APPENDIX B6: Imputed NEFSC scallop survey data for unsampled strata

Some strata were unsampled during 1979-2006 NEFSC scallop surveys, particularly in the Georges Bank region (Tables 1 and 2). In NEFSC (2004), these "holes" in the survey data for a particular year (*y*) were filled automatically in database retrieval software by borrowing data from the same survey strata collected during the previous (*y*-1) and/or next (*y*+1) annual surveys. Borrowed data were used to compute means for survey holes and stratified random means for larger areas in the normal manner. Borrowing was one-sided in cases where data from *y*-1 or *y*+1 were lacking, and in the most recent survey year in particular where data for year *y*+1 are never available.

The borrowing procedure and variance calculations are ad-hoc but have a number of advantages: 1) survey indices for year *y* do not change after year *y*+1; 2) a minimum of programming and staff time is required; 3) the most relevant data are used, and 4) the calculations (linear interpolation between adjacent surveys) are simple, objective and make few assumptions about spatial patterns in population dynamics. No allowance is made for measurement errors in borrowed data. However, scallop survey data are relatively precise and important strata with high scallop abundance were generally not missed.

A more complicated statistical model based procedure was used in this assessment to fill all of the holes in NEFSC scallop survey data. However, data for Georges Bank during 1979-1981 were not used in the assessment, even after holes were filled, because the number of unsampled strata was relatively high (Figure 1).

The new statistical model was fit to tow-by-tow survey data (number of 40+ mm SH sea scallops per tow) by maximum likelihood using the glm.nb() function in Splus with a log link and assuming that measurement errors in the survey data were from a negative binomial distribution. Years and "newstrata" (see below) were categorical variables in the model and separate models were used for each subregion and post-stratification scheme. Residuals plots indicated that the model used to predict strata means fit the data reasonably well (Figure 1).

Subregions and newstrata are specific to the post-stratification scheme employed in a particular database run. Newstrata are original survey strata split into open and closed management areas. Subregions are contiguous groups of newstrata that define areas of particular interest. Data used in models that fill holes and in calculating abundance indices are from random stations within the original survey strata so that statistical assumptions are not violated in splitting strata into newstrata. Post-stratification exacerbates problems with holes because sections of a stratum assigned to newstrata might not have been sampled during a particular survey even if the larger stratum was sampled.

After fitting, the statistical model was used to calculate and store predicted values for every combination of subregion, year and newstrata. Predicted survey length composition or each subregion, year and newstrata was calculated by applying the shell height composition (total numbers in each 5 mm bin) from tows in the same subregion during the same year to predicted total numbers per tow from the model. Survey database software automatically retrieves predicted values for each shell height group to fill holes, as required. Predicted biomass per tow was calculated in the survey database software in the normal manner by applying a shell height/meat weight relationship.

The standard error for predicted number or biomass per tow is used in database variance calculations for larger subregions and regions. The standard error for predicted catch per tow in a particular size bin was $\sqrt{p_L^2 s^2}$ where *s* is the standard error for predicted mean number per tow from the model (all sizes) and *p* is the observed proportion of mean numbers per tow for shell height bin *L*. Variances in the proportion p are not considered because the number of shell height measurements in a subregion is normally high.

The major benefit of the new modeling approach is that secondary holes in newstrata that occur after poststratification are automatically filled and that variance calculations have a better statistical basis. Differences in abundance and biomass indices between the complicated model based- and simple borrowing procedures were modest for Georges Bank as a whole and almost identical for the Mid-Atlantic Bight as a whole. Differences between model based and borrowing estimates were more substantial, however, for some subregions on Georges Bank.

APPENDIX Table B6-1. Numbers of random tows in NEFSC scallop surveys on Georges Bank by survey stratum and year (including tows by the F/V Tradition during 1999). Black areas indicate strata that were not sampled.

Year	Stratum									
	46	47	49	50	51	52	53	54	55	58
1979	4	9	5	4	7	3	3	5	7	2
1980		4	5	5	7	5		4	10	2
1981	5	9	5	5	8	5	5	6	9	2
1982	6	9	6	8	8	6	6	6	6	3
1983	6	9	6	12	11	6	6	6	6	4
1984	6	9	7	12	12	6	6	6	5	4
1985	6	10	9	11	12	7	7	7	7	4
1986	6		9	16	12	11	7	7	1	8
1987	6	12	9	16	11	11	7	7	9	8
1988	6	12	9	16	12	12	7	7	10	8
1989	6	12	8	15	12	12	7	6	10	8
1990	6	12	9	15	13	12	7	7	10	8
1991	6	12	9	16	12	12	7	7	10	8
1992	6	12	9	16	11	11	7	7	10	8
1993	6	12	9	13	9	10	7	7	10	8
1994	6	12	9	16	12	12	7	7	10	8
1995	6	12	9	16	11	12	7	7	10	8
1996	6	12	5	16	12	11	7	7	10	8
1997	6	13	7	16	12	14	9	10	10	8
1998	15	22	9	16	11	12	7	7	10	8
1999	6	15		5	6	14	11	15	14	8
2000	6	12	7	13	9	9	6	7	10	8
2001	6	14	9	15	14	14	15	11	12	6
2002	6	14	6	13	14	13	16	11	12	6
2003	6	13	9	14	10	14	15	13	10	6
2004	4	18	9	12	12	11	15	20	10	4
2005	5	20	10	11	12	12	12	19	10	4
2006	4	18	7	14	10	16	13	17	14	4

APPENDIX Table B6-1 continued

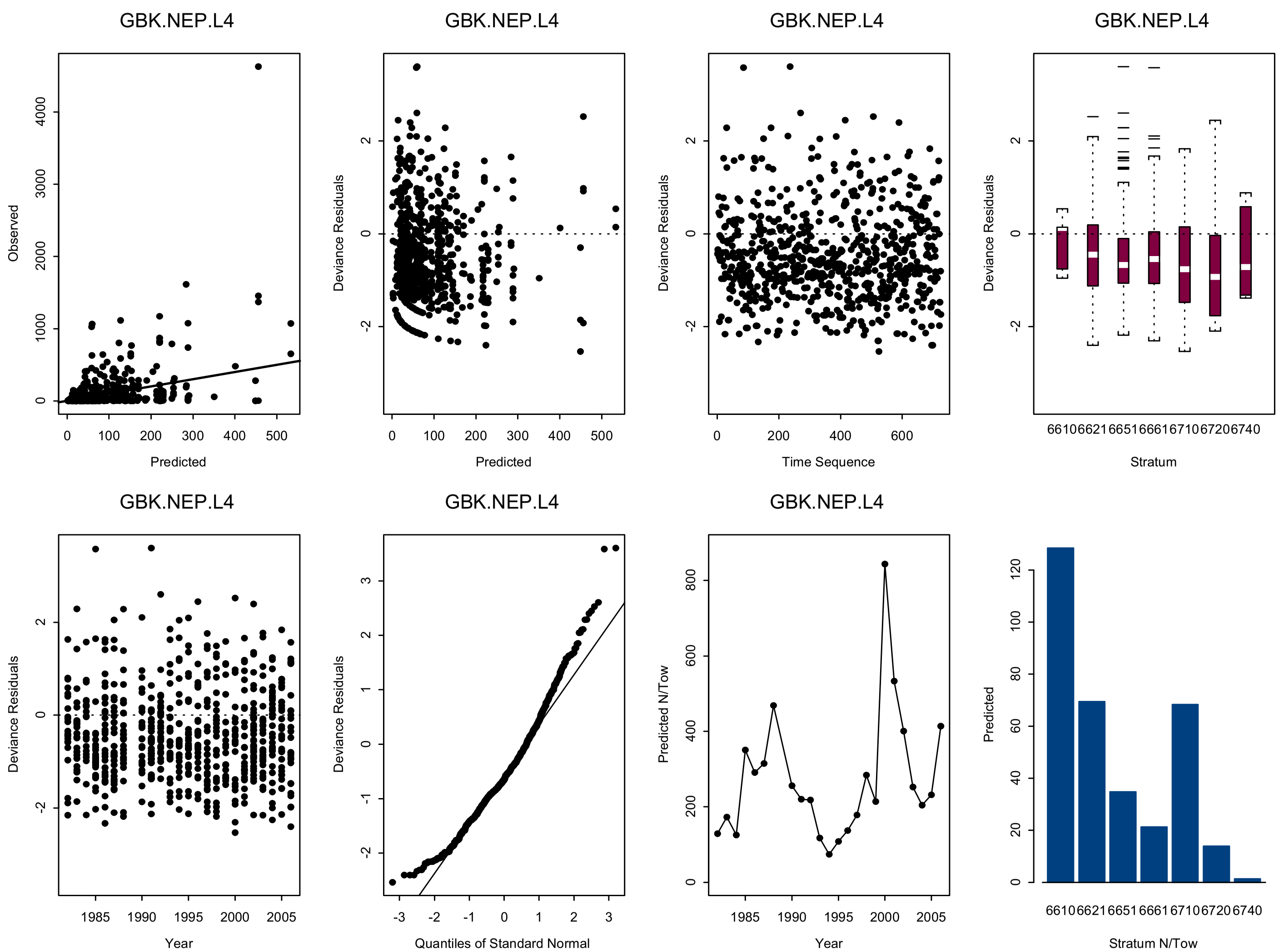
Year	Stratum									
		59	60	61	62	63	65	66	71	72
1979	10	8	[redacted]							
1980	10	8	[redacted]							
1981	9	8	[redacted]							
1982	10	9	7	9	4	6	6	4	5	
1983	8	8	7	8	5	9	8	4	4	
1984	8	8	7	8	3	9	8	5	4	
1985	12	12	8	12	7	10	10	6	6	
1986	12	12	8	13	7	12	12	6	6	
1987	12	12	8	12	7	12	12	5	6	
1988	12	12	8	12	6	11	12	6	6	
1989	12	12	[redacted]							
1990	12	12	8	12	7	12	12	6		
1991	12	12	8	12	7	12	12	6	6	
1992	12	12	8	12	7	11	12	6	6	
1993	12	12	8	12	7	10	10	6	6	
1994	12	12	8	12	7	12	12	6	6	
1995	12	12	8	12	7	12	12	6	6	
1996	12	12	8	12	7	12	12	6	6	
1997	12	12	8	12	7	15	14	8	5	
1998	11	11	8	12	7	12	10	6	6	
1999	12	12	8	14	6	11	11	4	2	
2000	12	12	7	12	7	11	12	6	[redacted]	
2001	10	12	18	23	6	10	11	5	[redacted]	
2002	10	10	18	24	4	12	14	8	5	
2003	8	9	16	21	4	12	12	8	[redacted]	
2004	7	6	24	24	3	12	10	12	3	
2005	8	7	22	24	3	11	9	12	[redacted]	
2006	6	7	24	17	3	12	9	19	[redacted]	

APPENDIX Table B6-2. Numbers of random tows in NEFSC scallop surveys in the Mid-Atlantic Bight by survey stratum and year (including tows by the F/V Tradition during 1999). Black areas indicate strata that were not sampled.

Year	Stratum										
	6	7	10	11	14	15	18	19	22	23	24
1979	2	1	5	7	7	12	7	5	12	20	3
1980	1	2	5	7	7	12	7	5	12	20	3
1981	2	1	5	6	7	12	7	5	12	20	3
1982	4	3	8	6	6	12	7	5	12	20	6
1983	4	4	8	8	8	12	6	7	8	16	6
1984	5	4	8	8	10	12	6	8	8	16	6
1985	5	5	8	8	10	12	8	8	8	16	6
1986	5	5	8	8	12	12	10	13	8	16	6
1987	5	5	8	8	12	11	10	12	8	16	4
1988	6	4	8	8	12	12	10	12	8	16	6
1989	5	5	8	8	12	12	10	12	8	16	6
1990	3	3	8	8	12	12	10	12	8	16	5
1991	5	5	8	8	12	12	10	11	8	16	6
1992	5	5	8	8	12	12	10	12	8	16	6
1993	5	5	8	8	12	12	8	10	8	16	6
1994	5	5	8	8	12	12	10	12	8	16	5
1995	5	5	8	8	12	12	10	12	8	16	6
1996	5	5	8	8	12	12	8	10	8	16	6
1997	5	5	8	8	11	12	9	12	8	16	6
1998	5	5	8	8	12	12	10	12	8	16	6
1999	5	5	8	8	12	12	10	12	8	16	6
2000	5	5	8	8	12	12	10	13	8	16	6
2001	5	5	9	14	10	12	8	12	10	22	8
2002	5	5	9	12	10	12	8	11	12	22	8
2003	5	5	8	12	10	12	10	12	10	20	6
2004	3	2	8	12	14	16	24	21	14	25	10
2005	2	3	7	10	15	16	26	22	14	26	8
2006	3	2	6	10	14	20	20	25	14	25	5

APPENDIX Table B6-2 continued.

Year	Stratum									
	25	26	27	28	29	30	31	33	34	35
1979	4	8	12	2	8	14	24	2	4	7
1980	4	9	11	2	8	14	24	4	4	6
1981	5	8	12	2	8	14	24	4	4	6
1982	7	9	12	3	8	14	24	7	7	5
1983	6	13	10	7	6	15	24	10	10	5
1984	7	14	10	6	8	15	24	10	14	5
1985	4	14	12	6	6	15	24	10	10	6
1986	4	14	20	10	6	15	24	7	13	10
1987	4	14	20	10	6	15	24	10	14	10
1988	4	14	19	10	6	15	23	10	14	10
1989	4	14	20	10	6	15	24	10	29	10
1990	3	12	17	10	5	14	24	10	14	10
1991	5	14	20	10	6	15	24	10	14	10
1992	4	14	20	10	6	15	24	10	14	10
1993	4	14	20	10	6	15	22	7	10	8
1994	4	14	20	10	6	15	23	10	14	10
1995	4	12	20	10	6	15	24	10	14	10
1996	4	13	19	10	6	15	20	8	10	8
1997	4	14	20	10	6	14	24	10	13	10
1998	4	14	19	9	[redacted]	14	23	6	14	10
1999	4	14	20	10	6	15	24	7	14	10
2000	4	13	20	10	6	15	24	10	14	10
2001	8	14	20	8	6	12	18	8	10	8
2002	6	10	19	7	6	10	16	6	6	6
2003	6	10	20	8	4	9	16	6	6	6
2004	5	8	20	8	4	6	20	5	5	18
2005	5	7	21	7	4	6	21	5	6	10
2006	6	7	16	5	5	9	20	5	5	8



APPENDIX Figure B6-1. Model residuals and diagnostics for a negative binomial model used to impute NEFSC scallop survey mean numbers per tow (40+ mm SH) based on tow by tow survey data from the Northeast Peak subregion of Georges Bank. Goodness of fit results for the Northeast Peak were typical although spatial and temporal patterns in predicted abundance varied among subregions.