666	196	233	115	1,228	120	341	202	8,251	250	449	232
2019	3,810	181	227	136	1,190	106	175	81	5,913	237	387	258
2021	3,015	189	128	54	1,991	148	305	157	6,721	235	239	106

Figures

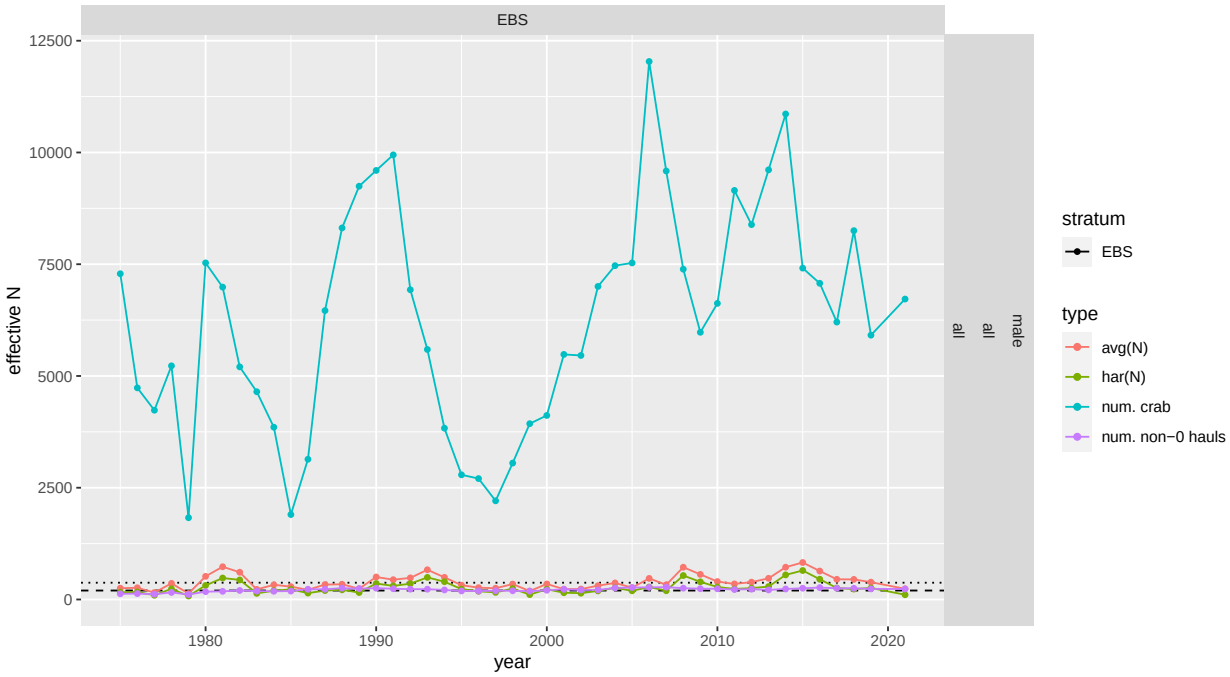

Figure 1: Comparison of numbers measured ("num. crab"), number of hauls with catches ("num. non-0 hauls"), and arithmetic mean effective N ("avg(N)") and harmonic mean effective N ("har(N)") from bootstrapping for male Tanner crab.

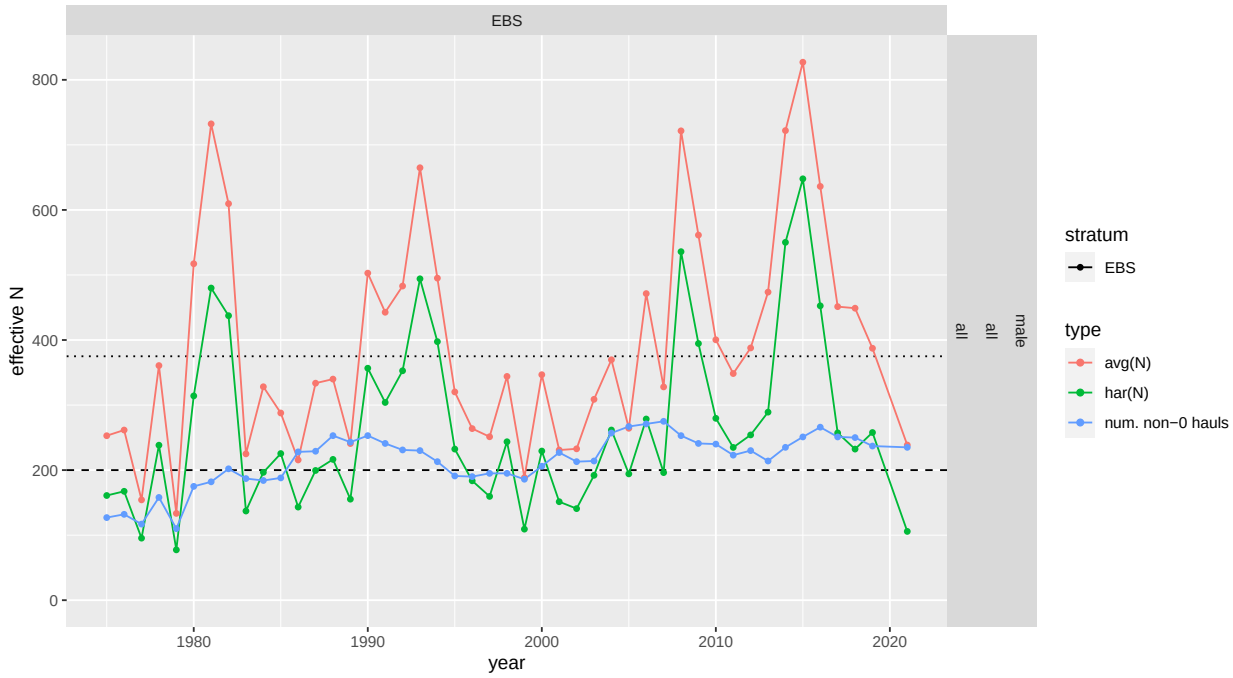

Figure 2: Comparison of number of hauls with catches ("num. non-0 hauls"), and arithmetic mean effective N ("avg(N)") and harmonic mean effective N ("har(N)") from bootstrapping for male Tanner crab.

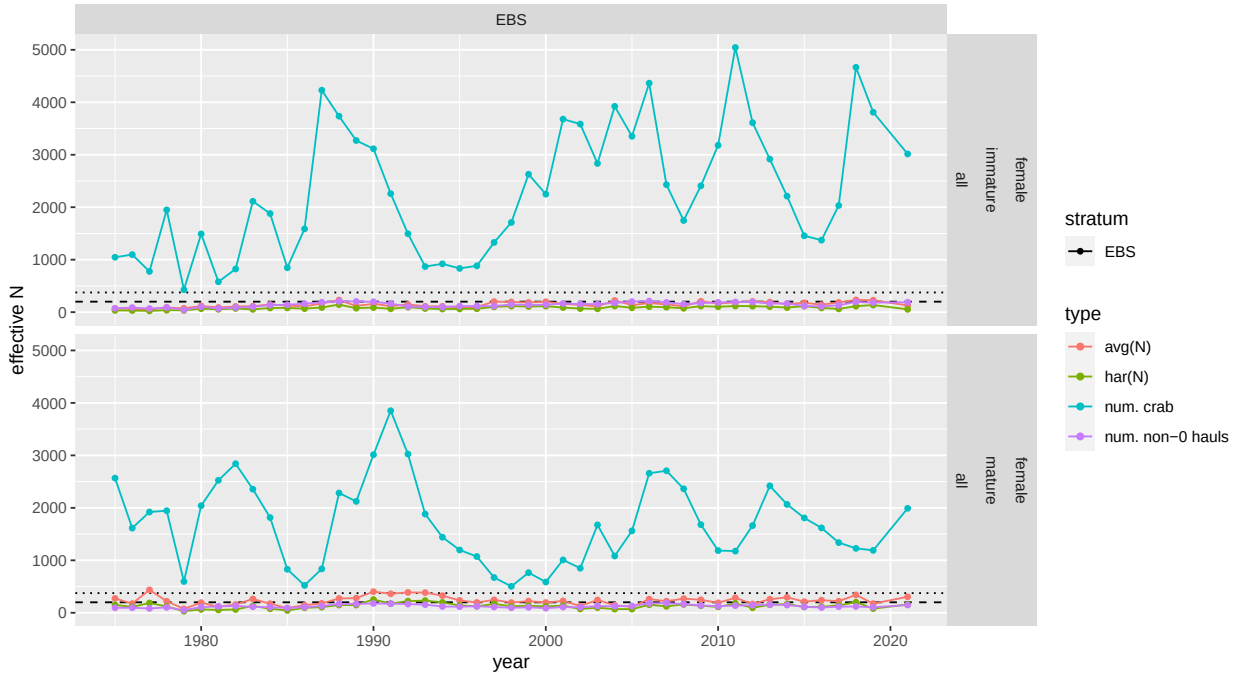

Figure 3: Comparison of numbers measured ("num. crab"), number of hauls with catches ("num. non-0 hauls"), and arithmetic mean effective N ("avg(N)") and harmonic mean effective N ("har(N)") from bootstrapping for female Tanner crab.

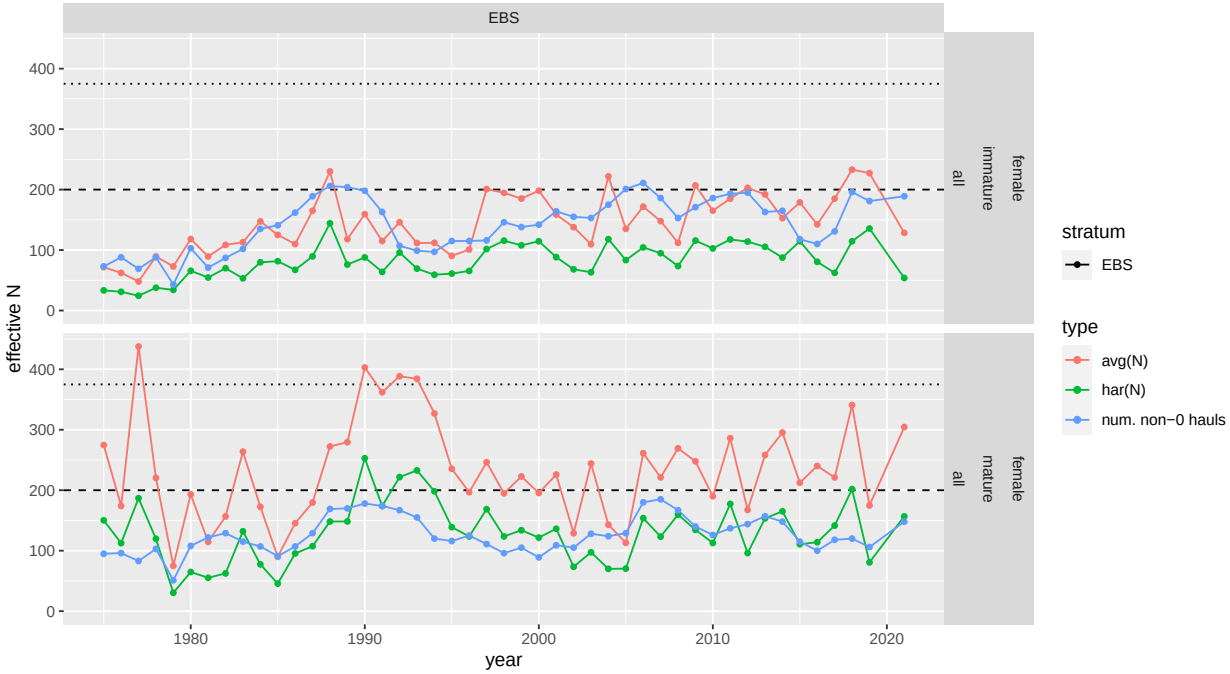

Figure 4: Comparison of number of hauls with catches ("num. non-0 hauls"), and arithmetic mean effective N ("avg(N)") and harmonic mean effective N ("har(N)") from bootstrapping for female Tanner crab.