Received: 5 October 2017
IOTC-2017-WPTT19-38

Standardisation of skipjack tuna CPUE for the EU purse seine fleet operating in the Indian Ocean.

Isidora Katara[1], Daniel Gaertner[2], Norbert Billet[2], Jon Lopez[3], Alain Fonteneau[2], Hilario Murua[3], Patrick Daniel[4], Jose Carlos Baez[5]

[1] Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Pakefield Road, Lowestoft, Suffolk, NR33 0HT, UK

[2] Institut de recherche pour le développement IRD-UMR MARBEC CRH, CS 30171, Av. Jean Monnet, 34203 Sete Cedex, France

[3] AZTI Tecnalia, Herrera Kaia, Portualde z/g, 20110 Gipuzkoa, Spain

[4] Commission européenne, Affaires maritimes et Pêche, Unité B3 - Accords bilatéraux et Contrôle des pêches dans les eaux internationales, J-99 02/63 - 1049 Bruxelles

[5] Instituto Español de Oceanografía (IEO), Centro Oceanográfico de Canarias. Vía Espaldón, dársena pesquera, Parcela 8 38180 Santa Cruz de Tenerife.

Abstract

The EU purse seine fleet catches of skipjack tuna (SKJ; *Katsuwonus pelamis*) from the Indian Ocean were standardized using the framework described in Katara et al (2016). The analysis was restricted to fishing sets related with floating objects (FOBs), due to the strong associative behaviour of species and the FOB-oriented strategy of the fleet. Two definitions for Catch per unit of Effort (CPUE) were explored, i) a more traditional catch per fishing hour and ii) an alternative catch per fishing set. The time series for both CPUEs were standardised for two different periods: one for the whole time series (1985-2016), and one for the more recent years (2004-2016), because the length of the time series of available covariates, differed. In the latter case (2004-2016), the lasso – least absolute shrinkage and selection operator- method was applied for data mining and model selection. The results are four standardised skipjack CPUE time series for floating object fishing, all of them showing similar trends. The time series for the two CPUEs based on different definitions of effort (fishing hours vs fishing set) are comparable, both showing a decreasing trend. The values are lower than the nominal CPUE values, possible due to accounting for unfished areas.

Keywords: CPUE standardization; purse seine fishery; FADs, mixed models

Introduction

This paper is the result of the *Workshop for the development of Skipjack indices of abundance for the EU tropical tuna purse seine fishery operating in the Indian Ocean*, held in the AZTI (Pasaia, Spain) on July 2017. The workshop aimed at the development of standardised CPUE time series to be provided to IOTC as an input for the upcoming stock assessment of skipjack tuna.

Following the recommendations of the 2016 workshop for *the development of indices of abundance for the EU tropical tuna purse seine fishery* (Gaertner et al., 2017), mixed generalised linear models (GLMM) were applied for the standardisation of skipjack CPUE. Mixed models allow us to account for: (1) the longitudinal nature of logbook data (Laird and Ware, 1982; Liang and Zeger, 1986), where observations are made over time on the same subjects (in this case the vessel and/or the skipper) and (2) sampling area variability, i.e. hyperstability (Cao et al., 2011). Furthermore, the covariates in the standardisation model comprised of a thorough list of available data to describe fishing strategies and the multitude of technological advances that characterise the evolution of the tropical tuna purse seine fishery (Gaertner and Pallares 2002, Scott and Lopez, 2014; Lopez et al. 2014, Gaertner et al, 2016).

The high value on the acquisition of data on covariates on fishing technology at a highly disaggregated scale has already been highlighted by Bishop (2006). Several authors have highlighted the need for monitoring of additional variables/ descriptors of the tropical tuna purse seine fishery, one of the most technologically advanced fleet in the world (Gaertner and Pallares 2002, Anonymous, 2012; Scott and Lopez, 2014; Lopez et al. 2014, Gaertner et al, 2016). Because the traditional effort metrics (i.e. search time), may not be useful as effort units for this fishery due to its FOB oriented activity, and technological creep, the development of standardized CPUEs has been difficult for this fleet ocean-wide. As such, very few stock assessments have been using standardized CPUEs of this fishery; as a result, a large part of the catch of tropical tunas is not standardized and hence not used as an abundance index. The present paper presents the development standardized CPUEs indices for skipjack tuna accounting for several factors regarding technology and strategy that directly affect catches and catchability.

Material and Methods

Data

The main source of information for the calculation of the CPUE was logbook data from the French and Spanish purse seine fleets targeting tropical tuna in the Indian Ocean. The database was subset into 2 datasets: i) free-school sets (FSC), and ii) FOB-related sets, and CPUE was calculated and standardised exclusively for FOB-related sets because this is where the majority of skipjack is caught (Dagorn et al., 2013). The logbook databases are managed by the Tuna Observatory and the IEO for the French and the Spanish fleets respectively. The logbooks and sales data were collected through the Data Collection Framework (Reg 199/2008 and 665/2008) funded by both IRD and the European Union. Species composition in logbooks is corrected based on port sampling. Logbook data have been collected since 1983, but the analysis was constrained to i) the period 1986-2016 because the fishery and its data collection system were not properly established before 1985; ii) areas (grid cells) that have fishing sets for at least 15 years, in all EEZs except for the Somalian, because other areas are considered outside the principal fishing grounds targeted by the EU purse seiners (i.e. opportunistic targets add little information and a lot of noise in the data).

CPUE was defined as i) a more traditional catch per fishing hour and ii) alternatively, as the catch per fishing set. Models were developed for both CPUE definitions in order to compare the resulting time series, as a fishing effort unit for the FOB-fishery has been difficult to determine.

Complementary data from other information sources were compiled, to be used as covariates in the CPUE standardization models. A full list of the covariates and their sources is given in table 1. As some of the variables were only available for the period 2004-2016, it was decided to run the models for 2 periods separately: for 1985-2016 with a subset of the covariates, and for 2004-2016 with all available covariates. Numeric variables were standardised (mean=0, and standard deviation=1) before using them in the models. More details on the datasets can be found in the annexed workshop report.

Models

Due to the skewed distribution of the response variable and in order to investigate the trend of the ratio of successful fishing effort (catch >0), a delta-lognormal model was considered for the 1986-2016 models. The model comprises of two sub-models that estimate (i) p, the probability of skipjack catch being positive (>0), i.e. a logistic model that detects structural vs sampling zeros and (ii) μ, the expected catch per unit of effort, conditional to it being positive. The product of the 2 estimates ($\mu\rho$), derived by fitting the models, gives the standardized CPUE. For the 2004-2016 models, linear regression was applied, with logarithmic transformation of the response variable; effort with zero catch was disregarded because of the very low number of zeros that remained constant throughout the time series. The schema of the analysis can be found in Fig 1.

Collinearity of covariates and model overfitting was examined. Model selection for the 1986-2016 models (i.e. using only a subset of the covariates) was performed in a step-wise manner (forward and backward; Bozdogan, 1987). For models using all the covariates, the LASSO regression (Tibshirani 1996, 2011) was applied to select the simplest of the "true" models. Variables with regression coefficients greater than 0 were selected for the final model fit. In conjunction with LASSO regression, we run some exploratory simple models and examined changes in the model fit, attributed to adding or excluding covariates as fixed or random effects. The final model was selected based primarily on the structure of the data and on expert knowledge of the fishery, complemented by the results of the aforementioned data mining techniques.

Unbiased estimates were derived from models fitted using ordinary least square (OLS; Friedman et al., 2009; Tibshirani, 1996, 2011), and predictions were made with *lsmeans* v2.27-2 (Lenth, 2014). The analysis was performed in R (v 3.4.1) (R Core Team, 2013).

Results

Catch per fishing hour

1985-2016

The selected logistic model for the binomial proportion of fishing hours with skipjack catch > 0, included year (p-value < 0.0001), quarter (p-value < 0.0001), fleet (p-value < 0.0001), vessel storage capacity class (p-value < 0.01), and vessel age (p-value > 0.05). The random effects included the interaction between year and grid cell, the vessel and the statistical area (data correction area). The fit of the model is shown in Fig 4.

The final log-normal model for catch per fishing hours conditional to skipjack catch > 0, included year (p-value < 0.0001), quarter (p-value < 0.0001), fleet (p-value < 0.0001), vessel storage capacity class (p-value <

0.0001), and vessel age (p-value < 0.01). The random effects structure is the same as the binomial model, i.e. it includes the interaction between year and grid cell, the vessel and the statistical area (data correction area). The plots of the residuals (Fig 11) shows that the model assumptions are not violated.

The time series resulting from the product of the fits of the two sub-models is shown in Fig 2 and the nominal CPUE in Fig 3.

2004-2016

The binomial proportion of fishing hours with skipjack catch > 0 remains stable and over 80% for this period. For this reason, the CPUE was estimated using a log-normal model for catch per positive fishing hour. The binomial proportion of fishing hours with skipjack catch > 0 was not expected to influence the trend of the suggested CPUE.

The selected log-normal model for catch per fishing hours conditional to skipjack catch > 0, included year (p-value < 0.0001), quarter (p-value < 0.0001), company (p-value < 0.0001), vessel storage capacity class (p-value > 0.05), vessel age (p-value < 0.0001), and percentage of echo-sounder buoys (p-value < 0.01). The random effects structure comprised of the interaction between year and grid cell, the vessel and the statistical area (data correction area). The plots of the residuals (Fig 11) shows that the model assumptions are complied with. The time series resulting from the fit of the model is shown in Fig 5.

Catch per fishing set

1985-2016

The selected logistic model for the binomial proportion of fishing sets with skipjack catch > 0, included the following fixed effects: year (p-value < 0.0001), quarter (p-value < 0.0001), fleet (p-value < 0.0001), and vessel storage capacity class (p-value < 0.01). The random effects included the interaction between year and grid cell – accounting for annual changes of the fished area -, the vessel – due to the longitudinal nature of the data -, and the statistical area (data correction area) – because, as mentioned above, species composition ratios are corrected per sampling area.

The selected log-normal model for catch per positive fishing set, included year (p-value < 0.0001), quarter (p-value < 0.0001), fleet (p-value < 0.0001), and vessel storage capacity class (p-value < 0.0001). The random effects structure is identical to that of the binomial model, comprising of the interaction between year and grid cell, the vessel and the statistical area (data correction area). The plots of the residuals (Fig 11) shows compliance to model assumptions.

The time series resulting from the fit of the two sub-models and their product are shown in Figs 6-9.

2004-2016

The binomial proportion of fishing sets with skipjack catch > 0 is stable and high (over 80%) for this period and was not expected to influence the trend of the CPUE. Thus, the CPUE was here defined as catch per positive fishing set.

The final log-normal model for catch per positive fishing set included year (p-value < 0.0001), quarter (p-value < 0.0001), vessel company (p-value < 0.0001), vessel storage capacity class (p-value > 0.05), vessel age (p-value = 0.05), and percentage of buoys equipped with echo-sounder (p-value < 0.05). The random effects structure comprised of the interaction between year and grid cell, the vessel and the sampling area (data correction area). The plots of the residuals (Fig 11) shows no violation of the model assumptions. The time series resulting from the fit of the model can be seen in Fig 10.

Discussion

The framework for CPUE standardization for the tropical tuna purse seine fisheries accounts for the hierarchical structure of the data, for the non-randomised sampling and, conditional to information availability, for technological developments and evolving fishing strategies. By including the interaction of year/grid cell in the standardisation models, the models essentially predict CPUE for unfished areas. As a result, the standardised CPUE was lower than the nominal CPUE, a phenomenon also observed by (Cao et al., 2011) in their study on the jigging squid fishery. It is also notable that the *company* the vessel belongs to is an important covariate, possibly because different companies develop divergent strategies that in turn drive skippers' decisions. These differences are significant between fleets (ie. Spanish and French companies), especially after the mid-2000s (Gaertner et al., 2017). Similarly, technological advances are captured by the time series of the percentage of buoys equipped with echo sounder, which shows how this type of buoys became a staple for the FOB fishery (Lopez et al 2014). Although the information on echo sounder buoys is highly aggregated it still useful for standardisation purposes, a fact that highlights the importance of covariates that describe technological advancements and their considerable impact on catchability. Nonetheless, to capture the complexity of the FOB fishery, highly disaggregated variables on the use of different types of buoys and FOBs would be desirable, as well as a consistent record of the technology used by each vessel. Currently, the spatial component of this information is largely missing and the data are aggregated at best at the fleet level, whereas information at the vessel level would allow us to disentangle the impact of the use of new technologies and strategies on catchability.

The tropical tuna purse seine fishery has been changing and evolving rapidly. The fishing strategy is intricate due to the complex nature of the skipper's decision-making process and the multitude of tools that skippers have at hand, to help them make informed decisions and maximise profit. Defining an appropriate unit of effort, and consequently catch per unit of effort, has been challenging due to the complexity of this fishery. We compared the standardised time series for two CPUEs one using the fishing hour as the unit of effort and the other using the fishing set. The two CPUEs gave comparable trends; therefore, they can be interchangeable and robust for stock assessment purposes. The CPUEs were also similar, in terms of trends, to the stock status indicators developed by Marsac et al. (2017).

The current exercise highlighted once more the need for high-quality covariates, in terms of relevance and resolution, to add to a better understanding of the use of FOBs and its significance in the fishing strategy and effort. Disaggregated, high-resolution information, possibly at the level of the unit of fishing effort (set/hour), on FOBs and their use is necessary for the improvement of abundance indices.

Acknowledgements

The authors would like to thank all the participants in the workshop *for the development of Skipjack indices of abundance for the EU tropical tuna purse seine fishery operating in the Indian Ocean*, and the AZTI laboratory (Pasaia, Spain) for hosting the event. We are thankful to the IRD - Ob7 – Observatoire des Ecosystèmes Pélagiques Tropicaux exploités for providing part of the data and managing related databases. The workshop and this study were supported by the European Commission – Maritime Affairs and Fisheries (DG MARE).

References

Anonymous, 2012. Report of the 2012 ISSF Stock Assessment Workshop: Understanding Purse Seine CPUE. ISSF Technical Report 2012-10 International Seafood Sustainability Foundation, Washington, DC, USA.

Bishop, J., 2006. Standardizing fishery-dependent catch and effort data in complex fisheries with technology change 16, 21–38.

Bozdogan, H., 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika 52, 345–370.

Cao, J., Chen, X., Chen, Y., Liu, B., Ma, J., Li, S., 2011. Generalized linear Bayesian models for standardizing CPUE: an application to a squid-jigging fishery in the northwest Pacific Ocean. Sci. Mar. 75, 679–689.

Dagorn, L., Holland, K.N., Restrepo, V., Moreno, G., 2013. Is it good or bad to fish with FADs? What are the real impacts of the use of drifting FADs on pelagic marine ecosystems? Fish Fish. 14, 391–415.

Friedman, J., Hastie, T., Tibshirani, R., 2009. glmnet: Lasso and elastic-net regularized generalized linear models. R Package Version 1.

Gaertner D, Ariz J, Bez N, Clermidy S, Moreno G, Murua H, Soto M, 2016. Objectives and first results of the CECOFAD project. Collect Vol Sci Pap ICCAT 72: 391-405.

Gaertner D, Katara I, Billet N, Fonteneau A, Lopez J, Murua H, Daniel P, 2017. Workshop for the development of Skipjack indices of abundance for the EU tropical tuna purse seine fishery operating in the Indian Ocean. 17-21 July 2017. AZTI, Pasaia (Spain).

Gaertner D, Pallares P, 2002. The European Union Research Project, Efficiency of Tuna Purse-Seiners and Effective Effort (ESTHER): Scientific report of project. Part I SCTB15-FTWG-3. SCTB15-FTWG-3.

Katara I., Gaertner D., Chassot E., Soto M., Abascal F., Fonteneau A., Floch L., Lopez J., Cervantes A. (2016). A framework for the standardisation of tropical tuna purse seine CPUE: application to the yellowfin tuna in the Indian Ocean. Victoria Mahé: CTOI, (IOTC-2016-WPTT18-24), 16 p. multigr. (Documents de Réunion; IOTC-2016-WPTT18-24). Working Party on Tropical Tunas, 18, Victoria (SEY), 2016/11/05-10.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 963–974.

Lenth, R., 2014. lsmeans: Least-squares means. R Package Version 2.

Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

Lopez J, Moreno G, Sancristobal I, Murua J 2014. Evolution and current state of the technology of echo-sounder buoys used by Spanish tropical tuna purse seiners in the Atlantic, Indian and Pacific Oceans. Fisheries Research 155: 127-137 doi http://dx.doi.org/10.1016/j.fishres.2014.02.033.

Marsac, F., Fonteneau, A., Lucas, J., Baez, J-C., Floch, L., 2017. Data-derived fishery and stocks status indicators for skipjack tuna in the Indian Ocean. IOTC–2017–WPTT19–43.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Scott GP, Lopez J 2014. The use of FADs in tuna fisheries. European Parliament Policy Department B: Structural and Cohesion Policies: Fisheries IP/B/PECH/IC/2013-123: 70p

Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 273–282.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol. 267–288.

Tables

Table 1 available covariates for CPUE standardisation models.

AVAILABLE COVARIATES	UNIT/FORMAT	TYPE	SOURCE
YEAR OF FISHING OPERATION	year	factor	Derived from ECD (corrected logbook data)
QUARTER OF OPERATION	quarter	factor	Derived from ECD (corrected logbook data)
1X1 GRID SQUARE / CELL	CWP grid	factor	Sheet CWP grid; http://www.fao.org/fishery/cwp/en
AREA OF CATCH CORRECTION	ZET polygon	factor	
UNIQUE VESSEL IDENTIFIER	ID (integer)	factor	c_quille from TURBOBAT
FLEET SEGMENT: FRA AND ESP	FRA or ESP	factor	Derived from 'pays' from TURBOBAT
INITIAL YEAR OF ACTIVITY	Year	numeric	year of activity / initial year of service (an_serv from TURBOBAT)
VESSEL AGE	years	numeric	Calculated from TURBOBAT
CAPACITY CLASS OF THE VESSEL (T)	8 classes	factor	c_cat_b from TURBOBAT
CUMULATED FISHING TIME SPENT BY THE VESSEL IN THE STRATUM	hour	numeric	v_tpec from table ACTIVITE of BALBAYA
CUMULATED NUMBER OF FISHING SETS BY THE VESSEL IN THE STRATUM	numeric		v_nb_calees from ECD
CUMULATED NUMBER OF SUCCESSFUL FISHING SETS BY THE VESSEL IN THE STRATUM	numeric		v_nb_calee_pos from ECD
CUMULATED NUMBER OF UNSUCCESSFUL FISHING SETS BY THE VESSEL IN THE STRATUM	numeric		v_nb_calee_neg from ECD
EXCLUSIVE ECONOMIC ZONE OF ORIGIN OF THE CATCH	ISO_3digit	factor	Shapefile VLIZ_EEZ; ABNJ = Areas Beyond National Jurisdiction;
ANNUAL ESTIMATED FOB VISITS (2003-2017; EXCLUDING FOBS WITH SETS)	numeric		
ANNUAL % OF BUOYS EQUIPPED WITH ECHOSOUNDER (2004-2016)	numeric		
ANNUAL TOTAL NUMBER OF SUPPORT VESSELS	numeric		
SUB-FLEETS BASED ON FISHING STRATEGIES FOR DIFFERENT COMPANIES			2 components in the French fleet, 2 sub-fleets. The French fleet is not homogenous, in terms of their target species. “Sashimi” vessels have different fishing strategies, leading to heterogeneity within the country-fleet.

Figures

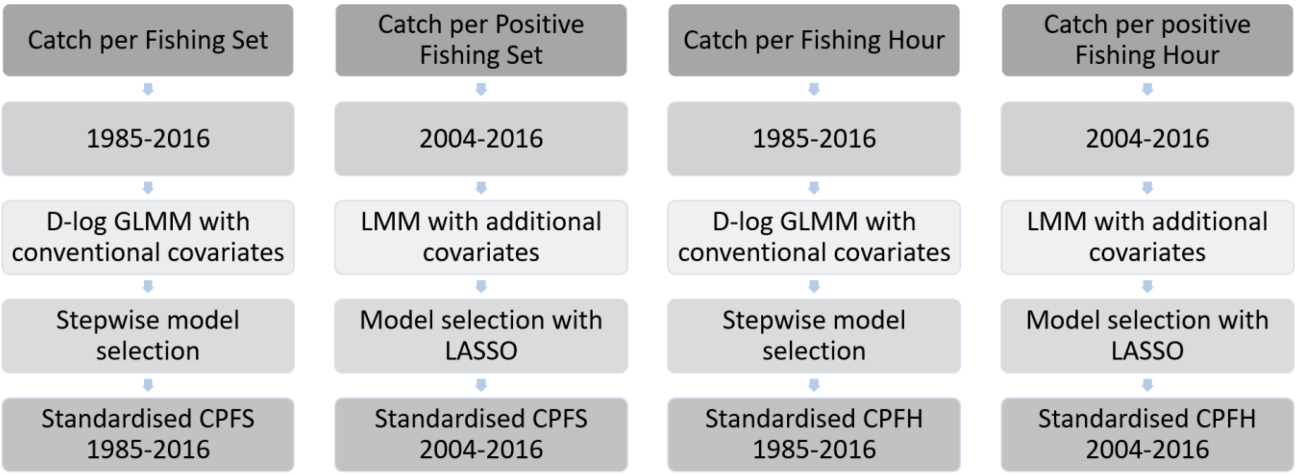

Figure 1 Simplified diagram of the steps taken to derive each of the time series for the standardised CPUE.

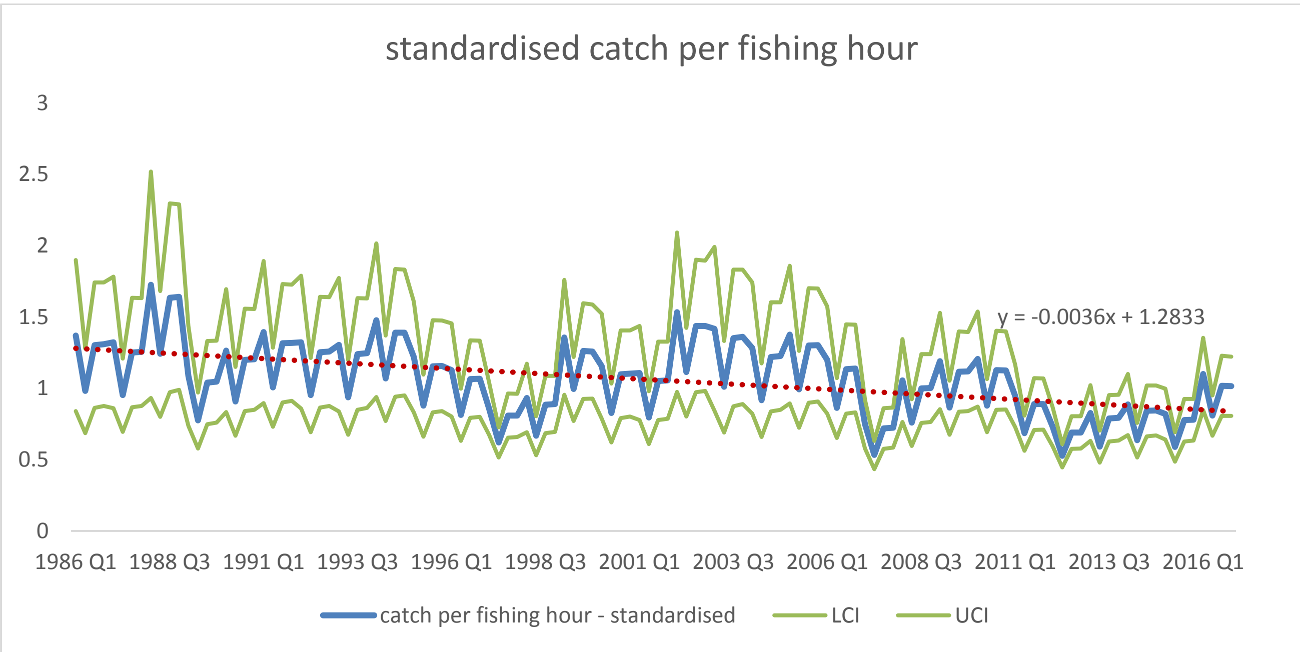

Figure 2 the product of the lsmeans fit of the 2 components of the delta-lognormal model and 95% confidence intervals (LCL: lower confidence limit; UCL: upper confidence limit) for skipjack catch per fishing hour.

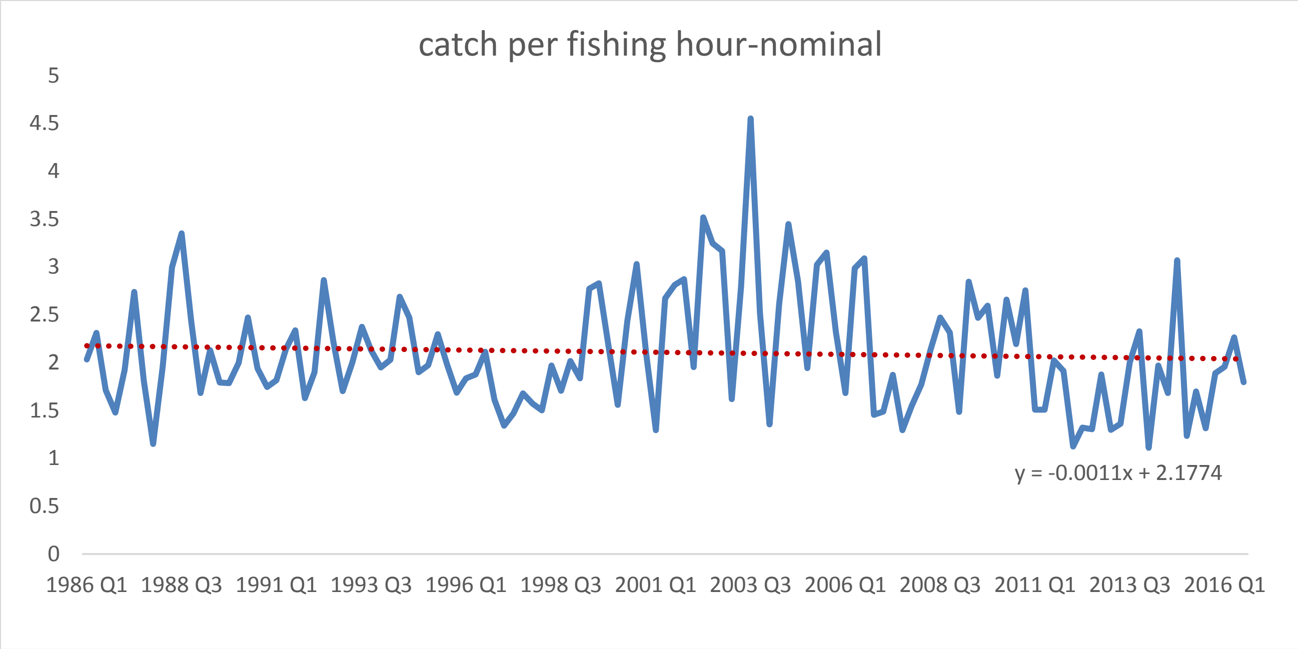

Figure 3 Nominal skipjack catch per fishing hour for the period 1986-2016.

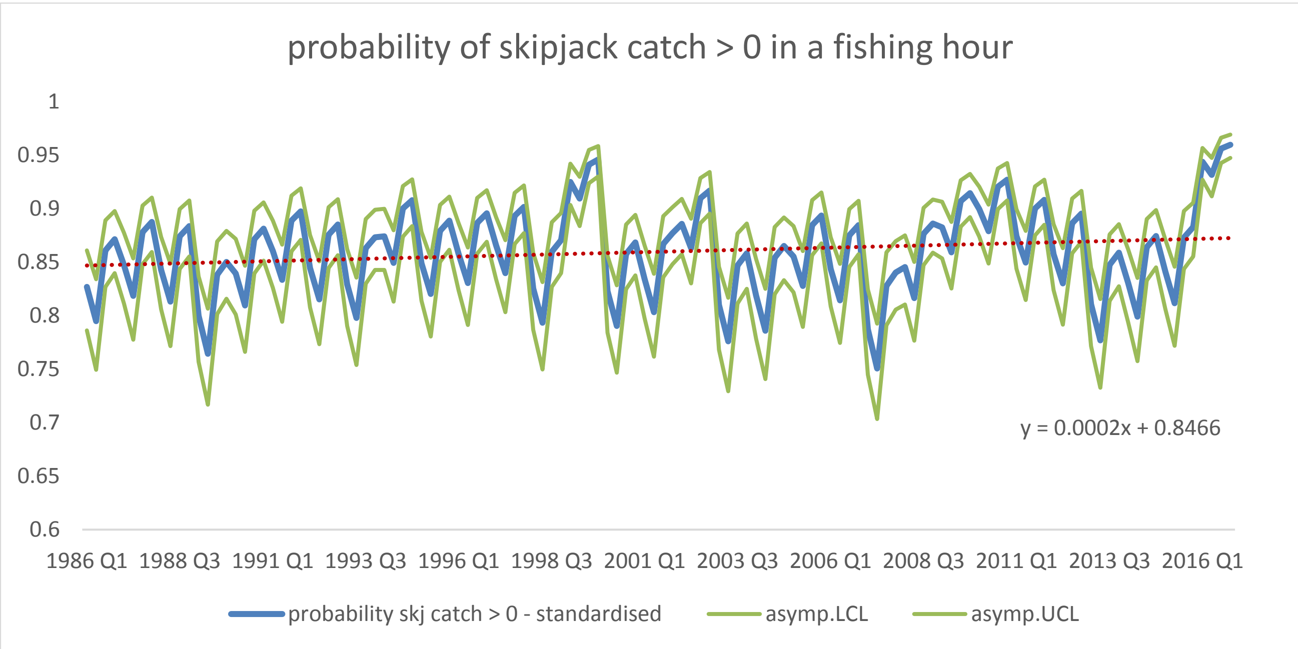

Figure 4 the lsmeans fit and 95% confidence intervals (LCL: lower confidence limit; UCL: upper confidence limit) of the model for the probability of skipjack catch in a fishing hour being > 0, for the period 1986-2016

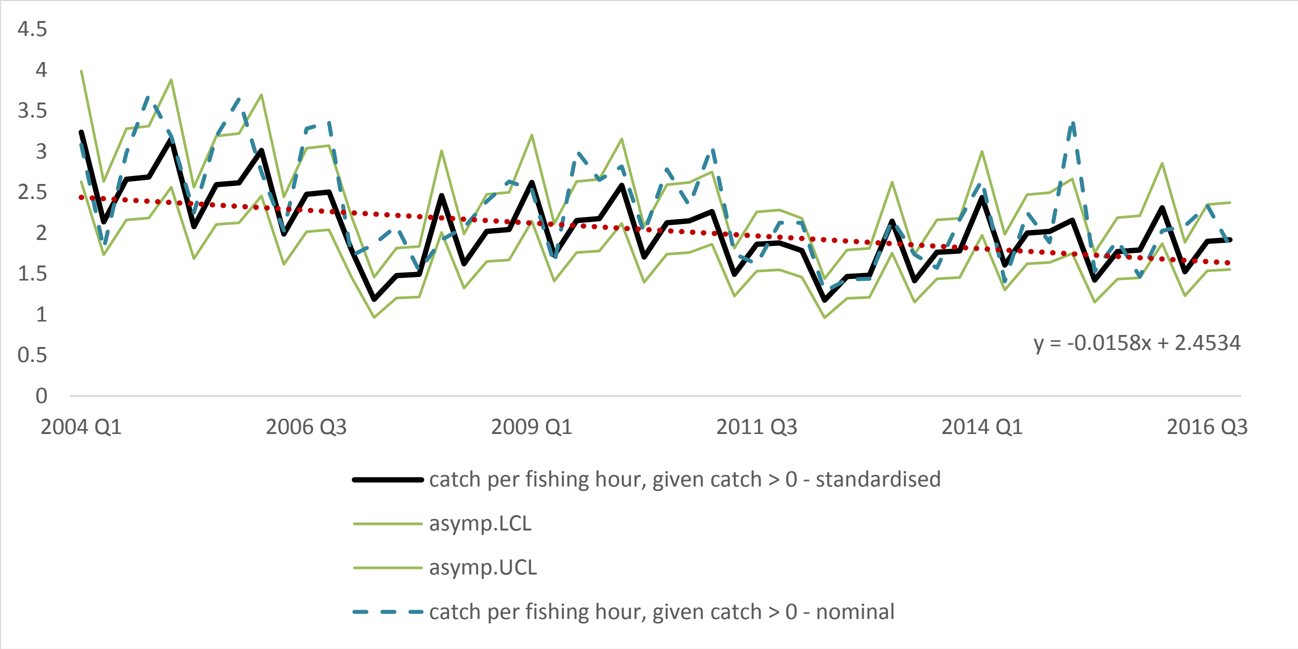

Figure 5 the lsmeans fit, 95% confidence intervals (LCL: lower confidence limit; UCL: upper confidence limit) of the lognormal model for catch per fishing hour, given catch > 0, for the period 2004-2016. Nominal values are also included.

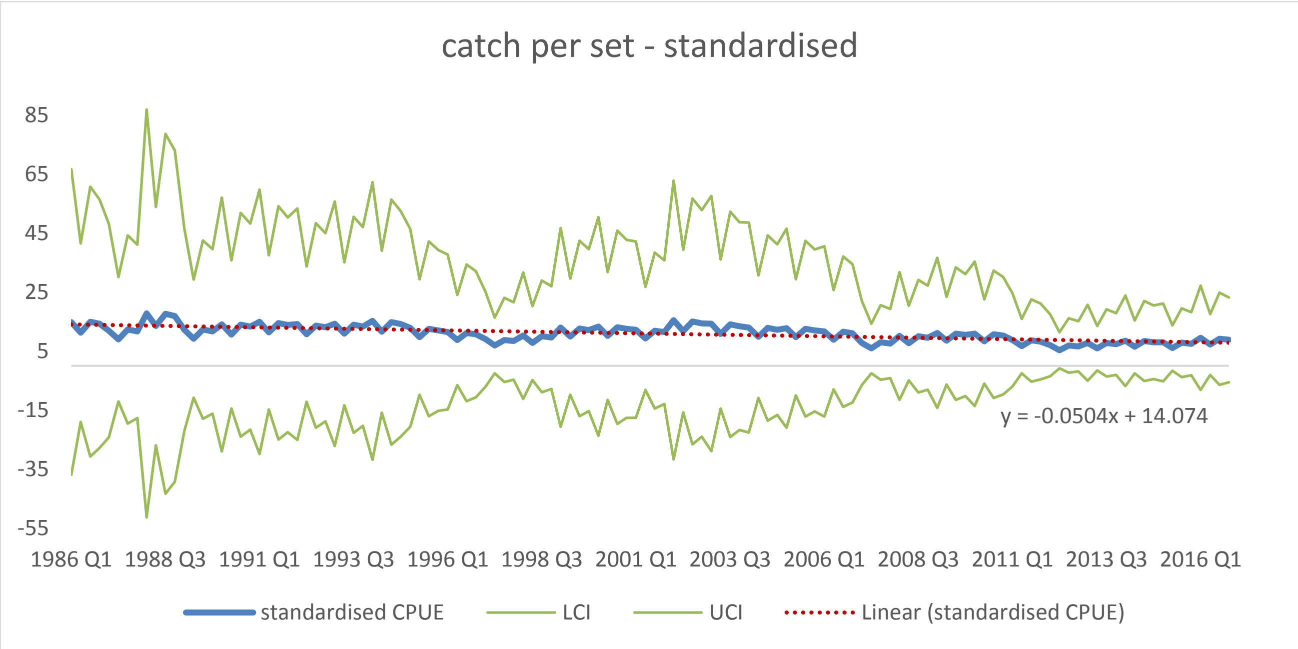

Figure 6 the product of the lsmeans fit of the 2 components of the delta-lognormal model and 95% confidence intervals (LCL: lower confidence limit; UCL: upper confidence limit) for skipjack catch per set.

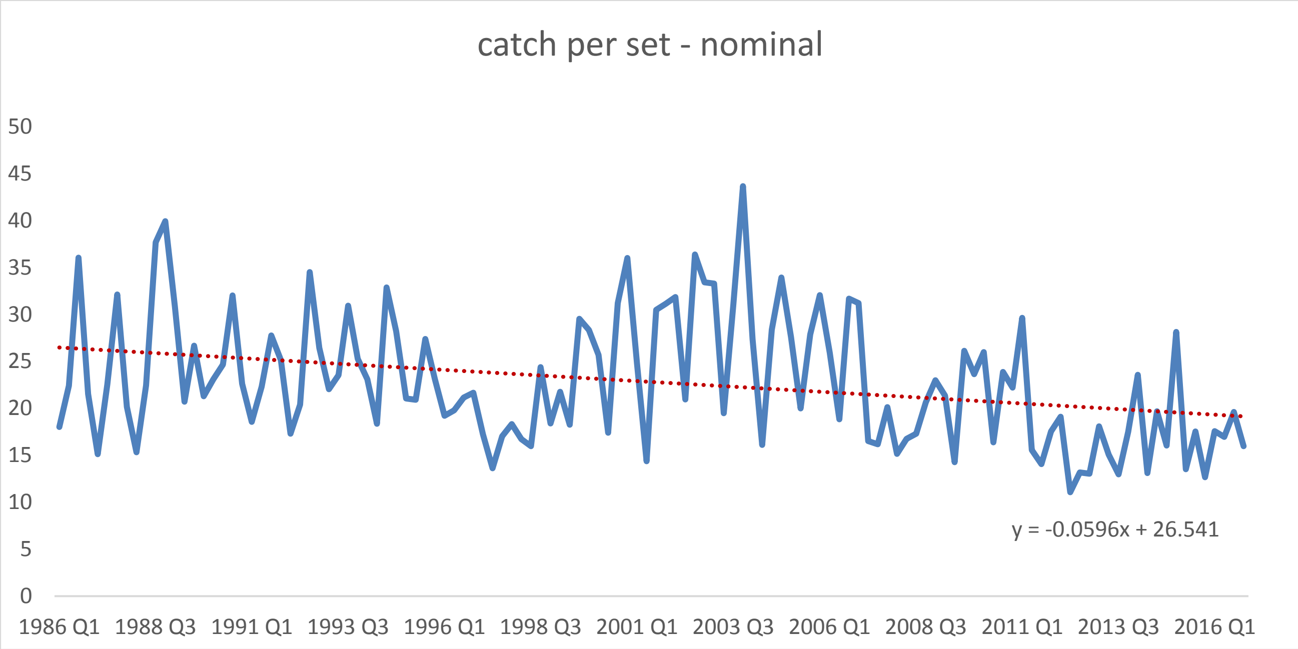

Figure 7 Nominal skipjack catch per fishing set for the period 1986-2016.

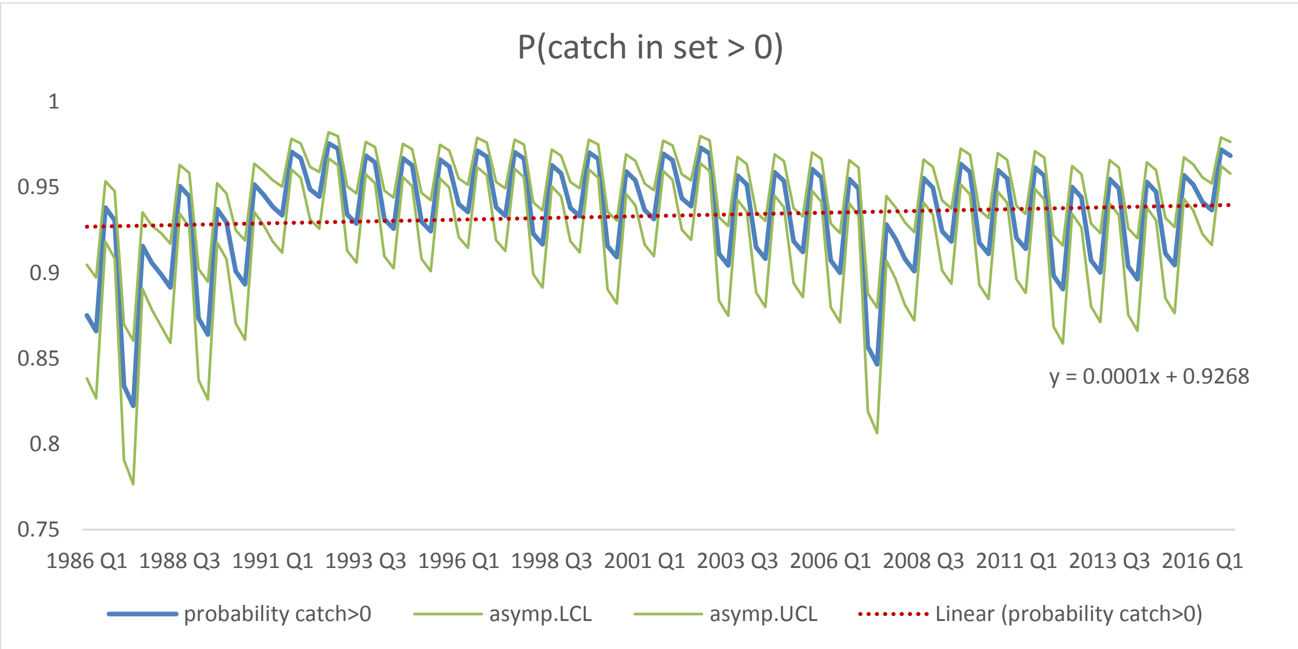

Figure 8 the lsmeans fit and 95% confidence intervals (LCL: lower confidence limit; UCL: upper confidence limit) of the model for the probability of skipjack catch in a set being > 0, for the period 1986-2016

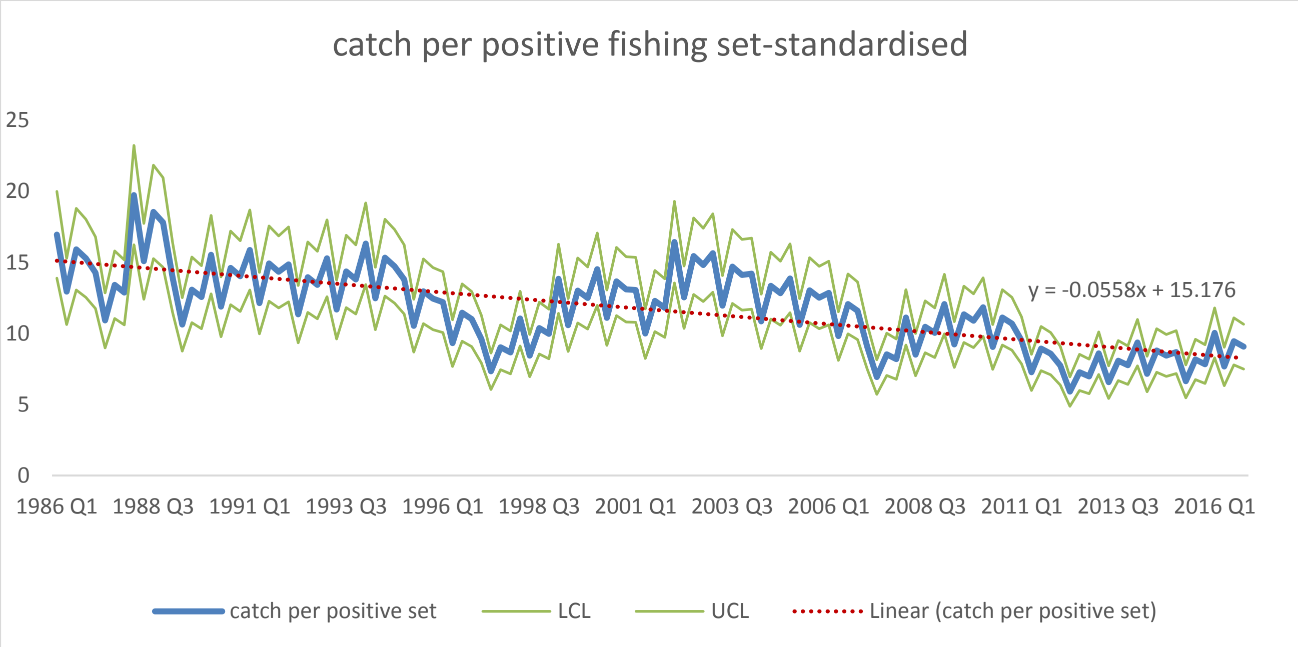

Figure 9 the lsmeans fit and 95% confidence intervals (LCL: lower confidence limit; UCL: upper confidence limit) of the lognormal model for catch per set, given catch > 0, for the period 1986-2016

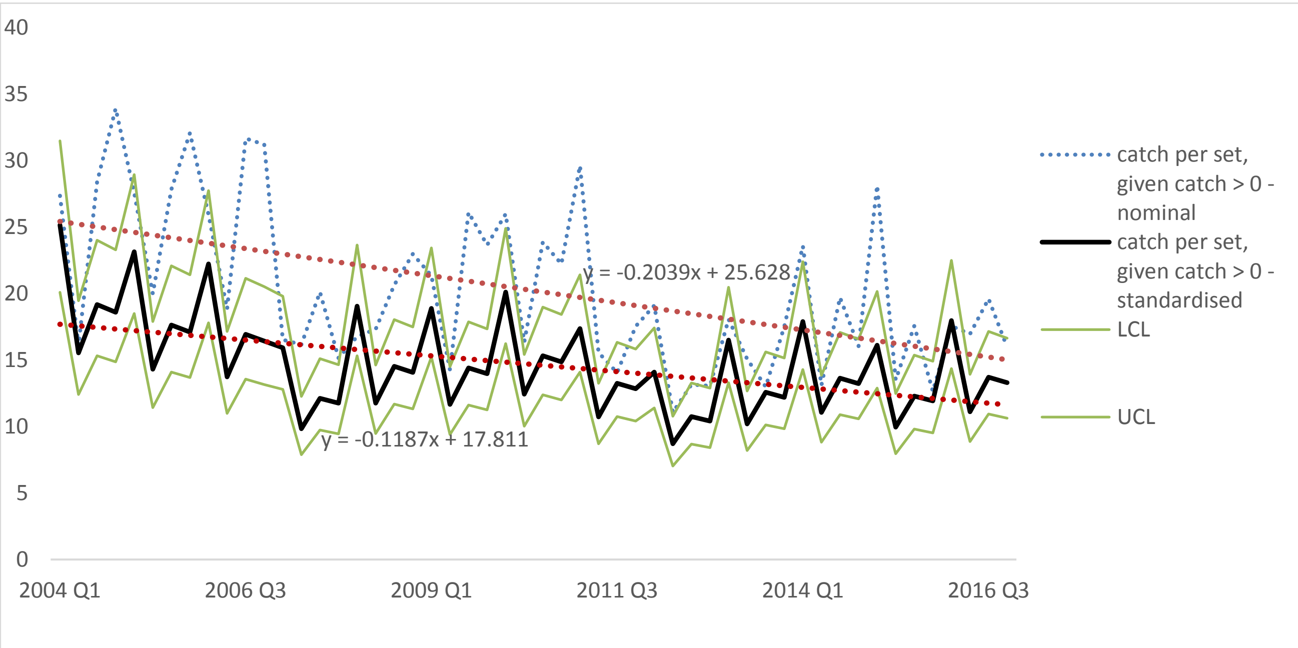

Figure 10 **the lsmeans fit, 95% confidence intervals (LCL: lower confidence limit; UCL: upper confidence limit) of the lognormal model for catch per set, given catch > 0, for the period 2004-2016. Nominal values are also included.**

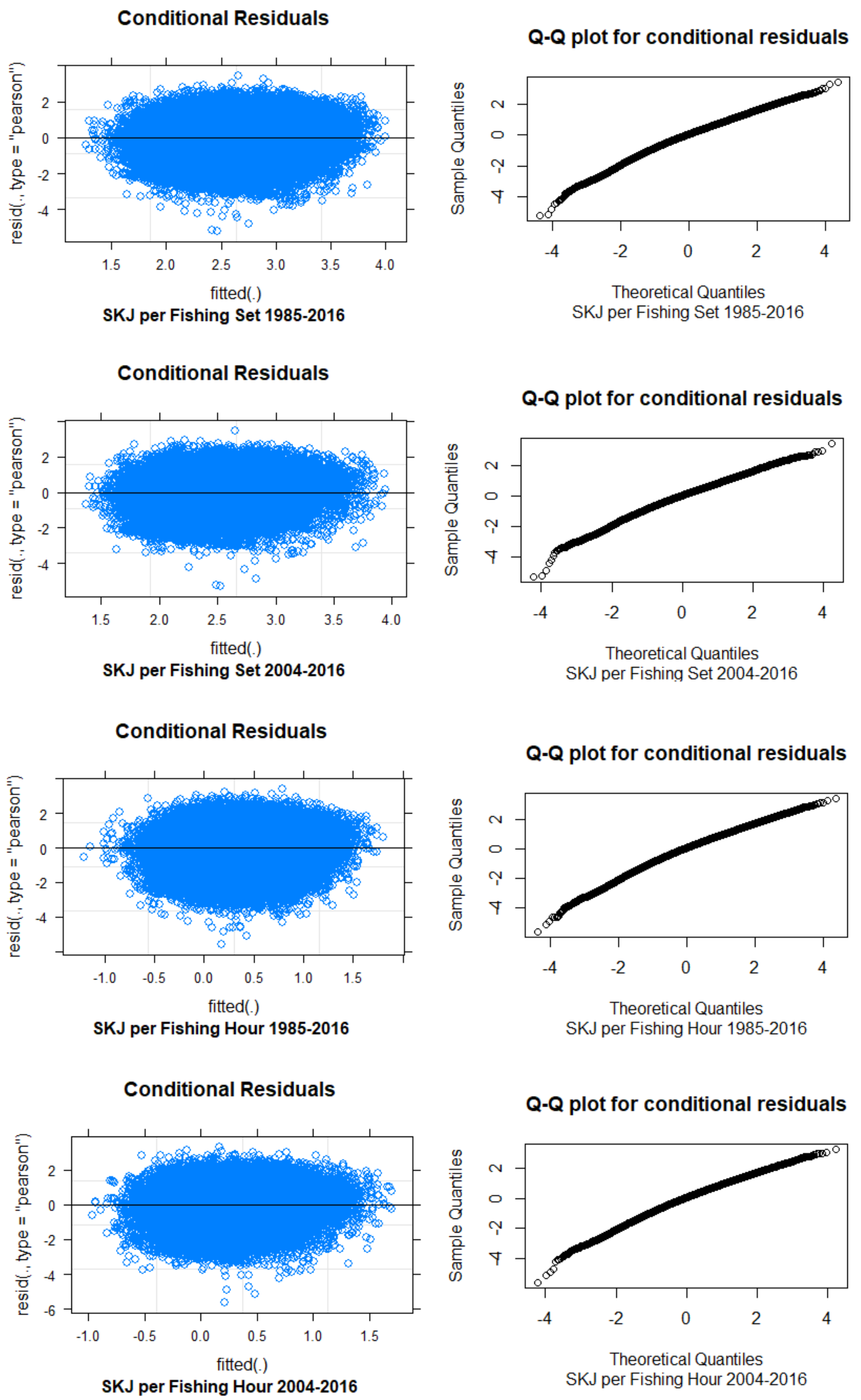

Figure 11 **Diagnostic plots for the log-normal models (from top to bottom): (i) catch per fishing set, conditional to catch > 0 for 1986-2016; (ii) catch per fishing set, conditional to catch > 0 for 2004-2016; (iii) catch per fishing hour, conditional to catch > 0 for 1986-2016; (iv) catch per fishing hour, conditional to catch > 0 for 2004-2016.**