Standardization of CPUE data from the Aleutian Islands Golden King Crab Fishery: Observer and Fish Ticket Data

M.S.M. Siddeek[1], J. Zheng[1] , and Doug Pengilly[2]

[1] Alaska Department of Fish and Game

Division of Commercial Fisheries

P.O. Box 115526, Juneau, Alaska 99811

[2] Alaska Department of Fish and Game

211 Mission Road, Kodiak, AK 99615

shareef.siddeek@alaska.gov; jie.zheng@alaska.gov; doug.pengilly@alaska.gov

Summary report prepared for the Crab Plan Team September 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 (GKC) assessment model (Siddeek et al., 2011). We presented an initial analysis to standardize the observer CPUE data using generalized linear model (GLM) at the May 2012 Crab Plan Team (CPT) meeting. The CPT and the Scientific and Statistical Committee (SSC) made a number of recommendations to improve the standardization procedure. The CPT also recommended that we compare the assessment results using new and old standardized data. In this draft report we focused only on a comprehensive CPUE standardization procedure for observer and fish ticket (retained catch) data and deferred their use in the assessment model as the next step. We used the fish ticket retained catch data for 1985/86–2010/11 and observer pot sample data for 1995/91–2010/11 to compute yearly CPUE indexes with confidence intervals. We computed the indexes for the whole Aleutian Islands region, east and west of 174°W.

Responses to some major CPT and SSC comments

1. *For the combined CPUE index authors should estimate confidence intervals.*
   We used the Jackknife procedure, by removing one vessel at a time, to determine the mean, standard error, and confidence interval (personal communication, Doonan, New Zealand) for the combined CPUE Index for the observer data.
2. *GAM routine should be tried to address the non-linearity of Soak time and Depth effects.*
   We used the STEP.GAM routine to select the explanatory variables and then used the selected variables in the GLM. This way we could use the statistical strength of GLM to explore various diagnostic statistics.
3. *The longline fashion of setting pots in the GKC fishery causes response variable observations to be non independent.*
   In the short term, it is difficult to identify observer-sampled pots by longlined pot set. Furthermore, the current observer sampling design ensures independence (personal communication, Gaeuman, Kodiak). A number of residual diagnostics shown in this document also support the assumption of independence.
4. *Year vs. Soak time, Soak-time vs. Gear, or various other interactions factors can be considered.*
   Year factor is an important explanatory variable related to abundance. If there is year interaction with other factors, it will cause a problem for interpreting the year and abundance relationship. The generalized variance inflation factor (GVIF) determined for the variables selected to the model did not show any collinearity. Hence we did not include interactions terms in the models.
5. *Lognormal fit of CPUE needs a bias correction factor (SSC comment).*
   Since we are computing CPUE ratio relative to base year as an index we presumed the bias is minimum.

Method

Preliminary data processing

Fish Ticket Data

The Aleutian Islands golden king crab fish ticket data totaled 28839 records for 1985/86–2010/11 after removing some incomplete records. There were no zero catches. Hence only a lognormal model was considered to analyze this data set. 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 is 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 when crab were caught. 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.

East/West = The east/west subdivision code, 1 for east and 2 for west of 174°W. 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.

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.

Captain Code = Identification code for Captain. This is another important variable because Captain contributes to capture efficiency. This is treated as a predictor factor variable in the model.

Stat Area = ADFG identification code for a fishing area. Crab abundance varies by area. This is treated as a predictor factor variable in the model.

Catch/Pot-lifts = CPUE= Number of crabs caught divided by the total number of pot lifts realized in the trip completed by the reporting vessel. This is the numerical response variable in the model.

Observer Pot Sample Data

The observer pot sample data of Aleutian Islands golden king crab totaled 101,627 records for 1990/91–2010/11 after removing some incomplete records. We used the 1995/96–2010/11 data in the analysis because of a gap in the data series (1994/95 fishery not sampled) for east of 174°W. The following variables from each record were considered in the model:

Year = Federal Fisheries Management Year. This is treated as a predictor factor variable in the model.

Month = A calendar month in a fishing year. This is treated as a predictor factor variable in the model.

Vessel = Identification code for a participating vessel. This is treated as a predictor factor variable in the model.

Captain Code = Identification code for Captain. This is treated as a predictor factor variable in the model.

Stat Area = ADFG identification code for a fishing area. This is treated as a predictor factor variable in the model.

Catch = Number of crabs caught. This is the response variable. We considered two types of crab catch as response variable: Legal males (retained and non retained legal size crab) and sublegal males. Because the stock assessment model is designed for males only, we did not consider female catch. This is treated as a numerical response variable in the model.

East/West = The east/west subdivision code, 1 for east and 2 for west of 174°. We subset the data to 1, 2 or 1 & 2 for model fit.

Depth = Depth in fathoms. 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 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 numerical predictor 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 (Records ranged from 0 to 384 soak days) observed in certain years, we arbitrarily selected records within 5% to 95% soak time (Table 2). 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 5 to 95% soak time range, the number of records reduced from 102,849 to 101,627.

There was a maximum of 162 vessel 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 1990/91–2010/11. 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.

CPUE Standardization

For fish ticket data, we used the stepwise generalized linear model (GLM) procedure 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. For observer data, we used the stepwise generalized additive model (GAM) procedure to select the best model and then used the selected explanatory variables in the GLM to estimate a time series of CPUE indexes based on the relationship between CPUE vs. available predictive factor and continuous variables. We will provide the GAM after the GLM model specifications.

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 set 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, including 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 an R^2 based on model deviance and 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.

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

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

$$A'_y = \frac{A_y}{\bar{A}} \qquad (9)$$

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.

We used the STEP.GAM procedure to select the explanatory variables to use in the final GLM model. GAM model is similar to GLM, but the functions of explanatory numerical variables can be expressed by smooth functions s() or lo() which can fit the data locally rather than globally as done by GLM.

$$Y_i \sim X1_i + X2_i + s(X3_i, df) + s(X4_i, df) + \varepsilon_i \qquad (10)$$

The advantage of GAM is that you need not specify the functional form of the numerical explanatory variable. The data determine the form of s() or lo() functions.

Jackknife statistics

First we generated a jackknife sample (Manly, 1997) which has the value x_i (e.g., i th vessel) removed and then computed the ith partial estimate of the Combined index, call Comb_i. We then turned this ith partial estimate into the pseudo value Comb_i*, using equation

$$Comb_i^* = n \times Comb - (n-1) \times Comb_i \qquad (11)$$

where *Comb* is the combined index estimated using the whole data set (not removing any vessel); n is the number of sample points. The mean, variance, and standard errors of the pseudo value (there are i values) are estimated using

$$mean\ (Comb_i^*) = \sum_1^n Comb_i^* /n \qquad (12)$$

$$var\ (Comb_i^*) = \sum_1^n (Comb_i^* - mean(Comb_i^*))^2/(n(n-1)) \qquad (13)$$

$$\mathrm{std}\ (\mathrm{Comb}_i^*) = \sqrt{\mathrm{var}\ (\mathrm{Comb}_i^*)} \qquad (14)$$

can be estimated using standard formulas. These are the mean, variance, and standard error of the Combined CPUE index.

Test for collineaity

We used the generalized variance inflation factor (GVIF) to test for collineaity among selected predictor variables. Following Fox and Weisberg (2011), the estimated sampling variance of the jth regression coefficient can be written as

$$\widehat{var}(\beta_j) = \frac{\hat{\sigma}^2}{(n-1)\widehat{var}(X_j)} \times \frac{1}{1-R_j^2} \qquad (15)$$

Where $\hat{\sigma}^2$ is the estimated error variance, R_j^2 is the multiple R^2 for the regression of X_j on the other X covariates.

$\frac{1}{1-R_j^2}$ is called the VIF for linear (or GVIF for generalized linear) model β_j. If there are p regressors (df) in a X term, then $GVIF^{\frac{1}{2df}}$ is a one dimensional expression of the decrease in precision of estimation due to collinearity. $GVIF^{\frac{1}{2df}}$ is expected to be closer to 1 if there are no collinearity among X variables.

Software use

We coded in R to process the data (Appendix A).

Results

Fish ticket data analysis

Because the fish ticket records provide none zero catch and effort and hence CPUE, we applied only the lognormal GLM. We used the forward step-wise selection procedure to select the best model. We assumed the null model to be

$$\ln(I_i) \sim Year_{y_i} + \varepsilon_i \qquad (16)$$

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

$$\ln(I_i) \sim Year_{yi} + Month_{mi} + Area_{ai} + Vessel_{vi} + Captain_{ci} + \boldsymbol{\epsilon}_i \qquad (17)$$

where I = number of males caught per pot lift (catch/effort) in i th record ($CPUE_i$); and all predictor variables are self explanatory by name, and subscript of small characters are factor levels. The factor levels are defined in the observer data analysis section.

For fish ticket data, the forward GLM selection procedure produced a suite of final models for the three different subsets of data (Table 3). All three subsets of data produced Year and Captain Code as the predictor variables in the final model. Tables 4–6 provide the analysis of deviance values for lognormal fit, respectively for combined, east of 174°W, and west of 174°W retained catch data for 1985/86 through 2010/11. The variable rows are in order of their selection. The variables selected to the final model have significant R^2 values (>0.01). Table 7 provides generalized variance inflation factor (GVIF) values for combined, east of 174°W, and west of 174°W fish ticket data. The GVIF is an indicator of collinearity among explanatory variables. If the values in the last column substantially exceed 1 then there is collinearity. Our results did not show any collinearity and thus we did not pursue including interaction terms in the model.

Figures 4, 7, and 10 provide QQ-plot and studentized residual plots for the best lognormal fit to retained crab while Figures 5, 8, and 11 provide observed vs. predicted response variable and Pearson residuals vs. predictor variables for the combined, east of 174°W, and west of 174°W areas, respectively. None of the diagnostic plots appear to drastically violate any model assumptions.

The CPUE trend plots from the respective area combinations are presented in Figures 6, 9, and 12. For all three figures the three CPUE indexes (Standardized Index (StdIndex), Base Index, and Arithmetic Index (ArithIndex)) tracked closely after 1993; however, the Arithmetic index was lower than the other two indexes until then and higher after 1993.

Observer pot sample data analysis

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

$$\ln(I_i) \sim Year_{y_i} + Month_{m_i} + Vessel_{vi} + Captain_{ci} + Area_{ai} + Gear_{gi} + Soak_{si} + Depth_{di} + \varepsilon_i \quad (14)$$

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

$$\ln(I_i) \sim (1 + Year_{y_i}) \sim \left(1 + Month_{m_i}\right) \sim (1 + Vessel_{vi}) \sim (1 + Captain_{ci}) \sim (1 + Area_{ai}) \sim \left(1 + Gear_{gi}\right) \sim (1 + Soak_{si}) + poly(Soak_{si}, 2) + s(Soak_{si}, 4) + s(Soak_{si}, 6) + s(Soak_{si}, 8) + s(Soak_{si}, 10)) \sim (1 + Depth_{di}) + poly(Depth_{di}, 2) + s(Depth_{di}, 4) + s(Depth_{di}, 6) + s(Depth_{di}, 8) + s(Depth_{di}, 10)) + \varepsilon_i \quad (18)$$

We used the orthogonal polynomial of degree 2 and s() function of varying degrees in the exploratory model. Each factor variable either goes into the model as linear or is ignored (1 refers to ignoring the factor variable). Note that although the golden king crab fishery is prosecuted with only a single gear type (pots), the gear configuration has changed over the years, so different gear factor levels were considered for the model.

The factor levels considered for each factor variable are:

Year: 1985 to 2010 for retained catch or 1995 to 2010 for observer pot sample data;

Month: 1 to 12;

Vessel: Vessel registration number;

Captain: Captain identified number code; and

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

Then we used the binomial GAM 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 the same as on the right hand side of equation (18). We used a binomial logit link function in the selection process.

The GAM step-wise selection procedure selected a suite of best (final) models for different subsets of observer data (Table 8). Table 9 provides AIC values for the start and final selection models for various sets of data. Tables 10–12 provide GVIF values for combined, east, and west of 174°W observer data, respectively. Each table lists results separately for legal and sublegal crabs. The values in the last column are near 1 and do not show any collinearity problem.

Figures 13, 19, and 25 provide QQ-plot and studentized residual plots for the best (final) lognormal fit to legal crab while Figures 14, 20, and 26 provide observed vs. predicted response variable and Pearson residuals vs. predictor variables for the various year and area combinations. Figures 16, 22, and 28 provide QQ-plot and studentized residual plots for the best lognormal fit to sublegal crab while Figures 17, 23, and 29 provide observed vs. predicted response variable and Pearson residuals vs. predictor variables for the various year and area combinations. None of the diagnostic plots appear to drastically violate any model assumptions.

The CPUE trend plots from the different year and area combinations are presented in Figures 15, 21, and 27 for legal crab and in Figures 18, 24, and 30 for sublegal crabs. Five CPUE indexes, Combined, Log normal (Ln), Binomial (Binom), Base, and Arithmetic (Arith), are shown. Base index considers only the Year effect disregarding the influence of all other factors or numerical variables. The combined index considers positive and zero catches in the calculation and hence is considered the best among all the indexes. The lognormal CPUE index trend matches the combined CPUE index trend well for legal crab. The binomial, base, and arithmetic index values are higher than combined index values since 2004 for legal crabs. This pattern is not clear for sublegal crab, but the binomial index values are higher than all other indexes since 2005-2006. Substantially large binomial index values for legal crabs were observed for combined area data in 2008 (Figure 15) and for east of 174°W data in 2009 (Figure 21). For Figure 24 (sublegal crab) the combined index values show a rising trend in recent years.

Figures 31 and 32 summarize the combined CPUE index plots for various areas for legal and sublegal crab, respectively. The trends of legal crabs index for combined and east of 174°W observe data tracked well throughout the years, and since 2006, the values for west of 174°W data are lower than those for the combined and east of 174°W data. The combined CPUE index trends of sublegal crabs systematically decreased until 2008 and thereafter moved in opposite directions for different subsets of data (Figure 32).

Discussion

We estimated the time series of standardized lognormal CPUE indexes for retained catch data (Fish Ticket data) and combined (lognormal and binomial) indexes for observer data. Both indexes are provided with confidence intervals determined using the jackknife procedure. These indexes with their standard errors will be used in the stock assessment model for the various year and area combinations. The GAM step-wise procedure on observer data did not exclude the initial set of variables offered, but selected particular functional forms from many provided in the scope. Although GAM did not discard the Captain Code factor variable, we had to exclude it from the model because examination of GVIF indicated collinearity.

Acknowledgements

Gretchen Bishop spent time discussing and reviewing this draft. Paul Starr provided insight into CPUE data and interpretations, volunteered to review this draft, and provided some R scripts. Andre Punt and Crab Plan Team reviewed various versions of this draft. We thank them for their time and 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).

Fox, J. and S. Weisberg. 2011. An R companion to applied regression. 2nd edition Sage Publications, Inc. 449p.

Manly, B.F.J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. 2nd edition Chapman & Hall. 399p.

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. Standardized CPUE analysis exploration: using the rock lobster voluntary logbook and observer catch sampling programmes. New Zealand Fisheries Assessment Report 2012/34, 75p.

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)

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

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	5%–95% Percentile range (days)
East and West Combined	Pre-rationalization (1990/91–2004/05)	2–18
	Post-rationalization (2005/06–2010/11)	6–39
East 174W	Pre-rationalization	2–10
	Post-rationalization	5–28
West 174W	Pre-rationalization	2–24
	Post-rationalization	9–41

Table 3. Step-wise model selection for various scenarios for the Aleutian Islands golden king crab fish ticket (retained catch). GLM routine was used for selection of variables and final fit. Fish ticket data for 1985/86–2010/11 was used.

Area	Crab category	Final model
East and West combined	Legal	Ln(CPUE)~ Year+Captain Code
East 174W	Legal	Ln(CPUE)~ Year+Captain Code
West 174W	Legal	Ln(CPUE)~ Year+Captain Code

Table 4. Analysis of deviance for stepwise lognormal model selection the Aleutian Islands golden king crab fishery. The response variable is retained catch CPUE. 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). Fish ticket data from combined east and west of 174°W for 1985/86–2010/11 were used.

Factor	df (difference from null)	Deviance	Residual df	Residual deviance	R^2
Year			26108	-51.89	0.113
Captain Code	-188	-375.95	25920	-427.83	0.167

Table 5. Analysis of deviance for stepwise lognormal model selection the Aleutian Islands golden king crab fishery. The response variable is retained catch CPUE. Fish ticket data from east of 174°W for 1985/86–2010/11 were used.

Factor	df (difference from null)	Deviance	Residual df	Residual deviance	R^2
Year			8308	-51.917	0.083
Captain Code	-130	-259.96	8178	-311.881	0.119

Table 6. Analysis of deviance for stepwise lognormal model selection the Aleutian Islands golden king crab fishery. The response variable is retained catch CPUE. Fish ticket data from west of 174°W for 1985/86–2010/11 were used.

Factor	df (difference from null)	Deviance	Residual df	Residual deviance	R^2
Year			17774	-51.86	0.137
Captain Code	-143	-285.94	17631	-337.80	0.202

Table 7. Test for collinearity of selected variables for the fish ticket retained catch data for the Aleutian Islands golden king crab fishery. $GVIF^{\frac{1}{2df}}$values are close to 1, indicating lack of collinearity . Fish ticket data for 1985/86–2010/11 was used.

	GVIF	Df	$GVIF^{\frac{1}{2df}}$
Combined east and west of 174°W:			
Year	267.3843	25	1.118
Captain code	267.3843	188	1.015
East of 174°W:			
Year	905.9636	25	1.146
Captain code	905.9636	130	1.027
West of 174°W:			
Year	1224.831	25	1.153
Captain Code	1224.831	143	1.025

Table 8. Step-wise model selection for various scenarios for the Aleutian Islands golden king crab observer data. Step.gam routine was used for selection of variables and GLM was used for final fit. Observer data for 1995–2010 was used.

Area	Crab category	Final model
East and West Combined	Legal	Ln(CPUE)~ Year+Month+Vessel+Area+Gear+s(Soak,10)+s(Depth,8)
		Binomial(Success)~ Year+Month+Vessel+Area+Gear+s(Soak,4)+s(Depth,4)
	Sublegal	Ln(CPUE)~ Year+Month+Vessel+Area+Gear+poly(Soak,2)+s(Depth,10)
		Binomial(Success)~ Year+Month+Vessel+Area+Gear+poly(Soak,2)+s(Depth,10)
East 174°W	Legal	Ln(CPUE)~ Year+Month+Vessel+Area+Gear+s(Soak,8)+s(Depth,10)
		Binomial(Success)~ Year+Month+Vessel+Area+Gear+s(Soak,8)+s(Depth,4)
	Sublegal	Ln(CPUE)~ Year+Month+Vessel+Area+Gear+s(Soak,6)+s(Depth,10)
		Binomial(Success)~ Year+Month+Vessel+Area+Gear+s(Soak,4)+s(Depth,10)
West 174°W	Legal	Ln(CPUE)~ Year+Month+Vessel+Area+Gear+s(Soak,10)+poly(Depth,2)
		Binomial(Success)~ Year+Month+Vessel+Area+Gear+s(Soak,8)+s(Depth,8)
	Sublegal	Ln(CPUE)~ Year+Vessel+Area+Gear+poly(Soak,2)+s(Depth,10)
		Binomial(Success)~ Year+Month+Vessel+Area+Gear+poly(Soak,2)+s(Depth,10)

Table 9. AIC (= -2*Max.LogLikelihood+2* number of parameters) statistics on step.gam selection process for observer data. There were 101,627 records for East and West combined. Observer data for 1995–2010 was used.

Area	Crab category	Start model	Final model
East and West	Legal:		
Combined	Log(CPUE)	191136	190792
	Binomial (success)	56895	56589
	Sublegal:		
	Log(CPUE)	193551	193501
	Binomial (success)	74814	74596
East 174°W	Legal:		
	Log(CPUE)	93715	93496
	Binomial (success)	22717	22633
	Sublegal:		
	Log(CPUE)	96489	96377
	Binomial (success)	32778	32717
West 174°W	Legal:		
	Log(CPUE)	94910	94662
	Binomial (success)	33637	33359
	Sublegal:		
	Log(CPUE)	93759	93699
	Binomial (success)	41654	41481

Table 10. Test for collinearity of selected variables for the observer data for the Aleutian Islands golden king crab fishery. $GVIF^{\frac{1}{2df}}$ values are closer to 1 indicating no collinearity. Observer data from combined east and west of 174°W for 1995–2010 was used.

	GVIF	Df	$GVIF^{\frac{1}{2df}}$
Legal:			
Year	18.615	15	1.102
Month	7.950	11	1.099
Vessel	11383.98	21	1.249
Area	3868.936	155	1.027
Gear	44.067	14	1.145
Poly(SoakDays, 2)	3.672	2	1.384
poly(Depth, 2)	1.692	2	1.140
Sublegal:			
FMPYear	18.257	15	1.102
PotSampMonth	7.917	11	1.099
ADFG	11476.81	21	1.249
Statarea	3405.653	151	1.027
Gear	46.9134	14	1.147
poly(SoakDays, 2)	3.613	2	1.379
poly(Depth, 2)	1.682	2	1.139

Table 11. Test for collinearity of selected variables for the observer data for the Aleutian Islands golden king crab fishery. $GVIF^{\frac{1}{2df}}$ values are closer to 1 indicating no collinearity. Observer data from east of 174°W for 1995–2010 was used.

	GVIF	Df	$GVIF^{\frac{1}{2df}}$
Legal:			
Year	96.033	15	1.164
Month	16.392	11	1.136
Vessel	5781.154	20	1.242
Area	400.340	43	1.072
Gear	48.597	14	1.149
Poly(SoakDays, 2)	3.694	2	1.386
poly(Depth, 2)	1.819	2	1.161
Sublegal:			
FMPYear	102.349	15	1.167
PotSampMonth	17.987	11	1.140
ADFG	6102.827	20	1.243
Statarea	355.003	42	1.072
Gear	56.751	14	1.155
poly(SoakDays, 2)	3.709	2	1.388
poly(Depth, 2)	1.826	2	1.162

Table 12. Test for collinearity of selected variables for the observer data for the Aleutian Islands golden king crab fishery. $GVIF^{\frac{1}{2df}}$ values are closer to 1 indicating no collinearity . Observer data from west of 174°W for 1995–2010 was used.

	GVIF	Df	$GVIF^{\frac{1}{2df}}$
Legal:			
Year	32.997	15	1.124
Month	5.666	11	1.082
Vessel	4947.594	19	1.251
Area	386.512	111	1.027
Gear	84.293	14	1.172
Poly(SoakDays, 2)	3.365	2	1.354
poly(Depth, 2)	1.498	2	1.106
Sublegal:			
FMPYear	15.469	15	1.096
ADFG	2716.002	18	1.246
Statarea	259.121	108	1.026
Gear	53.877	14	1.153
poly(SoakDays, 2)	3.199	2	1.337
poly(Depth, 2)	1.404	2	1.089

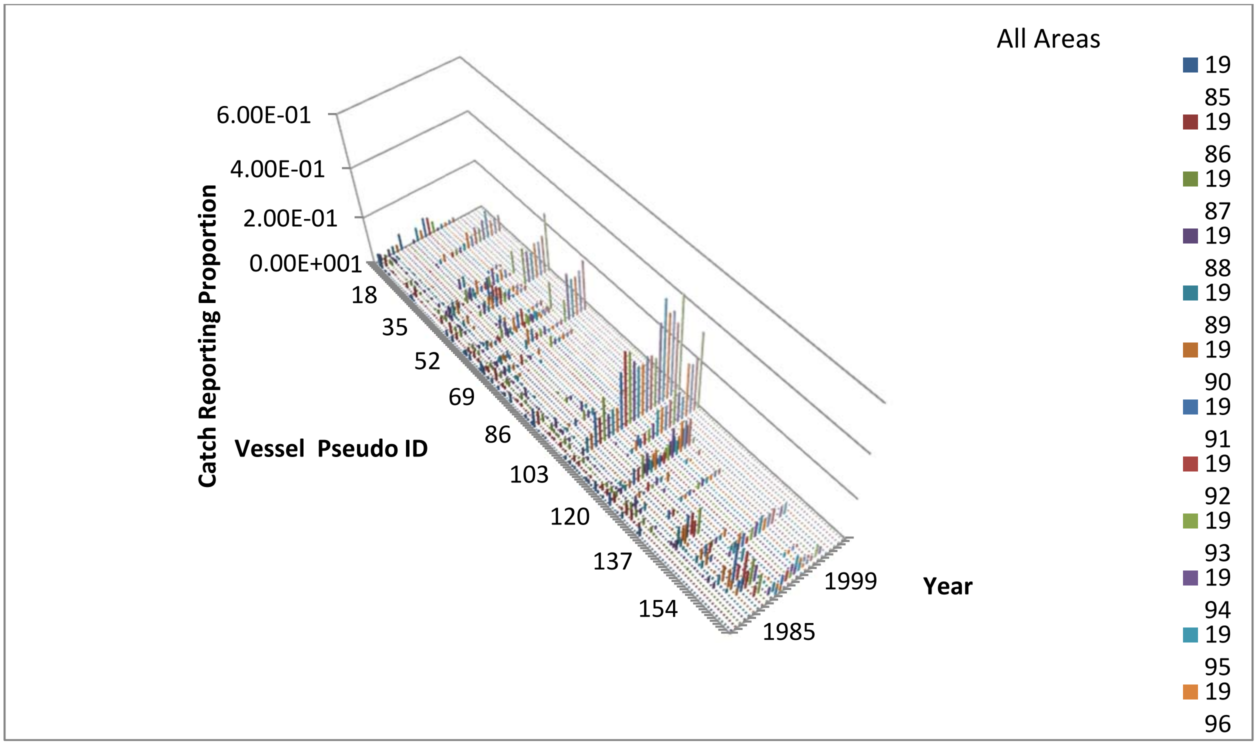


Figure 1. Golden king crab catch reporting frequency by vessel from Aleutian Islands. Fish ticket data from combined east and west of 174°W for 1985/86–2010/11 were used.

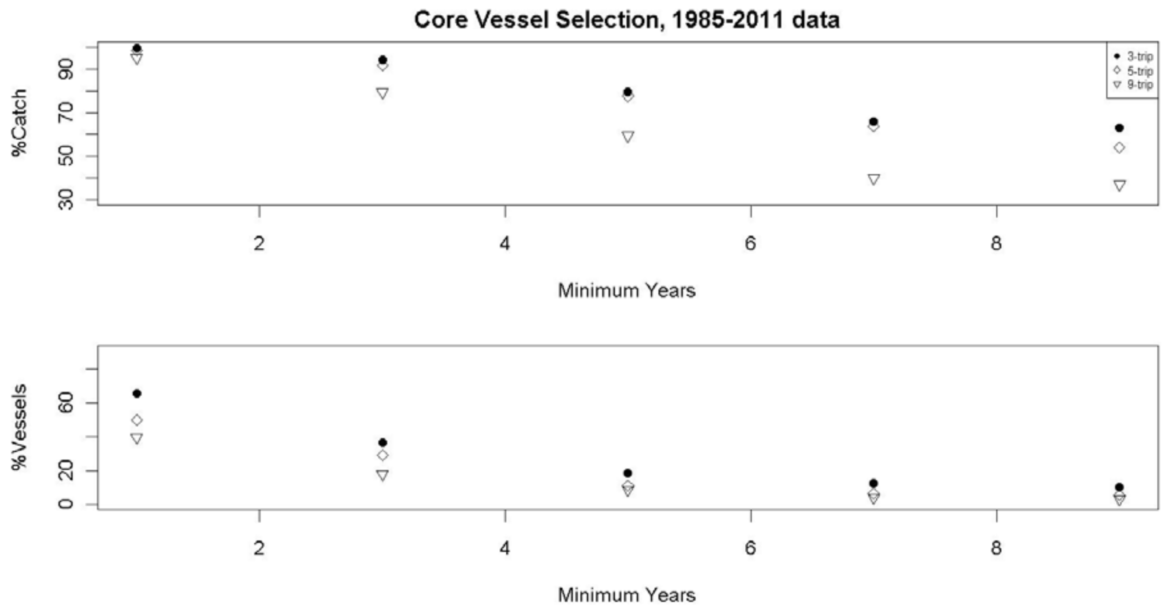


Figure 2. Core vessel selection for the Aleutian Islands golden king crab fishery. Fish ticket data from combined east and west of 174°W for 1985/86–2010/11 were used. 3-trip = three trips per year; 5-trip = five trips per year; and 9-trip = nine trips per year . The percentage catch and vessels dropped as the number of minimum years the vessels with those yearly reporting rates increased.

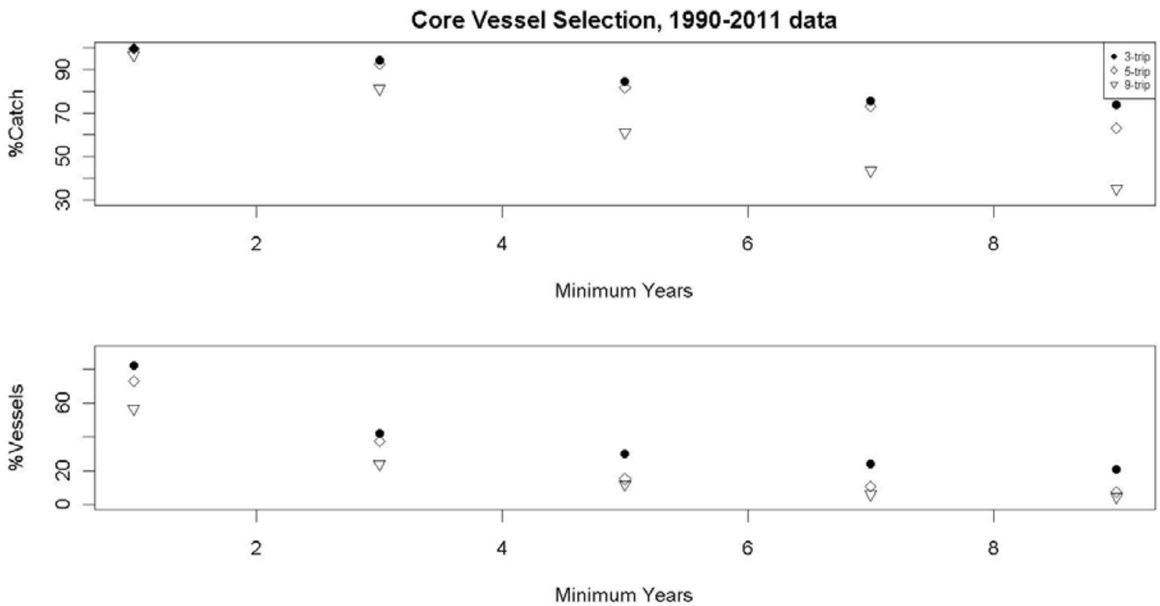


Figure 3. Core vessel selection based for the Aleutian Islands golden king crab fishery. Fish ticket data from combined east and west of 174°W for 1990/91–2010/11 were used. 3-trip = three trips per year; 5-trip = five trips per year; and 9-trip = nine trips per year. The percentage catch and vessels dropped as the number of minimum years the vessels with those yearly reporting rates increased.

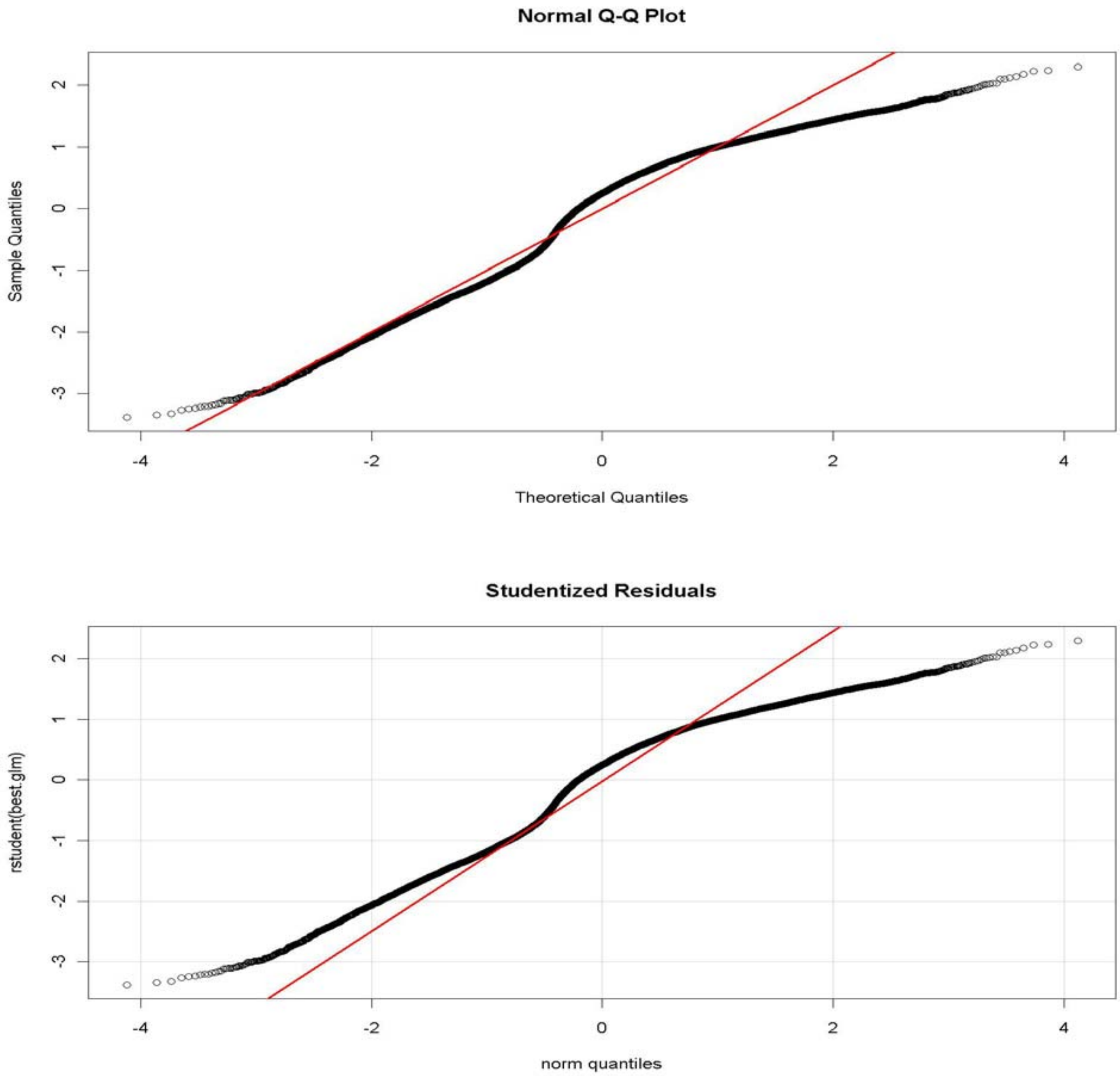


Figure 4. QQ and studentized residual plots of the best lognormal fit model for retained catch CPUE. Fish ticket data from combined east and west of 174°W for 1985/86–2010/11 were used.

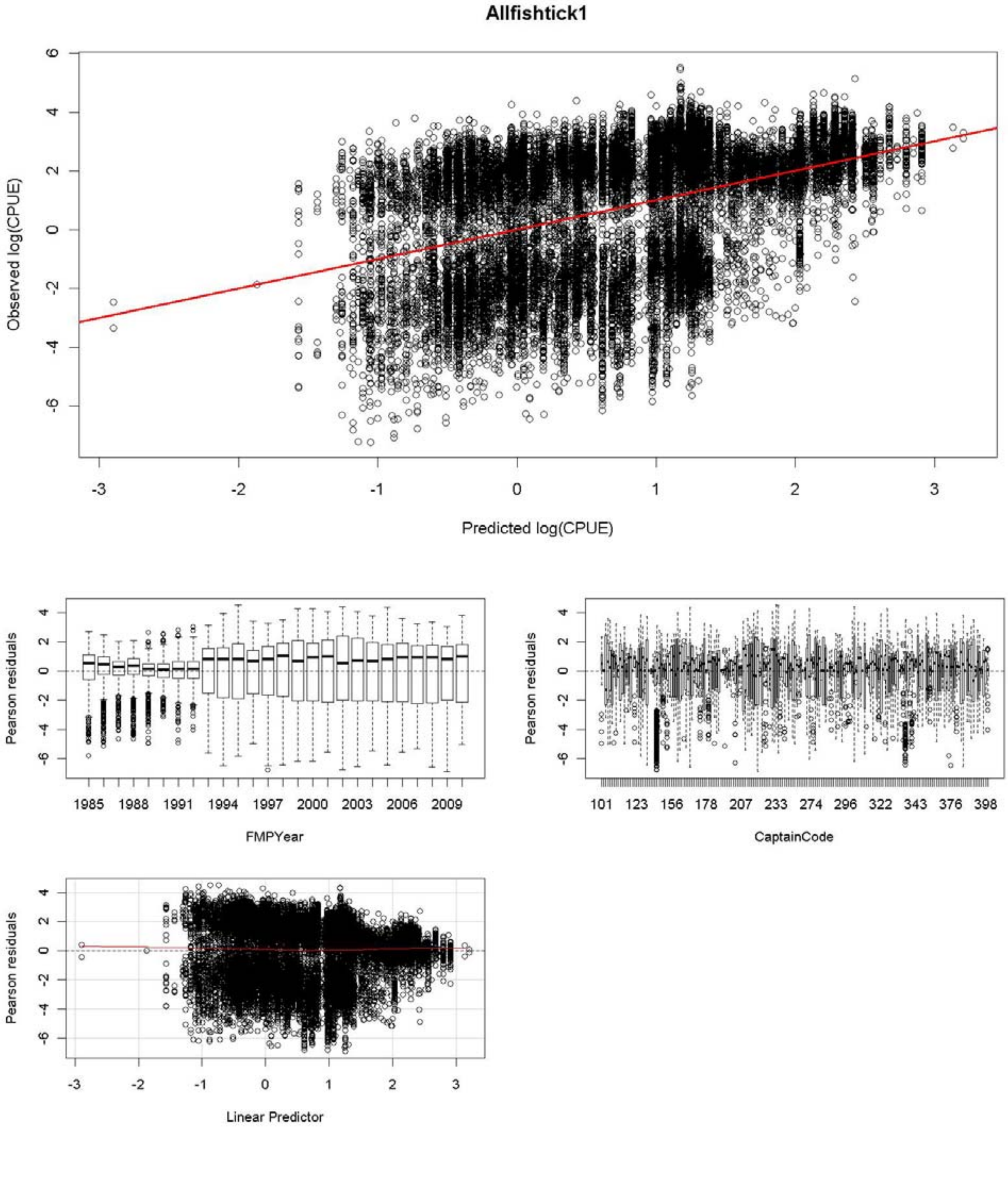


Figure 5. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for retained catch CPUE. Fish ticket data from combined east and west of 174°W for 1985/86–2010/11 were used.

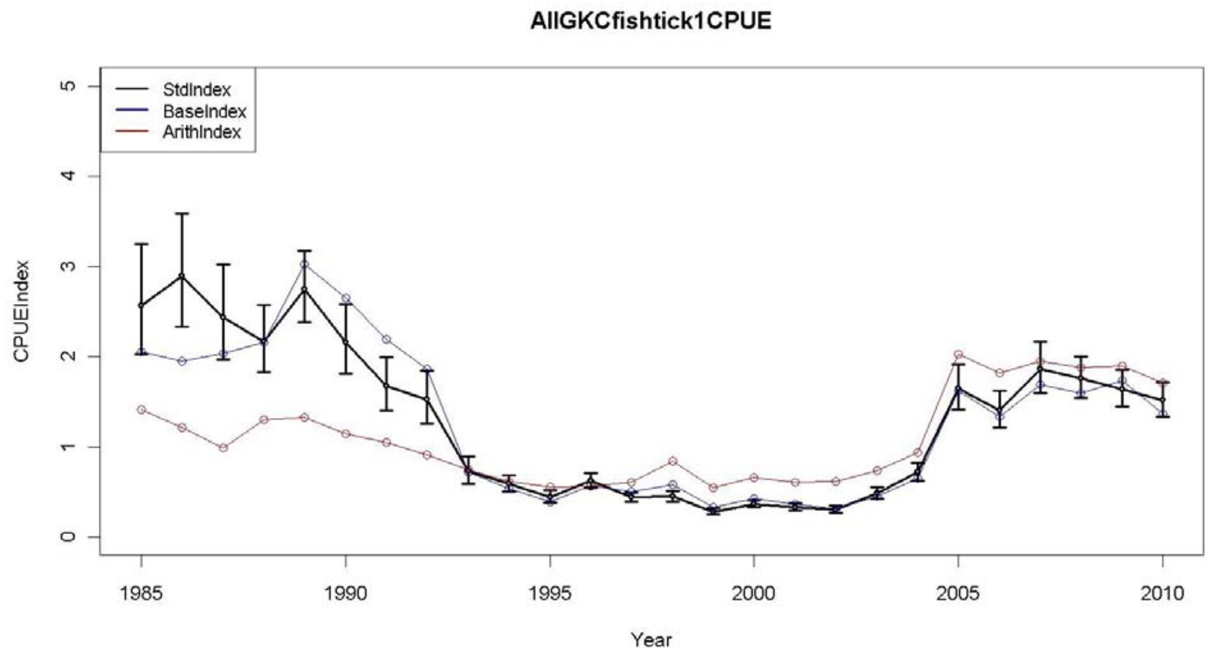


Figure 6. Trends in retained catch CPUE indexes for the Aleutian Islands golden king crab fishery. Standardized Index (Lognormal): black line with 2 standard errors; Base Index: blue line; and Arithmetic Index: red line. Fish ticket data from combined east and west of 174°W for 1985/86–2010/11 were used.

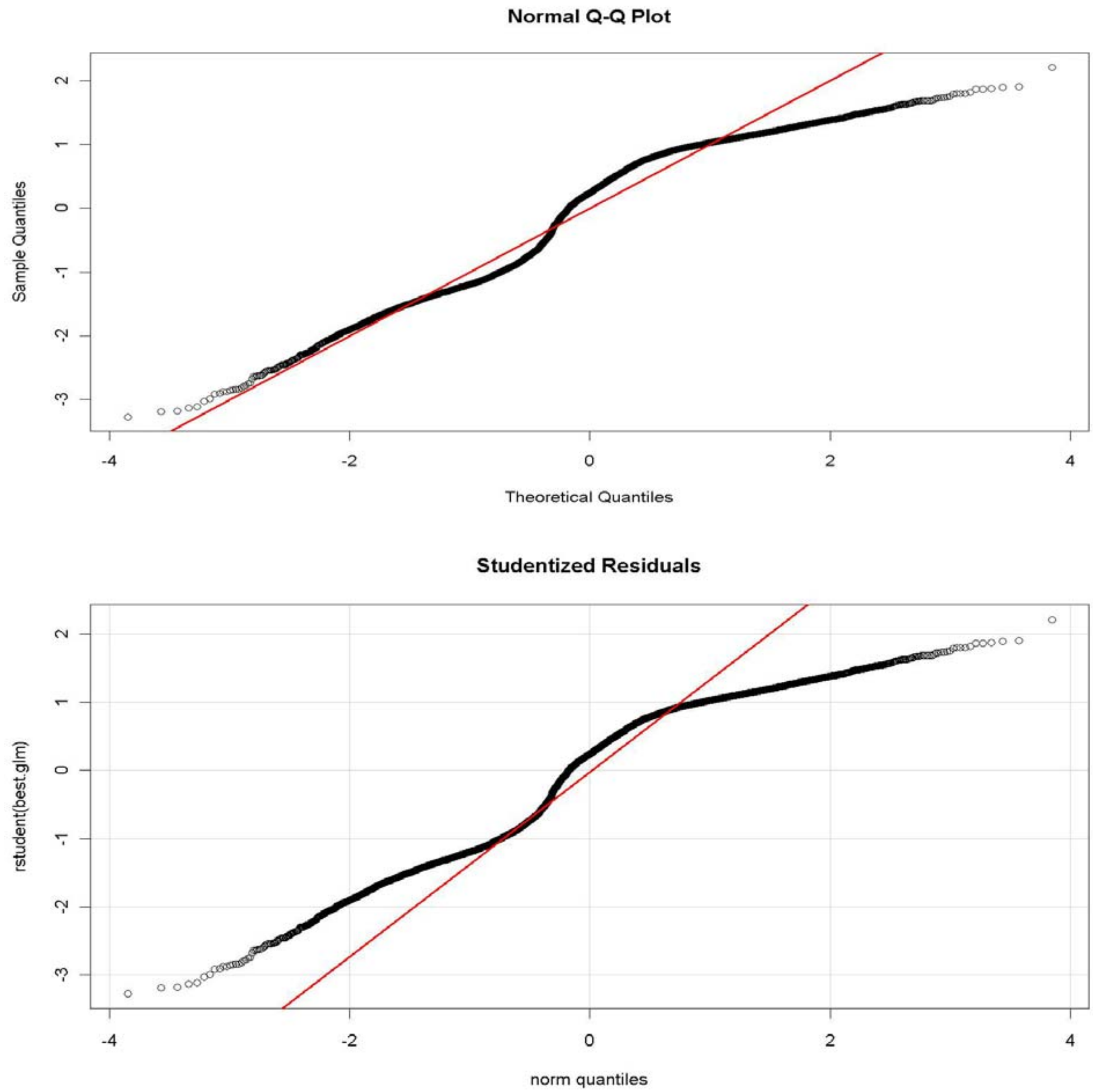


Figure 7. QQ and studentized residual plots of the best lognormal fit model for retained catch CPUE. Fish ticket data from east of 174°W for 1985/86–2010/11 were used.

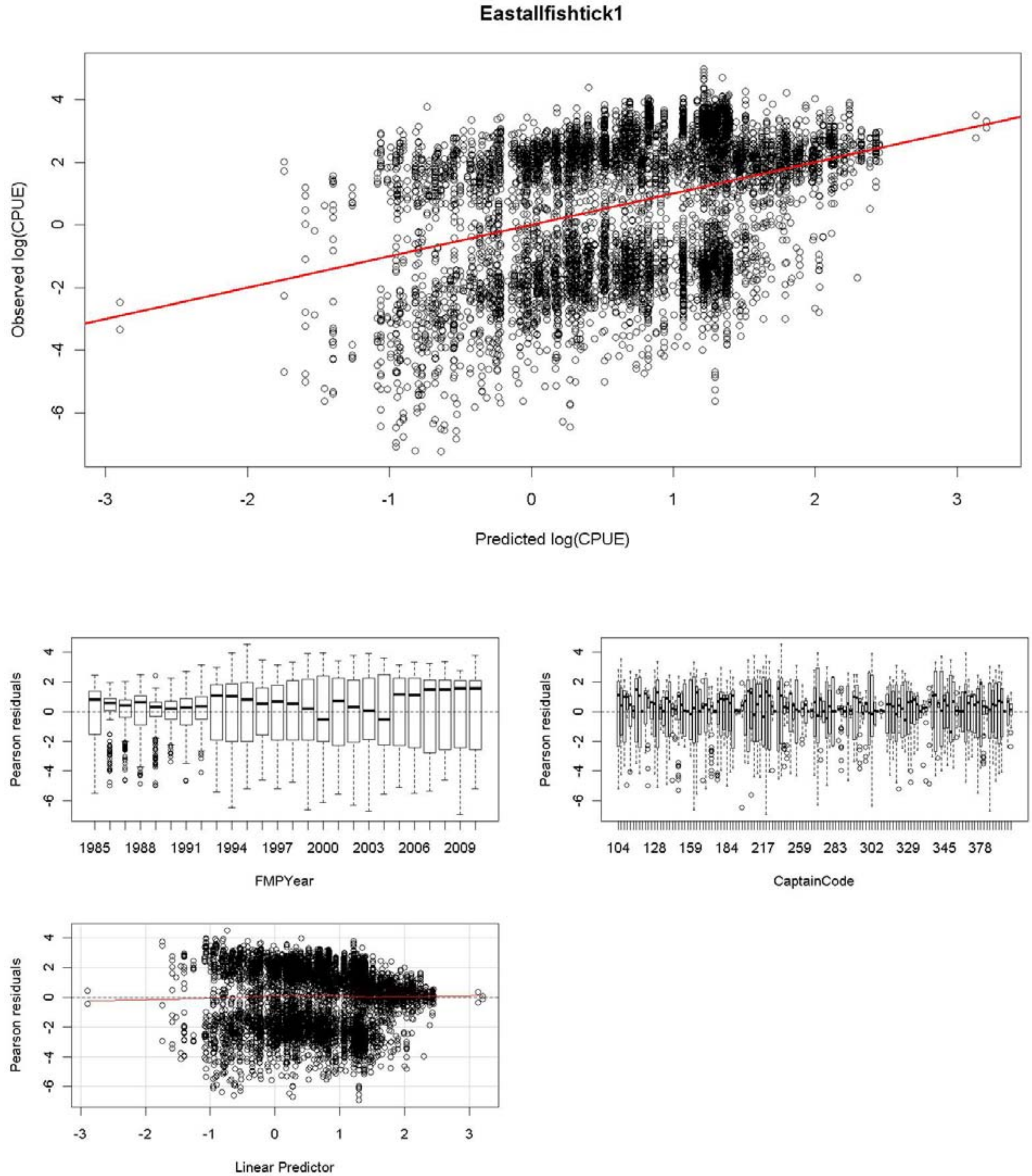


Figure 8. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for retained catch CPUE. Fish ticket data from east of 174°W for 1985/86–2010/11 were used.

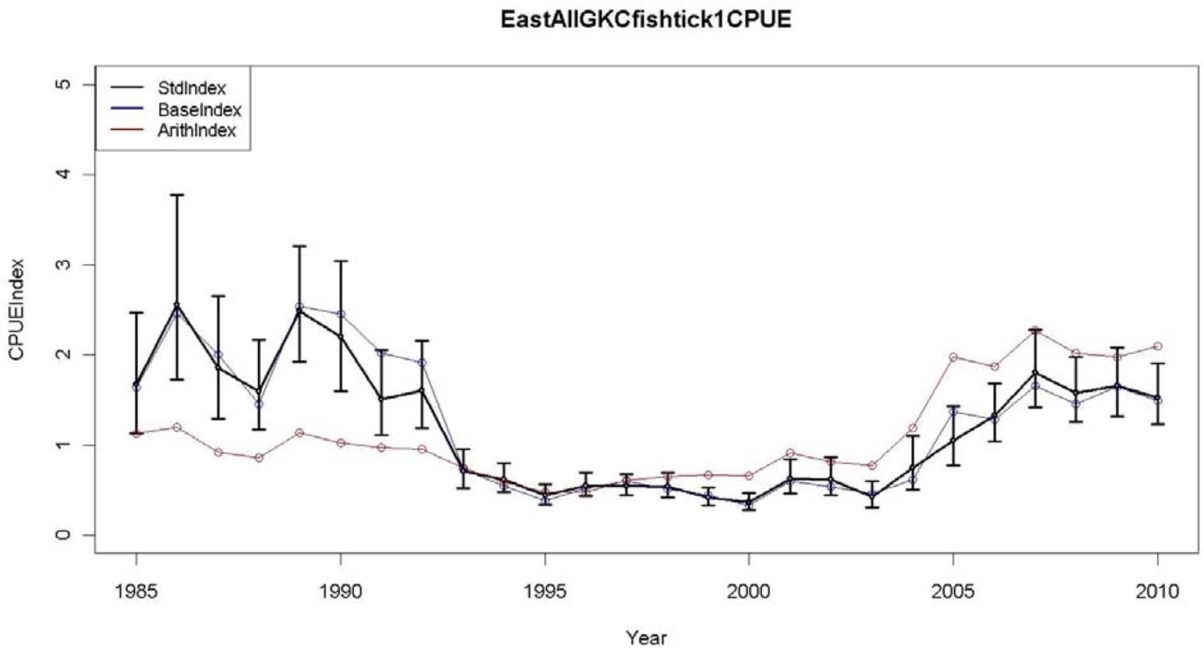


Figure 9. Trends in retained catch CPUE indexes for the Aleutian Islands golden king crab fishery. Standardized Index (Lognormal): black line with 2 standard errors; Base Index: blue line; and Arithmetic Index: red line. Fish ticket data from east of 174°W for 1985/86–2010/11 were used.

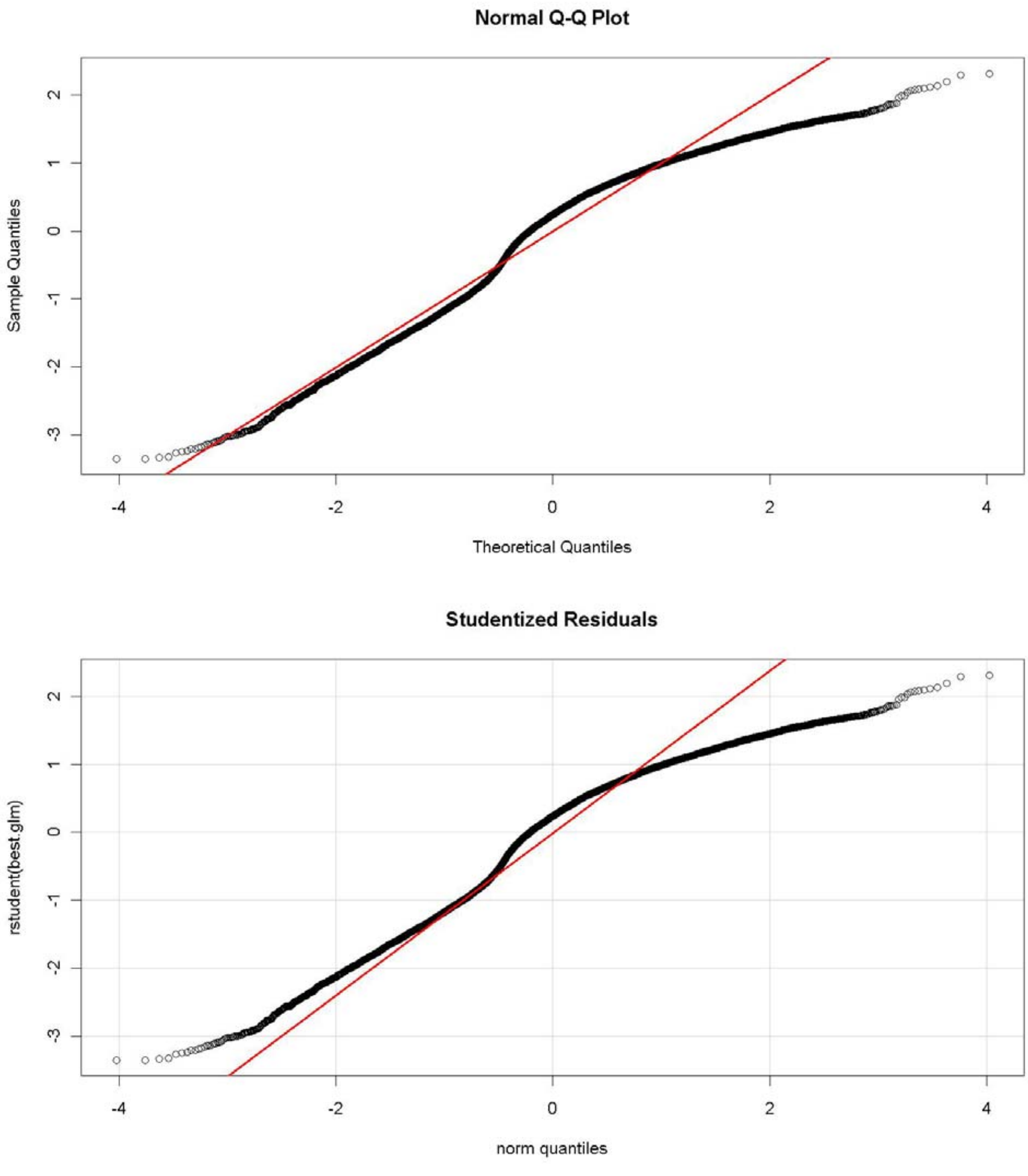


Figure 10. QQ and studentized residual plots of the best lognormal fit model for retained catch CPUE. Fish ticket data from west of 174°W for 1985/86–2010/11 were used.

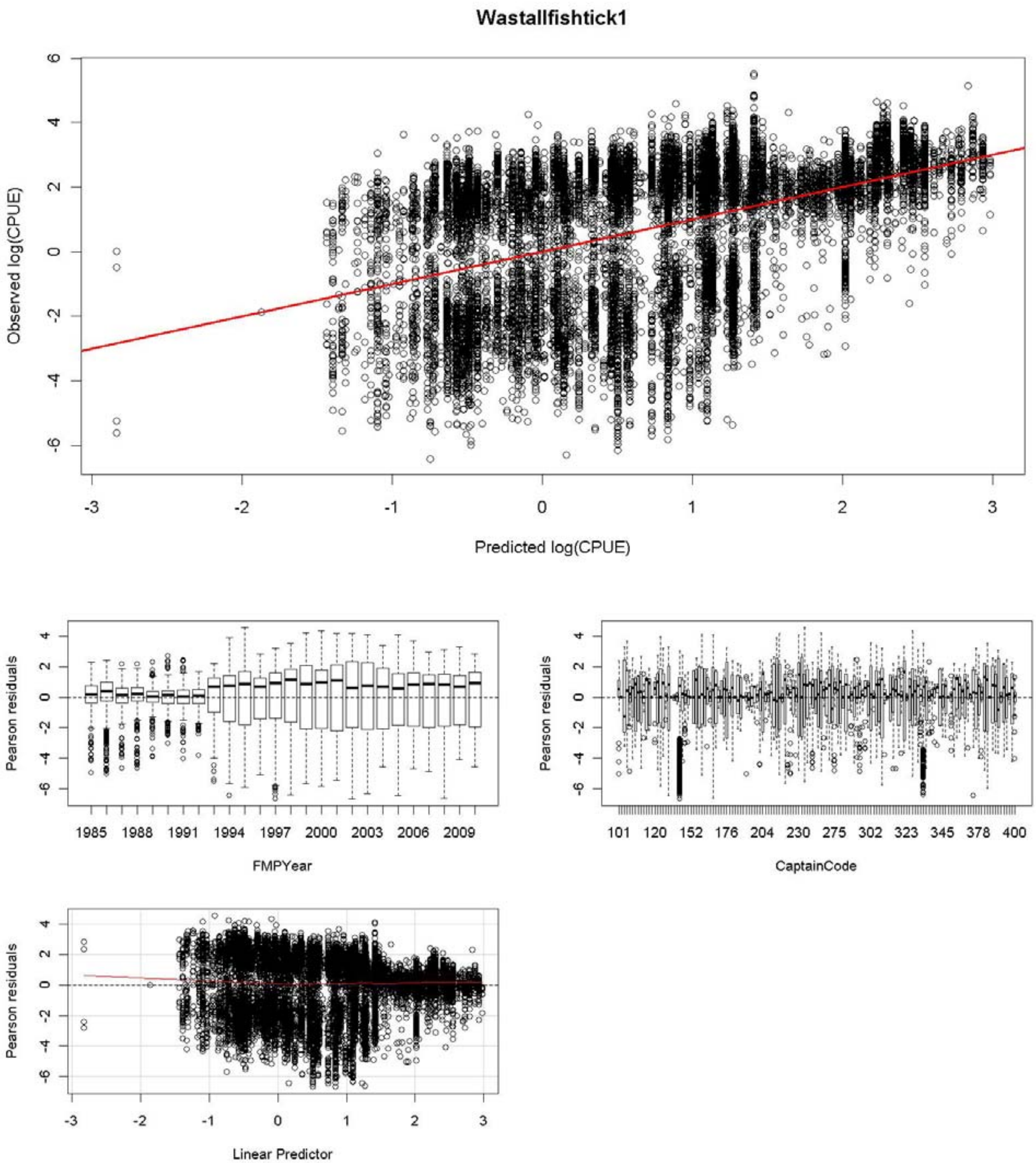


Figure 11. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for retained catch CPUE. Fish ticket data from west of 174°W for 1985/86–2010/11 were used.

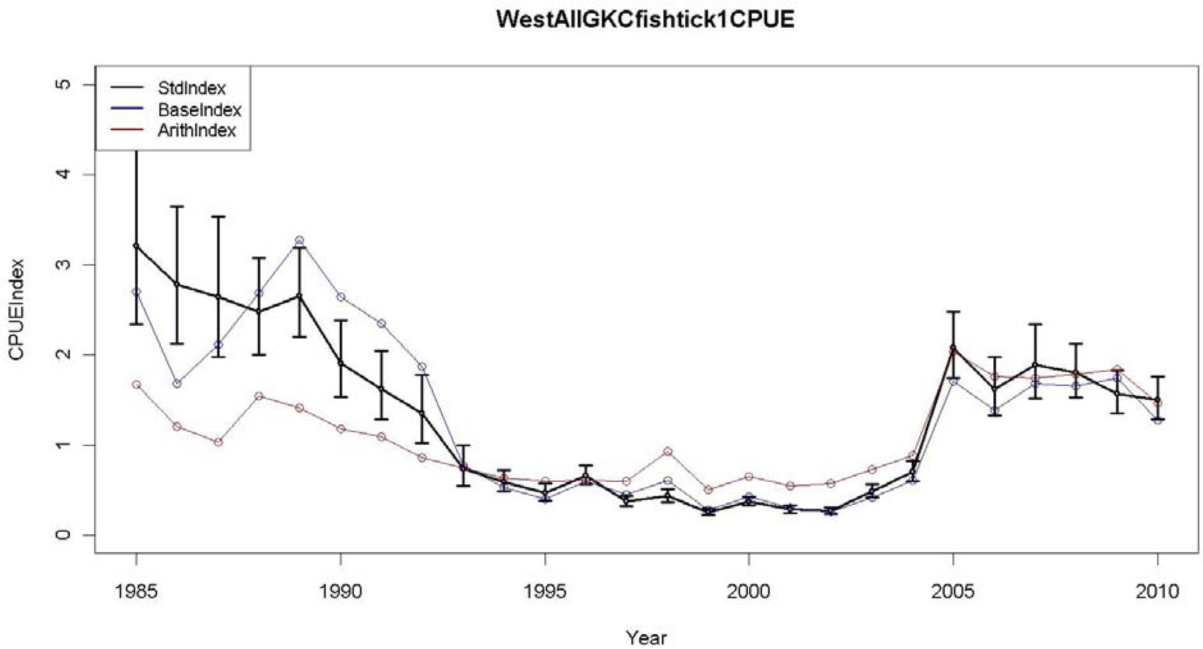


Figure 12. Trends in retained catch CPUE indexes for fish ticket data for the Aleutian Islands golden king crab fishery. Standardized Index (Lognormal): black line with 2 standard errors; Base Index: blue line; and Arithmetic Index: red line. Fish ticket data from west of 174°W for 1985/86–2010/11 were used.

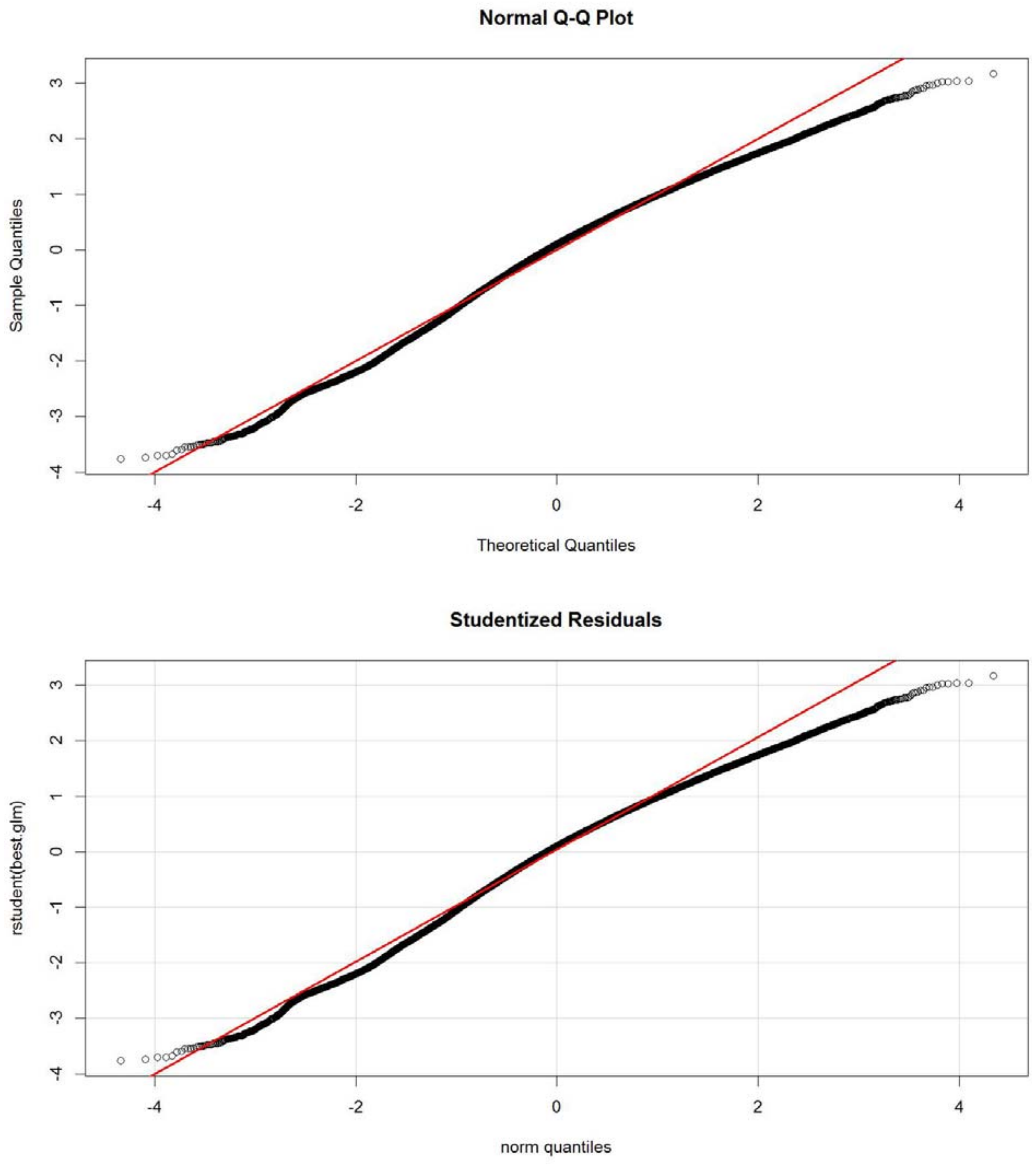


Figure 13. QQ and studentized residual plots of the best lognormal fit model for legal CPUE. Observer data from combined east and west of 174°W for 1995/96–2010/11 were used.

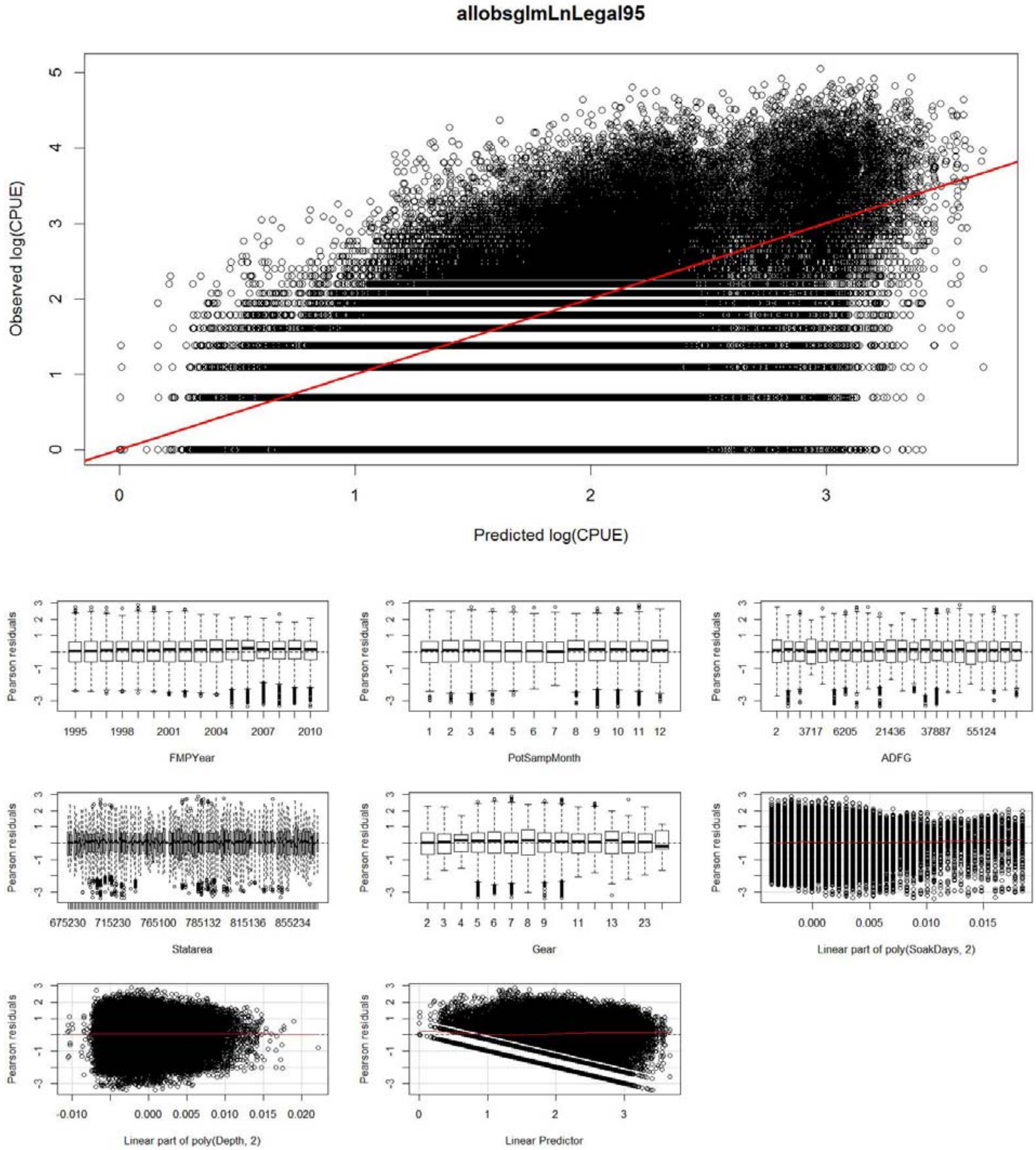


Figure 14. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for legal CPUE. Observer data from combined east and west of 174°W for 1995/96–2010/11 were used.

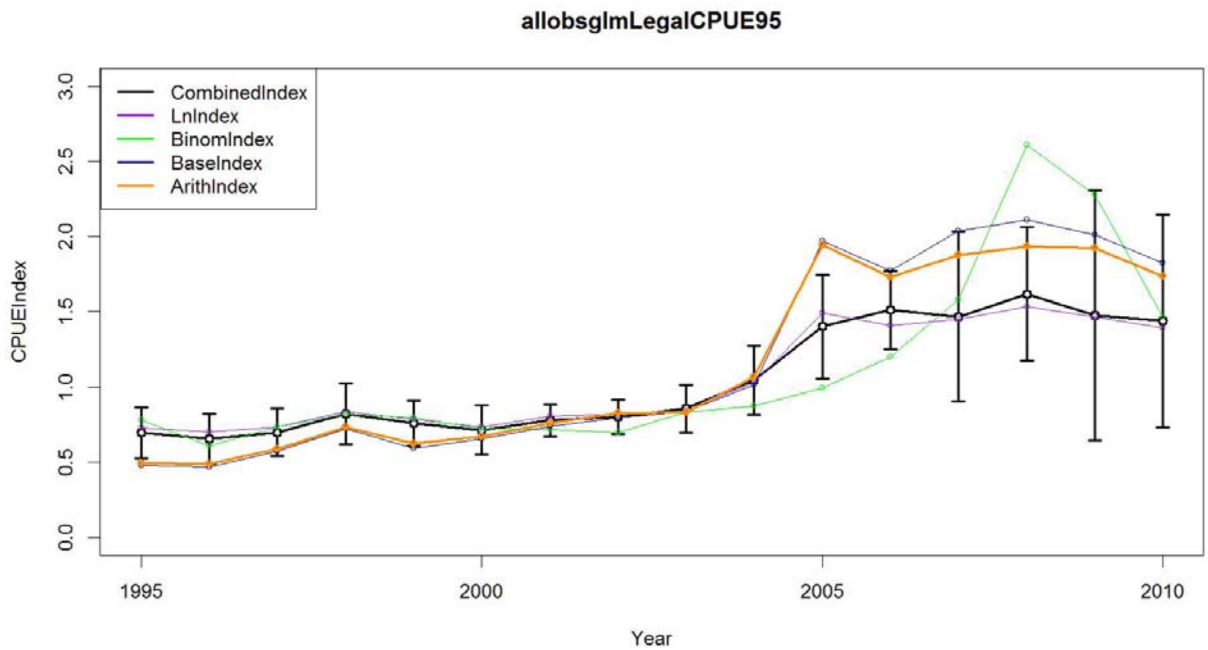


Figure 15. Trends in legal CPUE indexes for observer data for the Aleutian Islands golden king crab fishery. Combined Index: black line with 2 standard errors; LnIndex (Lognormal): purple line; BinomIndex (Binomial): green line; Base Index: blue line; and Arithmetic Index: orange line. Observer data from combined east and west of 174°W for 1995/96–2010/11 were used.

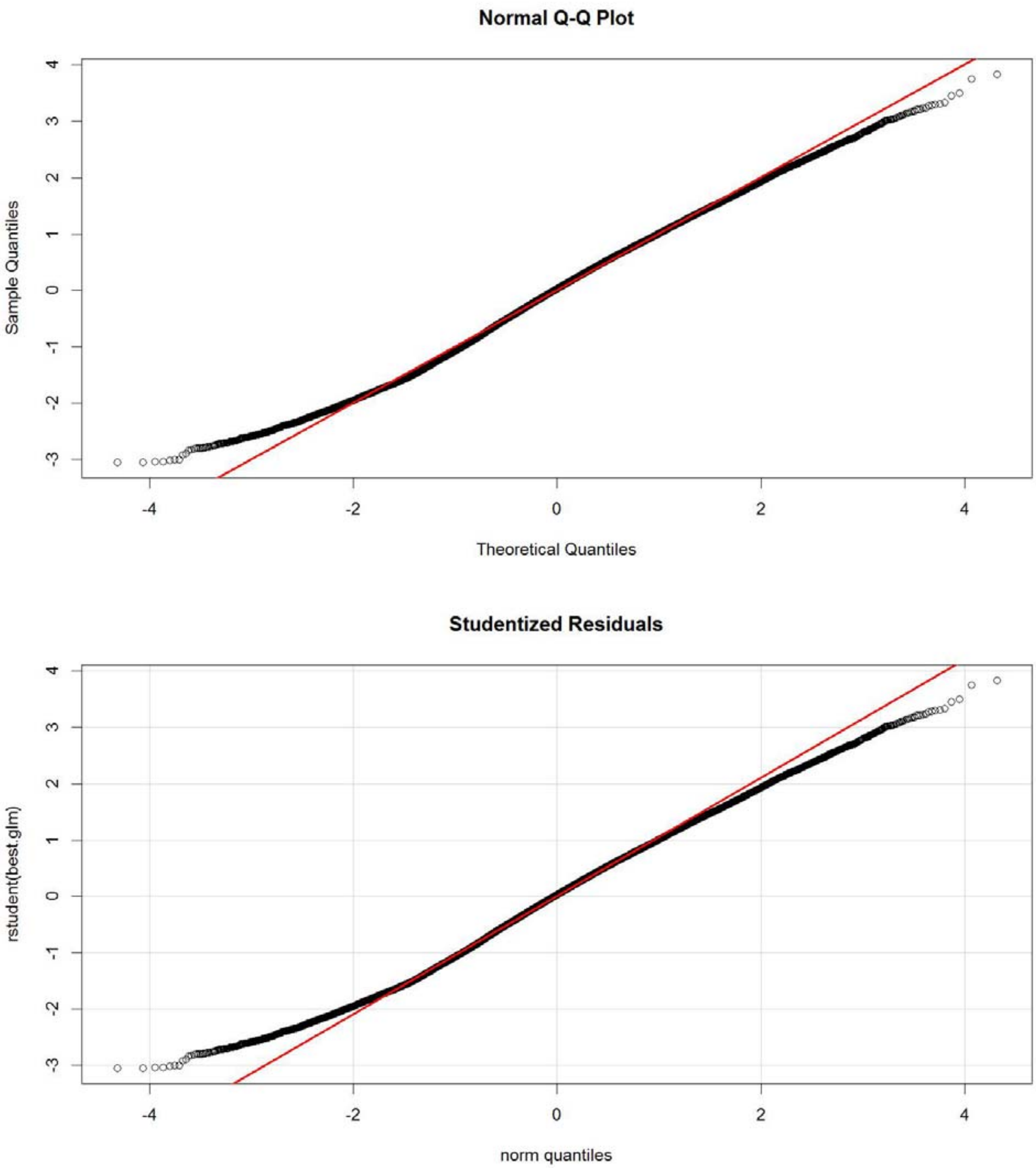


Figure 16. QQ and studentized residual plots of the best lognormal fit model for sublegal CPUE. Observer data from combined east and west of 174°W for 1995/96–2010/11 were used.

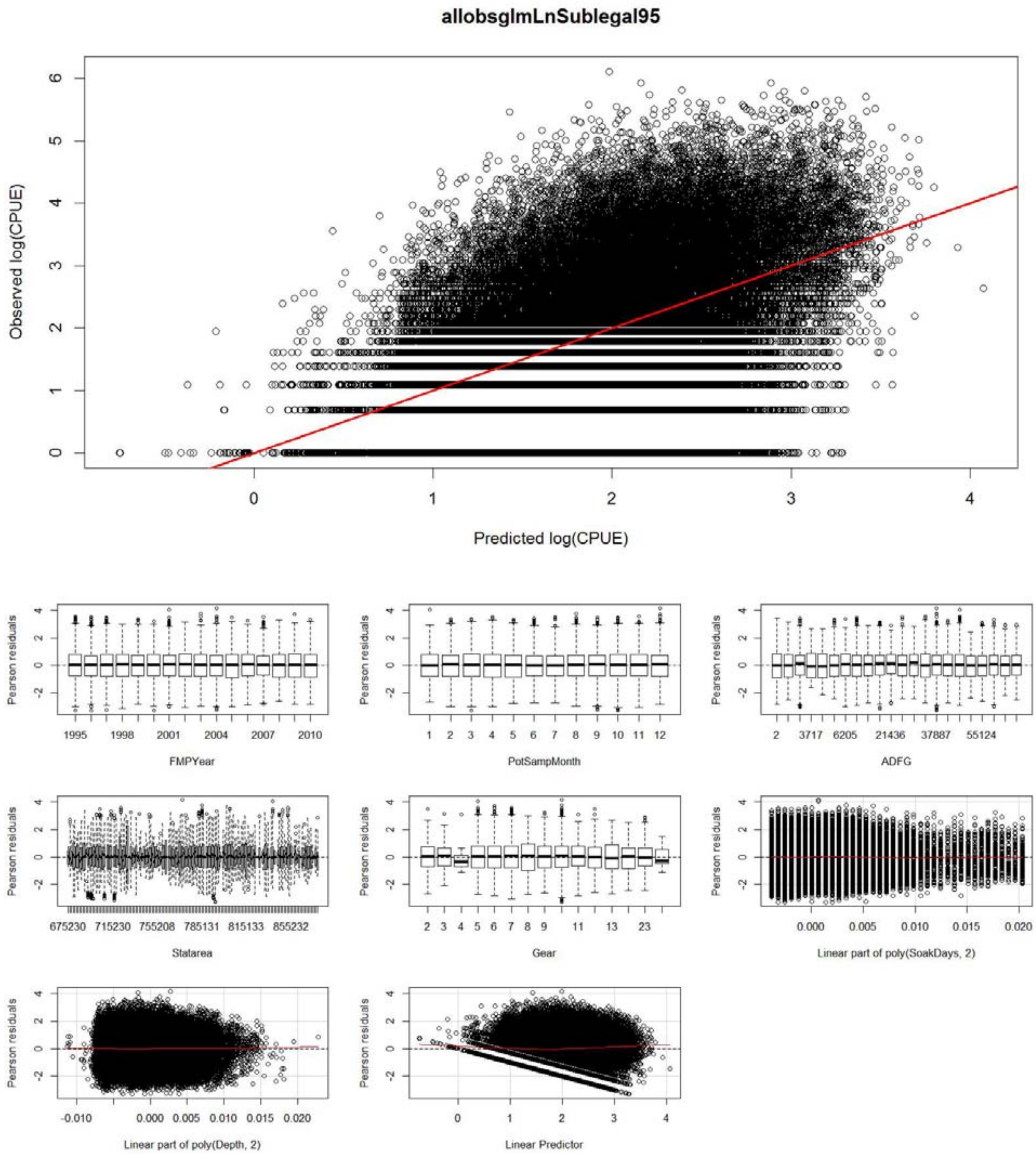


Figure 17. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for sublegal CPUE from combined east and west of 174 W. Observer data from combined east and west of 174°W for 1995/96–2010/11 were used.

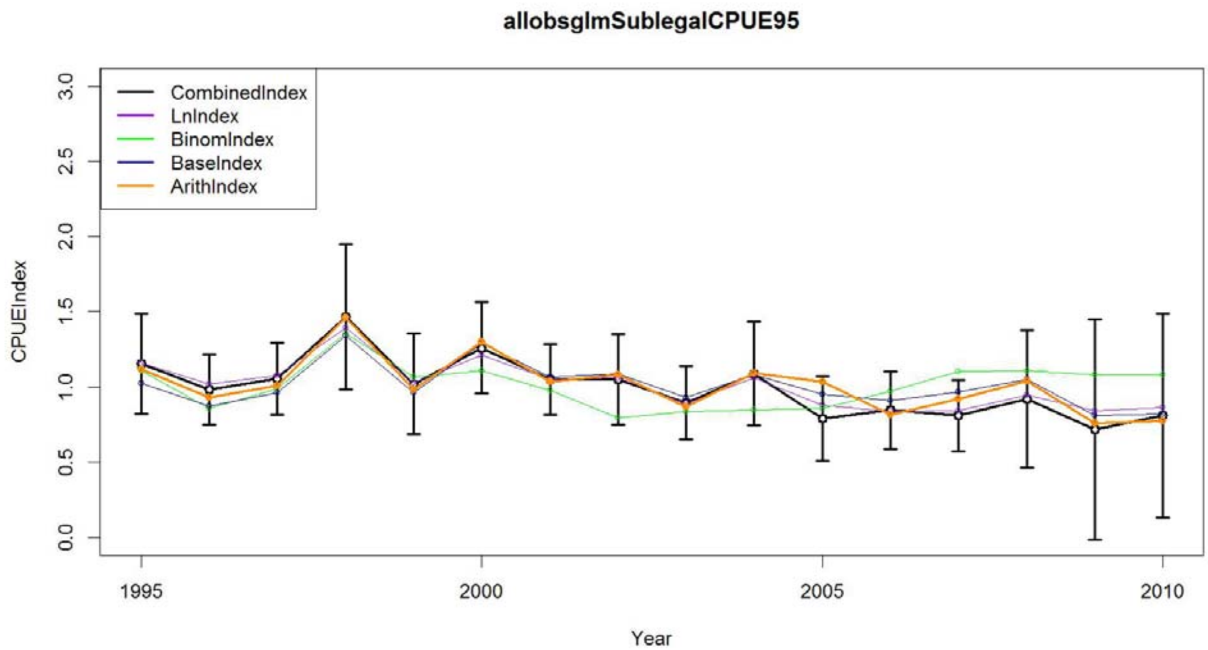


Figure 18. Trends in sublegal CPUE indexes for observer data for the Aleutian Islands golden king crab fishery. Combined Index: black line with 2 standard errors; LnIndex (Lognormal): purple line; BinomIndex (Binomial): green line; Base Index: blue line; and Arithmetic Index: orange line. Observer data from combined east and west of 174°W for 1995/96–2010/11 were used.

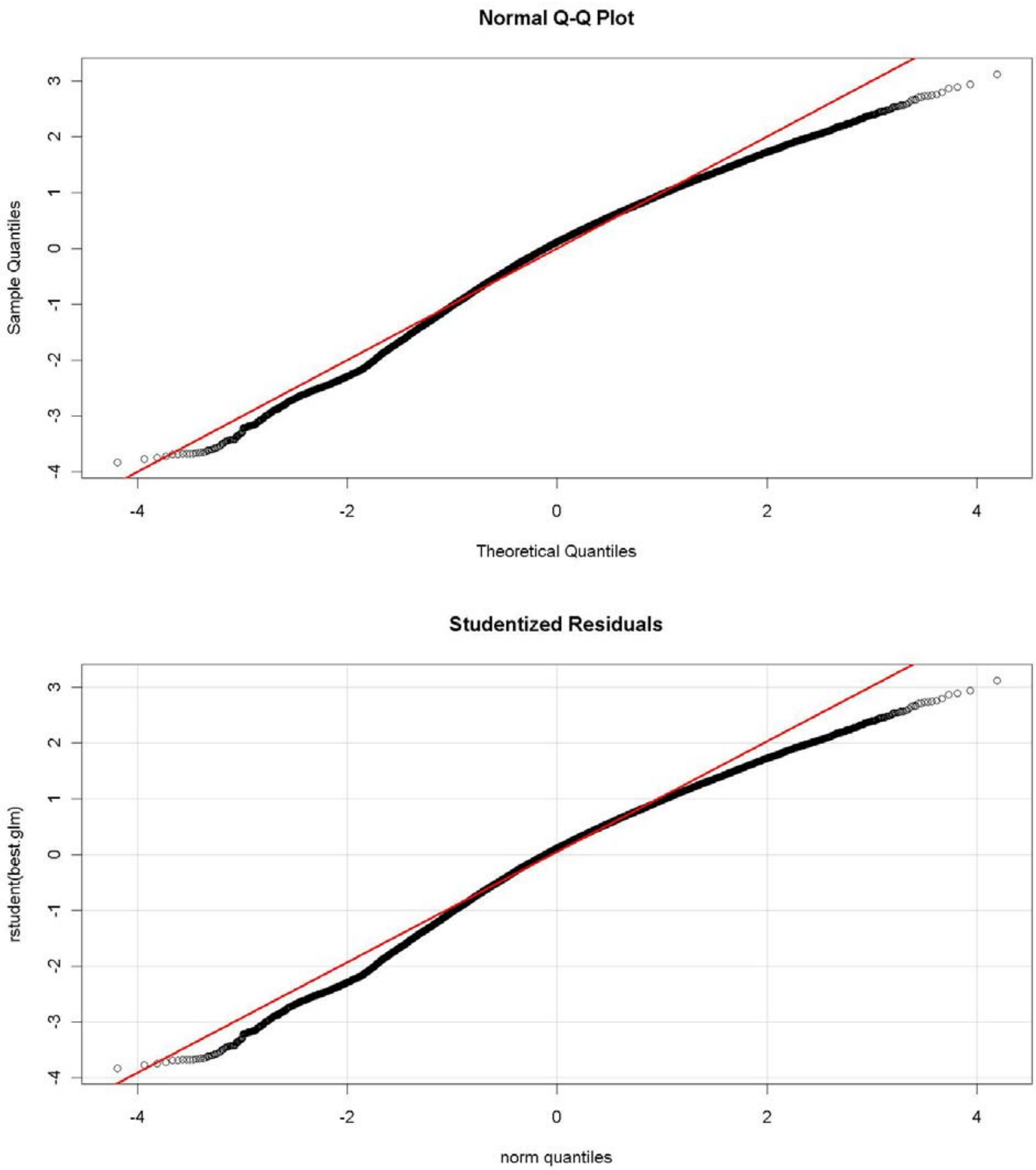


Figure 19. QQ and studentized residual plots of the best lognormal fit model for legal CPUE. Observer data from east of 174°W for 1995/96–2010/11 were used.

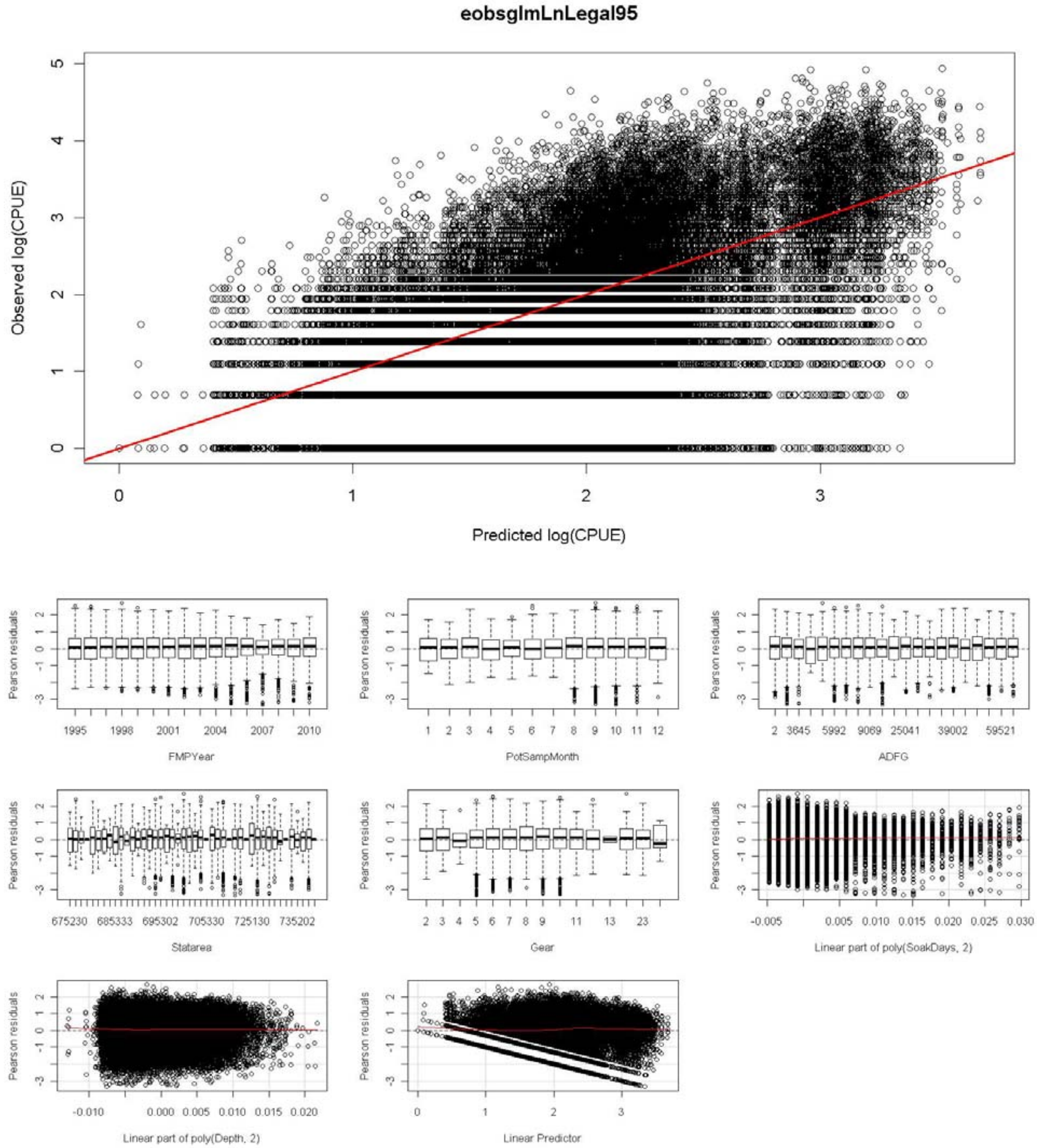


Figure 20. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for legal CPUE. Observer data from east of 174°W for 1995/96–2010/11 were used.

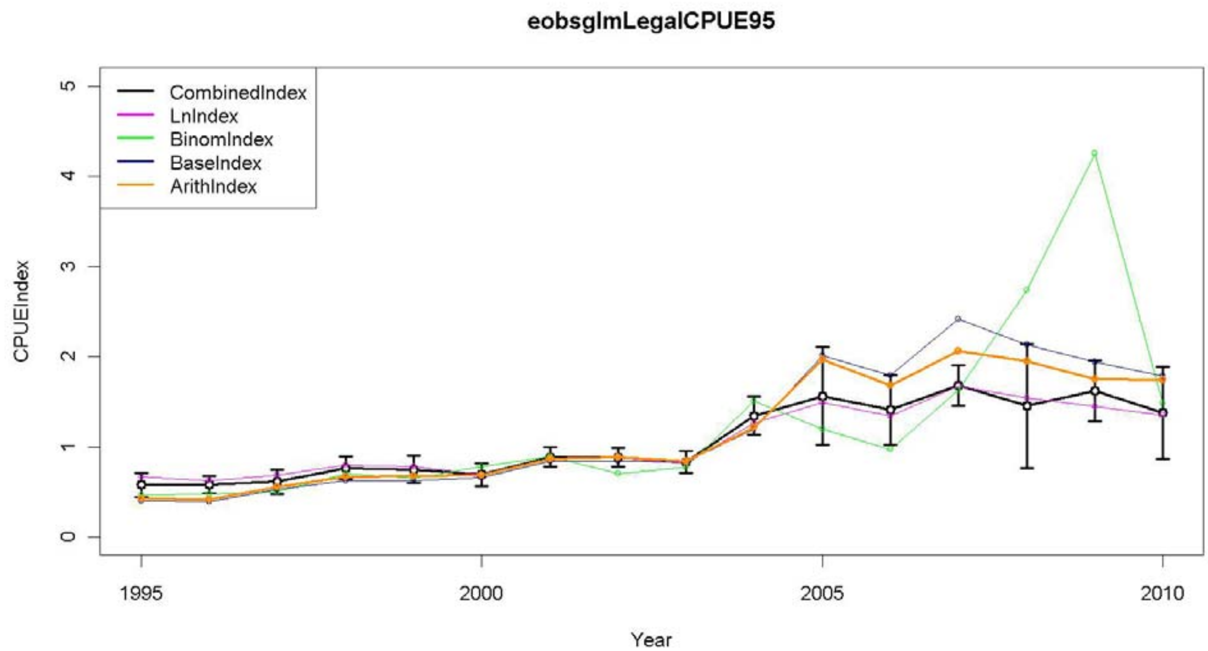


Figure 21. Trends in legal CPUE indexes for observer data for the Aleutian Islands golden king crab fishery. Combined Index: black line with 2 standard errors; LnIndex (Lognormal): purple line; BinomIndex (Binomial): green line; Base Index: blue line; and Arithmetic Index: orange line. Observer data from east of 174°W for 1995/96–2010/11 were used.

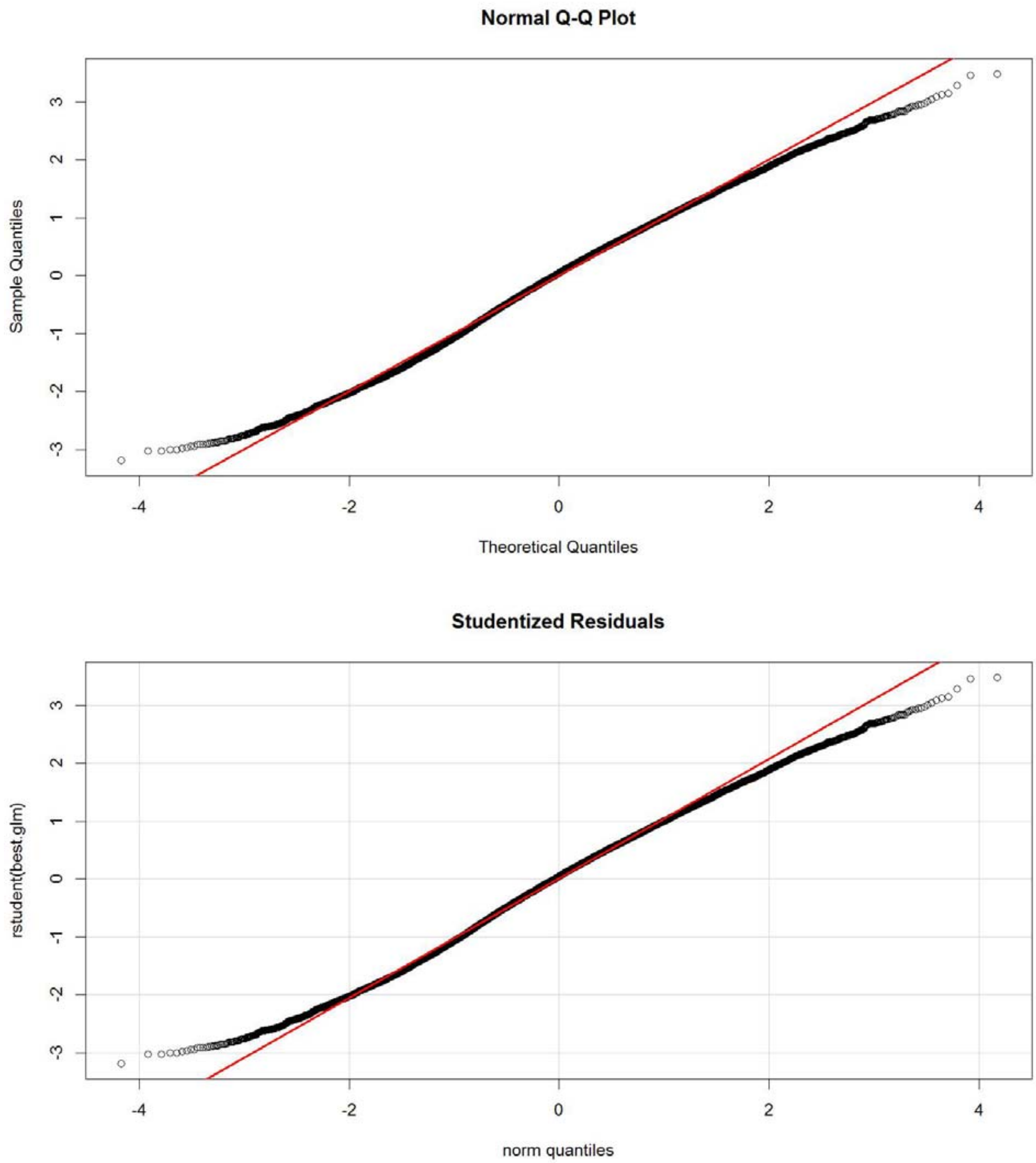


Figure 22. QQ and studentized residual plots of the best lognormal fit model for sublegal CPUE. Observer data from east of 174°W for 1995/96–2010/11 were used.

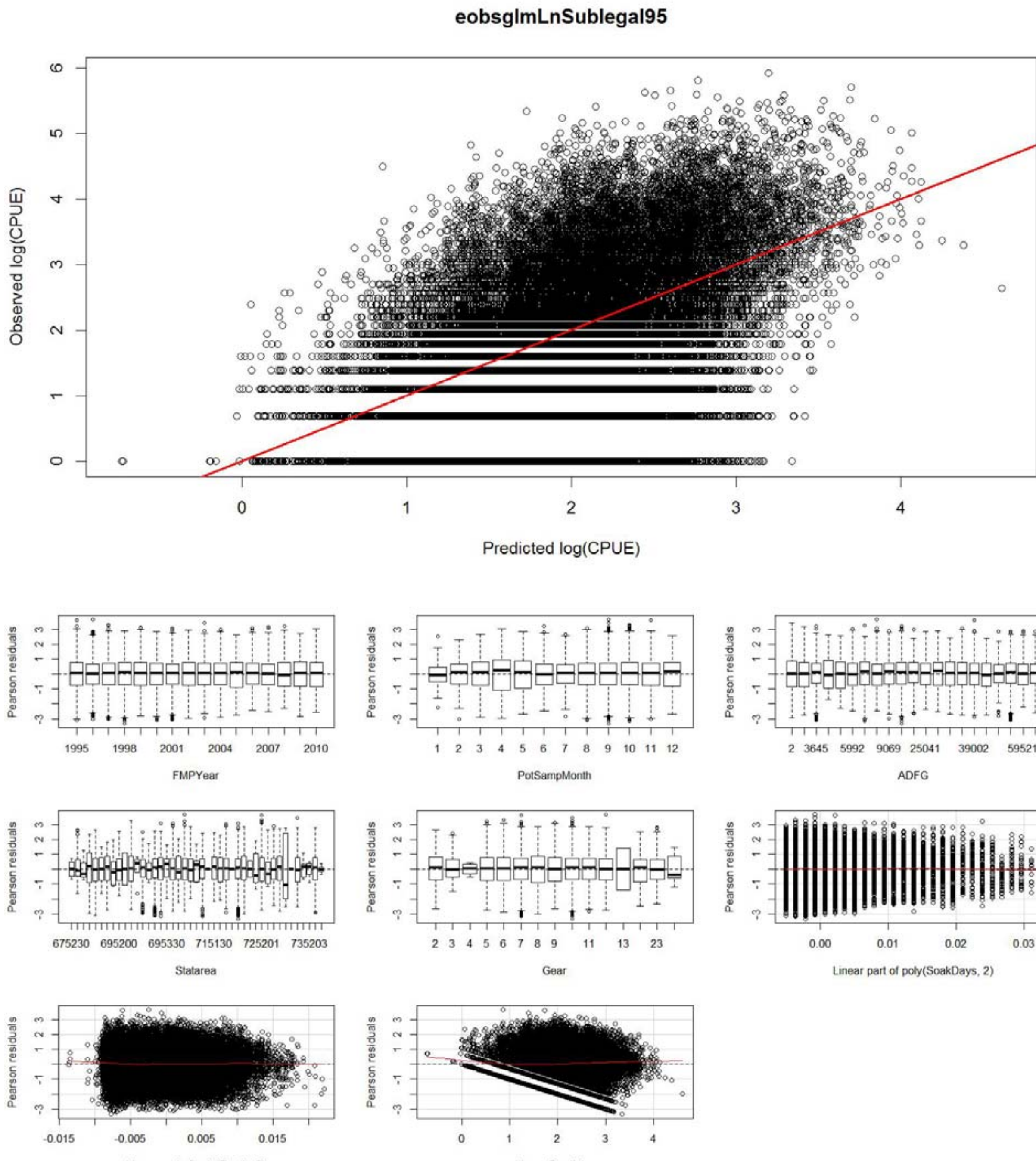


Figure 23. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for sublegal CPUE. Observer data from east of 174°W for 1995/96–2010/11 were used.

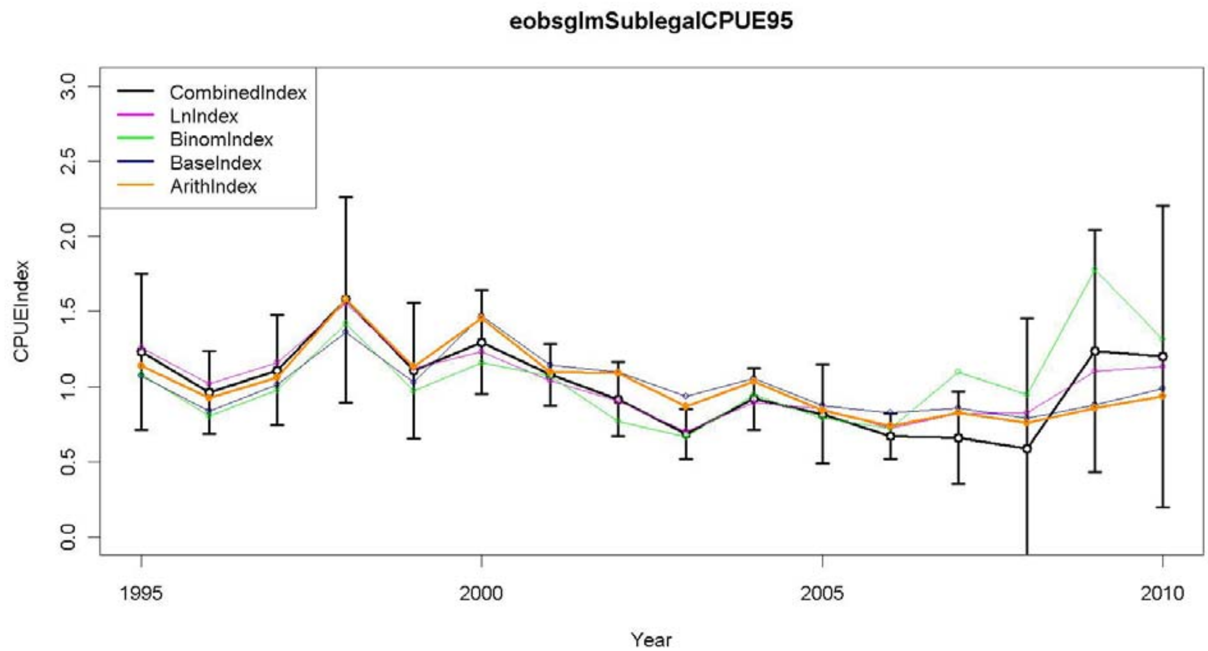
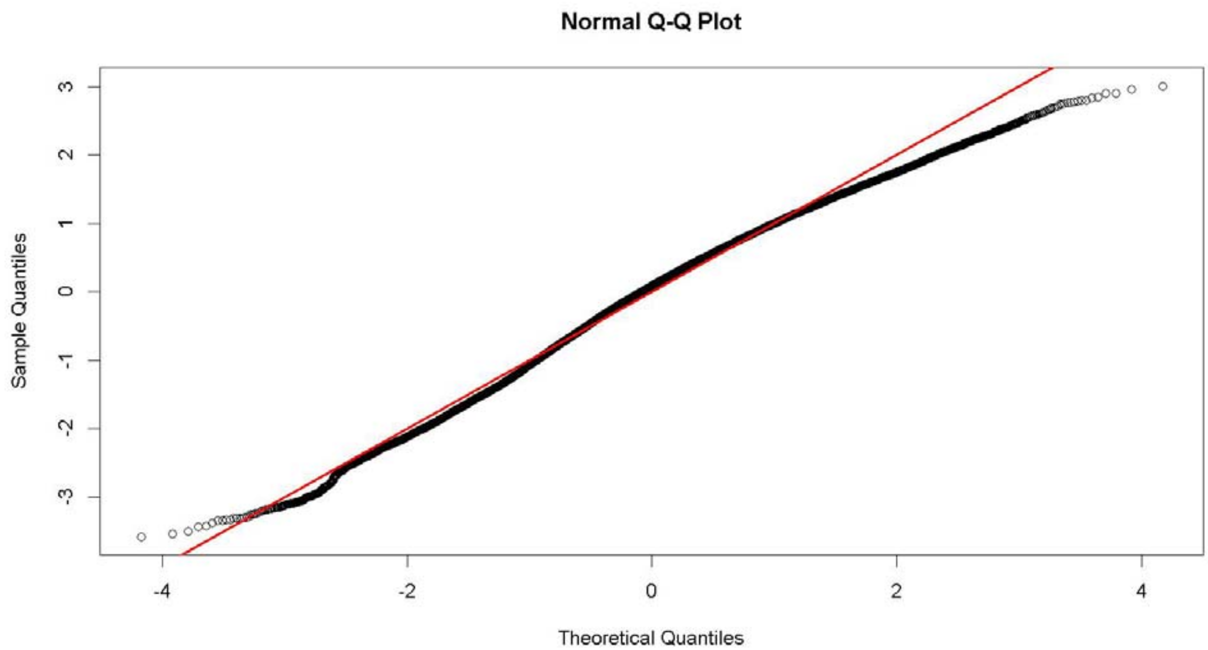


Figure 24. Trends in sublegal CPUE indexes for observer data for the Aleutian Islands golden king crab fishery. Combined Index: black line with 2 standard errors; LnIndex (Lognormal): purple line; BinomIndex (Binomial): green line; Base Index: blue line; and Arithmetic Index: orange line. Observer data from east of 174°W for 1995/96–2010/11 were used.



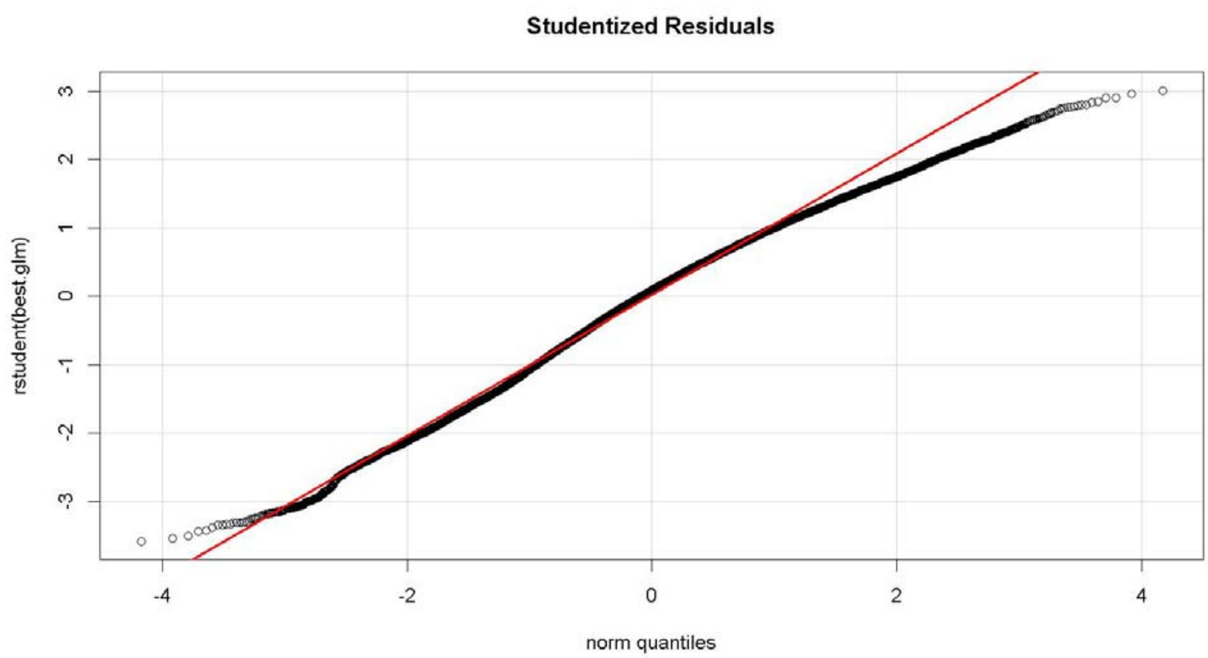


Figure 25. QQ and studentized residual plots of the best lognormal fit model for legal CPUE. Observer data from west of 174°W for 1995/96–2010/11 were used.

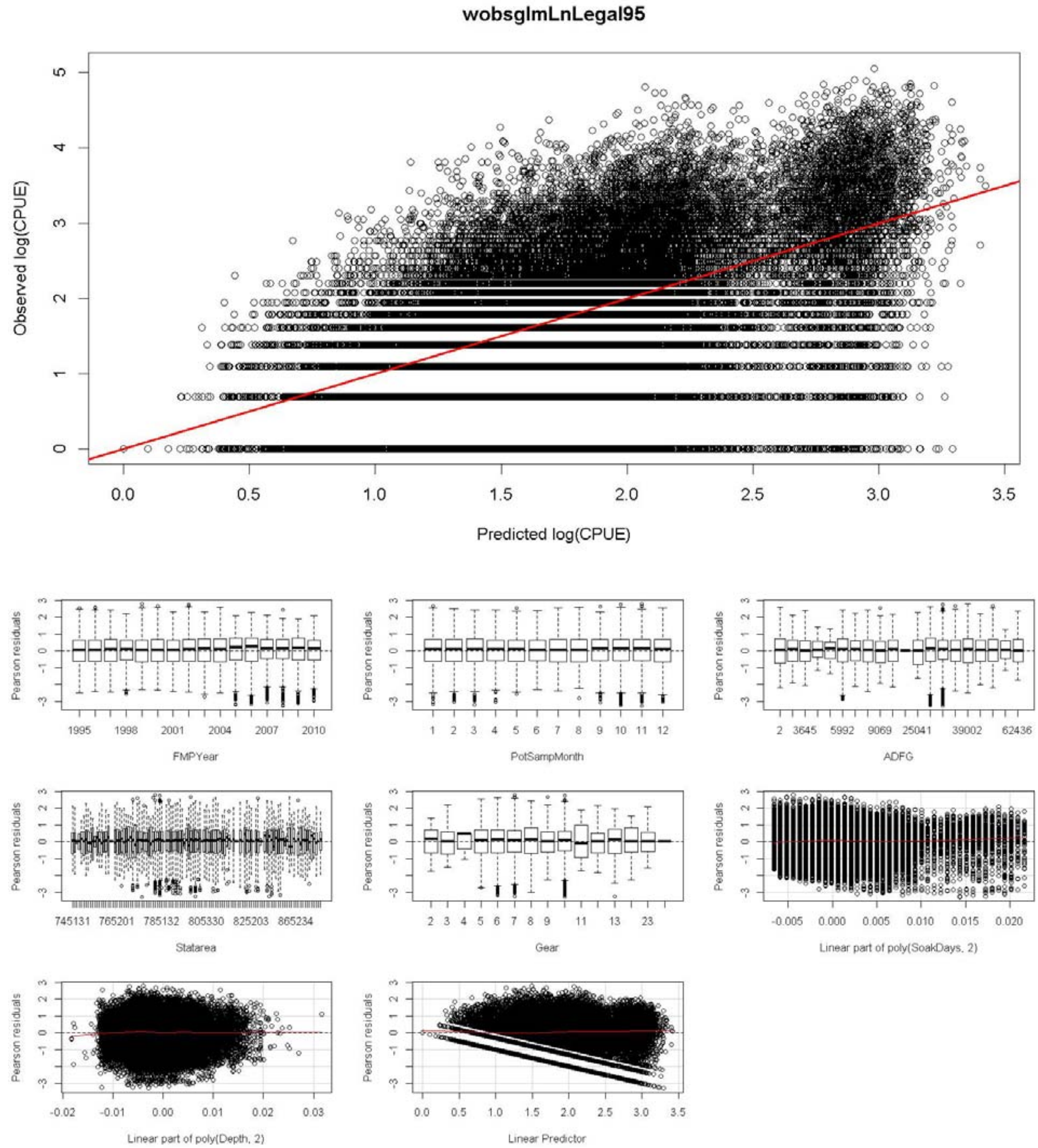


Figure 26. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for legal CPUE. Observer data from west of 174°W for 1995/96–2010/11 were used.

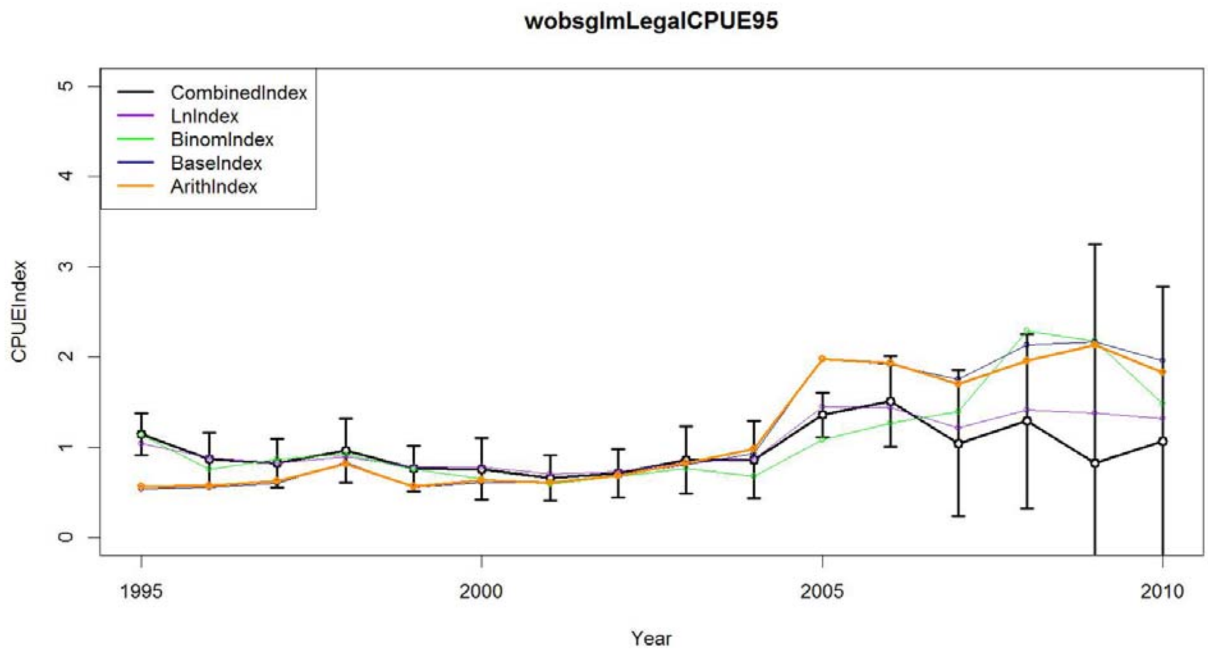


Figure 27. Trends in legal CPUE indexes for observer data for the Aleutian Islands golden king crab fishery. Combined Index: black line with 2 standard errors; LnIndex (Lognormal): purple line; BinomIndex (Binomial): green line; Base Index: blue line; and Arithmetic Index: orange line. Observer data from west of 174°W for 1995/96–2010/11 were used.

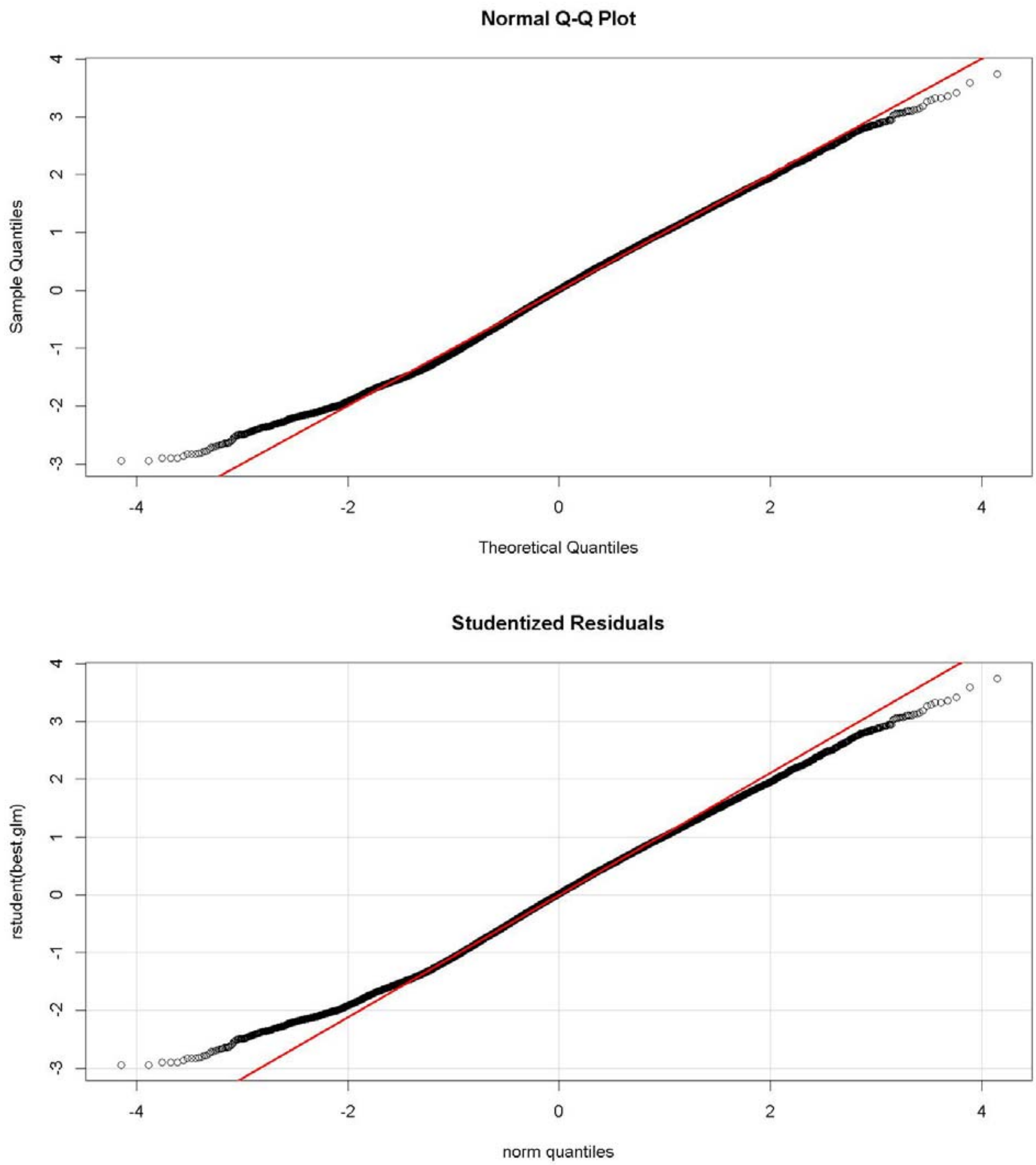


Figure 28. QQ and studentized residual plots of the best lognormal fit model for sublegal CPUE. Observer data from west of 174°W for 1995/96–2010/11 were used.

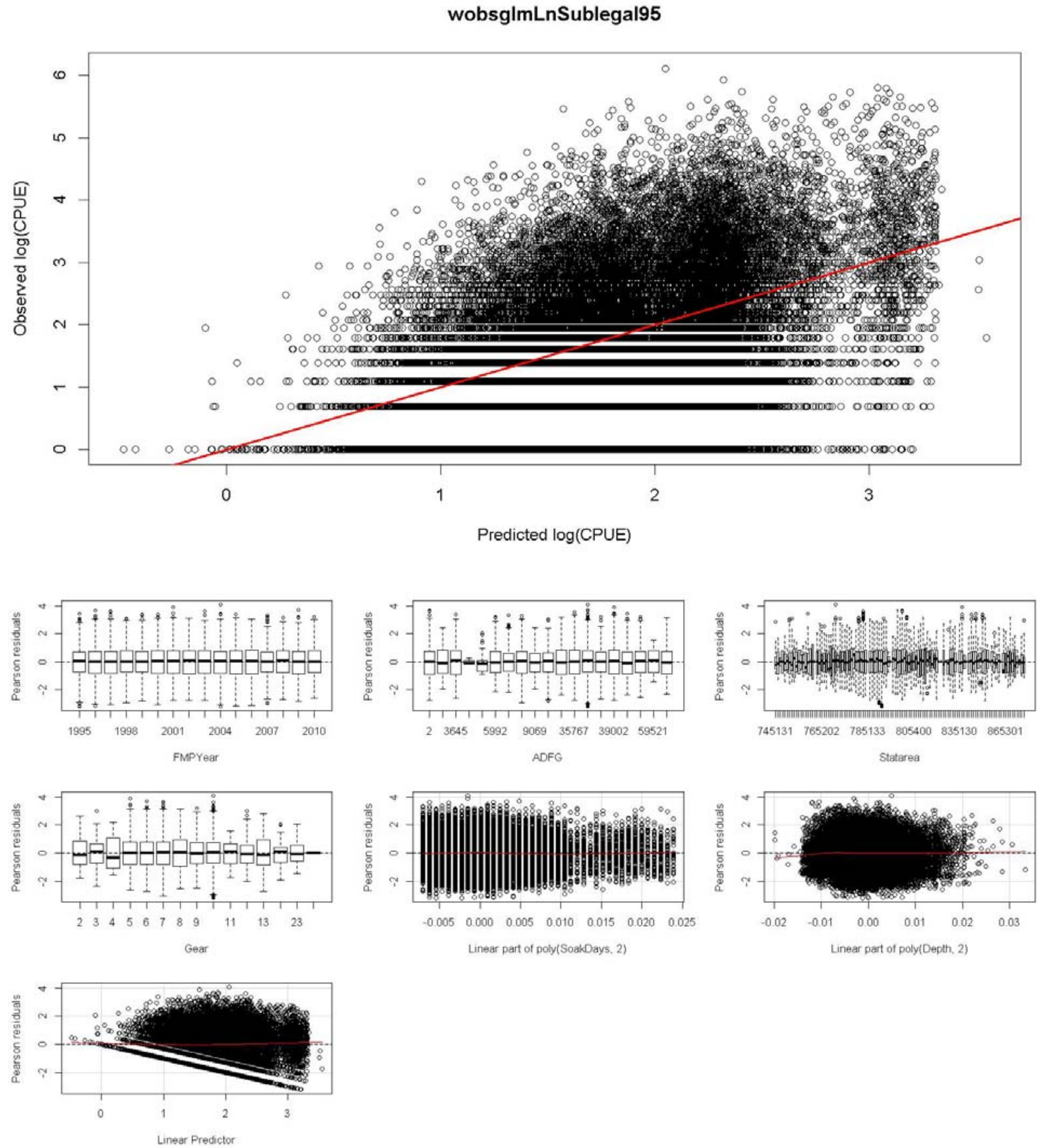


Figure 29. Predicted vs. observed ln(CPUE), Pearson residuals vs. explanatory and response variables of the best lognormal fit model for sublegal CPUE. Observer data from west of 174°W for 1995/96–2010/11 were used.

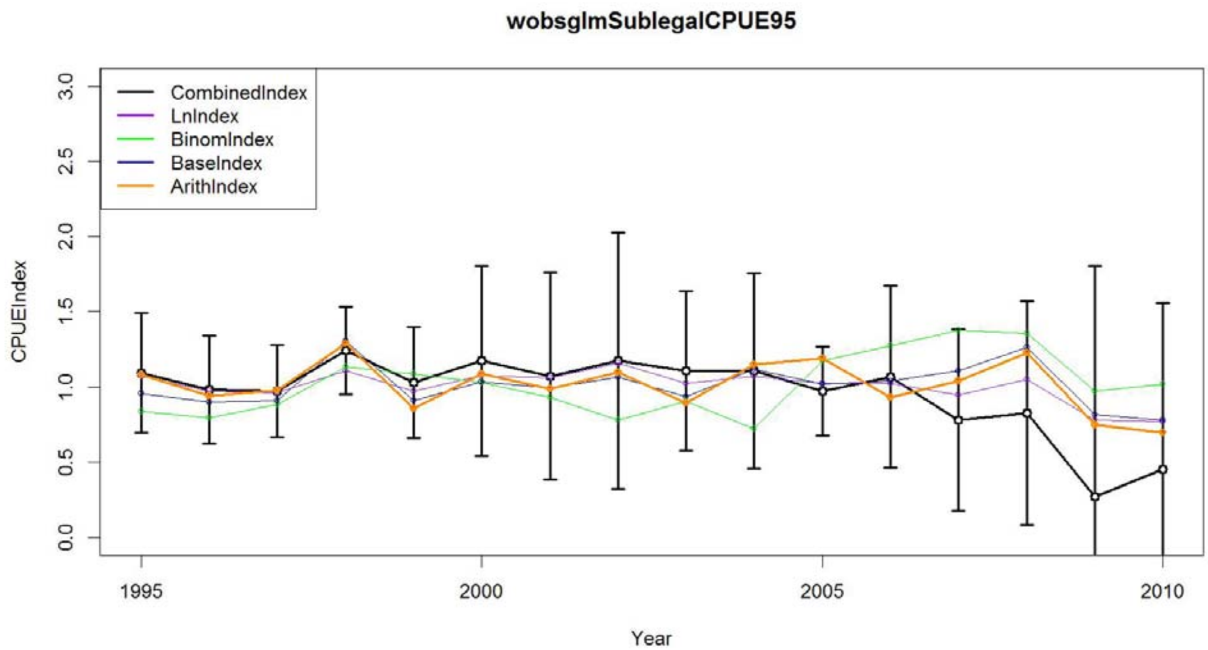


Figure 30. Trends in sublegal CPUE indexes for observer data for the Aleutian Islands golden king crab fishery. Combined Index: black line with 2 standard errors; LnIndex (Lognormal): purple line; BinomIndex (Binomial): green line; Base Index: blue line; and Arithmetic Index: orange line.Observer data from west of 174°W for 1995/96–2010/11 were used.

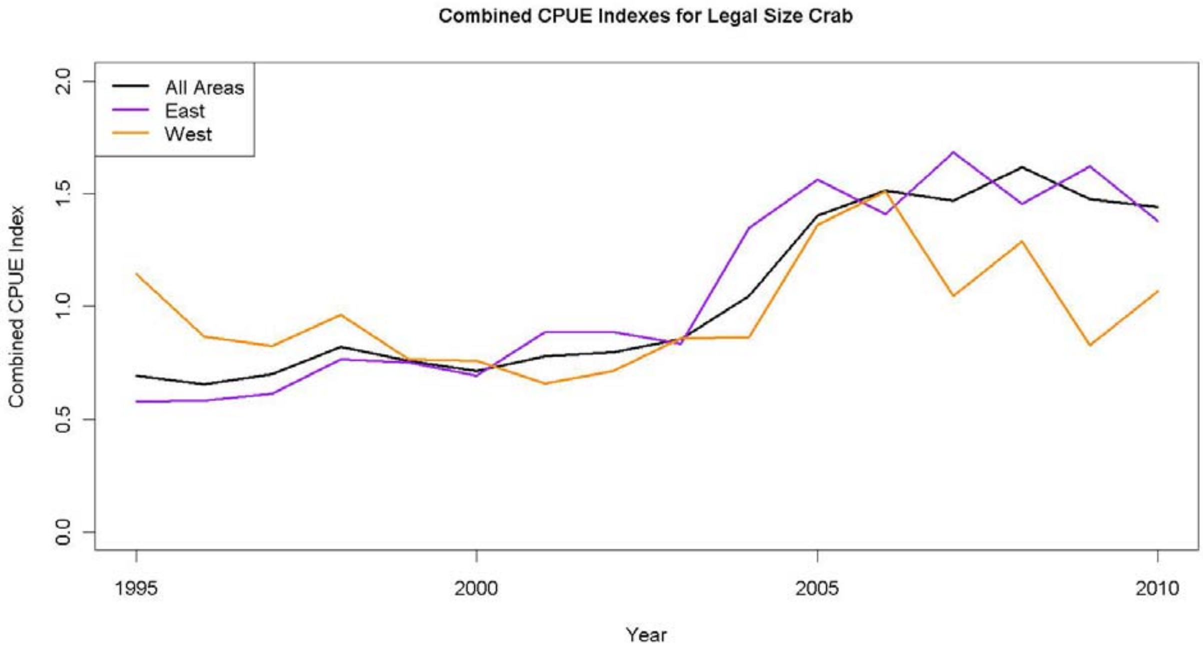


Figure 31. Trends in combined CPUE index of legal crabs for observer data for the Aleutian Islands golden king crab fishery. All areas: black line; East of 174°W: purple line; West of 174°W: orange line. Observer data for 1995/96–2010/11 were used.

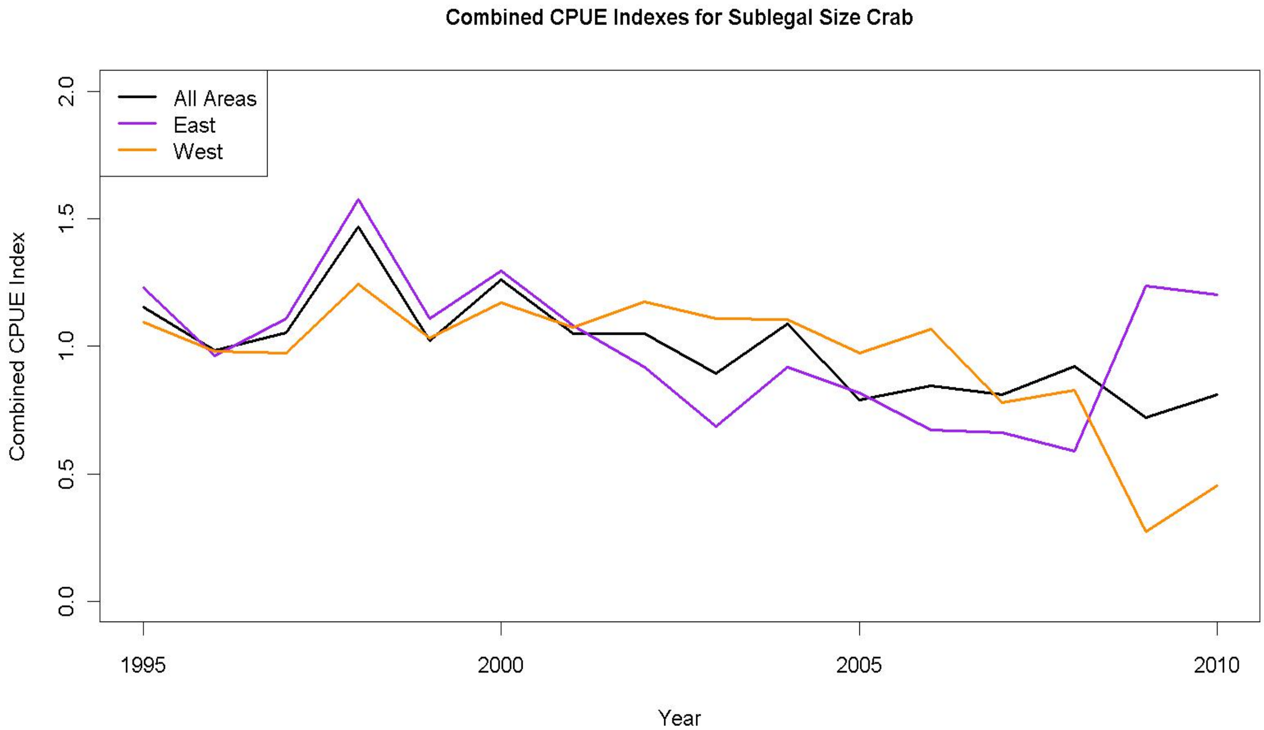


Figure 32. Trends in combined CPUE index of sublegal crabs for observer data for the Aleutian Islands golden king crab fishery. All areas: black line; East of 174°W: purple line; West of 174°W: orange line. Observer data for 1995/96–2010/11 were used.

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 variable setting

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

  options(object.size =100000000)

#
```

 # Read the observer data file

```
  allpotsample<- read.csv("c:/WorkRSep12/allpotsampleaddYrsof5Tr.csv", header=TRUE)

#
```

 # Divide into pre-post rationalization periods

```
#

   preallpotsample<- allpotsample[allpotsample$FMPYear<2005,]

 postallpotsample<- allpotsample[allpotsample$FMPYear>=2005,]

#
```

 # 5% and 95% percentile Trim by pre- and post-rationalization periods and combined them

```
preallpotsamplecut <- preallpotsample[preallpotsample$SoakDays>1 & preallpotsample$SoakDays<19,]

 postallpotsamplecut <- postallpotsample[postallpotsample$SoakDays>5 &
postallpotsample$SoakDays<40,]

 prepostallpotsamplecut<- rbind(preallpotsamplecut, postallpotsamplecut)

  allpotsampletrim<- prepostallpotsamplecut

#
```

 # Change some data frame variables to factors

```
#
  allpotsampletrim$FMPYear<- as.factor(allpotsampletrim$FMPYear)
  allpotsampletrim$PotSampMonth<- as.factor(allpotsampletrim$PotSampMonth)
  allpotsampletrim$ADFG<- as.factor(allpotsampletrim$ADFG)
  allpotsampletrim$Statarea<- as.factor(allpotsampletrim$Statarea)
  allpotsampletrim$EastWest<- as.factor(allpotsampletrim$EastWest)
  allpotsampletrim$Gear<- as.factor(allpotsampletrim$Gear)
  allpotsampletrim$CaptainCode<- as.factor(allpotsampletrim$CaptainCode)#
#
# Add a (binomial) variable to the data set to reflect success or failure
#
allpotsampletrim$success[allpotsampletrim$Legals>0]<- 1
  allpotsampletrim$success[allpotsampletrim$Legals==0]<- 0#
# Select core data
#
datacore<- allpotsampletrim[allpotsampletrim$Yrsof5Tr>=3,]
#
# Calculate the series of proportions zero (unsuccessful)
#
  prop.zero<- (table(datacore$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 gam and glm
```

#

library(gam)

gam.object1<-
gam(log(Legals)~FMPYear+PotSampMonth+ADFG+CaptainCode+Statarea+Gear+SoakDays+Depth,dat
a=datacore1)

allobsgamout1<-
step.gam(gam.object1,scope=list("FMPYear"=~(1+FMPYear),"PotSampMonth"=~(1+PotSampMonth),"
ADFG"=~(1+ADFG),

"CaptainCode"=~(1+CaptainCode),"Statarea"=~(1+Statarea),"Gear"=~(1+Gear),

"SoakDays"=
~(1+SoakDays+poly(SoakDays,2)+s(SoakDays,4)+s(SoakDays,6)+s(SoakDays,8)+s(SoakDays,10)),

"Depth"= ~(1+Depth+poly(Depth,2)+s(Depth,4)+s(Depth,6)+s(Depth,8)+s(Depth,10))),trace=TRUE)

best.glm<-
glm(log(Legals)~FMPYear+PotSampMonth+ADFG+Statarea+Gear+poly(SoakDays,2)+poly(Depth,2),d
ata=datacore1)

#

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

sumglm<-summary(best.glm)

coefsglm <- exp(as.numeric(c(0, sumglm$coefficients[2:16,1])))

#Get canonical lognormal indices

cpue1.glm<-getCPUE(best.glm,2:16, 1995:2010)

write.csv(cpue1.glm,"C:/WorkRSep12/allobsglmYearlyLnCPUEIndex95.csv",row.names=F)

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

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

sumglm1<-summary(base.glm)

coefsbaseglm <- exp(as.numeric(c(0, sumglm1$coefficients[2:16,1])))

#Get canonical lognormal indices for the base index

cpue2.glm<-getCPUE(base.glm,2:16, 1995:2010) write.csv(cpue2.glm,"C:/WorkRSep12/allobsglmBaseYearLnCPUEIndex95.csv",row.names=F)

#

#Find the best binomial model

gam.object2<-gam(success~FMPYear+PotSampMonth+ADFG+CaptainCode+Statarea+Gear+SoakDays+Depth,family=binomial(link=logit),data=datacore)

allobsgamout2<-step.gam(gam.object2,scope=list("FMPYear"=~(1+FMPYear),"PotSampMonth"=~(1+PotSampMonth),"ADFG"=~(1+ADFG),

"CaptainCode"=~(1+CaptainCode),"Statarea"=~(1+Statarea),"Gear"=~(1+Gear),

"SoakDays"=~(1+SoakDays+poly(SoakDays,2)+s(SoakDays,4)+s(SoakDays,6)+s(SoakDays,8)+s(SoakDays,10)),

"Depth"=~(1+Depth+poly(Depth,2)+s(Depth,4)+s(Depth,6)+s(Depth,8)+s(Depth,10))),family=binomial(link=logit),trace=TRUE)

best2.glm<-glm(success ~ FMPYear+PotSampMonth+ADFG+Statarea+Gear+poly(SoakDays,2)+poly(Depth,2),family=binomial(link=logit),data=datacore)

###################################

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

sumglm2<-summary(best2.glm)

coefsbin <- exp(as.numeric(c(0, sumglm2$coefficients[2:16,1])))

#Get canonical binomial indices

cpue3.glm<-getCPUE(best2.glm,2:16, 1995:2010)

write.csv(cpue3.glm,"C:/WorkRSep12/allobsglmYearlyBinomCPUEIndex95.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[2:16,1])))

#Get canonical binomial indices for the base index

```
cpue4.glm<-getCPUE(base3.glm,2:16, 1995:2010)
write.csv(cpue4.glm,"C:/WorkRSep12/allobsglmBaseBinomCPUEIndex95.csv",row.names=F)
#################################
```

#Calculate combined indices

```
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:/WorkRSep12/allobsglmCombCPUEIndex95.csv",row.names=F)
###########################
```

 # Positvie observed CPUE from core data

```
write.csv(datacore1,"C:/WorkRSep12/allobsglmCatches95.csv",row.names=T)
```

 # Arithmetic CPUE index

```
RCPUE<- tapply(datacore1$Legals,datacore1$FMPYear,mean)
GMRCPUE<- exp(mean(log(RCPUE)))
RCPUEdash<- RCPUE/GMRCPUE
write.csv(RCPUEdash,"C:/WorkRSep12/allobsglmScaledArithCPUEIndex95.csv",row.names=F)
##################################
```

 # Combine various CPUEs into a data frame after running Jackknife for creating CPUE figure

```
combined95<- read.csv("C:/WorkRSep12/jackallobsLegalGamGlmcpue95.csv",header=T)
```

allobsglmLegal95.cpue<- as.data.frame(cbind(cpue1.glm$Year,cpue2.glm$Index,RCPUEdash,cpue1.glm$Index,cpue3.glm$Index,combined95$CombInd,combined95$UpperComb,combined95$LowerComb))

names(allobsglmLegal95.cpue)<- list("Year", "BaseInd","ArithInd","LnInd", "BinInd","CInd","CUpper","CLower")

write.csv(allobsglmLegal95.cpue,"C:/WorkRSep12/allobsglmLegalcpue95.csv",row.names=F)

 # Plot various CPUE

 # Read the file

allobsglmLegal95.CPUE<- read.csv("C:/WorkRSep12/allobsglmLegalcpue95.csv",header=T)

 # Load the gplots package

library(gplots)

attach(allobsglmLegal95.CPUE)

jpeg(file="c:/WorkRSep12/allobsglmLegalCPUEIndex95Fig1.jpg",width=1600,height=900,res=150)

plotCI(x=Year,y=CInd,ui=CUpper,li=CLower,xlim=c(1995,2010),ylim=c(0,5),type="o",lty="solid",cex=1.0,lwd=2,pch=21,gap=0,sfrac=0.005, xlab="Year",ylab="CPUEIndex", main="allobsglmLegalCPUE95")

plotCI(x=Year,y=LnInd, ui=NULL,li=NULL,cex=0.75,type="o",lty="solid",lwd=1,col="darkviolet",add=TRUE)

plotCI(x=Year,y=BinInd,ui=NULL,li=NULL,cex=0.75,type="o",lty="solid",lwd=1,col="darkred",add=TRUE)
plotCI(x=Year,y=BaseInd,ui=NULL,li=NULL,cex=0.75,type="o",lty="solid",lwd=1,col="darkblue",add=TRUE)
plotCI(x=Year,y=ArithInd,ui=NULL,li=NULL,cex=0.75,type="o",lty="solid",lwd=1,col="darkorange",add=TRUE)
legend("topleft",c("CombinedIndex","LnIndex","BinomIndex","BaseIndex","ArithIndex"),lty=c("solid","solid","solid","solid","solid"),col=c("black","darkviolet","darkred","darkblue","darkorange"),lwd=2,cex=1.0)

dev.off()

detach(allobsglmLegal95.CPUE)

detach(package:gplots)

##

```
# Collect model fit diagnostic values

  Yhat<- best.glm$fitted.values   # predicted log(Legals)

  Ytemp<- datacore1$Legals

  Yobs<- log(Ytemp)    # observed log(Legals)

# Scatter plot of Yobs vs Yhat lognormal

#

  jpeg(file="c:/WorkRSep12/allobsglmLegalObsYPredY95Fig2.jpg",width=1600,height=900,res=150)

  plot(Yobs~Yhat,xlab= "Predicted log(CPUE)",ylab="Observed
log(CPUE)",main="allobsglmLnLegal95")

  abline(a=0,b=1,col="red",lwd=2)

   dev.off()

#

  Yhat1<- best2.glm$fitted.values   # predicted binomial success

  Yobs1<- datacore$success          # observed success

# Scatter plot of Yobs vs Yhat for binomial

  jpeg(file="c:/WorkRSep12/allobsglmLegalObsYPredY95Fig3.jpg",width=1600,height=900,res=150)

  plot(Yobs1~Yhat1,xlab= "Predicted success",ylab="Observed
success",main="allobsglmsuccessLegal95")

   dev.off()

########################################

# Plot residuals of log(CPUE)

   library(car)

    par(mfrow=c(3,3),cex=0.8,ps=8)
jpeg(file="c:/WorkRSep12/allobsglmLegalCPUEIndexManyRes95Fig4.jpg",width=1600,height=900,res
=150)

   residualPlots(best.glm,~.,fitted=TRUE,id.method="o") # against all predictors and fitted values

     dev.off()
```

```
  par(mfrow=c(1,1))
```

 # QQPlot of studentized residuals from car package

```
jpeg(file="c:/WorkRSep12/allobsglmLegalCPUEIndexStudentRes95Fig5.jpg",width=1600,height=900,res=150)

   qqPlot(rstudent(best.glm),envelope=FALSE,main="Studentized Residuals")

   dev.off()

#
```

 # qqnorm plot of residuals from stat package

```
#

jpeg(file="c:/WorkRSep12/allobsglmLegalCPUEIndexNormRes95Fig6.jpg",width=1600,height=900,res=150)

   qqnorm(rstandard(best.glm))

   abline(a=0,b=1,col="red",lwd=2)

   dev.off()

#####
```

 # Test for collinearity log (cpue) variables

```
   allobs95mod.vif<- vif(best.glm)

   write.csv(allobs95mod.vif,"C:/WorkRSep12/allobsLegal95VIF.csv",row.names=T)

  detach(package:car)

#
```

 # Check adding interactions

```
#   bestglm.add<- add1(best.glm, ~.^2,test="Chisq")

#   write.csv(bestglm.add,"C:/WorkRSep12/allobsglmaddinteractions95.csv",row.names=TRUE)
```