A Preliminary Analysis of Virginia Shark Long-line Survey Data 1974-2004.

Daniel S. Ha and John A. Musick
LCS Evaluation Workshop 2005
Panama City, FL

DO NOT CITE WITHOUT PERMISSION

Field Methods

Elasmobranchs for this study are collected by bottom long-line sampling starting in 1974, and ending (for this study) in 2004. Long-lines are fished once a month May through October at each of seven standard stations. However, varying levels of support and changing research goals led to certain years being under-sampled. In addition, weather or vessel constraints did prevent sampling certain stations or certain months.

The VIMS bottom long-line survey is performed with gear standard to the industry at the inception of the study, and the gear has remained the same throughout the survey. They are virtually identical to this date (2004) to those described by Musick et al. (1993), but they will be reiterated here. The long-line itself is a 6.4 mm (1/4 inch) hard-laid and tarred nylon mainline anchored at both ends with 3-5 m gangions spaced at approximately 20 m intervals. Buoys are set at twenty-gangion intervals, and ends are marked with radar reflectors raised approximately 3 m above sea level. Standard gangions used are of the type termed "Yankee gangions," these being a heavy-duty quick-snap (also termed a tuna clip) with an 8/0 swivel, 1-3 m of 3 mm (1/8 inch) hard-laid and tarred nylon line, and 8/0 swivel connecting 1-2 m of 1.6 mm (1/16 ") 1X7 or 7X7 stainless steel wire, and a 9/0 hook. Musick et. al (1993) reported that sonar surveys of the long-line indicated it dropped to no deeper than 80 m, and thus, for most stations, it targeted semi-demersal species at or near the bottom. Soak times in this survey range from 2 to 17 hours, with the majority 3-4 hours long. Bait used include many coastal teleost fishes such as croaker, spot, menhaden, bluefish and mackerel. Surveys since 1998 have used entirely menhaden as bait. Bait pieces are 0.10 to 0.25 kg each.

Statistical Methods

Since the data set is not heterogeneous in terms of data collected, and given the number of instances in which a species is not collected in a set of the long-line, several different techniques for analysis of these data are required. First, however, the variables to be analyzed and the constraints they place on the form of analysis must be described.

This long-line survey collects several types of data, and some of these vary over the course of the study. Data about the set is recorded, and can include date, number of hooks, location, time of deployment of the gear (set time), time of retrieval of the gear (haul time), and the duration of the set (soak time). Physical data are also collected, and can include bottom depth, surface temperature, and bottom temperature. Data about the catch, if any, may include species, pre-caudal length, fork length, total length, sex, and gonad state.

Not all of these data are present for each set, however, leaving gaps in the data. Out of 957 recorded stations, 1974-2003, 201 are missing bottom temperature data and 151 are missing haul time data, to cite the worst. The missing data is skewed to the earlier years, making them less available to analysis. Analyses in the following chapters exclude catch records for sets missing variables being analyzed, and thus the n varies with the number and type of effects in the individual model. Later, in the 1990 field season, other gangion/hook types were deployed on sets with standard gangions for a hooking efficiency study (Branstetter & Musick 1993). In the 1993 field season, CTD casts were made for each station, adding salinity and dO data, and increasing the temperature data available. In addition, in this field season, the systematic recording of batoids caught was begun.

The main dependent variable that is available for analysis since the inception of this study is the number of sharks caught per long-line set per hour of soak time, which is defined as CPUE for this study. Catch per unit effort is available as a total for all species, or broken down by species, or further to species age groups. The cpue data can also be aggregated in various ways, such as by NMFS management groups

Data from a 1961 shark long-line survey using virtually identical gear as the current VIMS long-line survey have been made available by NMFS. Data were obtained through a long-line survey off the coats of New Jersey. The long-line gear used in this survey was identical to current VIMS long-line gear in every way except the hooks used (Musick, personal communication). The 1961 survey used a Japanese tuna hook, and these were used in our resampling program. In the summer of 2005, a majority of the stations sampled in 1961 was resampled by the VIMS longline survey. Comparison of the samples from these two dates may provide valuable insights and possible corroboration for the findings of the Virginia long-line survey.

Methods:

Before analyses could begin, the requirements of the statistical methodologies had to be met. The preferred statistical method for this analysis was a generalized additive model (GAM) performed with the SAS/STAT statistical package (SAS Institute 2002). I transformed the data with an arc-sine/root transform (also know as the angular transform), the most appropriate transform for proportion data (Sokal and Rohlf 1995).

Some adjustments to the data were made in order to correct sampling problems. First, since only standard gangions have been used throughout the course of the long-line survey, only sharks caught on this gear were included in this analysis. Only standard stations sampled from the beginning of this survey were included in the following analyses.

One limitation of GAM models is that they can be limited in their cross-product terms, also termed interaction effects. The statistical program used for these analyses (Proc GAM: SAS Institute 2002) does not allow such cross-product terms, possibly due to the greater computational resources such effects would require. Venables and Dichmont (2004) discussed this issue, and concluded that the best way to handle such cross-product terms, short of including them in the model, is to choose variables that one would not expect interactions between to be very large. One way in which I've had to do this in this analysis is to try to eliminate any interaction effect of stations and years. Especially early in the program, funding and logistics led to some standard stations being under-represented in some years. For this reason, some of these under-represented years were combined with adjacent years into year categories, so that each standard station is as equally represented as possible in each year category. These year categories are not necessarily of equal intervals, and were represented in analyses as a continuous independent variable by their midpoints. Figure 1 shows the Shannon-Weiner diversity index (H) for standard stations by year. To calculate this, I substituted stations for species in the formula, and number of sets at each station for the number of individuals of each species. This modified index I calculated for each year of the survey. The figure shows clearly the unevenness in coverage of stations over the years. Figure 2 shows the results in modified index values of combining years with low coverage of standard stations with adjacent years so as to increase station diversity, while having as little impact on resolution as possible. The figure shows a much more even distribution of stations over the year groupings. The continuous year term in the model was then represented by the mean year of all stations in that year category.

Thus the year category 1974-76 is 1974.85, 1978-79 is 1978.6, 1981-84 is 1982.04, 1986-89 is 1987.39, and 1993-94 is 1993.11, while all others are simply the year. For simplicity, I refer to year categories by their mean year in the following.

Bottom temperature was haphazardly recorded during the course of this study, and Figure 3 shows the number of stations with no bottom temperature data recorded by year. The missing data is skewed towards earlier sets, with only 2 stations having a record bottom temperature before 1980. Due to this pattern of missing data, I excluded bottom temperature from the analyses in this chapter, as it would reduce the power to detect trends early in the time series, and the two early points might have an undue influence on the results.

GAMs allow for parametric and non-parametric curve fitting. I used the SAS/STAT statistical package to perform these analyses (SAS Institute 2002). I fitted a semi-parametric model, modeling the dependent variable against all main effects available since the start of the long-line survey (Year, station, month, mean water depth, time out, time in, and surface temperature). Since month and station are categorical variables, they were fit to parametric terms. I fit a lowess smoothing curve to the other variables, using generalized cross validation to determine degrees of freedom. This model was iterated removing the least significant term (significance defined as $p<0.05$) at each iteration until all terms in the model were significant.

Since the species classified as large coastal species (LCS) have changed over time, I have analyzed both LCS as originally defined (NMFS 1992), which I will refer to as LCS (1992), and as currently defined (NMFS 2003), which I will refer to as LCS (2003).

I analyzed the data for categories that could not be normalized in two ways. First, I transformed it into presence/absence data by changing all positive catches into 1, to represent presence. I analyzed these data with a logistic analysis, using the LOGISTIC procedure in SAS/STAT (SAS Institute 2002) to determine trends. Second, I deleted all sets with catches of 0, leaving only positive catches. I then analyzed these data with a GAM model in SAS/STAT (SAS Institute 2002), as above.

Figures:

Figure 1: The Shannon-Weiner diversity index calculated for standard stations across years, showing unevenness in standard station coverage over the course of this survey.

Figure 2: The Shannon-Weiner diversity index calculated for groupings of standard station sets across years, showing the combinations of years that produces the most even coverage while sacrificing the least resolution in time.

Figure 3: The number of sets with and without recorded bottom temperature data, showing an early lack of bottom temperature data.

Figure 4: Significant Lowess-smoothed trends in transformed cpue data fitted with a GAM model for the large coastal species group and the species within that group. The LCS group is shown both as defined originally in 1992 (NMFS 1992) and as defined currently (NMFS 2003). The current definition excludes the white, sand tiger, bignose and night sharks.

Figure 5: Results of t-tests of species collected in 1961 and 2005 survey of New Jersey stations.

Branstetter S, Musick JA. 1993. Comparisons of shark catch rates on longlines using rope/steel (Yankee) and monofilament gangions. Marine Fisheries Review 55(3):4-9.

Musick JA, Branstetter S, Colvocoresses JA. 1993. Trends in shark abundance from 1974 to 1991 for the Chesapeake Bight region of the U.S. Mid-Atlantic coast. In: Branstetter S, editor. NOAA Technical Report NMFS 115 Conservation Biology of Elasmobranchs. Silver Spring MD: NOAA. p 1-18.

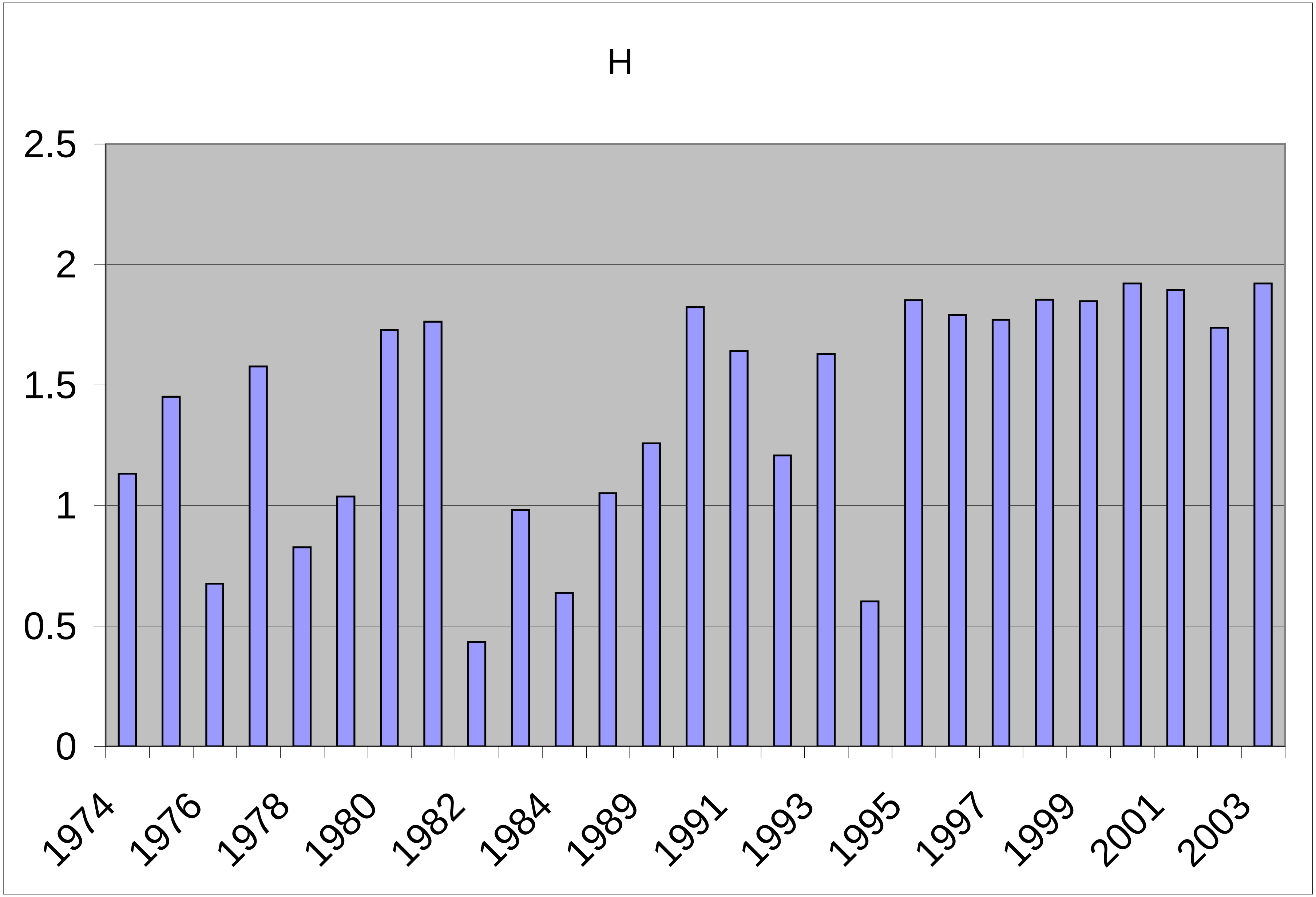
NMFS. 1992. Fishery Management Plan for Sharks of the Atlantic Ocean. Washington, DC: National Marine Fisheries Service. 160 p.

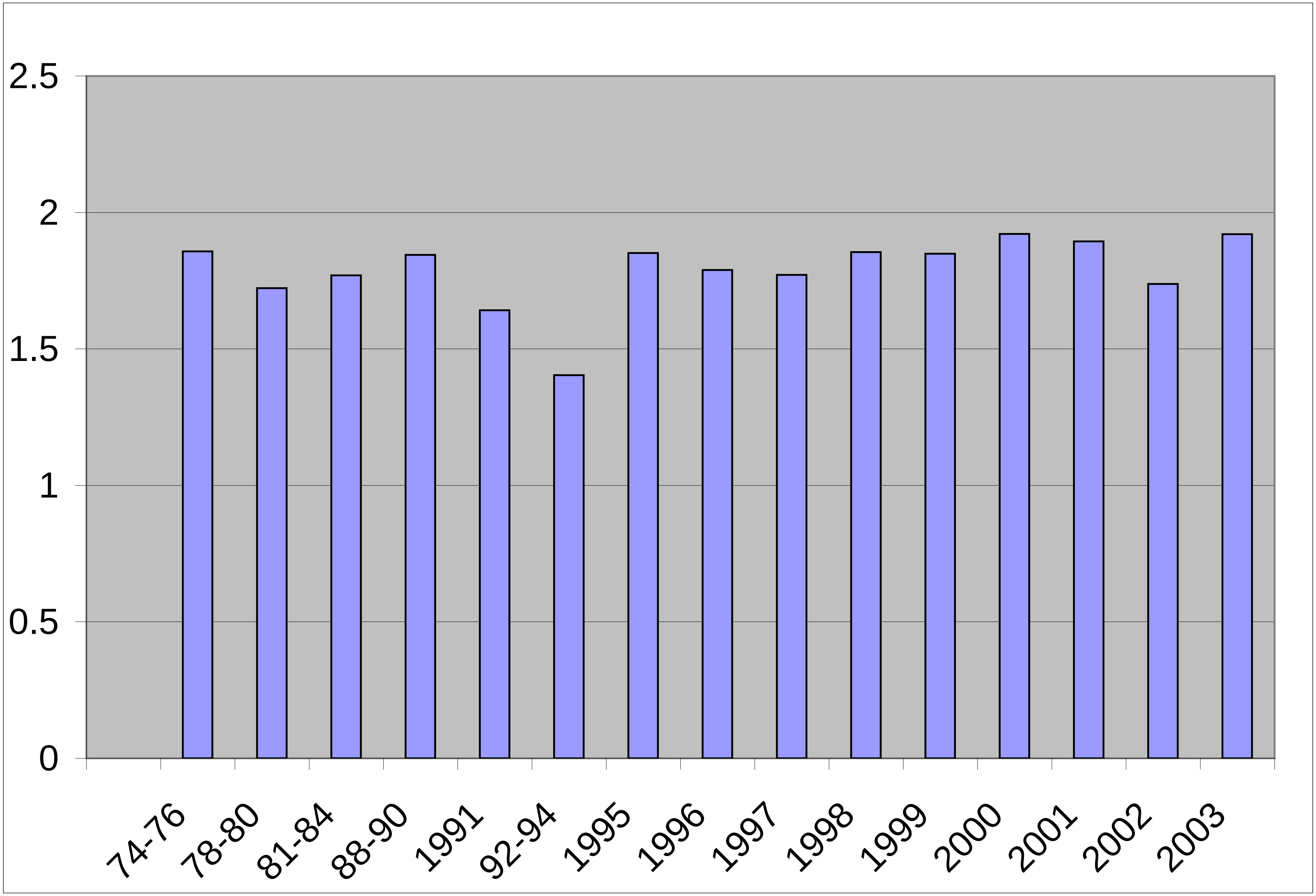
NMFS. 2003. Final Amendment 1 to the Fishery Management Plan for Atlantic Tunas, Swordfish, and Sharks. Silver Spring, Md: National Marine Fisheries Service.

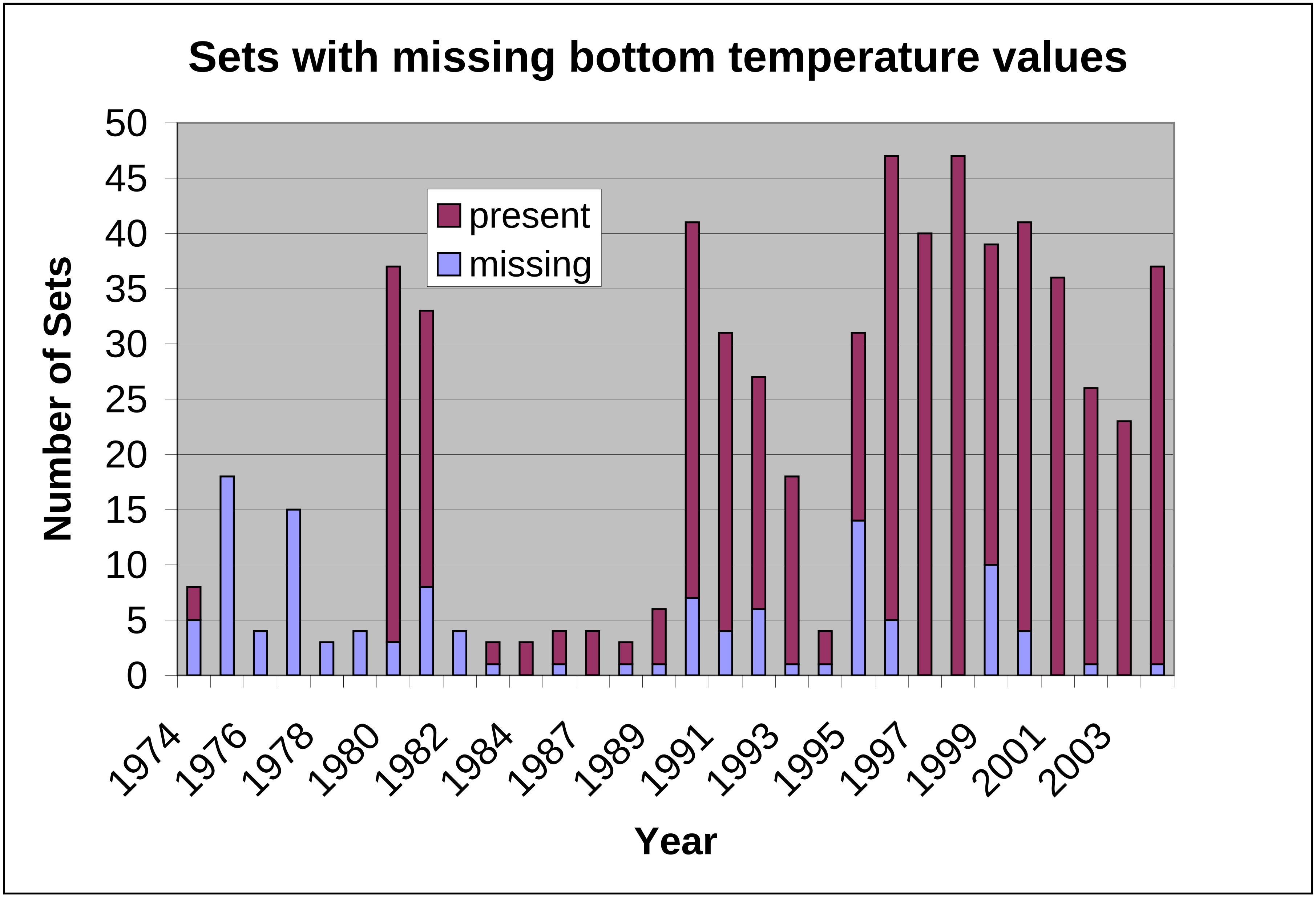
SAS Inc. 2002. SAS/STAT software, Version 9.0 of the SAS System for Windows. Version 9.0. Cary, NC: SAS Institute Inc. SAS and all other SAS Institute Inc. Product or service names are registered trademarks or trademarks of SAS Institute Inc.

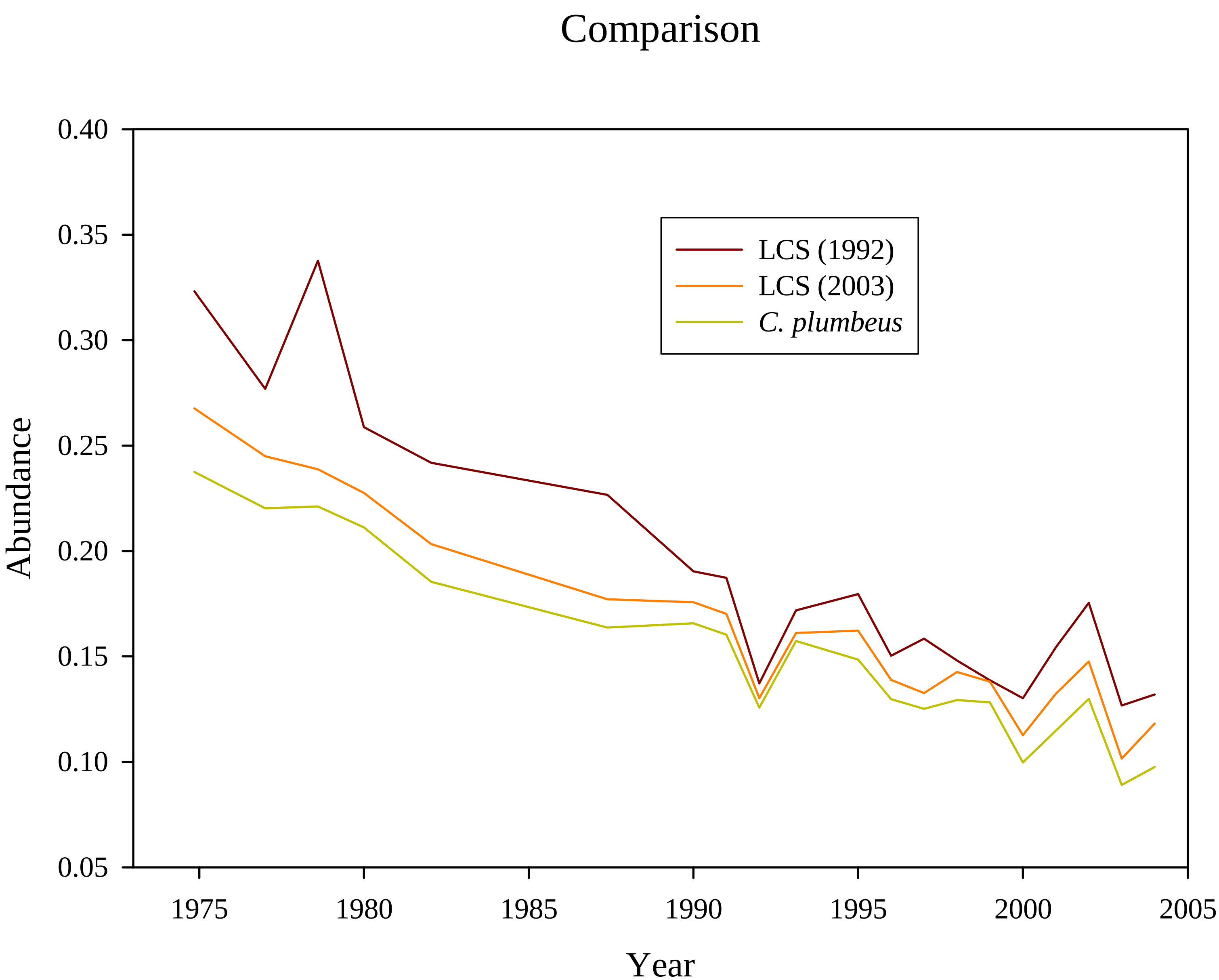
Sokal RR, Rohlf FJ. 1995. Biometry. New York, NY: W. H. Freeman and Company. 887 p.

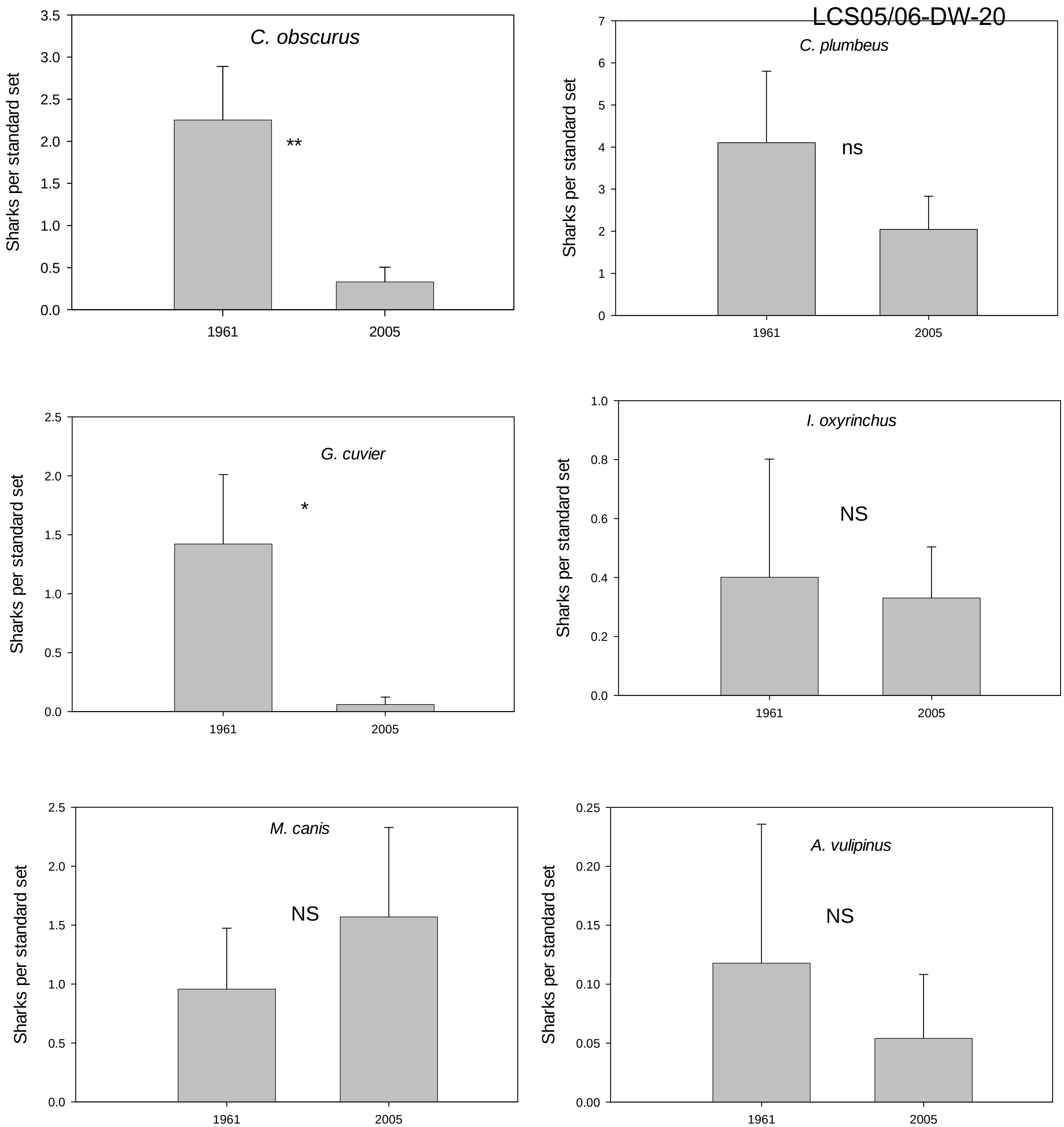
Venables WN, Dichmont CM. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fisheries Research 70:319-337.


H
2.5
2
1.5
1
0.5
0
1974
1976
1978
1980
1982
1984
1989
1991
1993
1995
1997
1999
2001
2003


2.5
2
1.5
1
0.5
0
74-76
78-80
81-84
88-90
1991
92-94
1995
1996
1997
1998
1999
2000
2001
2002
2003


Sets with missing bottom temperature values
Number of Sets
50
45
40
35
30
25
20
15
10
5
0
present
missing
1974
1976
1978
1980
1982
1984
1987
1989
1991
1993
1995
1997
1999
2001
2003
Year


Comparison
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
Abundance
LCS (1992)
LCS (2003)
C. plumbeus
1975
1980
1985
1990
1995
2000
2005
Year


C. obscurus
**
Sharks per standard set
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1961
2005
C. plumbeus
ns
Sharks per standard set
7
6
5
4
3
2
1
0
1961
2005
G. cuvier
*
Sharks per standard set
2.5
2.0
1.5
1.0
0.5
0.0
1961
2005
I. oxyrinchus
NS
Sharks per standard set
1.0
0.8
0.6
0.4
0.2
0.0
1961
2005
M. canis
NS
Sharks per standard set
2.5
2.0
1.5
1.0
0.5
0.0
1961
2005
A. vulipinus
NS
Sharks per standard set
0.25
0.20
0.15
0.10
0.05
0.00
1961
2005

In order to develop standardized indices of annual average CPUE for sandbar shark and LCC, a delta-lognormal model, as described by Lo et al. (1992), was employed. This index is a mathematical combination of yearly CPUE estimates from two distinct generalized linear models: a binomial (logistic) model which describes proportion of positive CPUE values (i.e., presence/absence) and lognormal model which describes variability in only the nonzero CPUE data. The GLMMIX and MIXED procedures (Patetta, 2002) in SAS were employed to provide yearly index values for both the binomial and lognormal sub-models, respectively. A backward stepwise selection procedure was used to develop each sub-model. The parameters tested for inclusion in each sub-model were year, month, station, surface temperature and depth, and separate covariance structures were developed for each survey year. For the binomial models, a logistic-type mixed model was employed for all areas for LCC. The fit of each model was evaluated using the fit statistics provided by the GLMMIX macro. Initially, several model types were used to describe the nonzero CPUE data. These included lognormal, Poisson and negative binomial. Based on analyses of residual scatter and QQ plots, the lognormal model was more fitting than the others in describing the variability in the nonzero data in most of the models. The following tables and graphs summarized the results for sandbar and then the LCC. –Walter Ingram

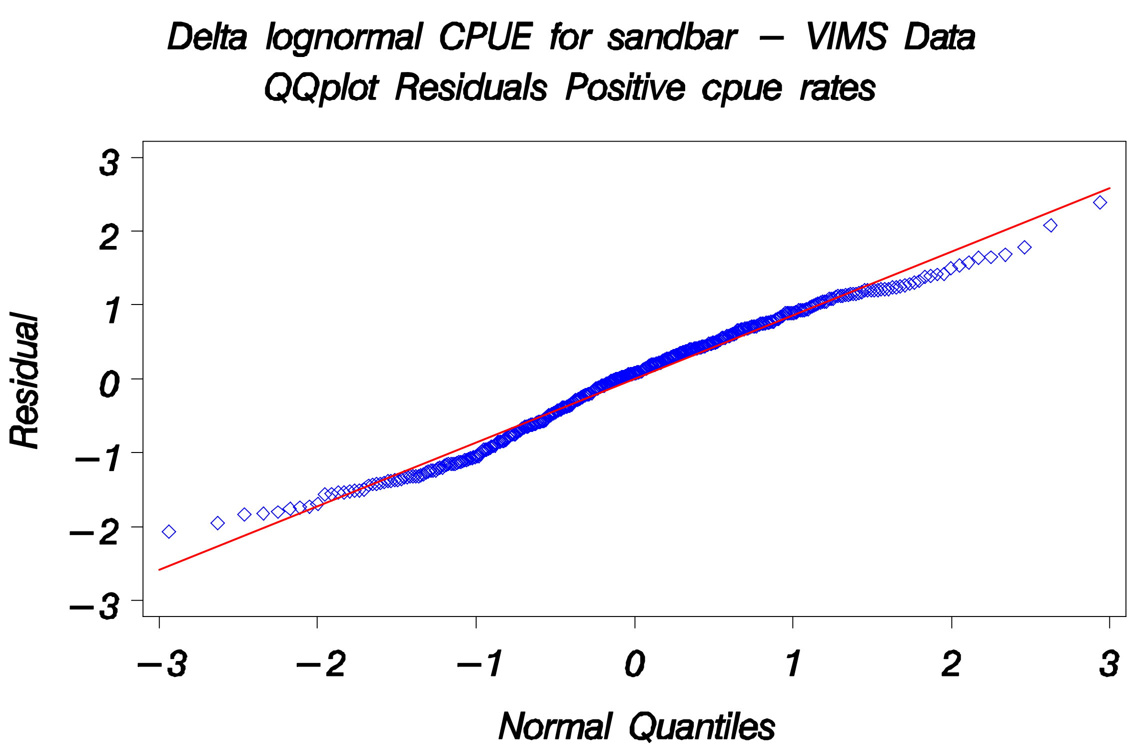
Sandbar Shark

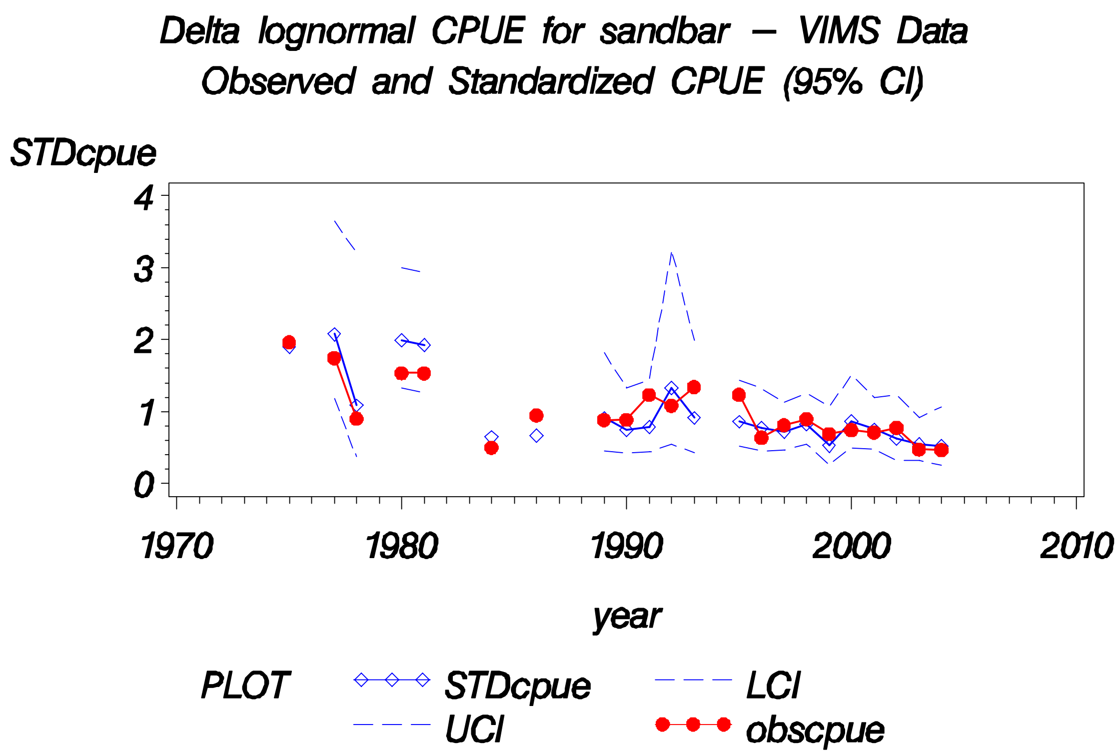
Binomial Model (of occurrence data of sandbar sharks) Variable Results

Type 3 Tests of Fixed Effects						
Effect	**Num DF**	**Den DF**	**Chi-Square**	**F Value**	**Pr > ChiSq**	**Pr > F**
Year	21	49.2	33.78	1.29	0.0382	0.2254
Station	6	450	45.19	7.53	<.0001	<.0001
Surface Temp	1	410	18.47	18.47	<.0001	<.0001

Lognormal Model (of nonzero data of sandbar sharks) Variable Results and QQ Plot of Residuals

Type 3 Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Year	21	345	4.37	<.0001
Station	6	345	7.95	<.0001
Month	5	345	5.12	0.0002





Sandbar Index Output

SurveyYear	Frequency	N	LoIndex	StdIndex	CV	LCL	UCL
1975	0.83333	17	0.019822	1.90034	0.23271	1.20050	3.00817
1977	0.66667	15	0.021667	2.07722	0.28711	1.18310	3.64706
1978	0.66667	3	0.011319	1.08517	0.58275	0.36799	3.20004
1980	0.86486	37	0.020808	1.99493	0.20558	1.32803	2.99671
1981	0.75758	33	0.020080	1.92508	0.21419	1.26035	2.94041
1984	0.66667	3	0.006747	0.64689	1.01363	0.12040	3.47583
1986	0.25000	4	0.006940	0.66536	1.08966	0.11340	3.90388
1989	0.66667	6	0.009499	0.91067	0.35817	0.45456	1.82441
1990	0.67500	40	0.007778	0.74565	0.29514	0.41832	1.32910
1991	0.61290	31	0.008215	0.78760	0.30447	0.43419	1.42866
1992	0.44444	27	0.013879	1.33059	0.46767	0.54670	3.23850
1993	0.55556	18	0.009539	0.91454	0.40248	0.42135	1.98498
1995	0.77419	31	0.008968	0.85983	0.26193	0.51365	1.43932
1996	0.57143	49	0.008029	0.76977	0.27439	0.44910	1.31943
1997	0.65000	40	0.007525	0.72143	0.22527	0.46231	1.12576
1998	0.59574	47	0.008615	0.82593	0.20952	0.54564	1.25020
1999	0.51282	39	0.005506	0.52783	0.36478	0.26031	1.07028
2000	0.51220	41	0.009022	0.86498	0.28108	0.49830	1.50148
2001	0.63889	36	0.007865	0.75404	0.23611	0.47324	1.20144
2002	0.65385	26	0.006529	0.62596	0.34985	0.31725	1.23509
2003	0.60870	23	0.005706	0.54703	0.26489	0.32496	0.92088
2004	0.46154	39	0.005415	0.51918	0.37114	0.25310	1.06498

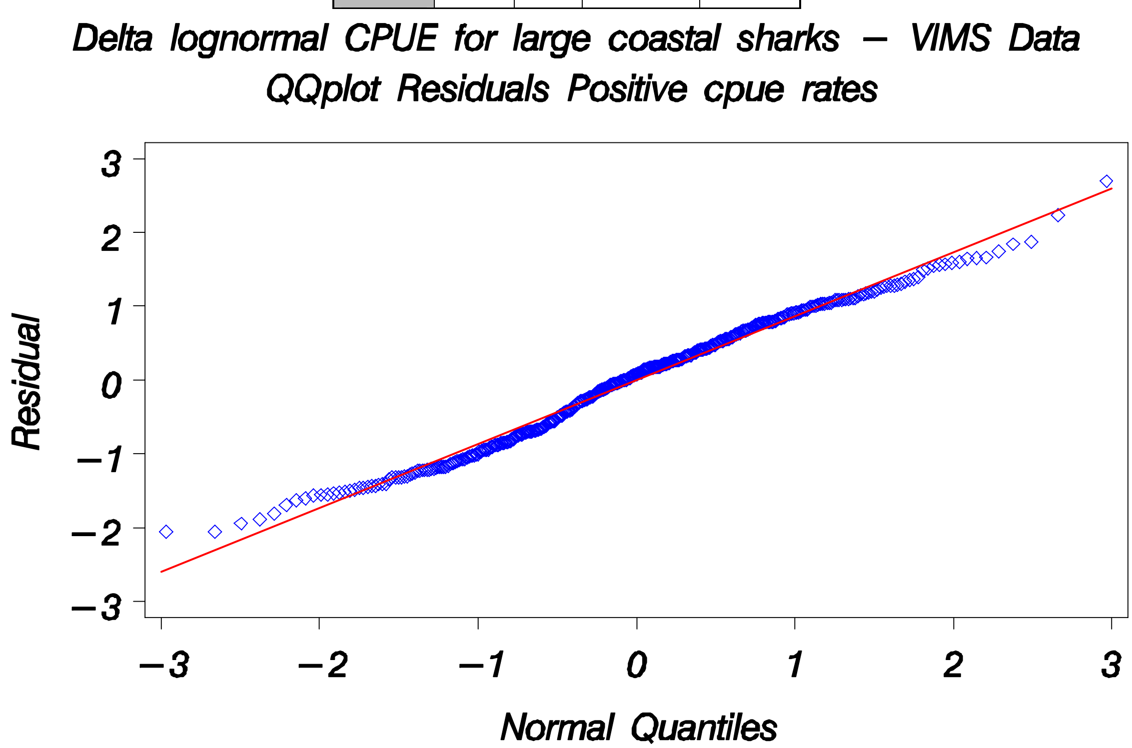
Large Coastal Sharks

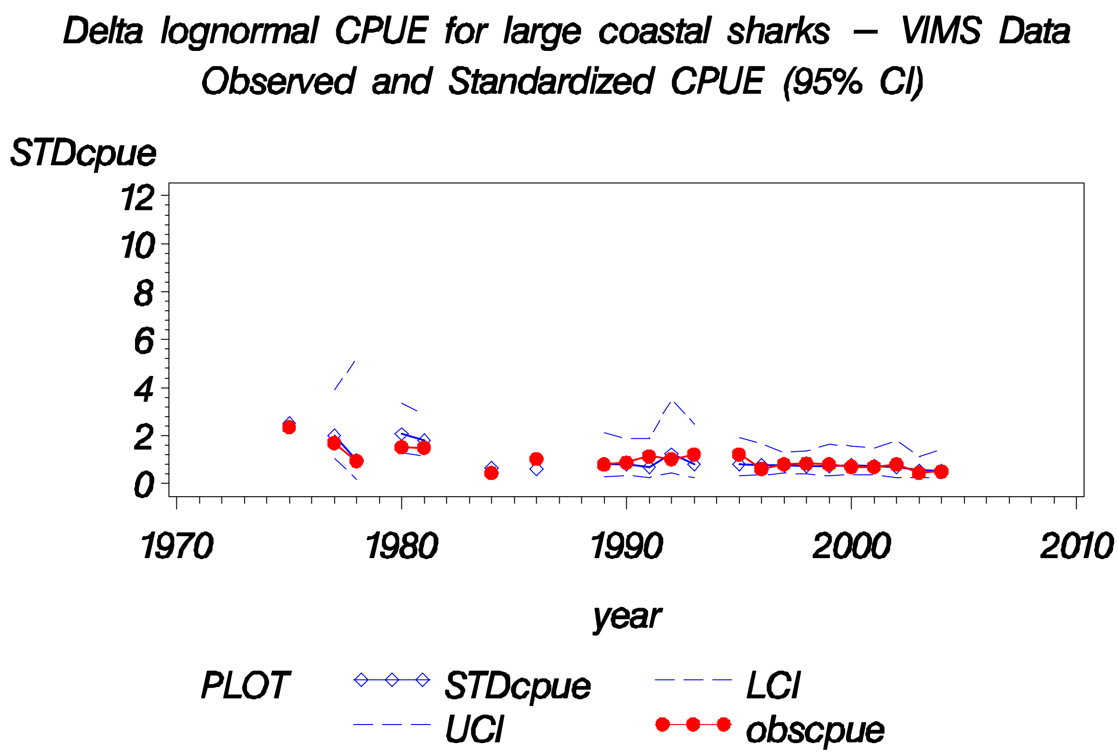
Binomial Model (of occurrence data of large coastal sharks) Variable Results

Type 3 Tests of Fixed Effects						
Effect	Num DF	Den DF	Chi-Square	F Value	Pr > ChiSq	Pr > F
Year	21	49.5	27.96	1.07	0.1413	0.4056
stn	6	449	18.90	3.15	0.0043	0.0049
st	1	399	39.44	39.44	<.0001	<.0001

Lognormal Model (of nonzero data of large coastal sharks) Variable Results and QQ Plot of Residuals

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Year	21	382	3.15	<.0001
stn	6	382	11.00	<.0001
month	5	382	5.30	0.0001





Large Coastal Sharks Index Output

SurveyYear	Frequency	N	LoIndex	StdIndex	CV	LCL	UCL
1975	0.83333	17	0.028510	2.50845	0.30721	1.37580	4.5736
1977	0.73333	15	0.022664	1.99409	0.34390	1.02176	3.8917
1978	0.66667	3	0.011076	0.97456	1.00632	0.18297	5.1910
1980	0.86486	37	0.023444	2.06270	0.24630	1.26953	3.3514
1981	0.75758	33	0.020404	1.79525	0.23730	1.12414	2.8670
1984	0.66667	3	0.007482	0.65829	1.61131	0.06851	6.3253
1986	0.25000	4	0.006952	0.61168	2.71511	0.03314	11.2892
1989	0.66667	6	0.008975	0.78964	0.52575	0.29399	2.1209
1990	0.77500	40	0.009262	0.81488	0.43699	0.35318	1.8802
1991	0.64516	31	0.007974	0.70161	0.52371	0.26210	1.8781
1992	0.51852	27	0.013996	1.23143	0.56022	0.43312	3.5011
1993	0.55556	18	0.009029	0.79443	0.61886	0.25437	2.4811
1995	0.83871	31	0.009213	0.81060	0.44792	0.34466	1.9064
1996	0.59184	49	0.008707	0.76612	0.40591	0.35082	1.6731
1997	0.75000	40	0.008556	0.75277	0.27634	0.43756	1.2950
1998	0.61702	47	0.008378	0.73712	0.31832	0.39600	1.3721
1999	0.69231	39	0.008067	0.70982	0.43748	0.30737	1.6392
2000	0.60976	41	0.008834	0.77722	0.36476	0.38332	1.5759
2001	0.69444	36	0.008381	0.73738	0.35622	0.36938	1.4720
2002	0.76923	26	0.007783	0.68481	0.50919	0.26213	1.7890
2003	0.65217	23	0.006209	0.54630	0.37332	0.26527	1.1251
2004	0.56410	39	0.006147	0.54084	0.51443	0.20521	1.4254