Aleutian Islands Golden King Crab CPUE Data Standardization

M.S.M. Siddeek[1] and J. Zheng[1]

Alaska Department of Fish and Game

1. Division of Commercial Fisheries

P.O. Box 115526, Juneau, Alaska 99811

Summary report prepared for the Crab Plan Team May 2012 meeting.

Preamble

Our primary task is to standardize the catch-per-unit-effort (CPUE) of observer pot sample and retained catch data to input to the Aleutian Islands golden king crab assessment model (Siddeek et al., 2011). The January 2012 modeling workshop also recommended to fully document the observer and retained catch data collection method and investigate the compatibility of observer pot sample with retained catch sample length frequencies. This latter task is dealt by Doug Pengilly (personal communication). This document focused on developing a standardization method based on generalized linear model (GLM) for estimating yearly CPUE index. For this report, we used the observer pot sample data for 1990/91-2010/11 for the whole Aleutian Islands region, east of 174W, and west of 174W to test the model selection and CPUE index estimation procedures.

We plan to improve on the estimated CPUE indexes for observer data, if there are any suggestions from the crab plan team (CPT) and carry forward the analysis to estimate retained catch CPUE index for the whole region, east of 174W, and west of 174W using 1985/86-2010/11 data.

Method

Preliminary data processing

The observer pot sample data of Aleutian Islands golden king crab totaled 102,965 records for 1990/91-2010/11. The following variables from each record were considered in the model:

Year = Federal Fisheries Management Year (July 1-June 30). This is the main focus of the analysis because abundance varies by year, but confounded with other fishery induced variables. This is treated as a predictor factor variable in the model.

Month = A calendar month in a fishing year. This is an important variable because the magnitude of CPUE changes as the season progresses during a fishing year. This is treated as a predictor factor variable in the model.

Vessel = Identification code for a participating vessel. This is another important variable because the type of vessel and the crew affect the capture efficiency. This is treated as a predictor factor variable in the model.

Catch = Number of crabs caught. This is the response variable. We considered different types of male catch as response variable: total males, legal males, and sublegal males. Because the stock assessment model is designed for male-only, we did not consider female catch. This is treated as a response numerical variable in the model.

Pot-lifts = Total number of pot lifts realized in the trip completed by the sampled vessel. We considered this as an important predictor variable because it affects the abundance available for catching. This is treated as a predictor numerical variable in the model.

Area = The area code is 1 for east of 174 W and 2 for west of 174. This is a management variable. Although this variable can be treated as a predictor factor variable we subset the data to 1, 2 or 1 &2 for model fit.

Depth = Depth in fathom. We considered this variable as an important predictor variable because crab abundance is not uniform by depth. This is treated as a predictor numerical variable in the model.

Soak Days = Soak time in number of days. We considered this variable as an important predictor variable because there were significant changes in soak-time duration between pre-and post-rationalization periods (Siddeek et al., 2011). This is treated as a predictor numerical variable in the model.

Gear = Identification code for different types of pot gear. Although a single gear (pot) is used in the fishery, the type and configuration varied over the years. Each type of pot has a unique number code (Table 1). We considered this variable as an important predictor variable because different gear configurations affect catching efficiency. This is treated as a predictor factor variable in the model.

Because of unusually high and low soak times (some records showed 384 days and some others showed 0 soak days) observed in certain years, we arbitrarily selected the records within 0.5% to 99.5% soak time (Table 2). The pre-rationalization (prior to the 2005/06 season) period soak time ranged 1-39 days and that for post-rationalization period ranged 4-54 days. We combined the trimmed data for the pre-and post rationalization period to obtain the total records for analysis. After removing missing information for variables considered in the model and trimming for 0.5-99.5% soak time range, the number of records reduced to 102,611.

There was a maximum of 162 vessels registration codes in the crab retained catch database during 1985/86-2010/11. The maximum number of vessels dropped to 67 when the period was restricted to 1990/11-2010/11. We used the number of catch delivery instances as surrogate for trips in each year for each vessel. They are plotted in Figure 1 to assess the overlap of different vessels' fishing activities over time. The percentage catch and vessel dramatically reduced when we considered vessels with 3, 5, and 9 trips/year over time (Figure 2 for 1985/86-2010/11 data series and Figure 3 for 1990/91-2010/11 data series). We considered the longer data series for analyzing retained catch and the shorter data series for analyzing observer pot sample data. We selected five delivery instances per year for at least three years as reasonable to select the core vessels. This set of core vessels produced nearly 92% and 93% of the maximum total catches and reduced the number of vessels to 29% and 37% of the maximum number of vessels for 1985/86-2010/11 and 1990/91-2010/11, respectively. This reduced the number of observer pot sample records to 88975.

CPUE Standardization

A stepwise generalized linear model (GLM) procedure was used to select the best model and estimate a time series of CPUE index based on the relationship between CPUE vs. available predictive factor and continuous variables. Following Quinn and Deriso (1999), the GLM based on lognormal distribution can be derived from the following:

$$U_{ijk} = U_0 \prod_i \prod_j P_{ij}^{X_{ij}} e^{\varepsilon_{ijk}} \qquad (1)$$

where U is the observed CPUE, U_0 is the reference CPUE, P_{ij} is a factor i at level j, and X_{ij} takes a value of 1 when the jth level of the factor P_{ij} is present and 0 when it is not. The random error ε_{ijk} for observation k is a normal random variable with 0 mean and standard deviation σ.
Taking the logarithm of equation (1) yields an additive generalized linear model for lognormal error distribution of U:

$$ln(U_{ijk}) = \ln(U_0) + \sum_{i=1}^{p} \sum_{j=1}^{n_j-1} X_{ij} \ln(P_{ij}) + \varepsilon_{ijk} \qquad \text{Or}$$

$$ln(U_{ijk}) = \beta_0 + \sum_{i=1}^{p} \sum_{j=1}^{n_j-1} X_{ij}\, \beta_{ij} + \varepsilon_{ijk} \qquad (2)$$

where β_0 is the intercept and $\beta_{ij} = ln(P_{ij})$.

The model described by equations 1 and 2 is over-parameterized. A common remedial solution is to setting a factor coefficient to zero, usually the first, whereupon the remaining n_j-1 coefficients of each factor i represent incremental effects relative to the reference level.

Coefficients obtained by fixing a factor level will differ with the choice of reference level. However, the relative differences among the estimated coefficients will not be affected by the choice of constraint. Following Francis (1999), coefficients for factor *i* were transformed to "canonical" coefficients over all levels *j* calculated relative to their geometric mean (Starr, personal communication, March 2012).

Geometric mean is calculated as ,

$$\bar{\beta} = \sqrt[n_j]{\prod_{j=1}^{n_j} \beta_{ij}} \qquad (3)$$

The canonical coefficient is

$$\beta_i{}' = \frac{\beta_i}{\bar{\beta}} \qquad (4)$$

As CPUE analysis is done in the non-log space, the non-log space canonical coefficient is equivalent to

$$b' = e^{\beta_i - \bar{\beta}}$$

A number of factors contribute to the variation in CPUE, which includes Year, Month, Vessel, Depth, Soak Time, Fishing Effort, etc. The year of capture is usually given special significance: variations between years in this factor are interpreted as relative changes in the annual abundance of the crab. **The resulting series of 'fishing year' canonical coefficients is termed as the "Standardized" annual CPUE index.**

For example, consider a model of the form

$$y_i = \ln(CPUE_i) = \beta'_0 + \beta'_1 x_1 + \beta'_2 x_2$$

If x_2 is a factor variable for year, then β'_2 would take on the values β'_{20} if the year is the reference year 0, and β'_{2i} if the year is some other year i. So, the CPUE index for year i relative to the reference year 0 is estimated as

$$CPUEindex_i = \frac{e^{y_i}}{e^{y_0}} = \frac{e^{\beta'_0 + \beta'_1 x_1 + \beta'_{2i} x_2}}{e^{\beta'_0 + \beta'_1 x_1 + \beta'_{20} x_2}} = e^{\beta'_{2i} - \beta'_{20}}$$

So, the relative year effects are calculated by dividing the inverse of the fitted model in year i by the inverse of the fitted model in the base year 0.

A selection procedure was applied to determine the relative importance of these factors in the model. The procedure involves a forward stepwise fitting algorithm which generates regression models iteratively, starting with the simplest model, ln(CPUE) = factor(Fishing year), and building in complexity subject to a stopping rule designed to include only the most important factors.

The following general procedure was used to fit the models:

1. Fit the GLM with each predictor variable from a maximum set of predictor (factor and non factor) variables against the natural log of CPUE (male total, legal, or sublegal catch per record).

2. Generate a R^2 based on model deviance and also number of degrees of freedom for each fit.

$$R^2 = \frac{(null\ model\ deviance - added\ parameter\ model\ deviance)}{null\ model\ deviance} \qquad (5)$$

where deviance = a constant-2Maximum log likelihood.

Select the predictor variable that has the highest R^2.

3. Repeat Steps 1 and 2, accumulating the number of selected predictor variables and increasing the model degrees of freedom, until the increase in residual deviance (as measured by R^2) for the final iteration is less than 0.01.

The log normal model is applicable for positive catch data. Zero catches are also encountered in observer samples. A GLM model based on a binomial distribution and using the presence/absence of crab (success = 1/0)) as the dependent variable was also fitted to the same set of data using the same set of explanatory variables. The binomial model will provide another series of standardized annual CPUE coefficients that is similar to the series estimated from the lognormal GLM. A combined model which integrates the two series of relative annual changes estimated by the lognormal and binomial models was estimated using the delta distribution which allowed zero and positive catches (Vignaux 1994; Starr, 2012).

$$Y_y^{Comb} = \frac{Y_y^{Ln}}{\left[1 - P_0\left[1 - \frac{1}{Y_y^{Binom}}\right]\right]} \qquad (6)$$

where

Y_y^{Comb} = combined CPUE index for year y

Y_y^{Ln} = lognormal CPUE index for year y

Y_y^{Binom} = binomial CPUE index for year y

P_0 = proportion of zeros for base year 0

For comparison with the standardized CPUE index, we also estimated the nominal CPUE (Arithmetic CPUE) and scaled to the level of standardized CPUE index.

$$A_y = \frac{\sum_{i=1}^{n_y} C_{iy}}{\sum_{i=1}^{n_y} E_{iy}} \qquad (7)$$

$$\overline{\mathrm{A}} = \sqrt[n_y]{\prod_{y=1}^{n_y} \mathrm{A_y}} \qquad (8)$$

$$A'_y = \frac{A_y}{\bar{A}} \tag{9}$$

where C_{iy} is the catch and E_{iy} is the effort for each record *i* in year *y*; $\bar{A}$ is the geometric mean of the Arithmetic CPUE; and A_y and A'_y are Arithmetic CPUE and scaled Arithmetic CPUE for year y, respectively.

Software use

We coded in R to process the data (Appendix A). We used two additional R scripts (Step CPUE function.R and getCPUE.R) obtained from Paul Starr which were slightly amended for our data for CPUE index calculation.

Results

Observer pot sample data analysis

To analyze the observer sample data we used first the lognormal GLM on non zero catches. We used the forward step-wise selection procedure to pick up the best model. We assumed the null model to be

$$\ln(I_i) \sim Year_{y_i} + \varepsilon_i \tag{10}$$

The maximum set of model terms offered to the stepwise selection procedure was:

$$\ln(I_i) \sim Year_{y_i} + Month_{m_i} + Vessel_{v_i} + Gear_{g_i} + f(SoakDays_i) + g(Depth_i) + PotLifts_i + \varepsilon_i \tag{11}$$

where I = number of males caught in i th record ($CPUE_i$) ; *f()* and *g()* are non-linear functions (we used a second order orthogonal polynomial functions) ; and all predictor variables are self explanatory by name, and subscript of small characters are factor levels. Note that although golden king crab fishery is executed with a single gear (pot), the gear configuration has changed over the years, so different factor levels for the gear were considered for the model.

The factor levels considered for each factor variable are:

Year: 1985 to 2010 for retained catch (not reported here) or 1990 to 2010 for observer pot sample data;

Month: 1 to 12;

Vessel: vessel registration number; and

Gear: Gear codes. They are provided in Table 1.

We explored the influence of each variable by fitting lognormal GLM to individual variable one-at-a time. For legal crab, Year factor produced the lowest Akike Information Criteria (AIC) whereas for sublegal crab, Vessel factor produced the lowest AIC (Table 3).

The forward selection procedure produced a suite of final models for different subsets of observer pot sample data. Gear and Vessel factors were predominant among the selected models (Table 4).

Tables 5 and 7 provide the analysis of deviance values for lognormal fit of legal and sublegal crabs, respectively for combined east and west of 174W data. The variable rows are in order of their selection. The corresponding results for east of 174 W data are in Tables 9 and 11 and that for west of 174W data are in Tables 13 and 15. The variables selected to the final model have significant R^2 values (>0.01).

Then we used the binomial GLM on catch success predictor variable (i.e., if catch>0, success=1, and if catch=0, success=0). The maximum set of model terms offered to the stepwise selection procedure was:

$$\text{success} \sim Year_{y_i} + Month_{m_i} + Vessel_{v_i} + Gear_{g_i} + f(SoakDays_i) + g(Depth_i) + PotLifts_i + \varepsilon_i \quad (12)$$

with a binomial logit link function.

The selection procedure produced a suite of best models for different subsets of data.
Tables 6 and 8 provide the analysis of deviance values for binomial fit of legal and sublegal crabs, respectively for combined east and west of 174W data. The corresponding results for east of 174 W data are in Tables 10 and 12 and that for west of 174W data are in Tables 14 and 16.

Figure 4 provides the residual distribution and qq-plot for the best lognormal fit to legal crab for the combined east and west of 174W data. Figure 5 compares various CPUE indexes for the combined east and west of 174W data for legal crab. Base CPUE index considers only the Year effect disregarding the influence of all other factors or numerical variables. It overestimated the recent trend in Arithmetic index. However, both indexes are not appropriate. The combined index represents the best CPUE index. The lognormal CPUE index trend matches the combined CPUE index trend well.

Figure 6 depicts the residual distribution and qq-plot for the best lognormal fit to sublegal crab for the combined east and west of 174W data. Figure 7 compares various CPUE indexes for the combined east and west of 174W data for the sublegal crab. All CPUE index trends closely match since 1997.

Figure 8 shows the residual distribution for the best lognormal fit to legal crab for the east of 174W data. Figure 9 compares various CPUE indexes for the east of 174W data for the legal crab. We used the 1995/96 -2010/11 data because there was a gap in the data series prior to 1995/96 (1994/95 data missing). The CPUE indexes match closely.

Figure 10 shows the residual distribution for the best lognormal fit to sublegal crab for the east of 174W data. Figure 11 compares various CPUE indexes for the east of 174W data for the sublegal crab. The CPUE indexes match closely.

Figure 12 compares various CPUE indexes for the east of 174W data for the legal crab when the soak time was capped at 7 days for pre- and 26 days for post-rationalization periods (these tmax values were estimated by fitting Zhou and Shirley model (Zhou and Shirley, 1997). The CPUE indexes trends were similar to that shown in Figure 9.

Figure 13 depicts the residual distribution and qq-plot for the best lognormal fit to legal crab for the west of 174W data. Figure 14 compares various CPUE indexes for the west of 174W data for the legal crab. The CPUE index dipped during post-rationalization period because of the soak-time influence.
Figure 15 depicts the residual distribution and qq-plot for the best lognormal fit to sublegal crab for the west of 174W data. Figure 16 compares various CPUE indexes for the west of 174W data for the sublegal crab. The CPUE indexes match closely.

Discussion

We tested the model selection procedure using the observer pot sample data pooled for both management regions (east and west of 174W), east of 174W, and west of 174W and calculated CPUE indexes for each subset of data. In most cases the trends in sublegal CPUE indexes were flat and different indices matched. The soak time had an influence on west of 174W CPUE index. We also tested a data set from a New Zealand fishery provided by Paul Starr (April 2012) and our results for combined CPUE agreed with his results. Work is in progress to get the lognormal, binomial, combined, base, and Arithmetic CPUE indexes separately for whole Aleutian Islands, east of 174W, and west of 174W for retained catch data.

Once the CPUE indexes are finalized we will use the combined indexes in the assessment model.

Acknowledgements

Doug Pengilly assembled the data for this report. Doug Woodby and Doug Pengilly provided advice on a number of occasions and reviewed various draft of the document. Paul Starr provided insight into CPUE data and interpretations. We thank them for their help.

REFERENCES

Francis, R.I.C.C. 1999. The impact of correlations on standardized CPUE indices. New Zealand Fishery Assessment Research Document 1999/42. 30 p. (Unpublished report held in NIWA library, Wellington, New Zealand).

Quinn, T.R. and R.B. Deriso. 1999. Quantitative Fish Dynamics. Oxford University Press. 542 p.

Siddeek, M.S.M., D. Pengilly, and J. Zheng. 2011. Aleutian Islands Golden King Crab (*Lithodes aequispinus*) Model Based Stock Assessment in Fall 2011. Presented at the Fall 2011 Crab Plan Team meeting, AFSC, Seattle.

Starr, P.J. 2012 (in press). Standardized CPUE analysis exploration: using the rock lobster voluntary logbook and observer catch sampling programmes. New Zealand Fisheries Assessment Report 2012/xx, January 2012.

Vignaux, M. 1994: Catch per unit effort (CPUE) analysis of west coast South Island and Cook Strait spawning hoki fisheries, 1987–93. N.Z. Fisheries Assessment Research Document 94/11. 29 p. (Unpublished report held in NIWA library, Wellington, New Zealand)

Zhou, S. and T.C. Shirley. 1997. A model expressing the relationship between catch and soak time for trap fisheries. North American Journal of Fisheries Management, 17:482-487.

Table 1. Gear code assigned to different types of pot gear by observers during the 1990/91-2010/11 seasons. The yellow highlighted gear are infrequent and not considered as factor levels. (Pengilly, personal information).

Gear Code	Description	Total pot samples
-9	#N/A - not recorded	66
1	Dungeness crab pot, small & round	2
2	pyramid pot, tunnel openings usually on sides, stackable	2,107
3	conical pot, opening at top of cone, stackable	1,998
4	4' X 4' rectangular pot	60
5	5' X 5' rectangular pot	16,198
6	6' X 6' rectangular pot	14,927
7	7' X 7' rectangular pot	22,242
8	8' X 8' rectangular pot	1,407
9	5 1/2' X 5 1/2' rectangular pot	6,339
10	6 1/2' X 6 1/2' rectangular pot	19,697
11	7 1/2' X 7 1/2' rectangular pot	375
12	round king crab pot, enlarged version of Dungeness crab pot	8,257
13	10' X 10' rectangular pot	466
14	9' X 9' rectangular pot	1
15	8 1/2' X 8 1/2' rectangular pot	1
17	8' X 9' rectangular pot	1
20	7' X 8' rectangular pot	232
22	snail pot	1
23	dome shaped pot, tunnel opening on top, often longlined in deepwater fisheries	6,755
80	Historical: Cod pot, any shape pot targeting cod, usually with tunnel fingers	711
81	Historical: Rectangular pot, unknown size, with escape rings	1,122
	Grand Total	102,965

Table 2. Percentile cutoff levels of soak time for excluding questionable data from the Aleutian Islands golden king crab observer database.

Area	Period	0.5%-99.5% Percentile Range (days)
East and West Combined	Pre-rationalization (1990/91-2004/05)	1-39
	Post-rationalization (2005/06-2010/11)	4-54
East 174W	Pre-rationalization	1-20
	Post-rationalization	3-41
West 174W	Pre-rationalization	1-42
	Post-rationalization	6-56

Table 3. AIC associated with lognormal fit of the CPUE to individual variable for the Aleutian Islands golden king crab observer pot sample data for east and west of 174W combined.

AIC = -2*Max.LogLikelihood+2* number of parameters. n= 68076 for Sublegal and 76588 for Legal crabs. The yellow highlighted values are the lowest in each category.

Predictors	Sublegal	Legal
Year	219060	215034
Soak Days	219566	222859
Gear	215701	220772
Vessel	215060	223560
Depth	219251	228224
Month	219148	224733
Pot Lifts	219765	225922

Table 4. Step-wise model selection for various scenarios for the Aleutian Islands golden king crab observer pot sample data.

Area	Crab Category	Final Model
East and West Combined	Legal	Ln(CPUE)~ Year+Gear+Month+Soak Days Binomial(Success)~ Year+Gear+Month+Vessel
	Sublegal	Ln(CPUE)~ Year+Vessel+Gear+Month Binomial(Success)~ Year+Vessel+Depth+Gear+Month
East 174W	Legal	Ln(CPUE)~ Year+Gear+Vessel Binomial(Success)~ Year+Gear+Vessel
	Sublegal	Ln(CPUE)~ Year+Vessel+Gear+Depth Binomial(Success)~ Year+Gear+Vessel+Depth
West 174W	Legal	Ln(CPUE)~ Year+Vessel+ Soak Days+Gear Binomial(Success)~ Year+Vessel+Gear+Soak Days
	Sublegal	Ln(CPUE)~ Year+Vessel+Gear Binomial(Success)~ Year+Vessel+Depth
East 174W	Legal: maximum soak time was set at 7days for pre- and 26 days for post-rationalization periods	Ln(CPUE)~ Year+Gear+Vessel Binomial(Success)~ Year+Gear+Vessel

Table 5. Analysis of deviance for stepwise lognormal model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). The response variable is **legal** crab CPUE.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			76567	-41.841	0.159
Gear	-15	-29.9570	76552	-71.798	0.202
Month	-11	-21.9769	76541	-93.775	0.225
Soak Days	-2	-3.9838	76539	-97.759	0.241

Selection process used R^2 difference > 0.01. Deviance = up to a constant, minus twice the maximized log-likelihood (constant is selected to make the deviance 0 for the saturated model).

Table 6. Analysis of deviance for stepwise binomial model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). The response variable is **legal** crab success catch.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			88954	-41.965	0.035
Gear	-15	-29.971	88939	-71.935	0.065
Month	-11	-21.980	88928	-93.916	0.084
Vessel	-24	-47.984	88904	-141.899	0.101

Table 7. Analysis of deviance for stepwise lognormal model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). The response variable is **sublegal** crab CPUE.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			68055	-41.987	0.013
Vessel	-24	-47.930	68031	-89.918	0.082
Gear	-15	-29.971	68016	-119.888	0.112
Month	-11	-21.985	68005	-141.873	0.127

Table 8. Analysis of deviance for stepwise binomial model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). The response variable is **sublegal** crab success catch.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			88954	-41.992	0.008
Vessel	-24	-47.965	88930	-89.957	0.043
Depth	-2	-3.977	88928	-93.935	0.065
Gear	-15	-29.985	88913	-123.920	0.080
Month	-11	-21.986	88902	-145.906	0.094

Table 9 Analysis of deviance for stepwise lognormal model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (east of 174 W). The response variable is **legal** crab CPUE.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			38586	-31.797	0.203
Gear	-14	-27.927	38572	-59.725	0.275
Vessel	-18	-35.984	38554	-95.709	0.291

Table 10. Analysis of deviance for stepwise binomial model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (east of 174 W). The response variable is **legal** crab success catch.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			43100	-31.956	0.044
Gear	-14	-27.932	43086	-59.888	0.112
Vessel	-18	-35.988	43068	-95.876	0.124

Table 11. Analysis of deviance for stepwise lognormal model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (east of 174 W). The response variable is **sublegal** crab CPUE.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			34713	-31.977	0.023
Vessel	-18	-35.876	34695	-67.854	0.146
Gear	-14	-27.962	34681	-95.816	0.184
Depth	-2	-3.988	34679	-99.803	0.197

Table 12. Analysis of deviance for stepwise binomial model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (east of 174 W). The response variable is **sublegal** crab success catch.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			43100	-31.988	0.012
Gear	-14	-27.940	43086	-59.928	0.072
Vessel	-18	-35.971	43068	-95.899	0.101
Depth	-2	-3.985	43066	-99.884	0.116

Table 13. Analysis of deviance for stepwise lognormal model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (west of 174 W). The response variable is **legal** crab CPUE.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			37366	-41.853	0.147
Vessel	-22	-43.945	37344	-85.798	0.202
Soak Days	-2	-3.976	37342	-89.774	0.226
Gear	-15	-29.984	37327	-119.758	0.242

Table 14. Analysis of deviance for stepwise binomial model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (west of 174 W). The response variable is **legal** crab success catch.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			45122	-41.963	0.037
Vessel	-22	-43.961	45100	-85.924	0.076
Gear	-15	-29.980	45085	-115.904	0.096
Soak Days	-2	-3.990	45083	-119.893	0.107

Table 15. Analysis of deviance for stepwise lognormal model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (west of 174 W). The response variable is **sublegal** crab CPUE.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			32750	-41.991	0.009
Vessel	-22	-43.916	32728	-85.907	0.093
Gear	-15	-29.985	32713	-115.892	0.108

Table 16. Analysis of deviance for stepwise binomial model selection for the observer pot sample data for the Aleutian Islands golden king crab fishery (west of 174 W). The response variable is **sublegal** crab success catch.

Factor	df (difference from null)	Deviance	Residual df	Residual Deviance	R^2
Year			45122	-41.988	0.012
Vessel	-22	-43.950	45100	-85.938	0.062
Depth	-2	-3.976	45098	-89.914	0.086

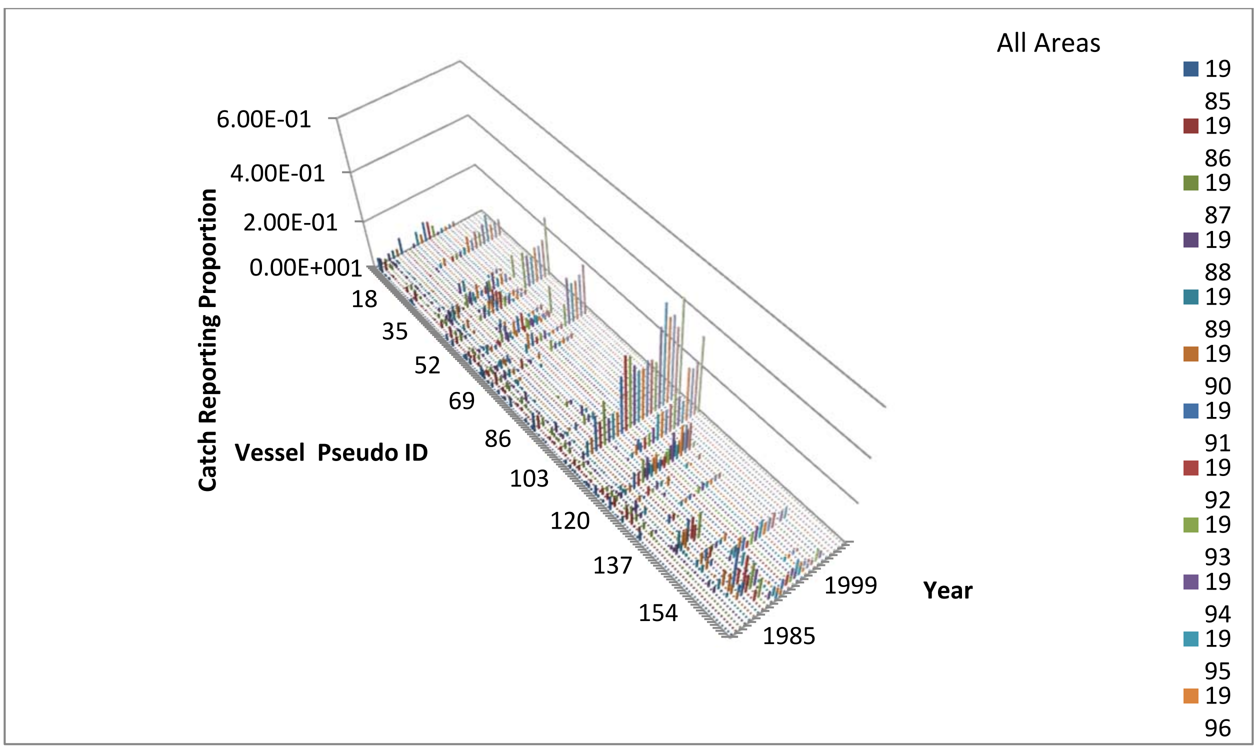


Figure 1. Golden king crab catch reporting frequency by vessel from both regions (combined east and west of 174 W) of Aleutian Islands.

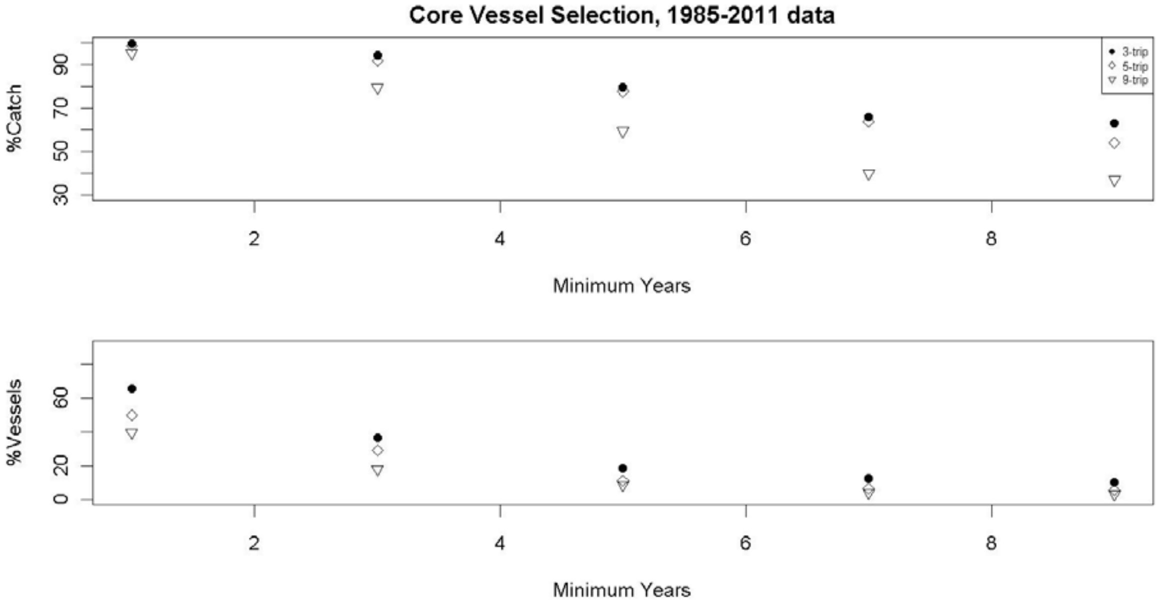


Figure 2. Core vessel selection based on 1985-2011 commercial fisheries data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). 3-trip = three trips per year; 5-trip = five trips per year; and 9-trip = nine trips per year . The percentage catch and vessels dropped as the number of minimum years the vessels with those yearly reporting rates increased.

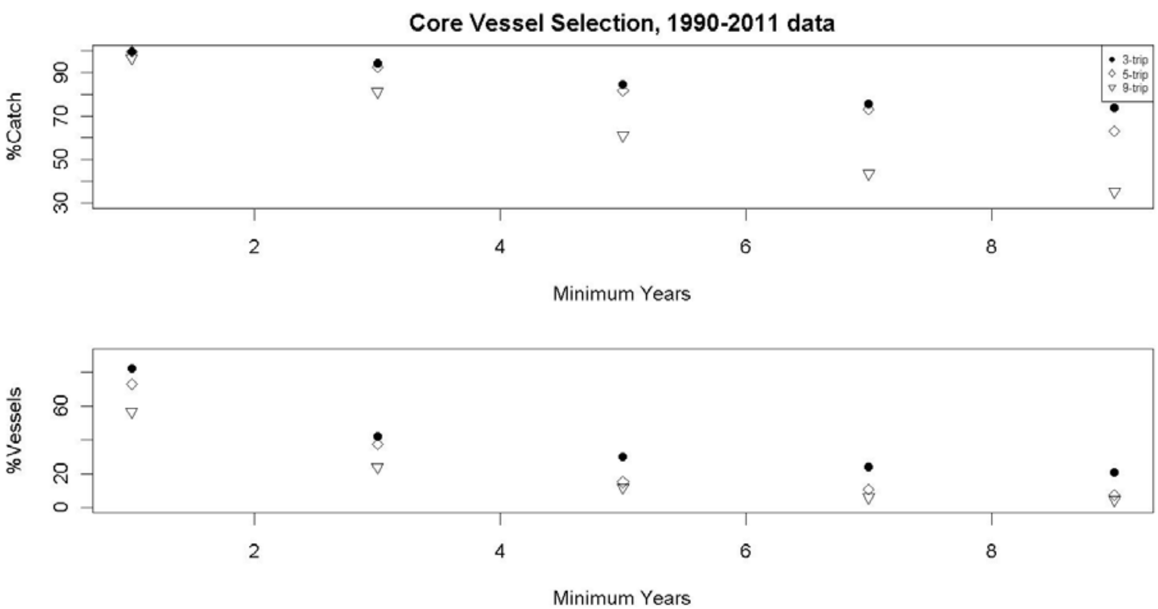
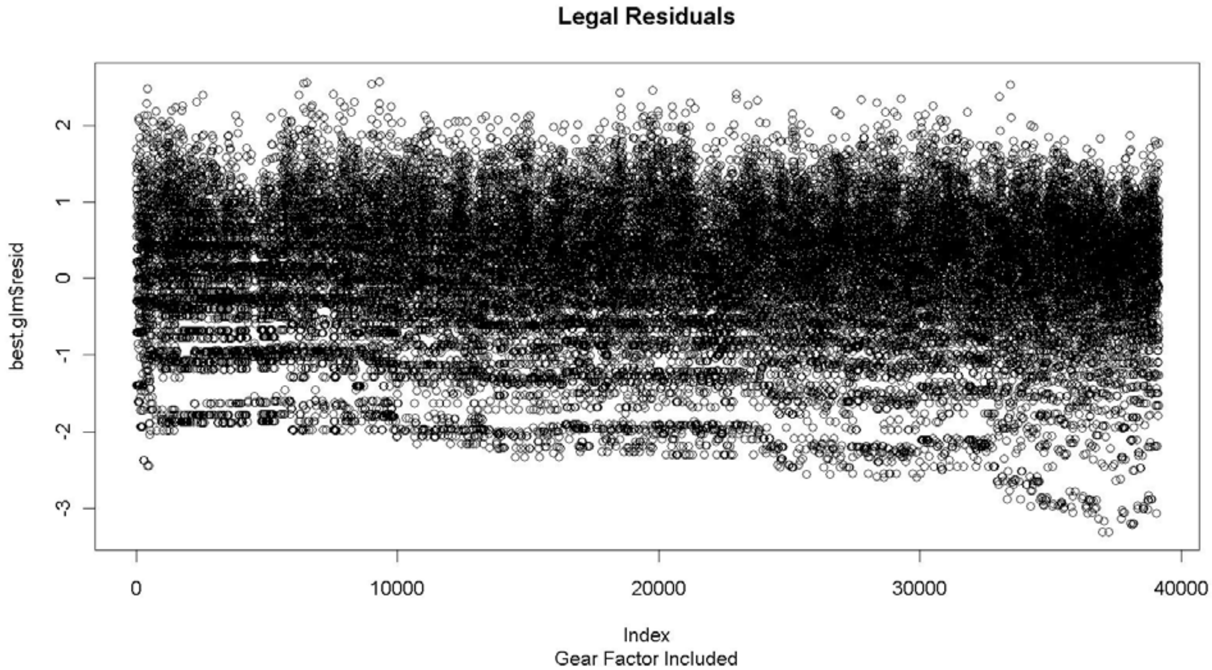


Figure 3. Core vessel selection based on 1990-2011 commercial fisheries data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). 3-trip = three trips per year; 5-trip = five trips per year; and 9-trip = nine trips per year. The percentage catch and vessels dropped as the number of minimum years the vessels with those yearly reporting rates increased.



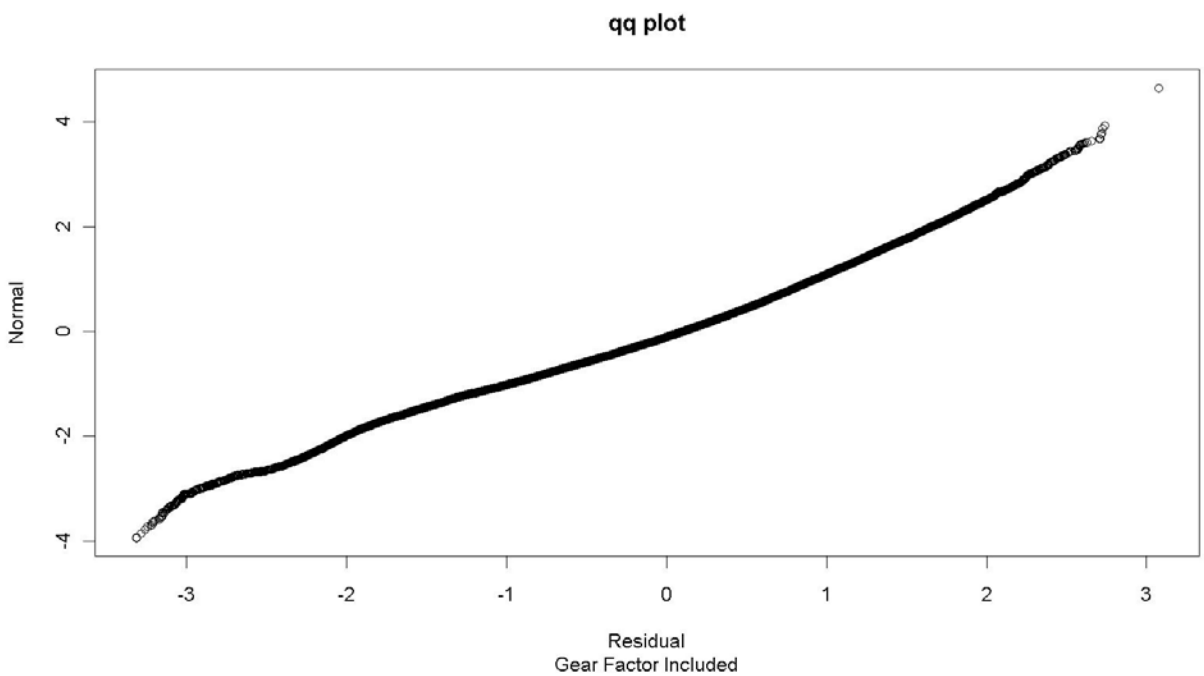


Figure 4. Residual and QQ plots of the best lognormal fit model for legal CPUE from combined east and west of 174 W. Observer pot sample data for 1990/91-2010/11 were used.

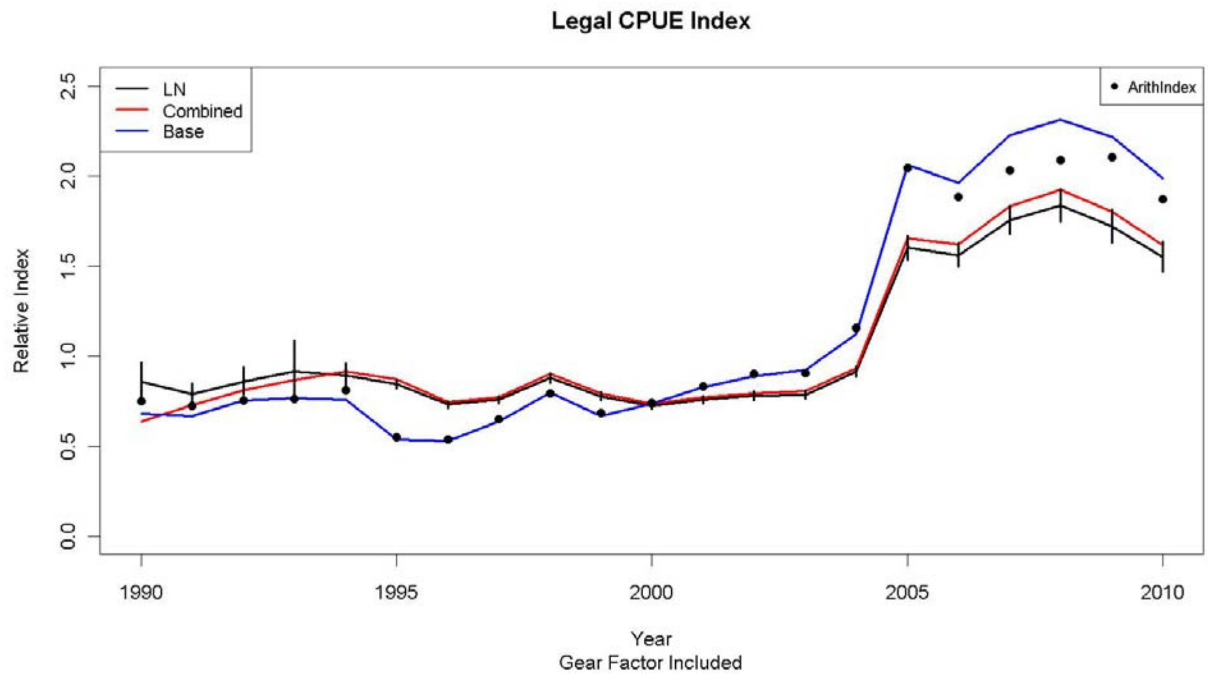
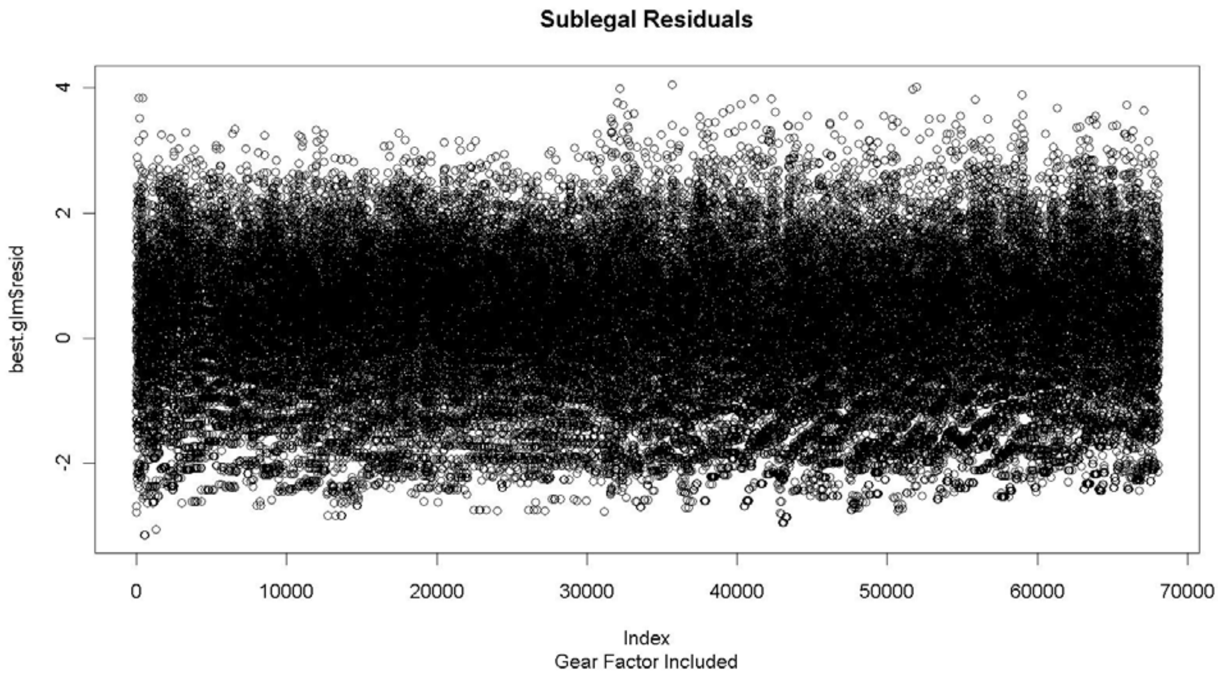


Figure 5. Trends in legal CPUE indexes for the observer pot sample data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). Lognormal: black line with 2standard errors; Combined: red line; Base: blue line; and Arithmetic: black filled circles.



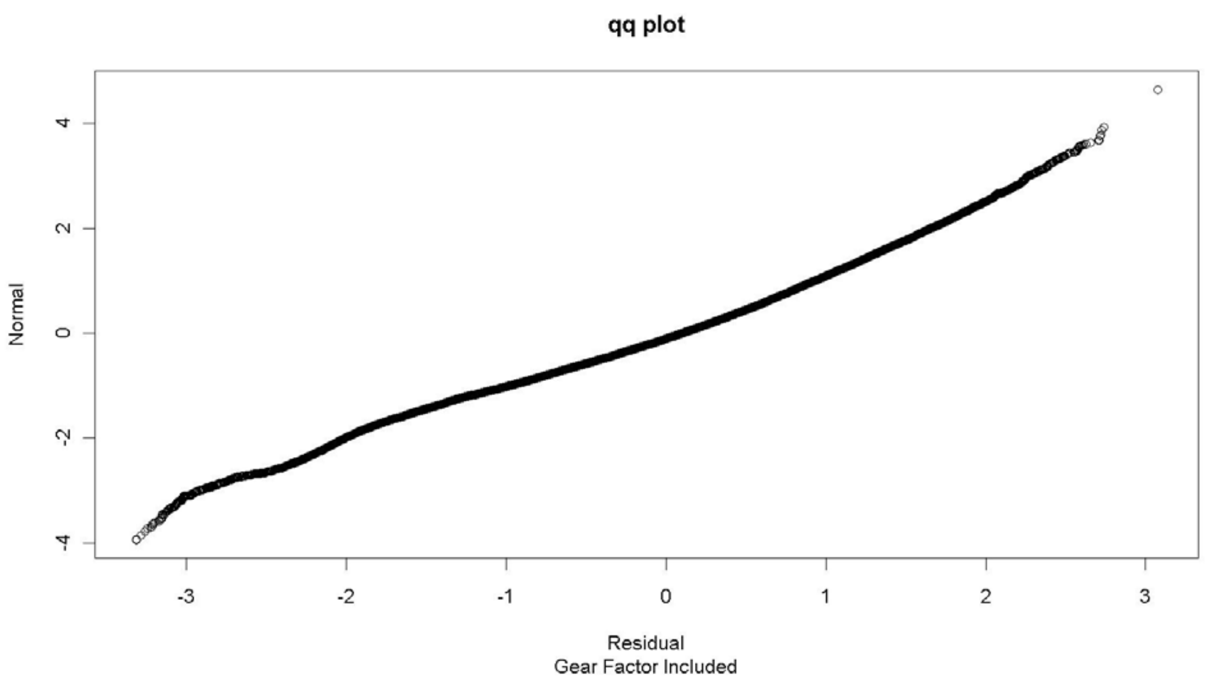


Figure 6. Residual and QQ plots of the best lognormal fit model for sublegal CPUE from combined east and west of 174 W. Observer pot sample data for 1990/91-2010/11 were used.

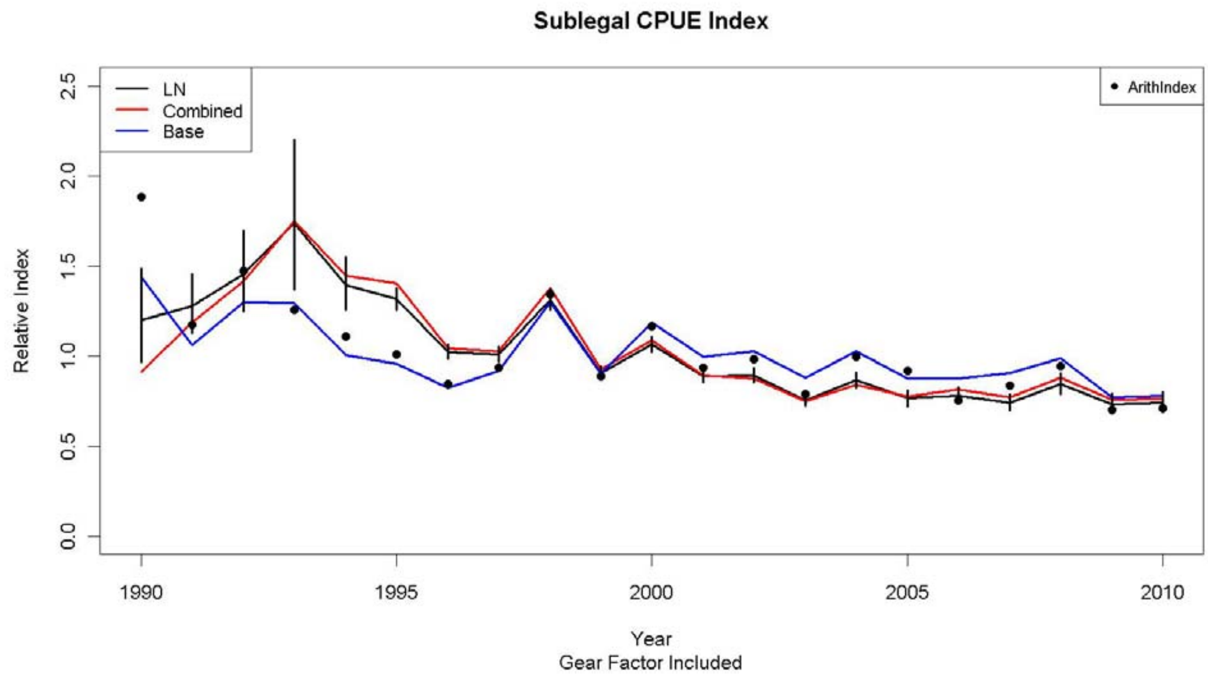


Figure 7. Trends in sublegal CPUE indexes for the observer pot sample data for the Aleutian Islands golden king crab fishery (combined east and west of 174 W). Lognormal: black line with 2standard errors; Combined: red line; Base: blue line; and Arithmetic: black filled circles.

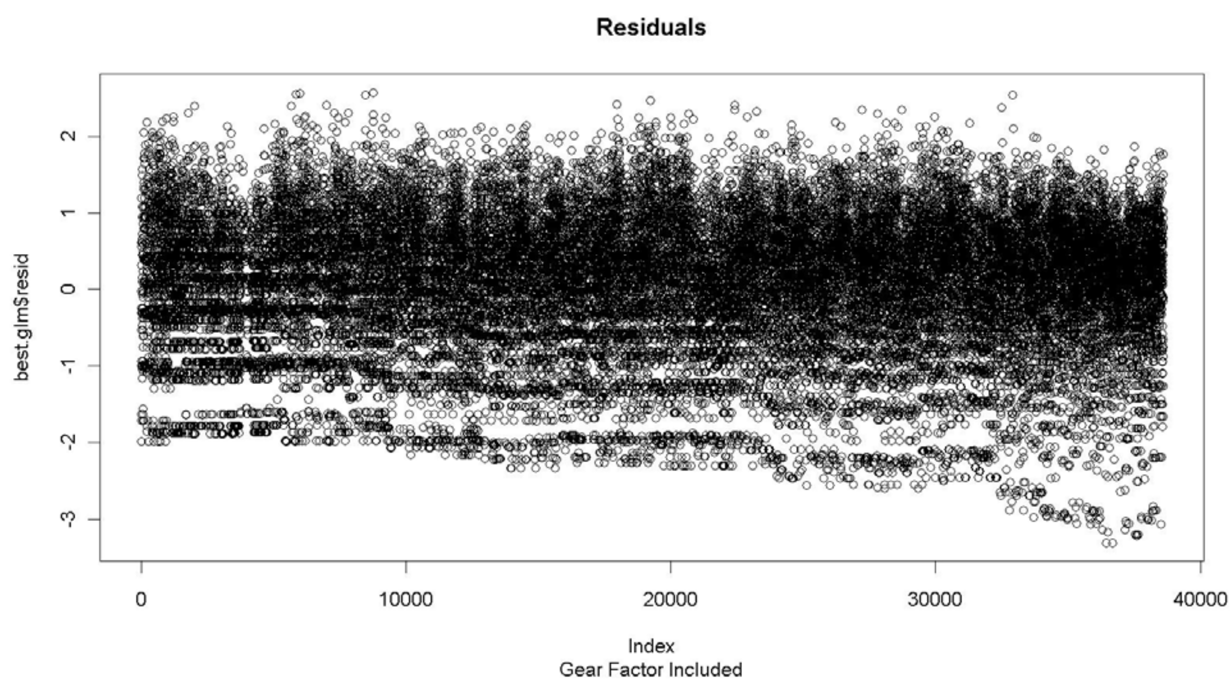


Figure 8. Residual plot of the best lognormal fit model for legal CPUE from east of 174 W. Observer pot sample data for 1995/96-2010/11 were used.

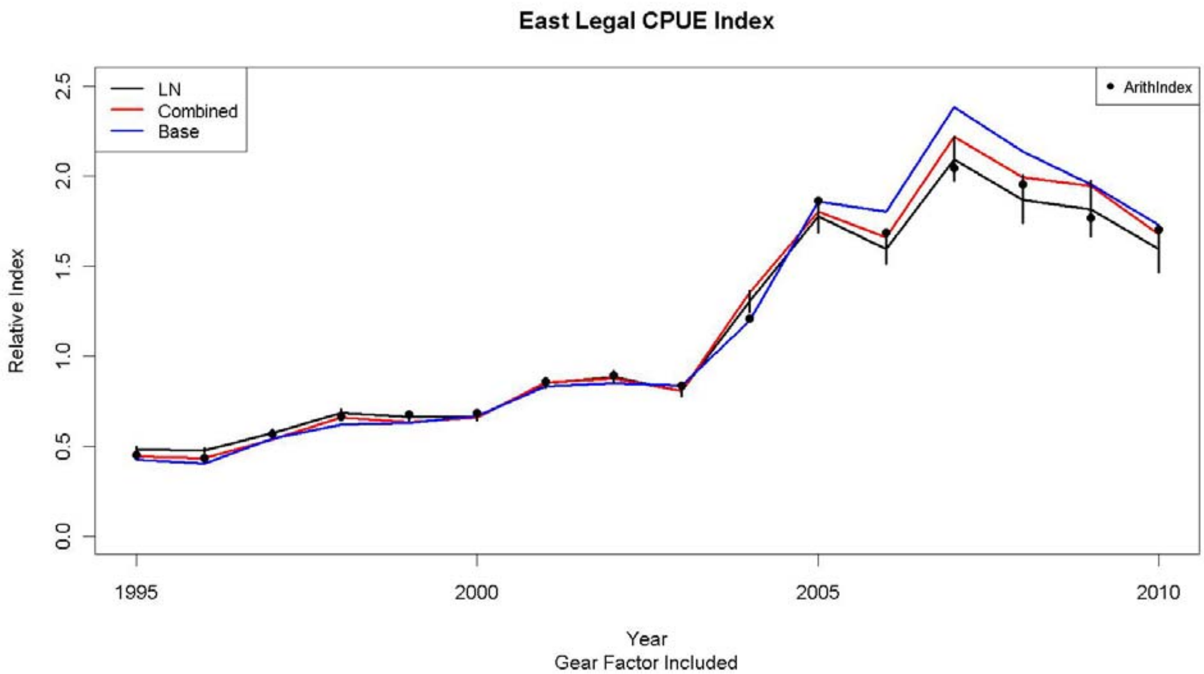


Figure 9. Trends in legal CPUE indexes for the observer pot sample data for the Aleutian Islands golden king crab fishery (east and of 174 W). Lognormal: black line with 2standard errors; Combined: red line; Base: blue line; and Arithmetic: black filled circles.

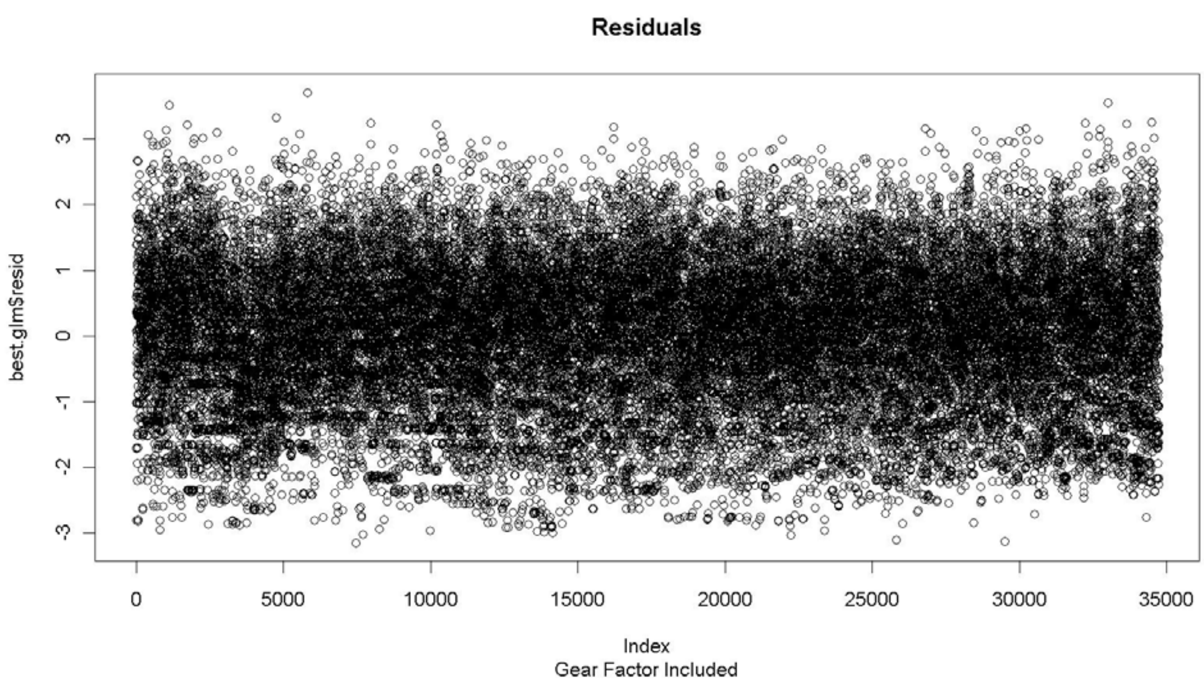


Figure 10. Residual plot of the best lognormal fit model for sublegal CPUE from east of 174 W. Observer pot sample data for 1995/96-2010/11 were used.

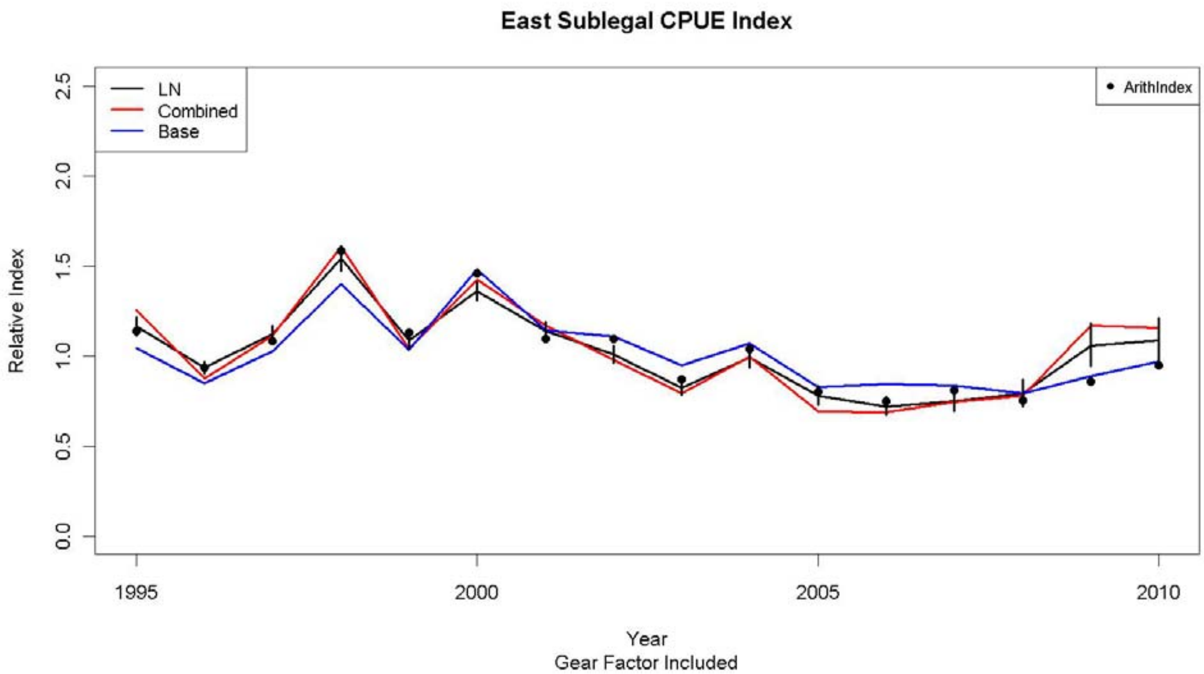


Figure 11. Trends in sublegal CPUE indexes for the observer pot sample data for the Aleutian Islands golden king crab fishery (east of 174 W). Lognormal: black line with 2standard errors; Combined: red line; Base: blue line; and Arithmetic: black filled circles.

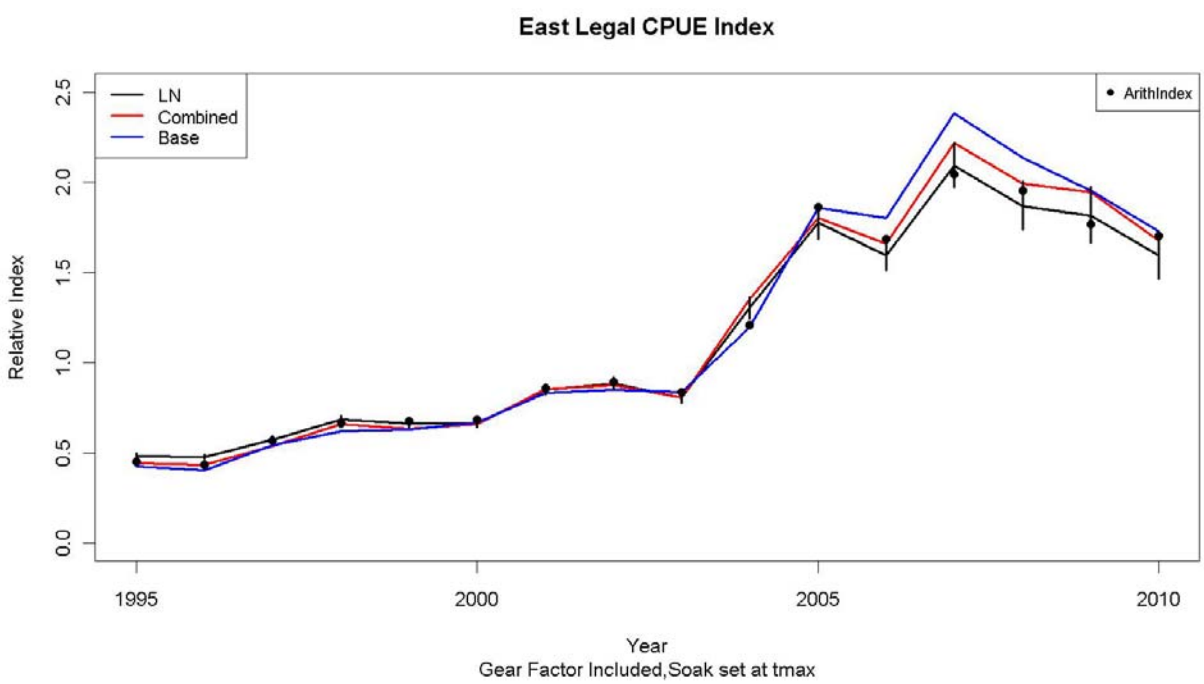
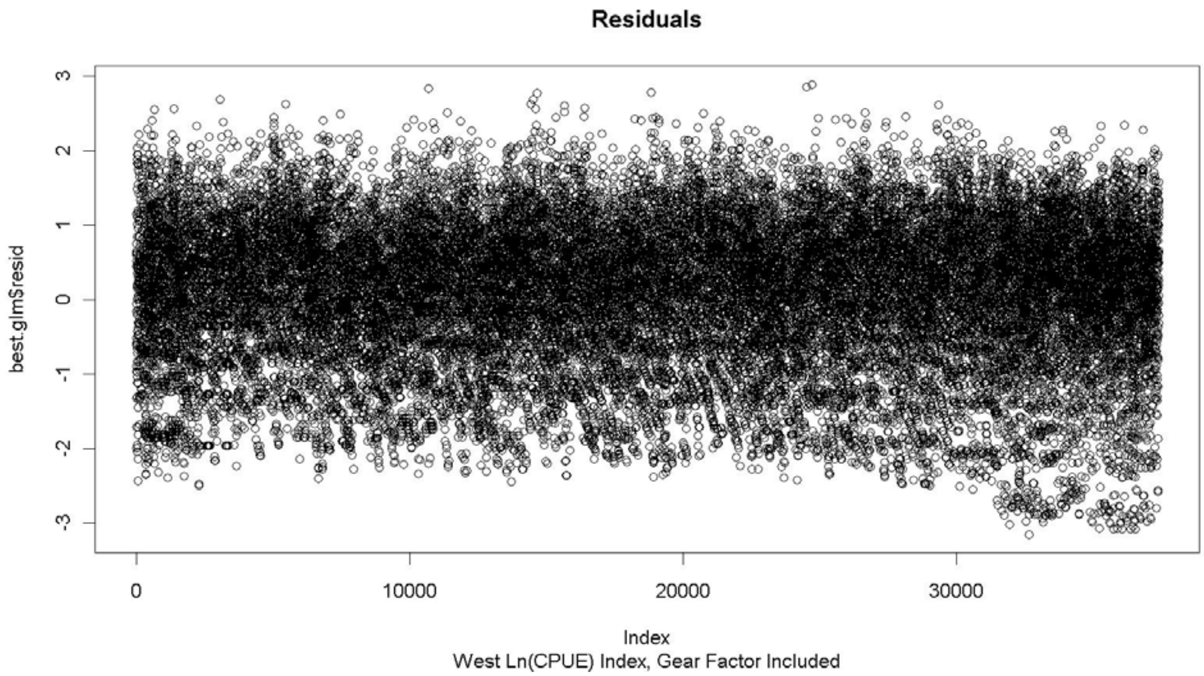


Figure 12. Trends in legal CPUE indexes for the observer pot sample data for the Aleutian Islands golden king crab fishery (east of 174 W). Lognormal: black line with 2standard errors; Combined: red line; Base: blue line; and Arithmetic: black filled circles. Soak time capped at maximum.



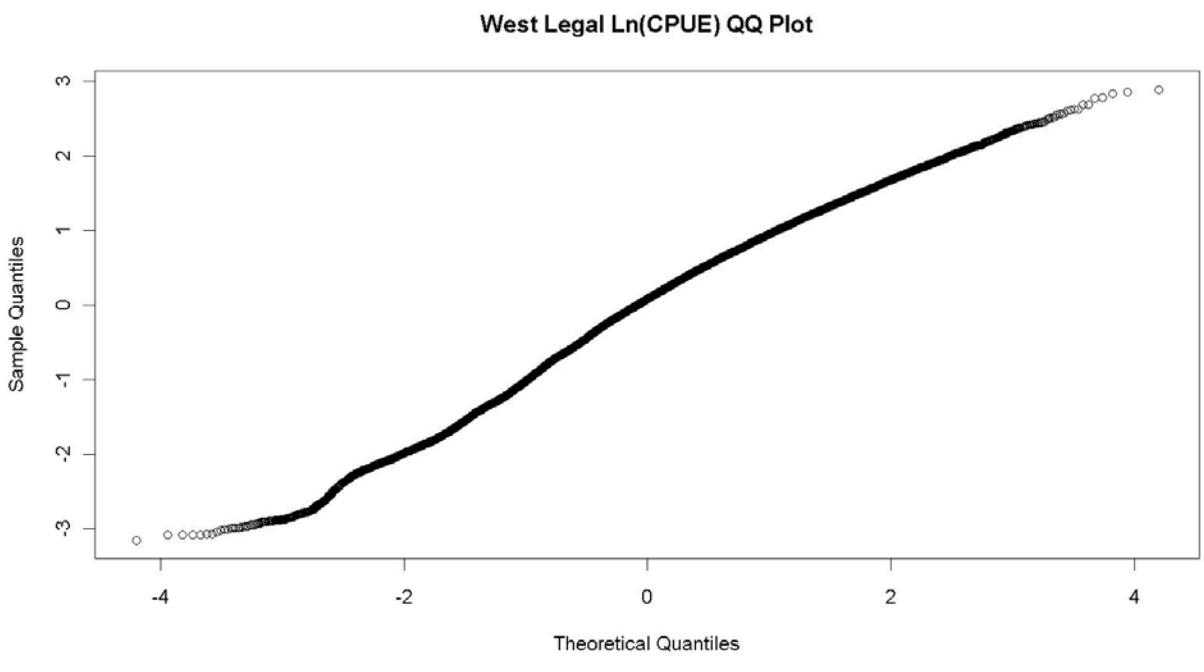


Figure 13. Residual and QQ plots of the best lognormal fit model for legal CPUE from west of 174 W. Observer pot sample data for 1990/91-2010/11 were used.

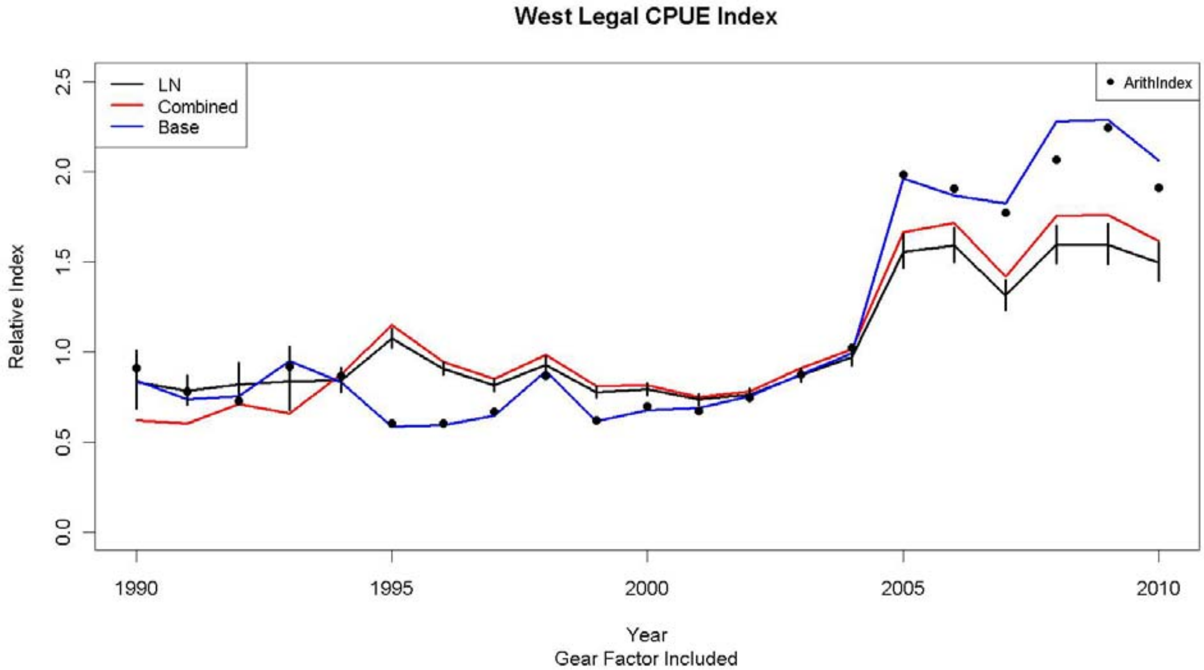
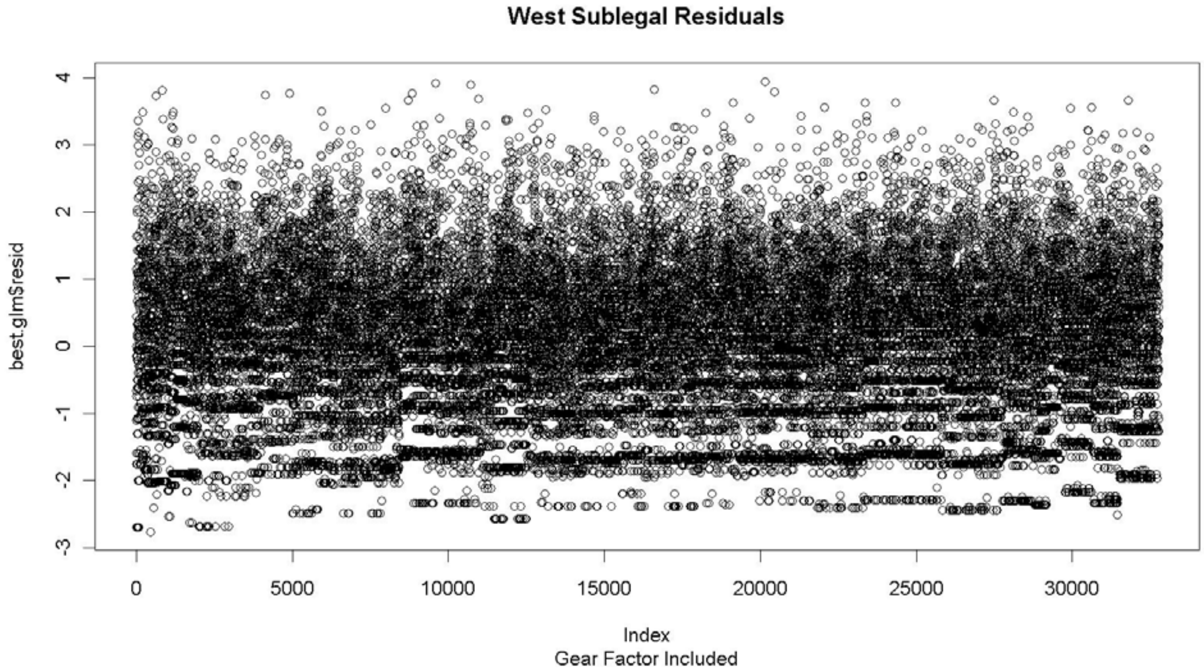


Figure 14. Trends in legal CPUE indexes for the observer pot sample data for the Aleutian Islands golden king crab fishery (west of 174 W). Lognormal: black line with 2standard errors; Combined: red line; Base: blue line; and Arithmetic: black filled circles.



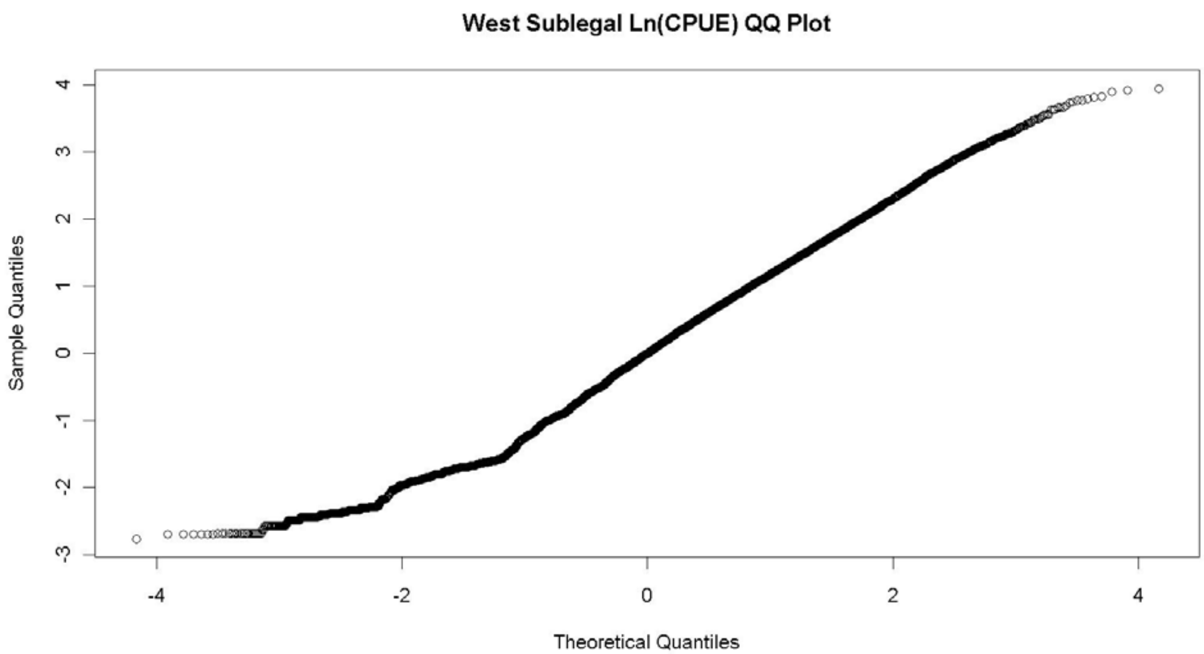


Figure 15. Residual and QQ plots of the best lognormal fit model for sublegal CPUE from west of 174 W. Observer pot sample data for 1990/91-2010/11 were used.

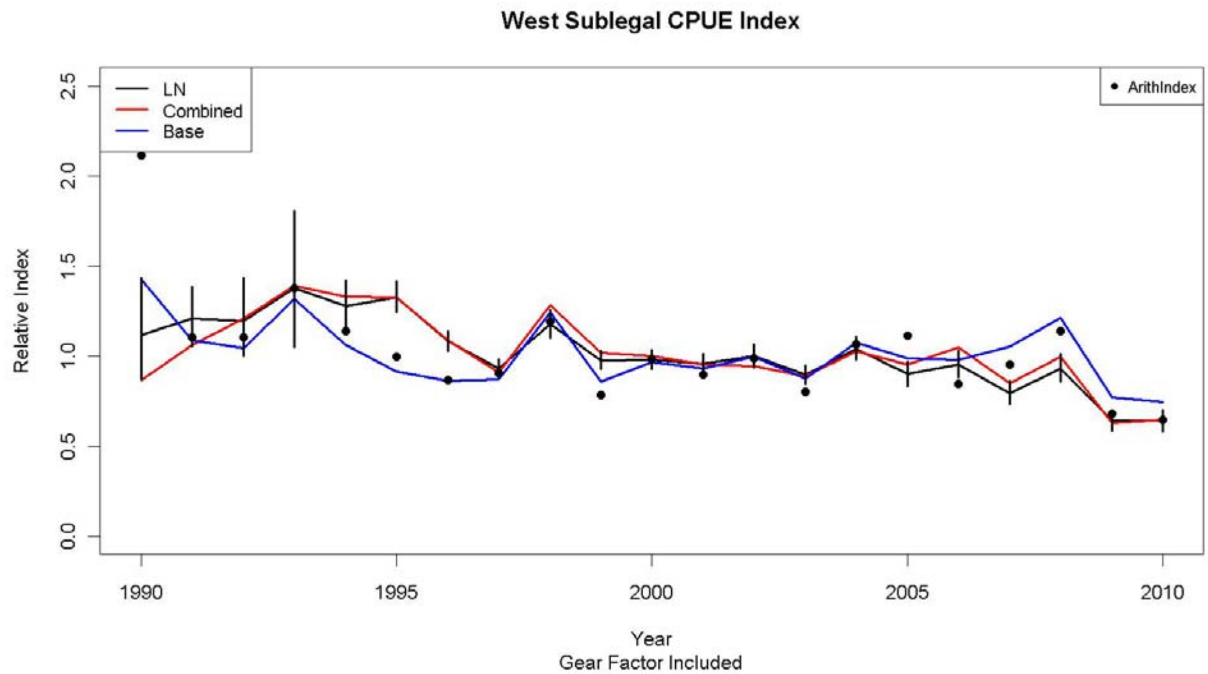


Figure 16. Trends in sublegal CPUE indexes for the observer pot sample data for the Aleutian Islands golden king crab fishery (west of 174 W). Lognormal: black line with 2standard errors; Combined: red line; Base: blue line; and Arithmetic: black filled circles.

Appendix

R script used in CPUE standardization.   The step CPUE (two R code files provided by Paul Starr) and the data file (restricted because of ADF&G privacy policy) are available with the first author.

 # initial environment setting

```
  options(contrasts=c("contr.treatment", "contr.poly"))

  options(object.size =100000000)

#
```

 # Read the observer data file

```
  allobsdat<- read.csv("c:/WorkRUpdate1/allobsbasedata_calc_trim1.csv", header=TRUE)

#
```

 # divide into pre-post rationalization periods

```
#

   preallobsdat<- allobsdat[allobsdat$FMPYear<2005,]

   postallobsdat<- allobsdat[allobsdat$FMPYear>=2005,]

#
```

 # Trim by percentile cutoff points of soak time

```
#

   preallobsdatcut <- preallobsdat[preallobsdat$SoakDays>0 & preallobsdat$SoakDays<40,]

   postallobsdatcut <- postallobsdat[postallobsdat$SoakDays>3 & postallobsdat$SoakDays<55,]

#
```

 # Combine the pre- and post- into combined data file

```
#

  prepostobsdatrim<- rbind(preallobsdatcut, postallobsdatcut)

  allobsdatrim<- prepostobsdatrim

#
```

 # Change some data frame variables to factors

```
#

 allobsdatrim$Date<- as.Date(allobsdatrim$Date,format="%m/%d/%Y")
```

```
 allobsdatrim$FMPYear<- as.factor(allobsdatrim$FMPYear)

 allobsdatrim$PotSampYear<- as.factor(allobsdatrim$PotSampYear)

 allobsdatrim$PotSampMonth<- as.factor(allobsdatrim$PotSampMonth)

 allobsdatrim$Trip<- as.factor(allobsdatrim$Trip)

 allobsdatrim$ADFG<- as.factor(allobsdatrim$ADFG)

 allobsdatrim$Area<- as.factor(allobsdatrim$Area)

 allobsdatrim$Gear<- as.factor(allobsdatrim$Gear)

#
```

 # add a (binomial) variable to the data set to reflect success or failure

```
#

 allobsdatrim$success[allobsdatrim$Mtotal>0]<- 1

 allobsdatrim$success[allobsdatrim$Mtotal==0]<- 0

#
```

 # select core data

```
#

  datacore<- allobsdatrim[allobsdatrim$Yrsof5>=3,]

#
```

 # Calculate the series of proportions zero (unsuccessful)

```
#

  prop.zero<- (table(datacore$FMPYear)-
table(datacore$FMPYear[datacore$success==1]))/table(datacore$FMPYear)

#
```

 # Subset core data by positive catch values for lognormal fit

```
#

  datacore1<- datacore[datacore$success==1,]

#
```

 # Find the best model from lognormal fit by glm and stepwise glm

```
#

  glm.object<- glm(log(Mtotal)~FMPYear,data=datacore1)
```

```
  obsdatout<- stepCPUE(glm.object,scope=list(upper=
~FMPYear+PotSampMonth+ADFG+Gear+poly(SoakDays,3)+poly(Depth,3)+FTPotlifts),lower=
~FMPYear,direction="forward",trace=9,r2.change=0.01)
```

#

 # Results from lognormal fit of the best model

#

```
  best.glm<-glm(log(Mtotal)~ FMPYear+Gear+PotSampMonth+ADFG,y=TRUE, data=datacore1)
```

#

#Get relative lognormal indices (with the base year =1)

```
      sumglm<-summary(best.glm)

      coefsglm <- exp(as.numeric(c(0, sumglm$coefficients[2:21,1])))
```

#get canonical lognormal indices

```
      cpue.glm<-getCPUE(best.glm,2:21, 1990:2010)

  write.csv(cpue.glm,"C:/WorkRUpdate1/allobsYearlyLnCPUEIndex.csv",row.names=F)
```

 # Get base year relative lognormal indices (with the base year =1)

```
  base.glm<-glm(log(Mtotal)~ FMPYear,y=TRUE, data=datacore1)

      sumglm1<-summary(base.glm)

      coefsbaseglm <- exp(as.numeric(c(0, sumglm1$coefficients[2:21,1])))
```

 # get canonical lognormal indices for the base index

```
      cpue1.glm<-getCPUE(base.glm,2:21, 1990:2010)

 write.csv(cpue1.glm,"C:/WorkRUpdate1/allobsBaseYearLnCPUEIndex.csv",row.names=F)
```

#

 # Find the best binomial model

```
  glm.object2<- glm(success~FMPYear,family=binomial(link=logit),data=datacore)

  obsdatout2<- stepCPUE(glm.object2,scope=list(upper=
~FMPYear+PotSampMonth+ADFG+Gear+poly(SoakDays,3)+poly(Depth,3)+FTPotlifts),lower=
~FMPYear,family=binomial(link=logit),direction="forward",trace=9,r2.change=0.01)
```

 # Results from binomial fit of the best model

```
  best2.glm<-glm(success ~
FMPYear+poly(Depth,3)+ADFG+PotSampMonth+Gear,family=binomial(link=logit),y=TRUE,data=datacore)
```

#Get relative binomial indices (with the base year =1)

```
        sumglm2<-summary(best2.glm)
        coefsbin <- exp(as.numeric(c(0, sumglm2$coefficients[2:21,1])))
# get canonical binomial indices
        cpue2.glm<-getCPUE(best2.glm,2:21, 1990:2010)
 write.csv(cpue2.glm,"C:/WorkRUpdate1/allobsYearlyBinomCPUEIndex.csv",row.names=F)
# Get base relative binomial indices (with the base year =1)
  base3.glm<-glm(success~ FMPYear,family=binomial(link=logit),y=TRUE, data=datacore)
        sumglm3<-summary(base3.glm)
        coefsbasebin <- exp(as.numeric(c(0, sumglm3$coefficients[2:21,1])))
#
# Calculate combined indices combining lognormal and binomial indexes
        n<-length(coefsglm)
        Comb<-rep(0,n)
        for(i in 1:n){
        Comb[i]<-coefsglm[i]/(1-prop.zero[1]*(1-1/coefsbin[i]))}
# get canonical combined indices
        Combined <- Comb/exp(mean(log(Comb)))
   write.csv(Combined,"C:/WorkRUpdate1/allobsCombCPUEIndex.csv",row.names=F)
#
# Arithmetic CPUE index
#
  RCPUE<- tapply(datacore1$Mtotal,datacore1$FMPYear,mean)
  GMRCPUE<- exp(mean(log(RCPUE)))
  RCPUEdash<- RCPUE/GMRCPUE
  write.csv(RCPUEdash,"C:/WorkRUpdate1/allobsScaledArithCPUEIndex.csv",row.names=F)
```