DRAFT

Skipjack CPUE series standardization by fishing mode for the European purse seiners operating in the Indian Ocean

Guery, L., Aragno, V., Kaplan, D., Grande M., Baez , J.C.,
Abascal, F., Urunga J., Marsac, F., Merino, G. and Gaertner, D.

08/06/2020

1 Introduction

2 Material and Methods

2.1 Conventional fishing data

Logbook data for the French and Spanish purse seine fleets targeting tropical tuna in the Indian Ocean from 1991 to 2018 were analysed to derive the standardised CPUEs. The logbook databases are managed by the Tuna Observatory (IRD-Ob7) and the IEO for the French and the Spanish fleets, respectively. For the calculation of the FAD density used in the FAD CPUE standardization, FAD data were provided by IRD-Ob7 for France and by AZTI for Spain. The raw logbook data (Level0) produced by the skippers were corrected in terms of total catch per set (to account for the difference between reported catch at sea and landed catch) and species composition (based on port size sampling and the T3 methodology – see Pallarès and Hallier 1997) to generate the Level 1 logbook database used in this paper. The free-school sets (FSC) dataset, i.e. non-associated school sets and whales' sets, was used to derive CPUE for the SKJ stock.

The analysis was restricted to:

- The period 1991-2018 for FSC sets to be able to have the larger coverage of days without set reported
- The area defined by all grid cells where SKJ were fished for at least 5 years over a period of no less than 15 years, to avoid areas that are not routinely fished
- Vessels with fewer activities than the 5% of the left hand distribution based on the cumulative number of days per boat (all activities confounded) were removed
- Entire days with at least one activity with problematic operations were removed
- All sets per boat and day were aggregated and attributed to the centroid of these set activities
- Distances between successive sets null-FSC/next-FSC for a boat is not significantly different from all other combinations: no need of buffer avoiding to count the same school several times
- Total number of sets per day per boat was filtered and days with unrealistic data were removed

Different time periods (1991-2018 and 2010-2018) were considered depending on the availability of the candidate variables.

For FAD sets on 1991-2018 period and 2010-2018 period, lognormal GLMMs were developed. They describe the catch conditional to positive set. Positive sets only were considered as 90-95% of FAD set are successful with no trend over time.

2.2 Modelling approach

2.2.1 Delta-lognormal GLMMs

For FSC sets on 1991-2018 period and 2010-2018 period, Delta-lognormal GLMMs were used (Guéry et al. ICCAT) It includes three sub-models:

- a poisson GLMM (component 1) that standardises the number of positive and negative sets
- a binomial GLMM (component 2) that takes into account the fraction of positive sets with SKJ
- a lognormal LMM (component 3) to describe the catch conditional to a positive set.

Available candidate variables are detailed in Table 1. We performed the Poisson GLMM (component 1) where the full model included the following fixed effects: fleet country, age of the vessel, number of sets on FOB, vessel storage capacity, piracy, year, quarter and 5ºx5º grid cell or Gulland's index. Gulland's index was tentalively used to depict changes over time in the spatial distribution of the fishing effort due to the capacity of purse seiners to locate the richest areas in free schools. The number of FOB sets per trip was included as a proxy for vessels' fishing strategy changes across time due to the increase of dFADs. The random structure of the model includes a vessel unique identifier. The time spent by searching centroid by day was calculated as (sun set time – sun rise time) – (number of set*median of setting time) and was used as an offset.

Component 1: num_sets_fsc ~ fleet country + age of the vessel + num_sets_fob + vessel storage capacity + piracy + year + quarter + cwp55_group or Gulland's index + (1| numbat) + offset(searching_centroid)

The full model for the binomial GLMM and the lognormal LMM (component 2 and component 3, respectively) included the following fixed effects: fleet country, vessel storage capacity, year, quarter, 5ºx5º grid cell or Gulland index, age of the vessel (only for lognormal LMM). The random structure of these models included a vessel unique identifier. The number of positive sets was used as an offset.

Component 2: SKJ_pos ~ fleet country + vessel storage capacity + year + quarter + cwp55_group or Gulland's index + (1 | numbat) + offset(nb of positive sets)

Component 3: log_capture ~ fleet country + age of the vessel + vessel storage capacity + year + quarter + cwp55_group or Gulland's index + (1|numbat) + offset(nb of positive sets)

2.2.2 Lognormal GLMM

For the 1991-2018 period, the lognormal LMM included the following fixed effects: fleet country, vessel storage capacity, year, quarter and 5ºx5º grid cell or Gulland index. The random structure of these models included a vessel unique identifier. The number of positive sets was used as an offset.

log_capture ~ fleet country + age of the vessel + vessel storage capacity + year + quarter + cwp55_group or Gulland's index + (1|numbat) + offset(nb of positive sets)

For the 2010-2018 period, the lognormal LMM included the following fixed effects: fleet country, vessel storage capacity, year, quarter, FAD density, Indice of supply vessel and 5ºx5º grid cell or Gulland index. The random structure of these models included a vessel unique identifier. The number of positive sets was used as an offset.

log_capture ~ fleet country + age of the vessel + vessel storage capacity + year + quarter + FAD density + Indice of supply vessel + cwp55_group or Gulland's index + (1|numbat) + offset(nb of positive sets)

2.2.3 LASSO regression

For multidimensional data, variable or model subset selection through stepwise selection becomes problematic. The number of possible models grows exponentially with the number of predictors and renders

computation infeasible. Moreover, when the number of observations is not much larger than the number of predictors, ordinary least squares may result in over-fitting. Penalized maximum likelihood methods allow regression modelling when the number of model parameters is high compared to the number of observations and prevent over-fitting (Tibshirani 1996). Models, consisting of all predictors, were fitted with LASSO, a popular technique that constrains regularizes) the coefficient estimates. Technically, Lasso minimizes the usual sum of squared errors, with a bound on the sum of the absolute values of the coefficients (Tibshirani, 1996, 2011). Model selection involved the use of the LASSO regression using algorithms that handle continuous explanatory 0) and grouped covariates (R package: grpreg; Breheny nd Breheny, 2018). Given a linear regression with standardized predictors X and centred response values y, the glmnet algorithm estimates the regression coefficients β to minimize:

$\frac{1}{N}||y - X\beta||_2^2 + \lambda||\beta||_1$ for $\beta \in R^p$

where λ covers a range of values. The tuning parameter λ is chosen through cross-validation. The LASSO procedure was followed by backward model selection for both the random and fixed effects of the mixed models using AIC and BIC. Finally, the selected model was refitted as an unrestricted GLMM (Rpackage: lme4; Bates et al., 2014) but not with LASSO, as LASSO estimated coefficients are known to be biased (Friedman et al., 2001). Finally, the standardized CPUEs were fitted using estimated means. All the statistical analyses were computed using the software R (v4.0.0; R Core Team, 2020).

3 Results

3.1 Lasso regression

Results are presented in Table 2

3.2 FSC sets (1991-2018 period)

3.2.1 Poisson GLMM (total number of free school sets, i.e. successful and unsuccessful)

Diagnostics are presented in fig(1, 2). Standardized time series are presented by year and year-quarter in fig(3, 4).

3.2.2 Binomial GLMM (fraction of positive set with SKJ)

Diagnostics are presented in fig(5, 6, 7 8). Standardized time series are presented by year and year-quarter in fig(9, 10).

3.2.3 Log-Normal GLMM (catch per hour conditional to SKJ catch > 0)

Diagnostics are presented in fig(11, 12, 13, 14). Standardized time series are presented by year and year-quarter in fig(15, 16).

3.2.4 Product of the 3 components (standardized CPUE)

Standardized time series are presented by year and year-quarter in fig(17, 18).

3.3 FSC sets (2010-2018 period)

3.3.1 Poisson GLMM (total number of free school sets, i.e. successful and unsuccessful)

Diagnostics are presented in fig(19, 20). Standardized time series are presented by year and year-quarter in fig(21, 22).

3.3.2 Binomial GLMM (fraction of positive set with SKJ)

Diagnostics are presented in fig(23, 24, 25, 26). Standardized time series are presented by year and year-quarter in fig(27, 28).

3.3.3 Log-Normal GLMM (catch per hour conditional to SKJ catch > 0)

Diagnostics are presented in fig(29, 30, 31, 32). Standardized time series are presented by year and year-quarter in fig(33, 34).

3.3.4 Product of the 3 components (standardized CPUE)

Standardized time series are presented by year and year-quarter in fig(35, 36).

3.4 FAD sets (1991-2018 period)

3.4.1 Log-Normal GLMM (catch per hour conditional to SKJ catch > 0)

Diagnostics are presented in fig(37, 38, 39, 40). Standardized time series are presented by year and year-quarter in fig(41, 42).

3.5 FAD sets (2010-2018 period)

3.5.1 Log-Normal GLMM (catch per hour conditional to SKJ catch > 0)

Diagnostics are presented in fig(43, 44, 45, 46). Standardized time series are presented by year and year-quarter in fig(47, 48).

4 Discussion

5 Acknowledgements

6 References

Table 1: Available variables for the calculation of CPUE and the development of the standardisation models

Variable	Description
Fleet country	France, Spain
Numbat	Unique vessel identifier
Vessel storage capacity	In cubic metre
Cwp 55 grid cell	Reference grid of the fishing area
Gulland's index	Measure the extent to which a fleet has concentrated its fishing effort in areas with higher than average catch rate
Number of sets on FOBs	Monthly resolution per gris cell
Number of positive set	Number of positive sets per boat per day per centroid
Year	Year activity
Quarter	Quarter of the year
Age of vessel	In year
FAD density	in CWP 11 gris cell
Indice of supply vessel	Grow with aditional supply vessel
Piracy	Zone of piracy
Searching time	In hours

Table 2: LASSO regression results

	FSC 1991-2018			FSC 2010-2018			FAD 1991-2018	FAD 2010-2018
	Component 1	Component 2	Component 3	Component 1	Component 2	Component 3	Component 3	Component 3
Lambda min	0,001	0,0011	0,0033	0,0008	0,0007	0,0005	0,0002	0,0005
Variable	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Intercept	2,5747	-1,1508	1,4657	4,607	-5,7471	0,6913	2,4678	2,3159
Country	0,2821	0,1602	0,1421	-0,0321	0.5164	0,5773	0,0821	0,2146
Capacity	-0,0403	-0,0611	0,0996	-0,0313	-0,1675	0,0872	0,1148	0,0909
Year	0,0033	-0,023	0,0086	-0,0317	0,029	-0,0184	-0,0165	0.0394
Quarter	-0,0624	0.123	0,0961	-0,0785	0,1009	0.0348	-0,0213	-0,0683
Cwp 55	0,0144	0.0298	-0,0095	0,0193	0,0123	-0,0083	-0,0118	-0,0054
Num bat	0,0034	-0,0064	0,0001	-0,0037	-0,0002	0.0085	0,001	0,002
Gulland	-0,1154	0.9082	0.221	-0,2414	1,2798	0,3549	0.0065	0,0475
Fad density	Not use	Not use	Not use	Not use	Not use	Not use	Not use	-0,0783
Ind supply	Not use	Not use	Not use	Not use	Not use	Not use	Not use	0,0593
Age	0,0158	Not use	0,0078	-0,0199	Not use	0,0872	0,0017	-0,0025
Piracy	-0,0596	Not use	Not use	-0,0259	Not use	Not use	Not use	Not use
Nbr FOB	-1,016	Not use	Not use	-1,0369	Not use	Not use	Not use	Not use
Searching T	-0,483	Not use	Not use	-0,5567	Not use	Not use	Not use	Not use
Nbr pos sets	Not use	0.0931	0.4369	Not use	3.1313	0,5265	0,5052	-0,115

Model with CWP55 position

Model with CWP55 position

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

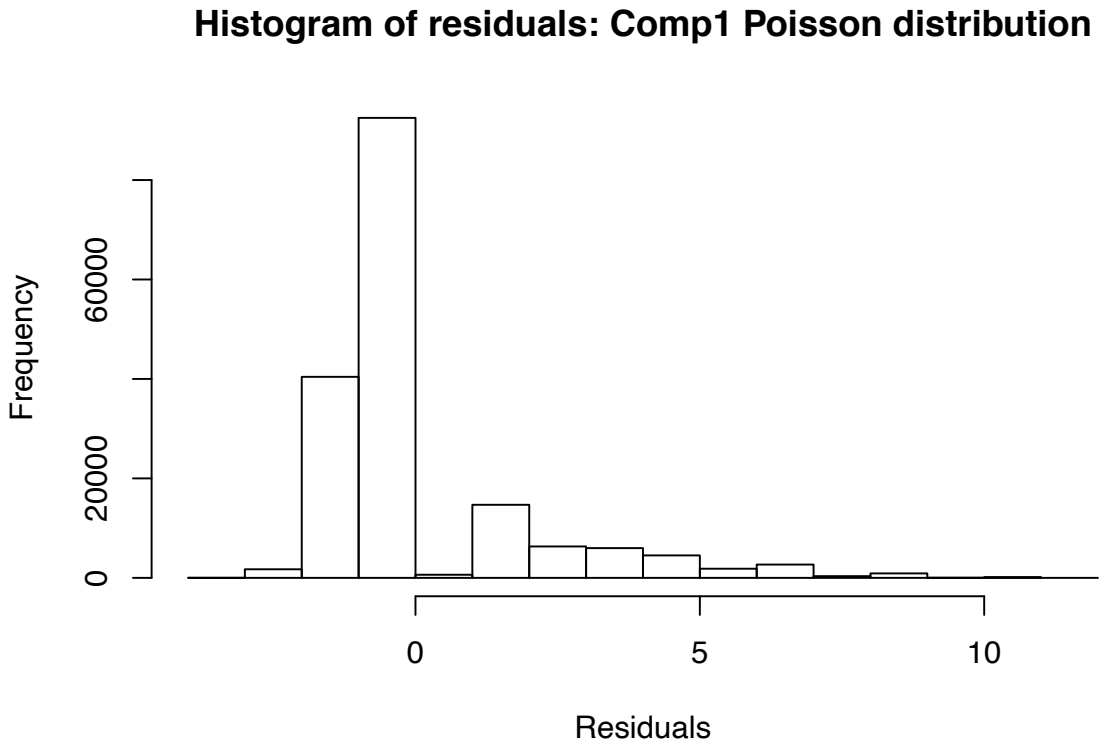

Figure 1: Histogram of residuals for component 1 (total number of free school sets > 0 and $=0$) for the 1991-2018 period. Model including Gulland index.

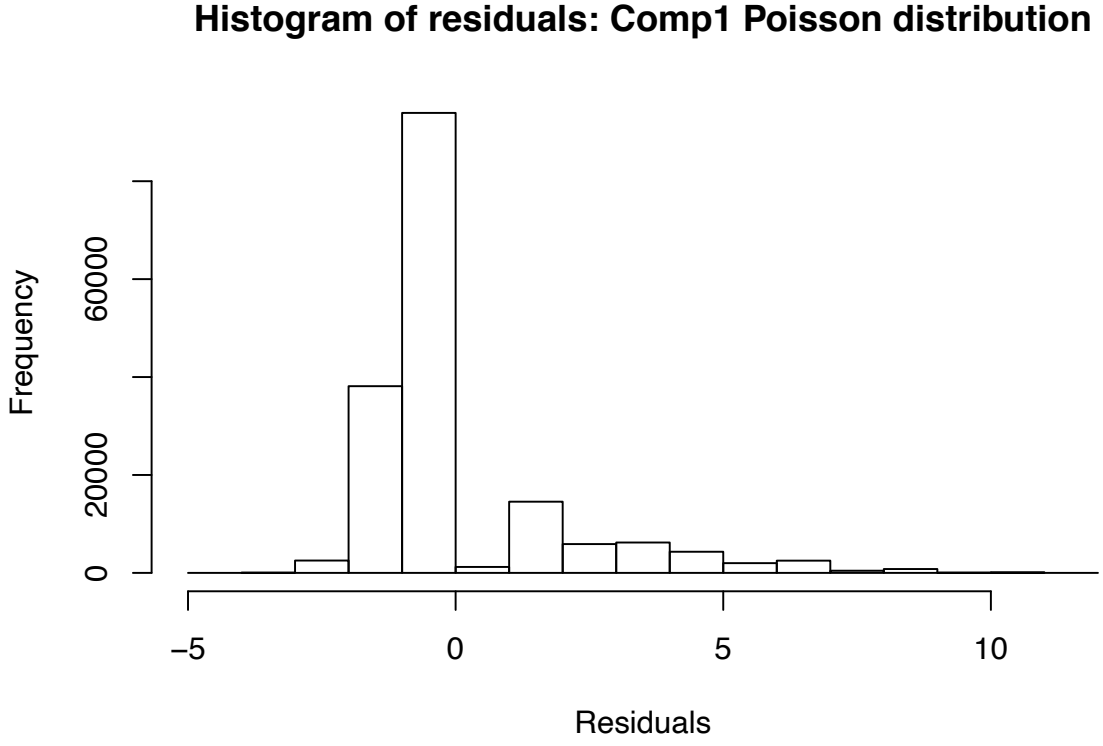

Figure 2: Residuals for component 1 (total number of free school sets > 0 and $=0$) for the 1991-2018 period. Model including spatial variable (cwp55).

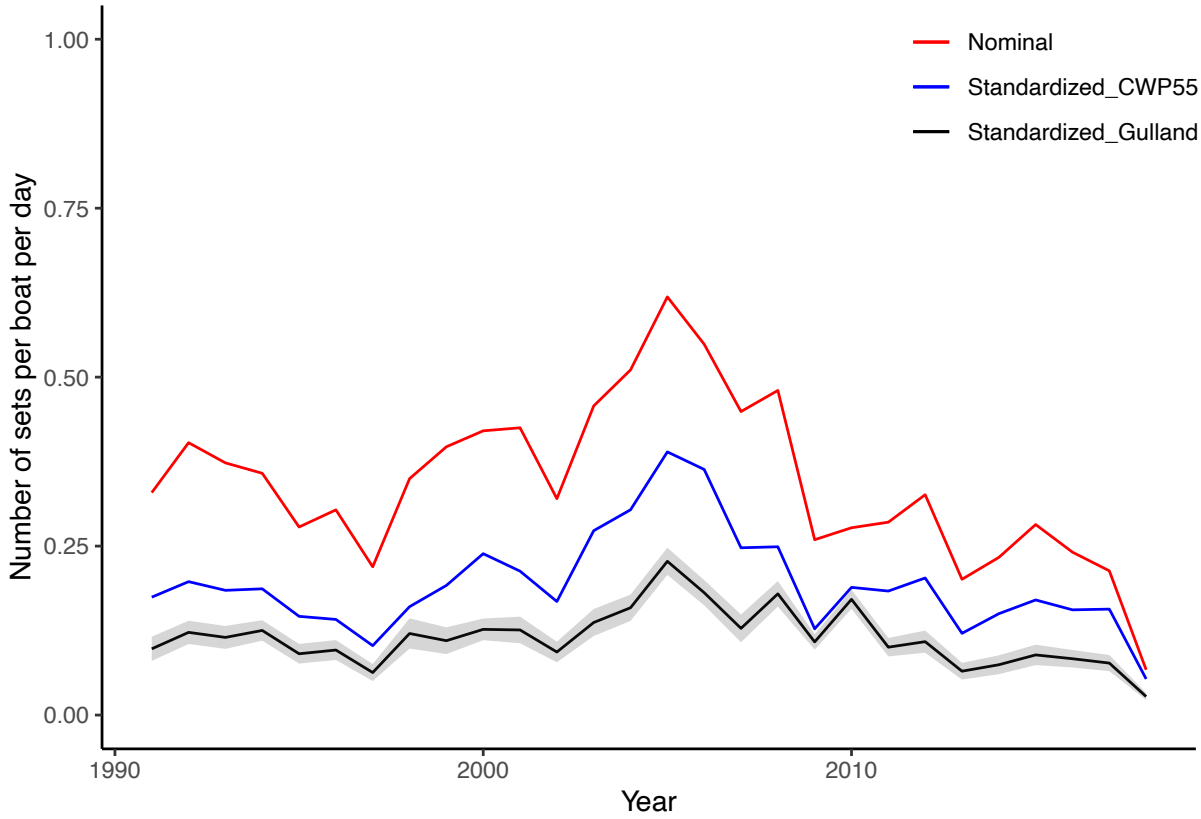

Figure 3: Standardised time series (1991-2018) by year of the number of free school sets > 0 and $=0$ (component 1) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

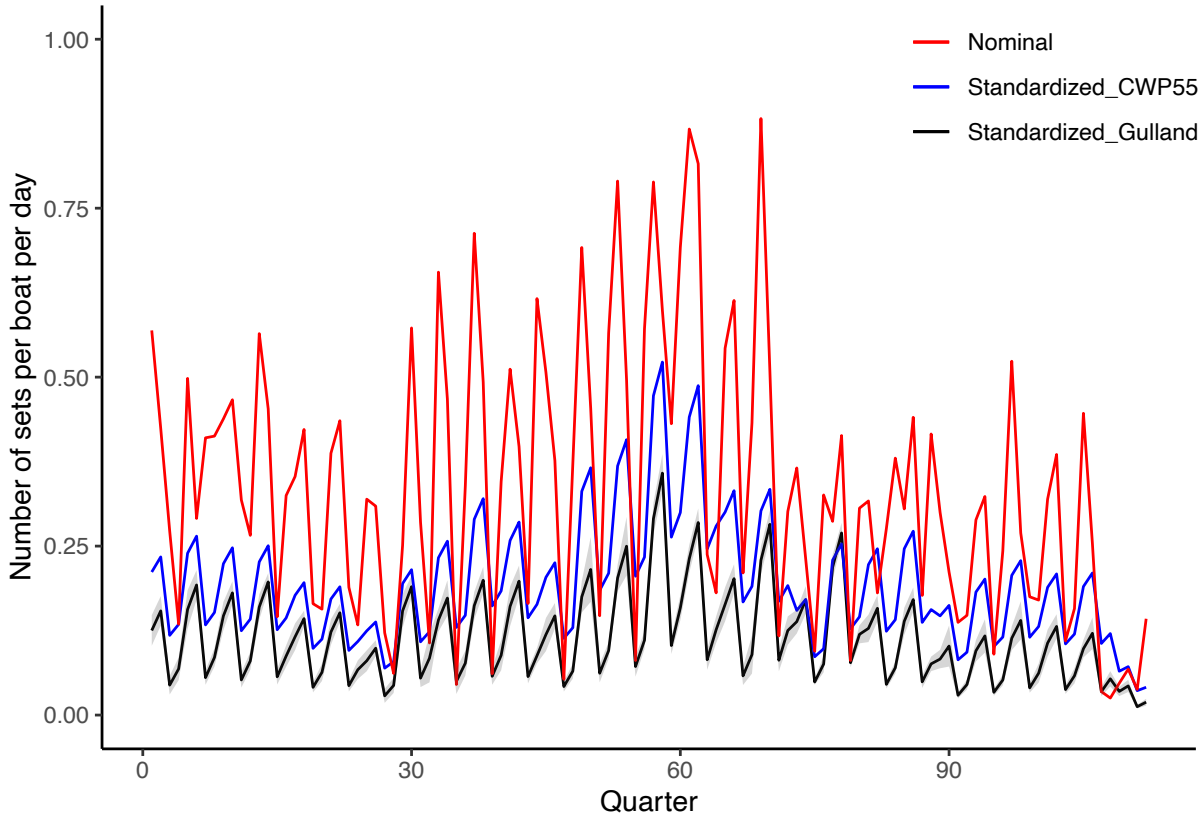

Figure 4: Standardised time series (1991-2018) by quarter of the number of free school sets > 0 and $=0$ (component 1) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

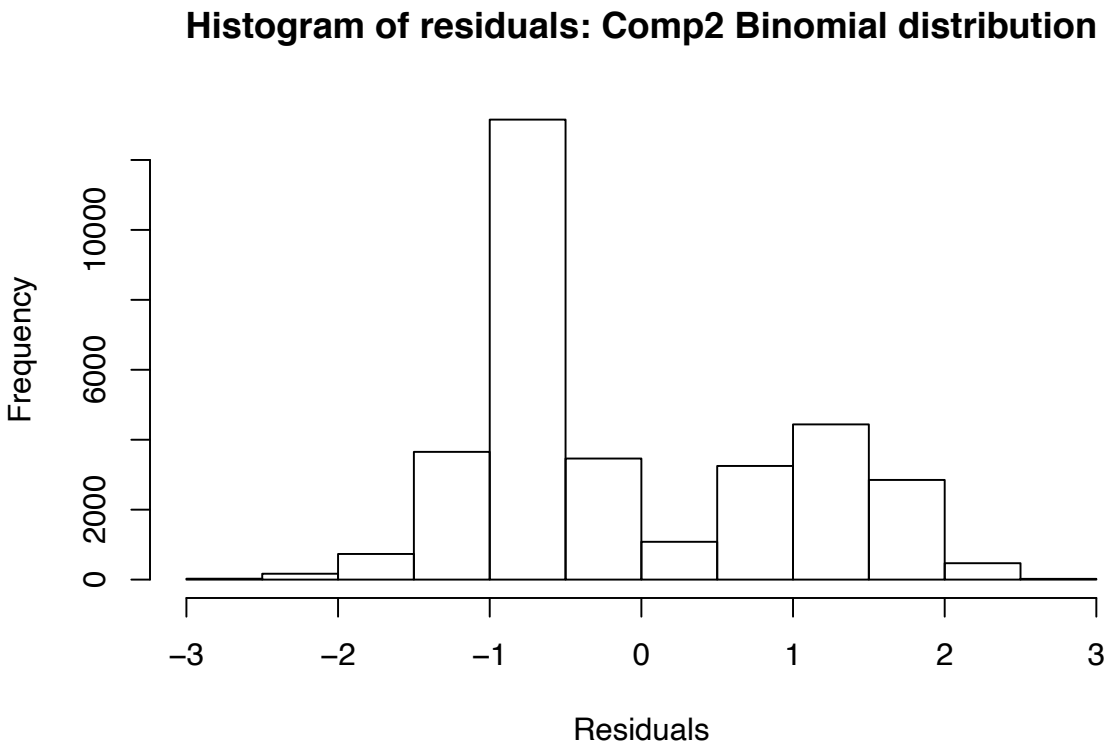

Figure 5: Histogram of residuals for component 2 (proportion of positive free school sets with SKJ catch) for the 1991-2018 period. Model including Gulland index.

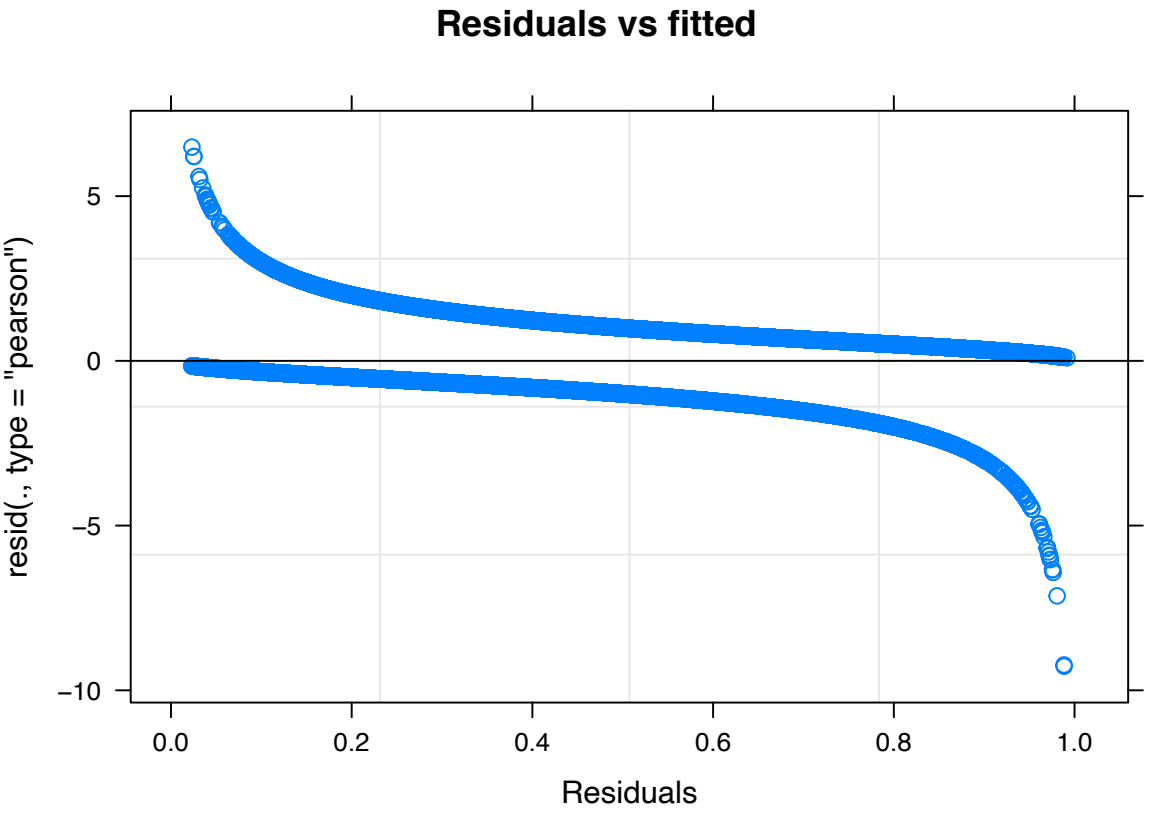

Figure 6: Plot of residuals versus fitted values for component 2 (proportion of positive free school sets with SKJ catch) for the 1991-2018 period. Model including Gulland index.

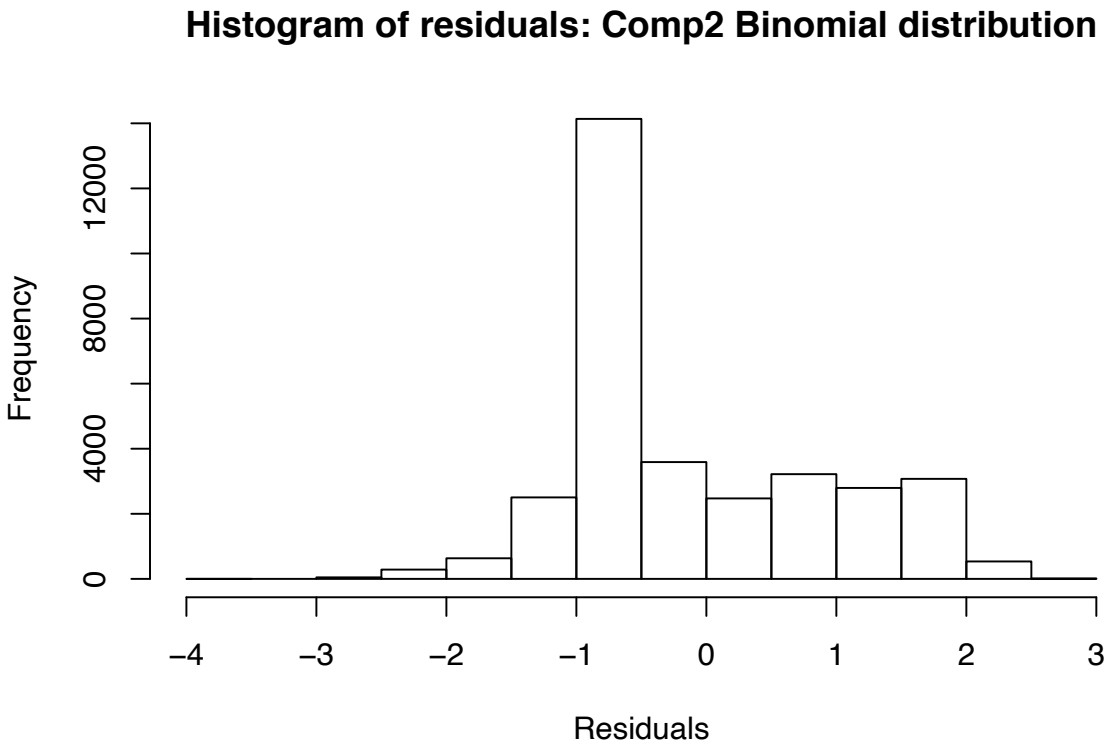

Figure 7: Histogram of residuals for component 2 (proportion of positive free school sets with SKJ catch) for the 1991-2018 period. Model including spatial variable (cwp55).

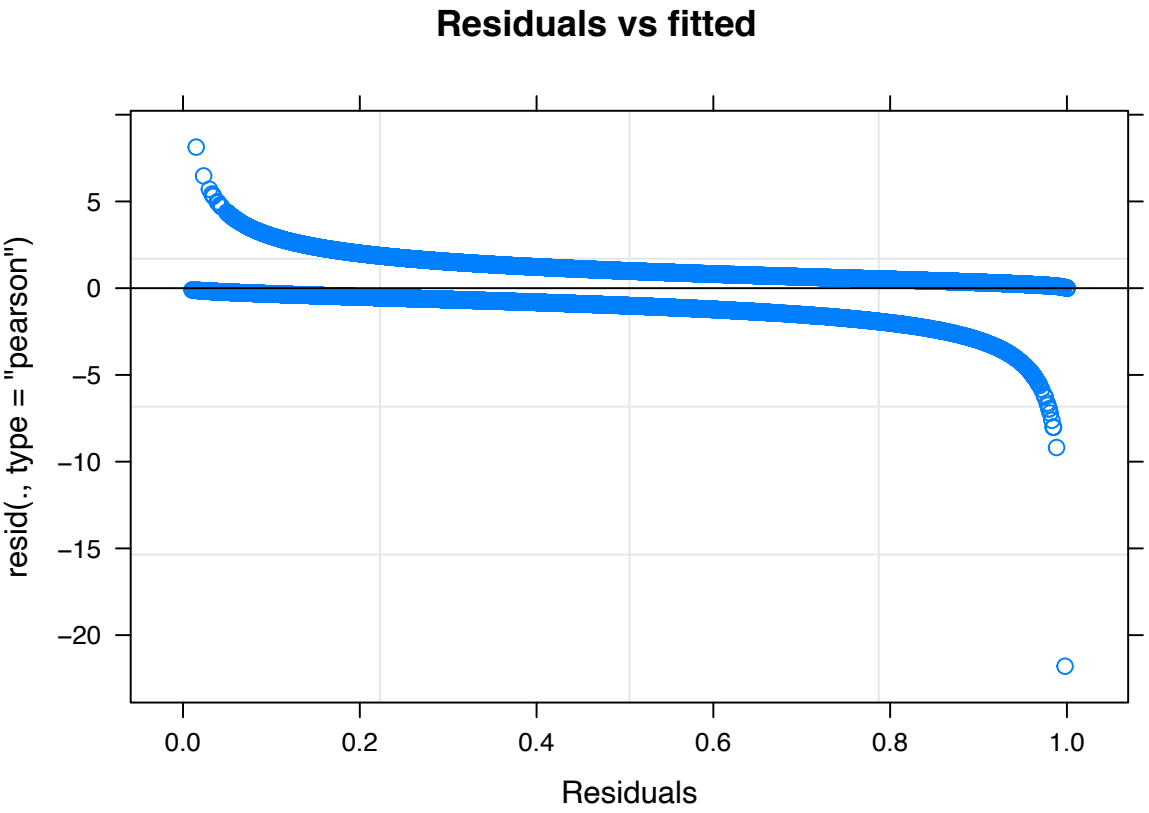

Figure 8: Plot of residuals versus fitted values for component 2 (proportion of positive free school sets with SKJ catch) for the 1991-2018 period. Model including spatial variable (cwp55).

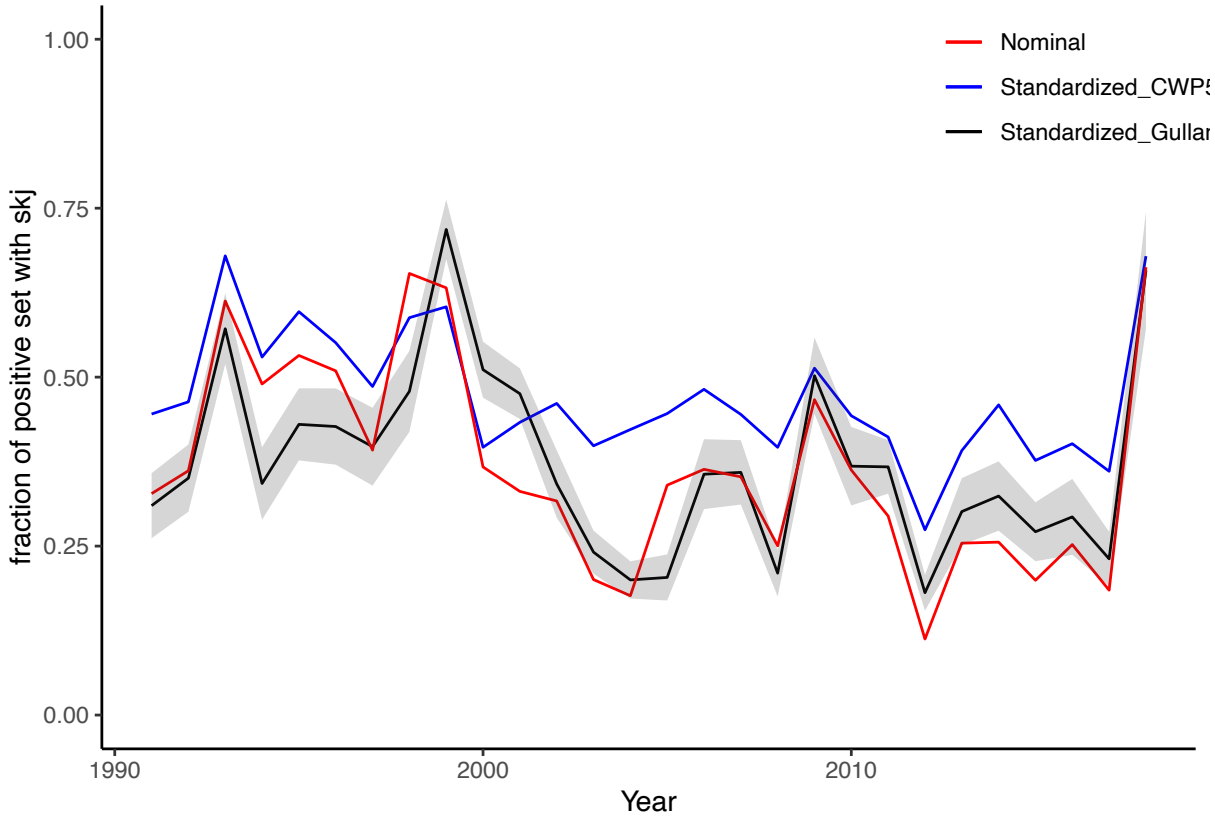

Figure 9: Standardised time series (1991-2018) by year of the proportion of successful free school sets with SKJ catch (component 2) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

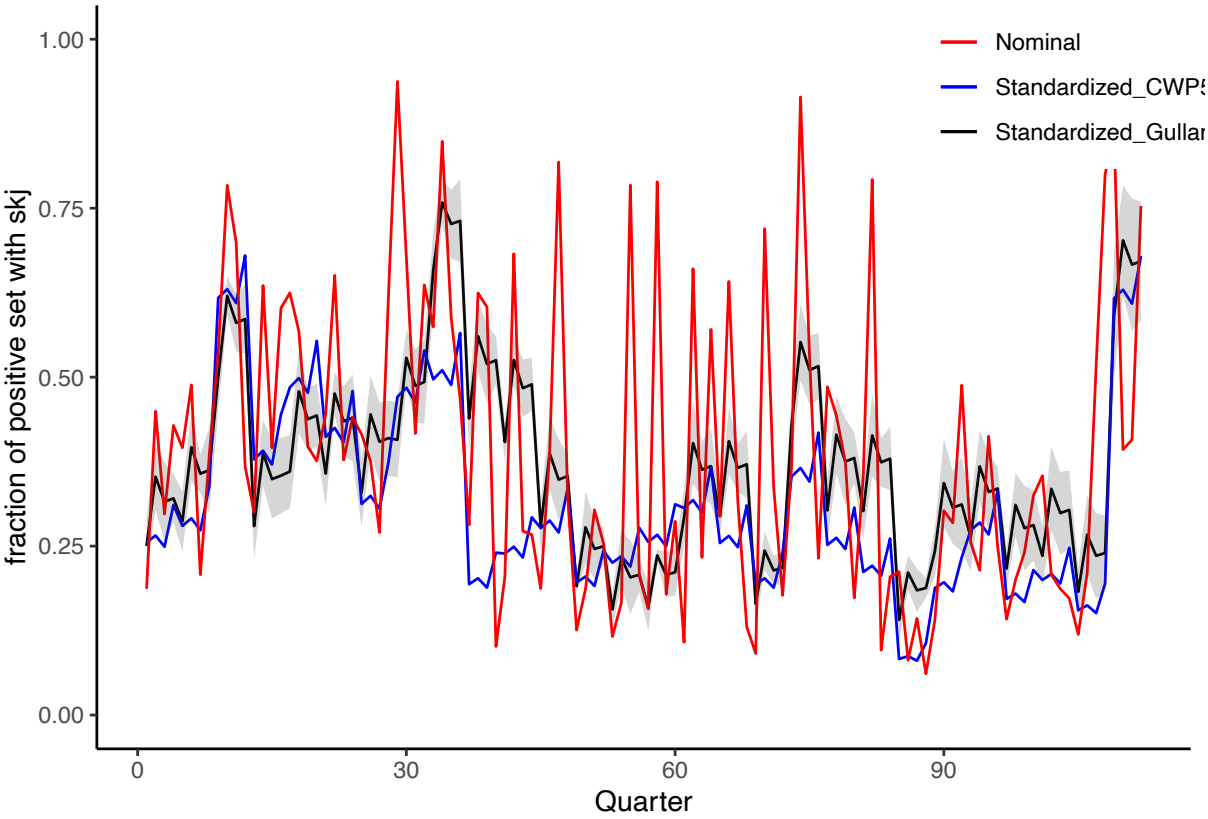

Figure 10: Standardised time series (1991-2018) by quarter of the proportion of successful free school sets with SKJ catch (component 2) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

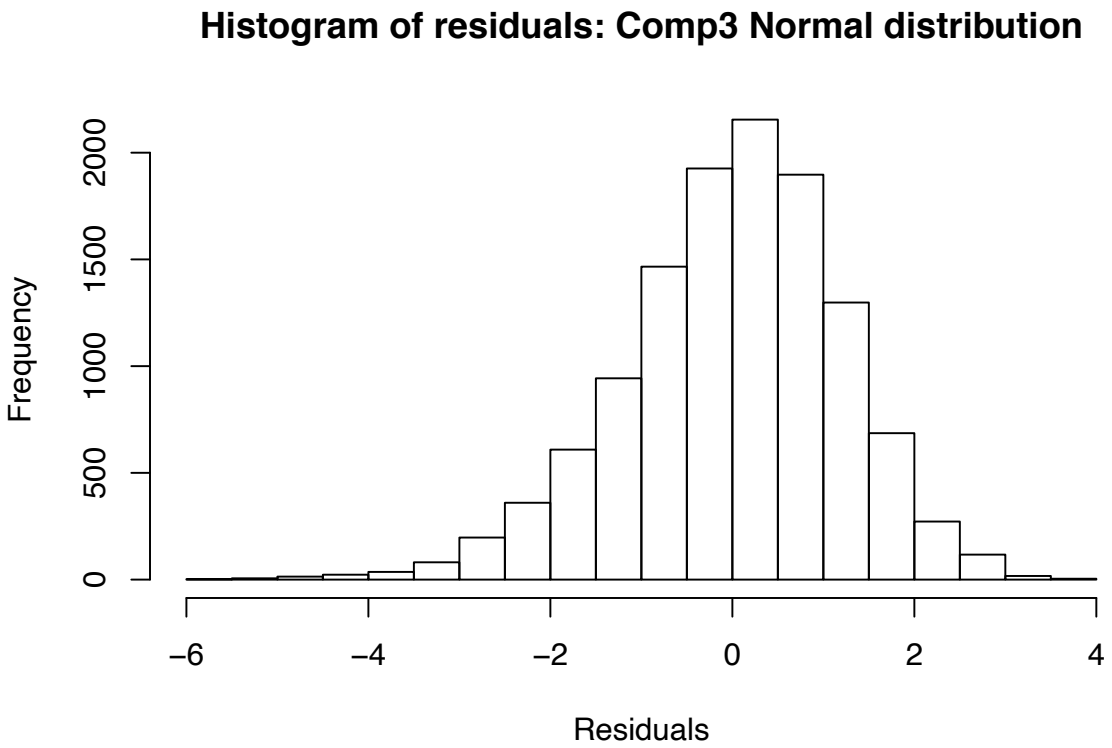

Figure 11: Histogram of residuals for component 3 (catch per positive free school set of SKJ) for the 1991-2018 period. Model including Gulland index.

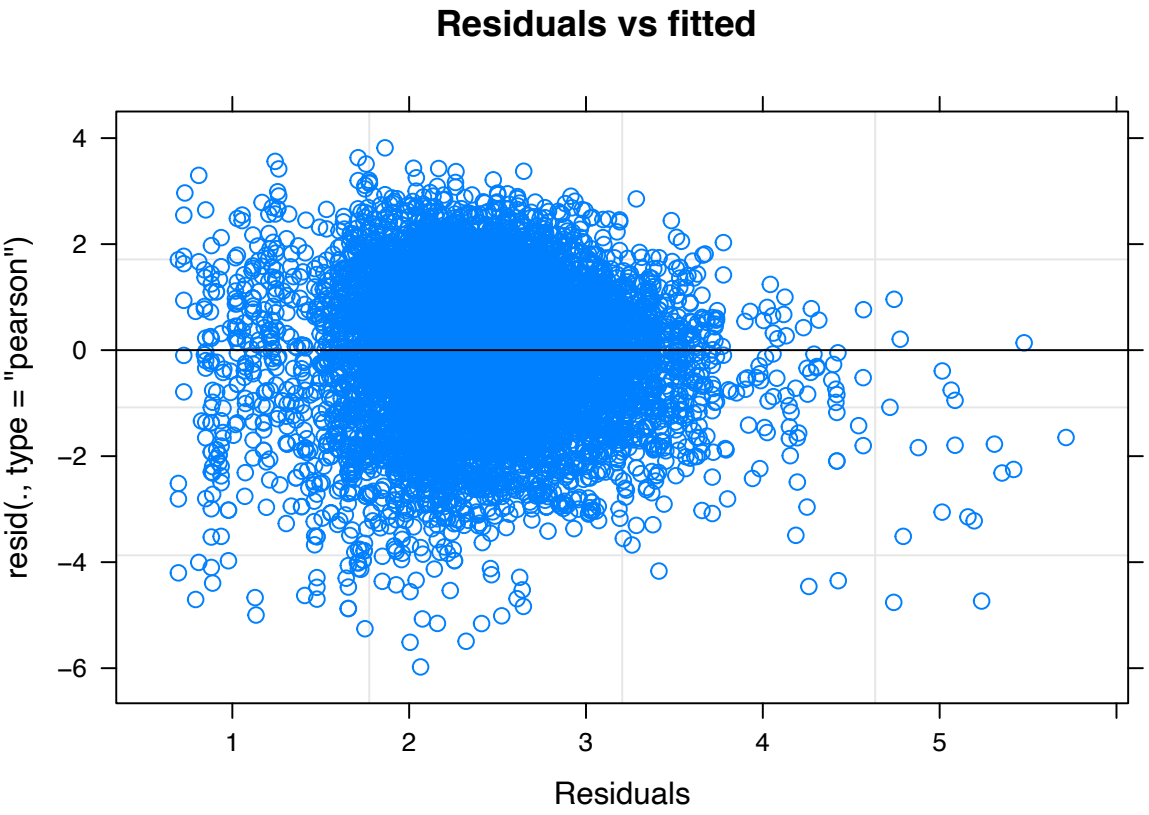

Figure 12: Plot of residuals versus fitted values for component 3 (Catch per positive free school set of SKJ) for the 1991-2018 period. Model including Gulland index.

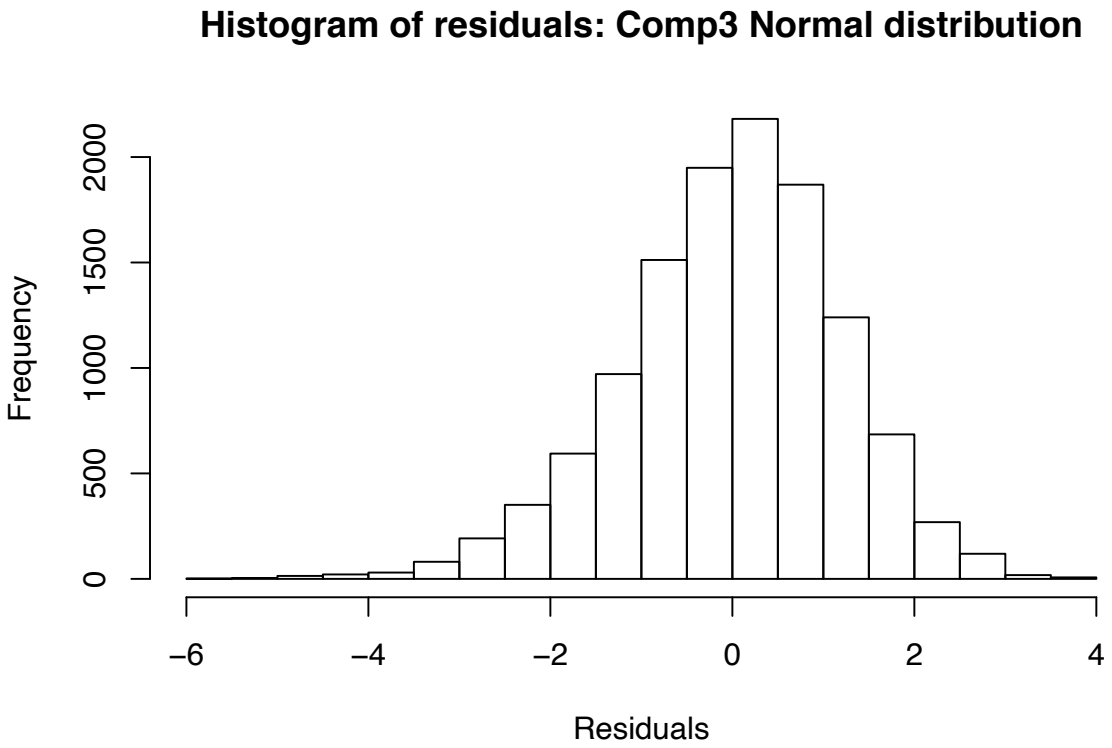

Figure 13: Histogram of residuals for component 3 (Catch per positive free school set of SKJ) for the 1991-2018 period. Model including a spatial variable (cwp55).

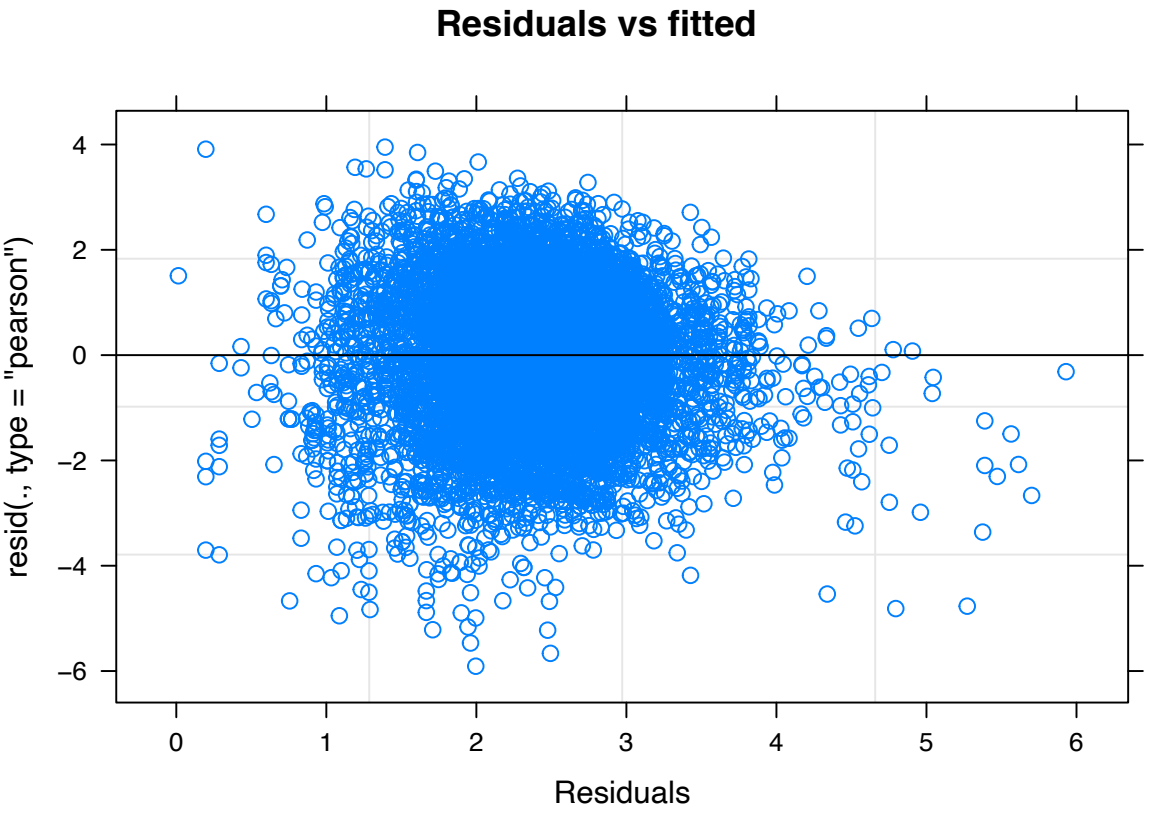

Figure 14: Plot of residuals versus fitted values for component 3 (Catch per positive free school set of SKJ) for the 1991-2018 period. Model including a spatial variable (cwp55).

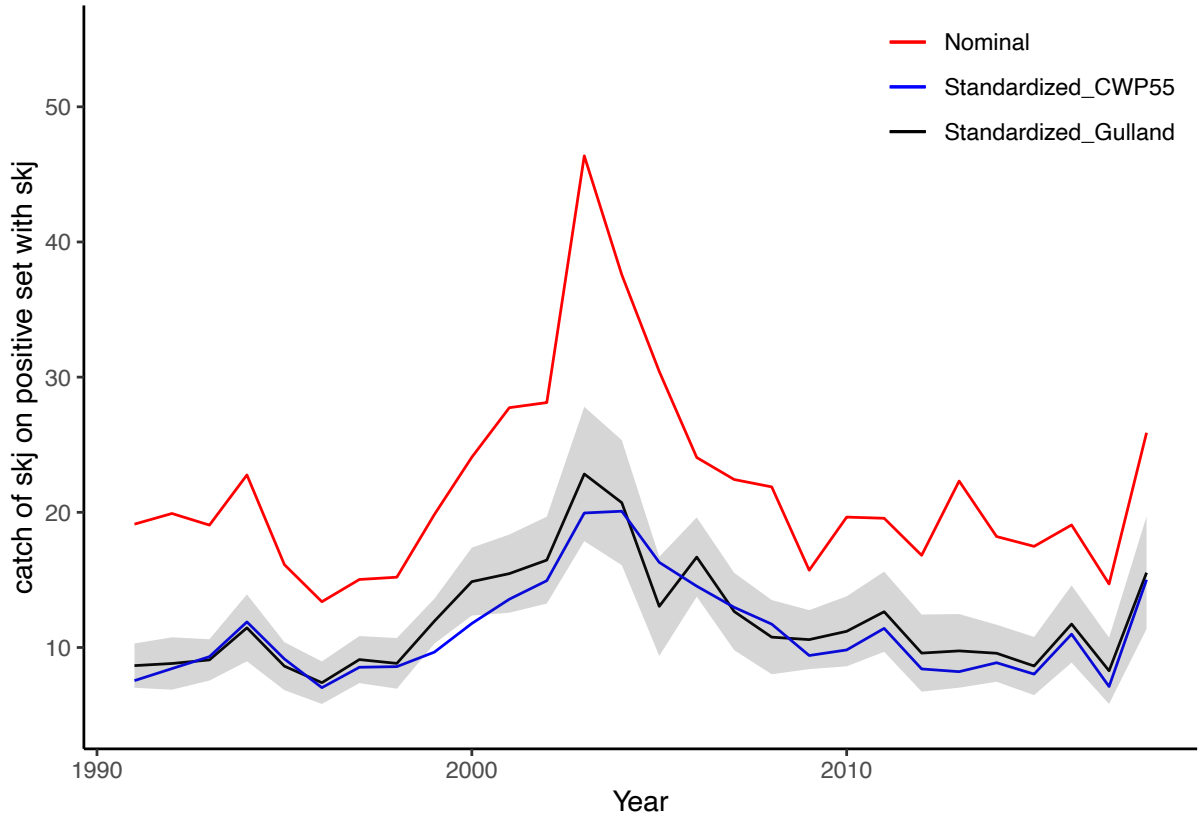

Figure 15: Standardised time series (1991-2018) by year of the catch of skj on positive FSC set (component 3) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

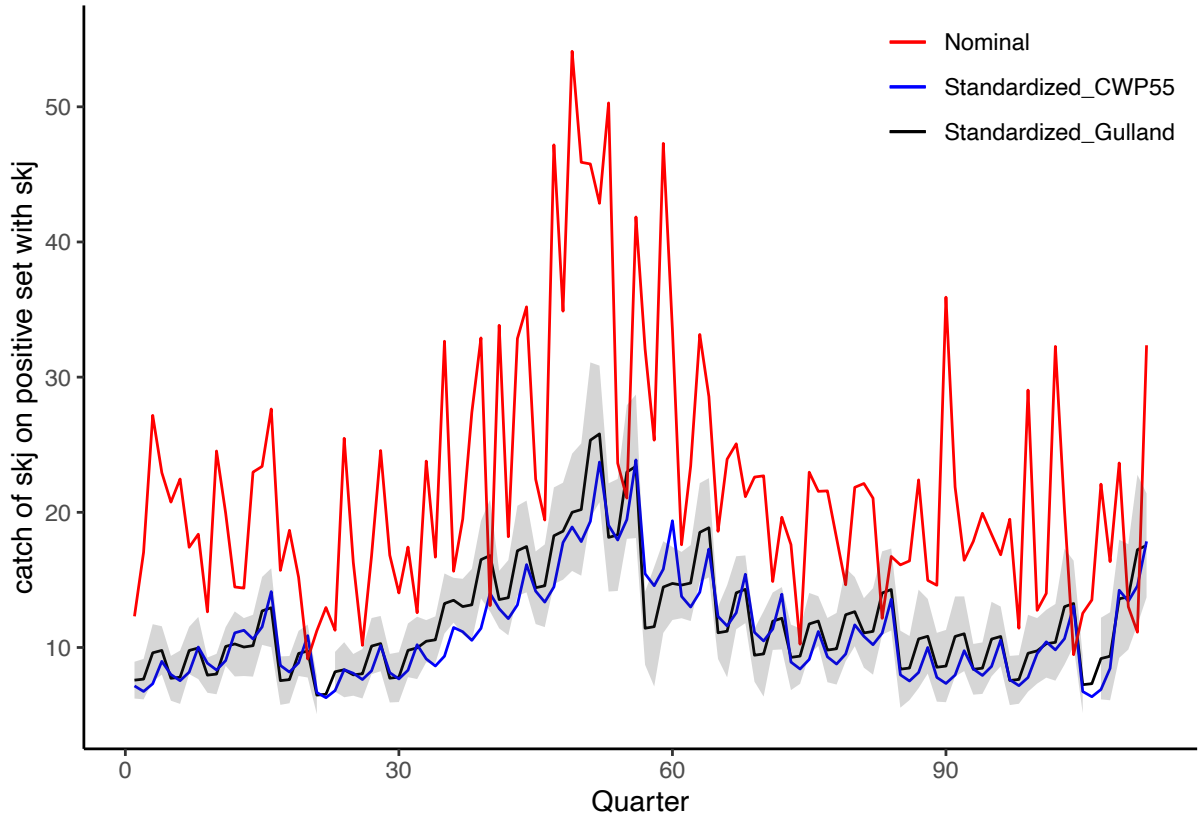

Figure 16: Standardised time series (1991-2018) by quarter of the catch of skj on positive FSC set (component 3) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

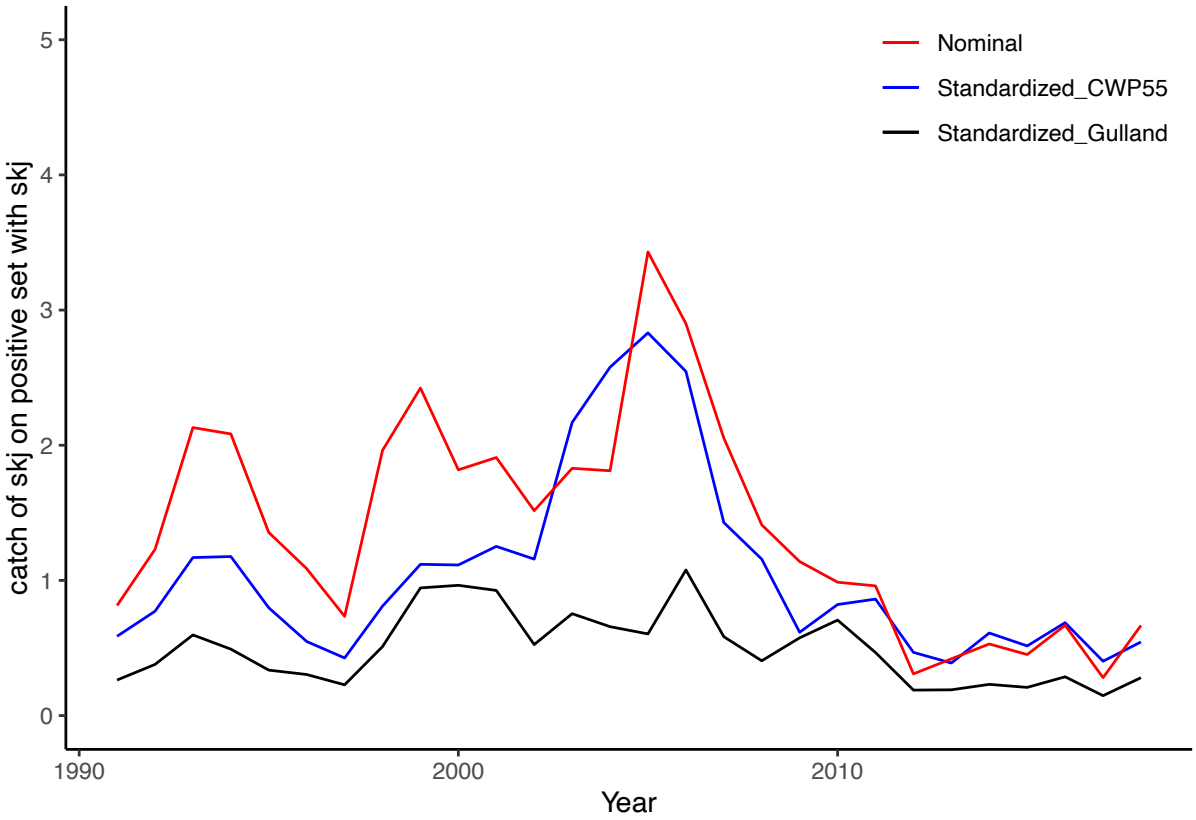

Figure 17: Standardised SKJ CPUE on FSC sets by year for the 1991-2018 period with Gulland Index (black), CWP55 position (blue) compared to nominal (red)

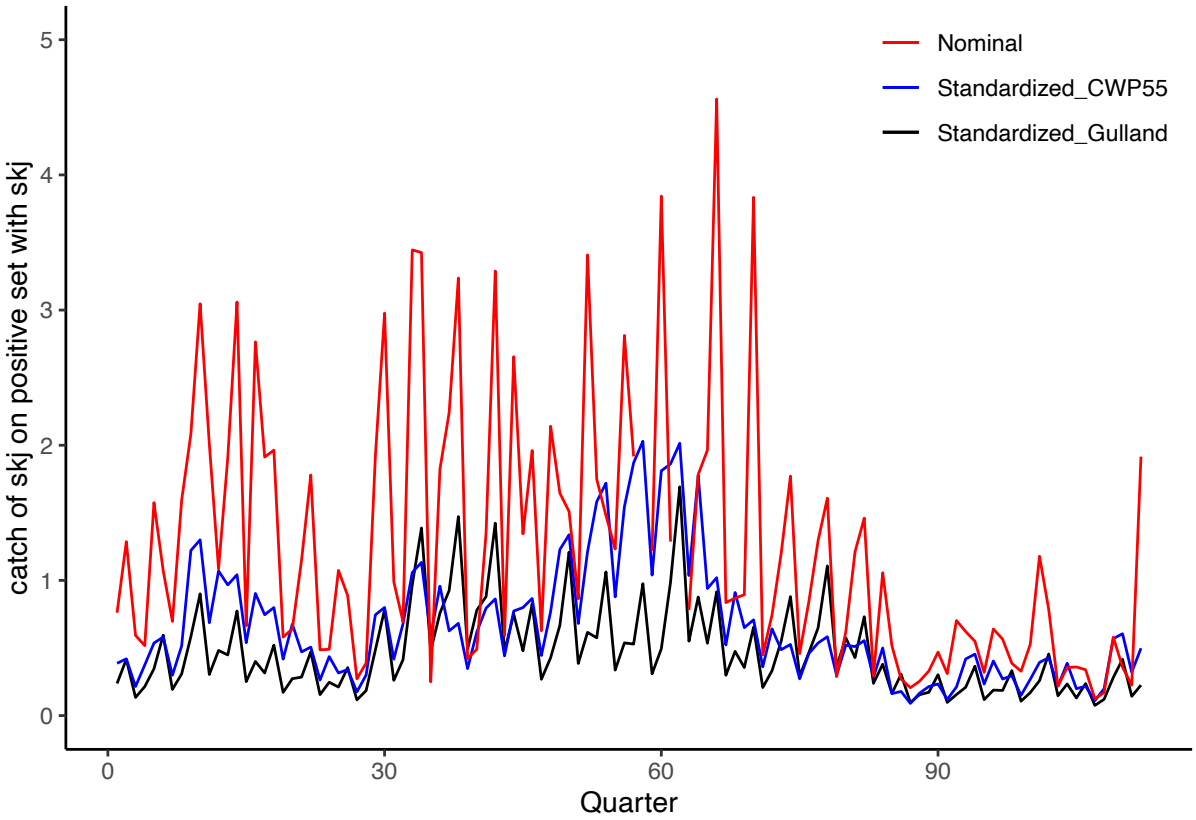

Figure 18: Standardised SKJ CPUE on FSC sets by quarter for the 1991-2018 period with Gulland Index (black), CWP55 position (blue) compared to nominal (red)

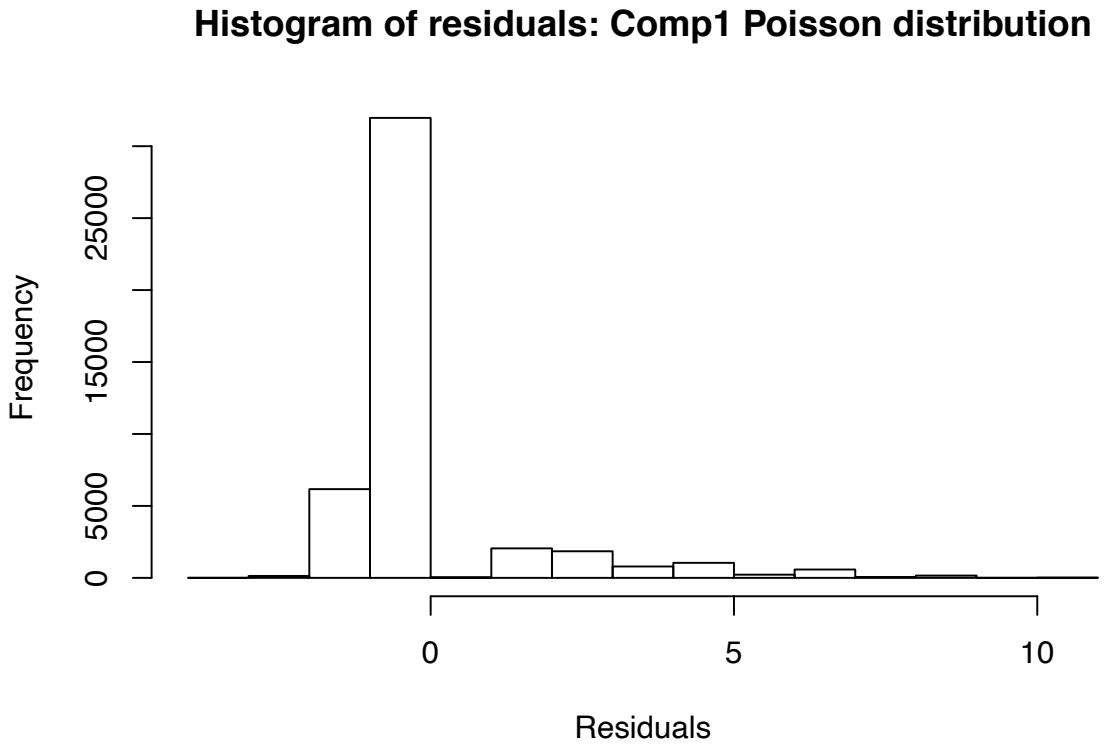

Figure 19: Histogram of residuals for component 1 (total number of free school sets > 0 and $=0$) for the 2010-2018 period. Model including Gulland index.

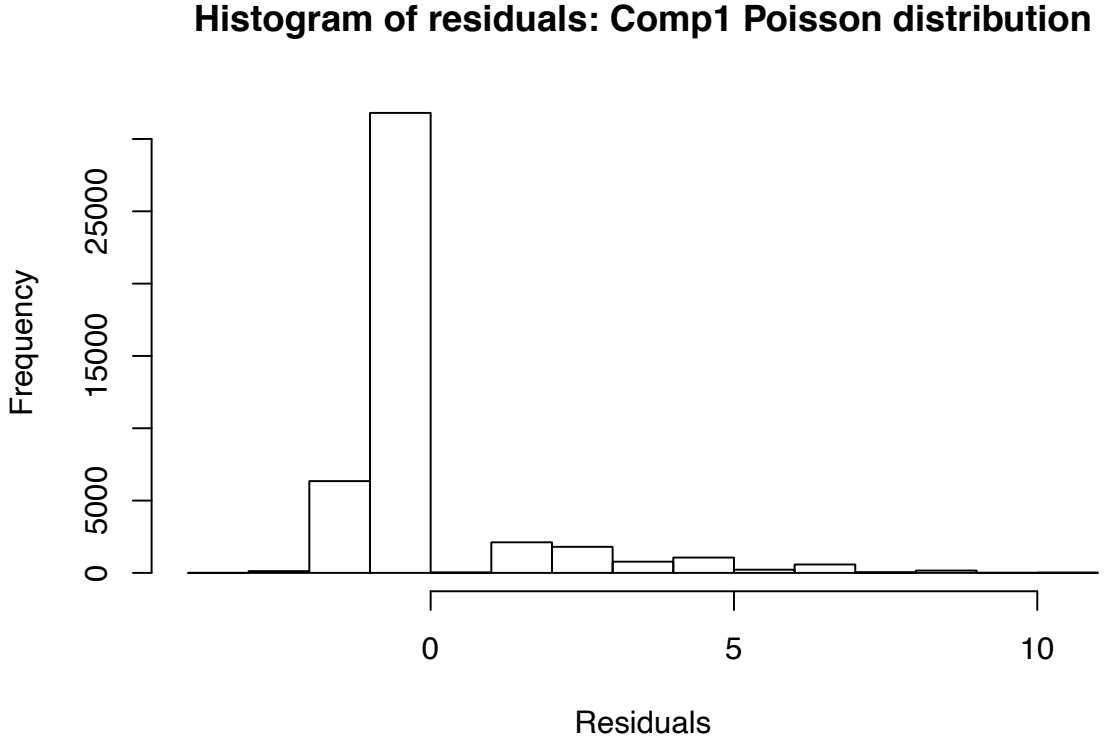

Figure 20: Histogram of residuals for component 1 (total number of free school sets > 0 and $=0$) for the 2010-2018 period. Model including a spatial variable (cwp55).

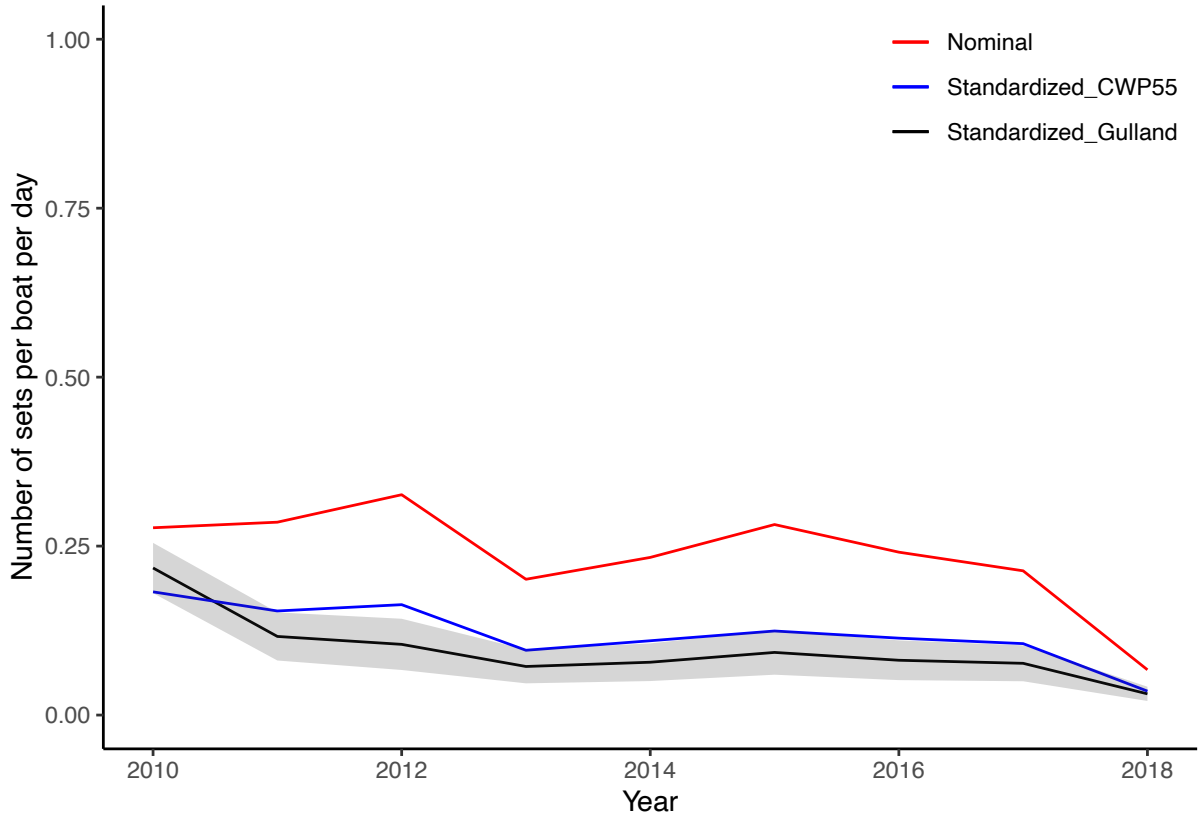

Figure 21: Standardised time series (2010-2018) by year of the number of free school sets > 0 and $=0$ (component 1) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

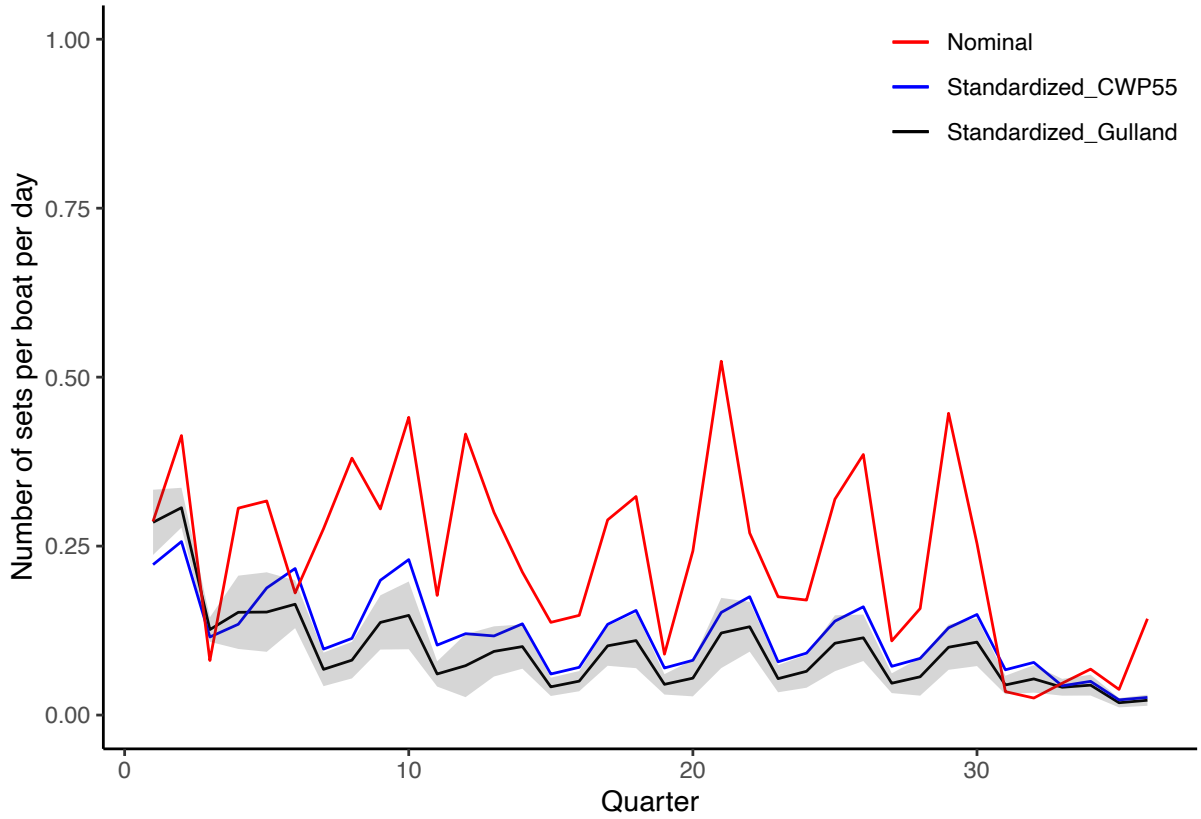

Figure 22: Standardised time series (2010-2018) by quarter of the number of free school sets > 0 and $=0$ (component 1) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

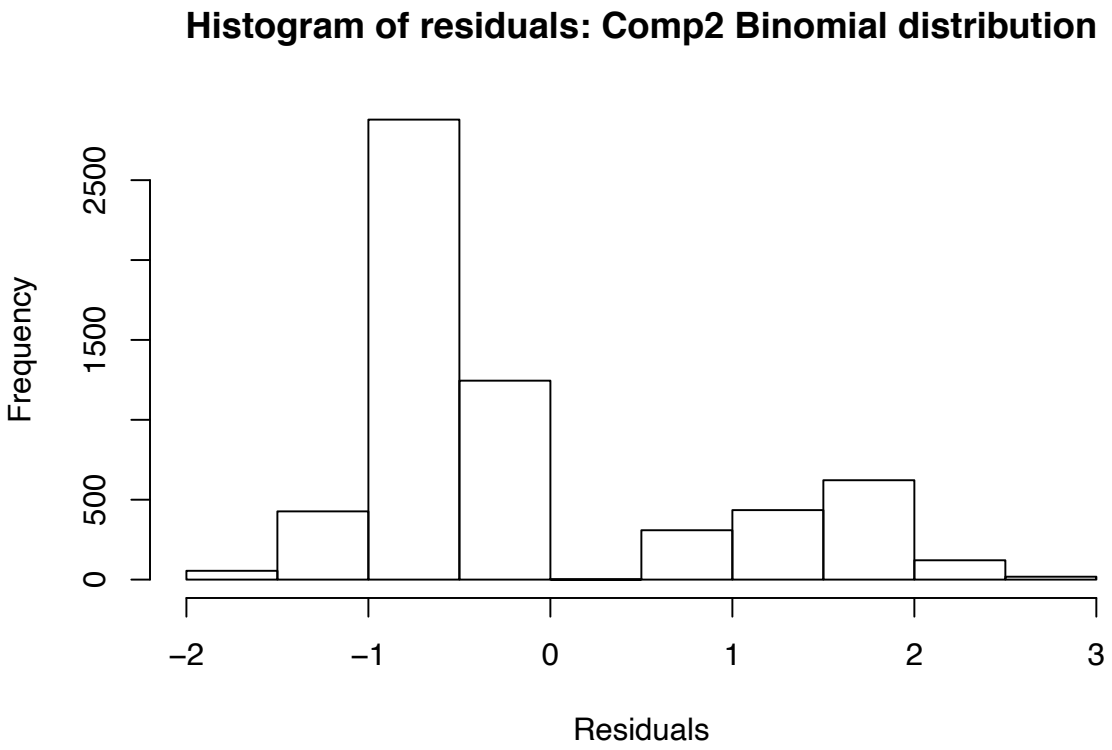

Figure 23: Histogram of residuals for component 2 (proportion of positive free school sets with SKJ catch) for the 2010-2018 period. Model including Gulland index.

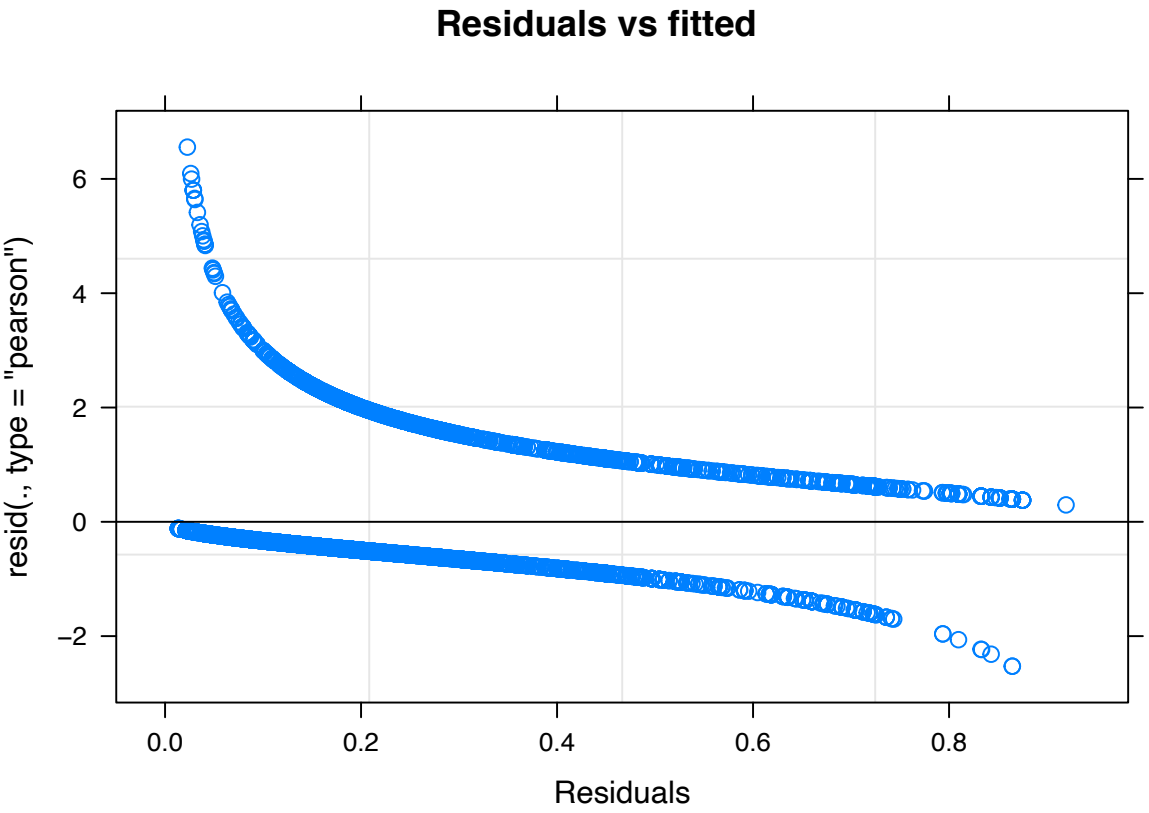

Figure 24: Plot of residuals versus fitted values for component 2 (proportion of positive free school sets with SKJ catch) for the 2010-2018 period. Model including Gulland index.

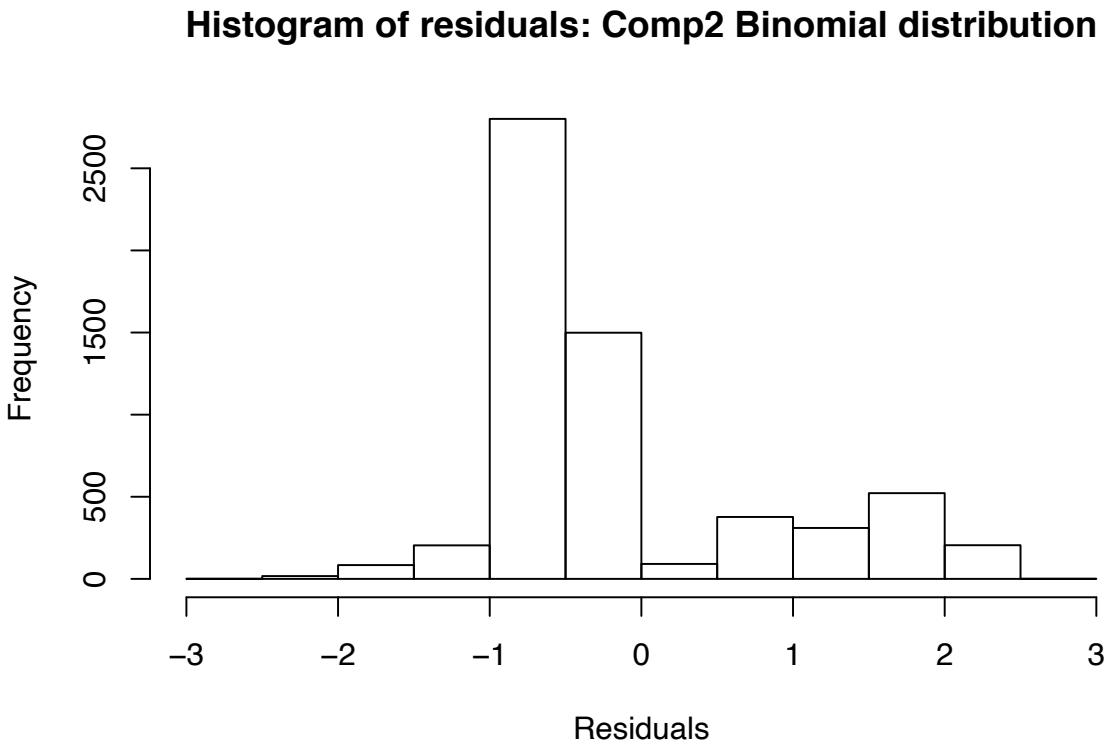

Figure 25: Histogram of residuals for component 2 (proportion of positive free school sets with SKJ catch) for the 2010-2018 period. Model including a spatial variable (cwp55).

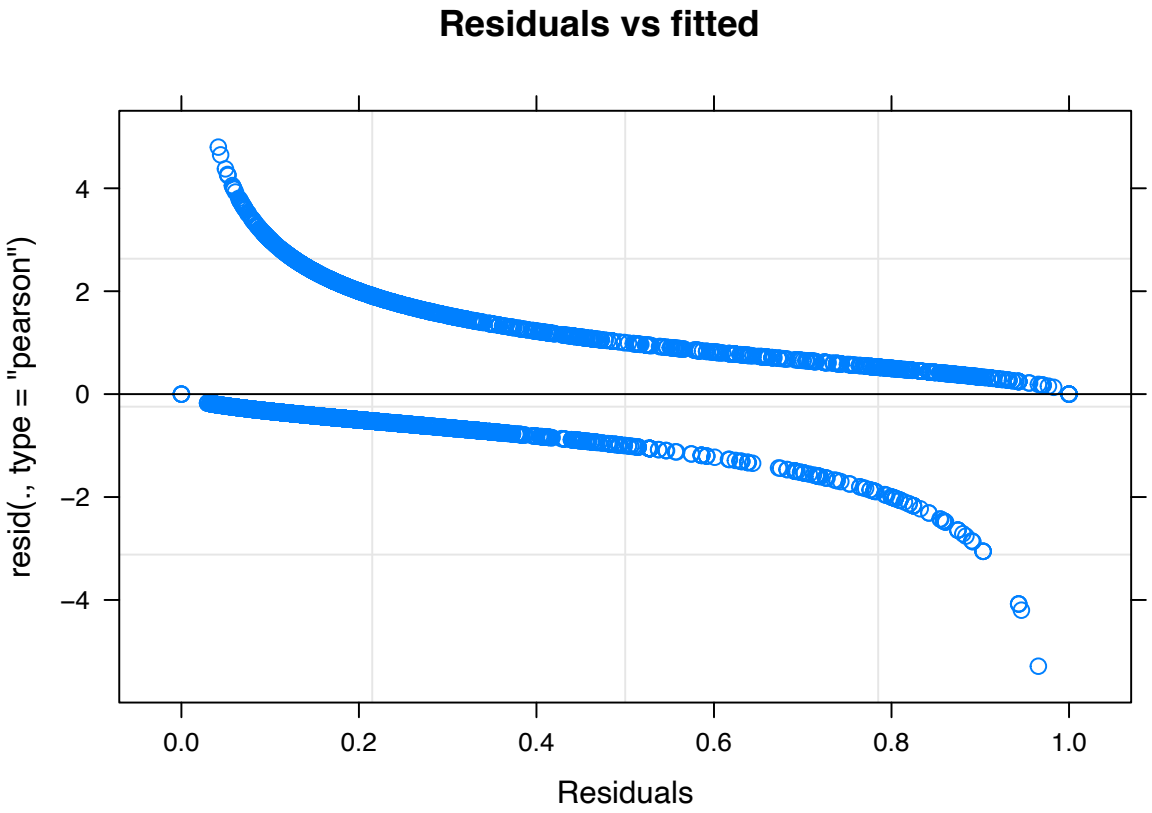

Figure 26: Plot of residuals versus fitted values for component 2 (proportion of positive free school sets with SKJ catch) for the 2010-2018 period. Model including a spatial variable (cwp55).

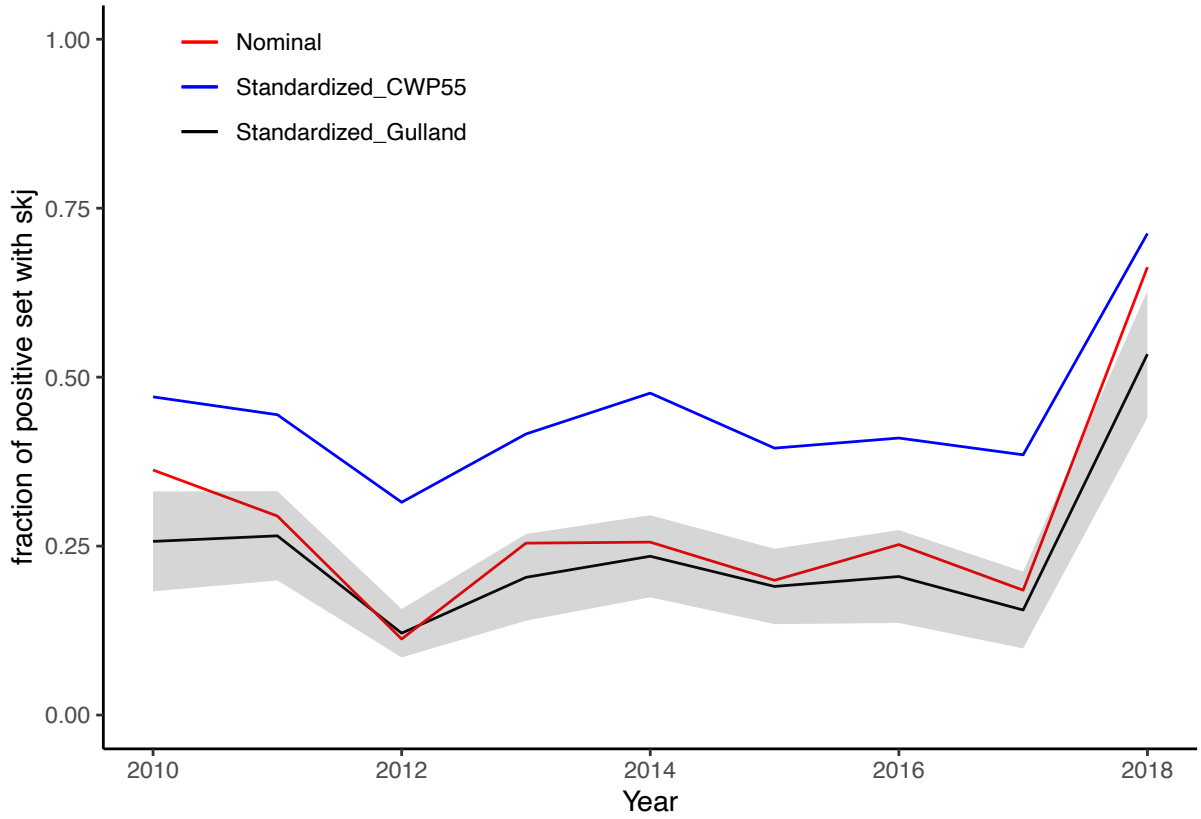

Figure 27: Standardised time series (2010-2018) by year of the proportion of successful free school sets with SKJ catch (component 2) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

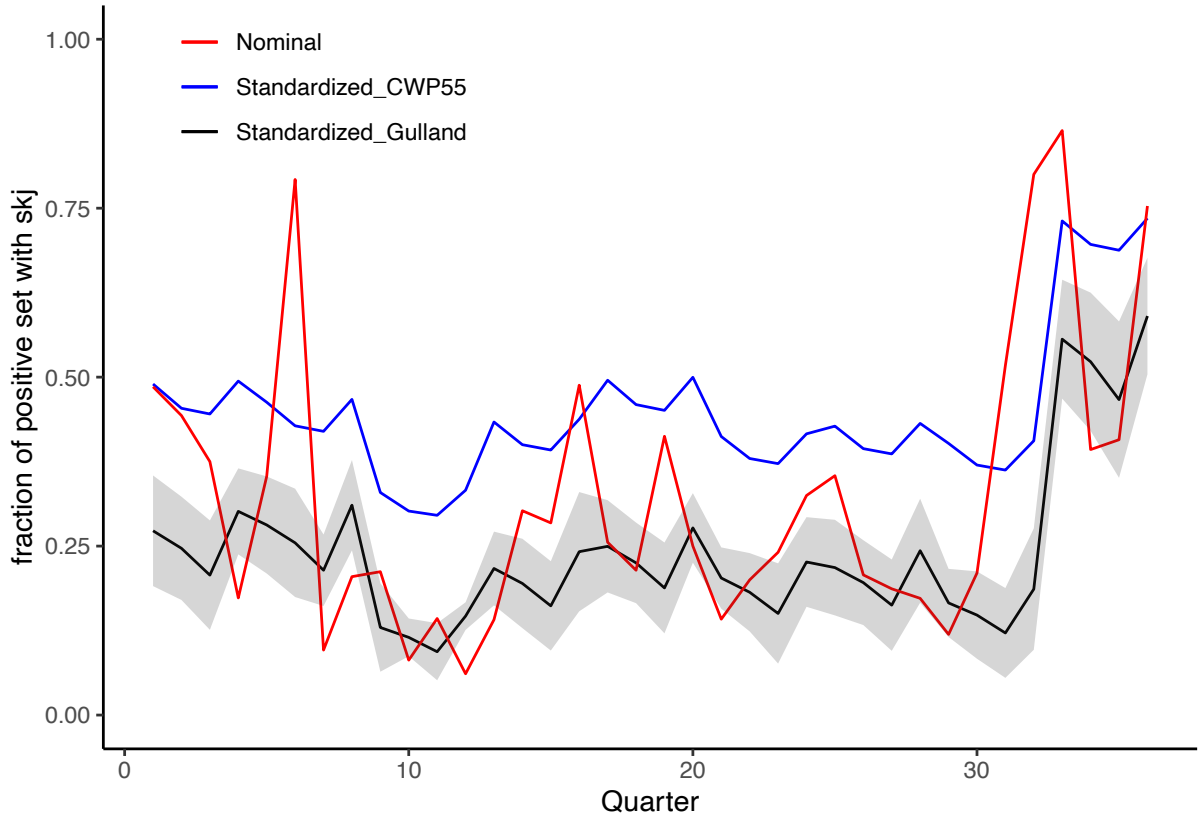

Figure 28: Standardised time series (1991-2018) by quarter of the proportion of successful free school sets with SKJ catch (component 2) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

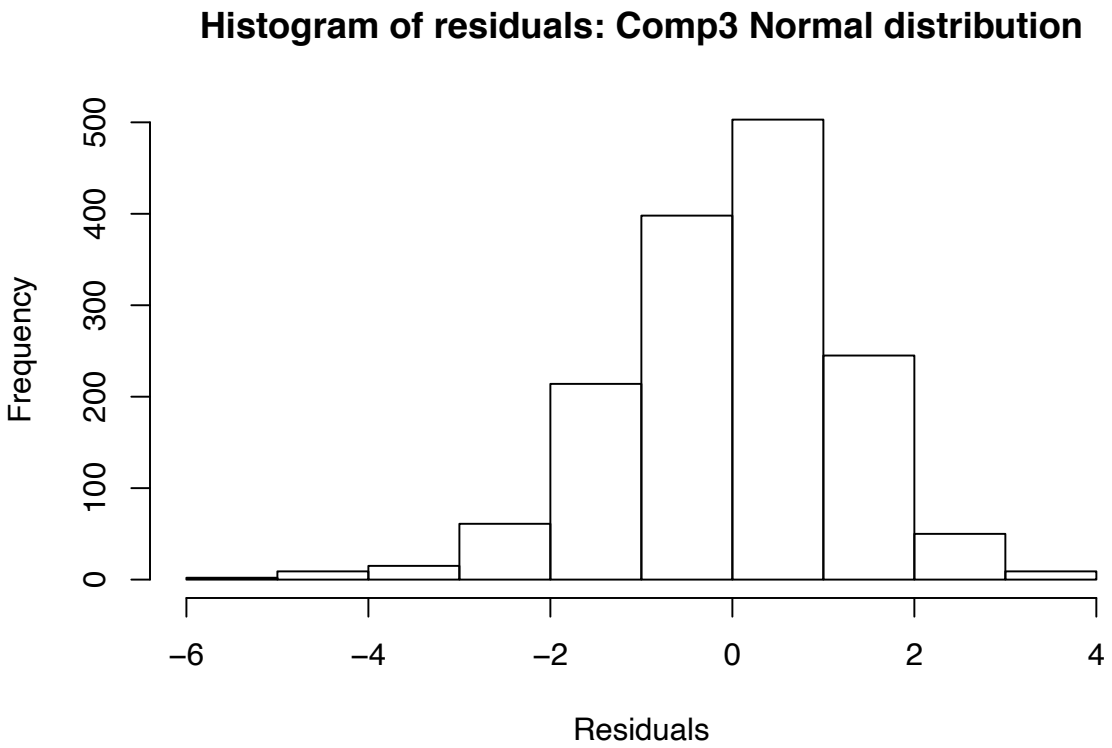

Figure 29: Histogram of residuals for component 3 (Catch per positive free school set of SKJ) for the 2010-2018 period. Model including Gulland index.

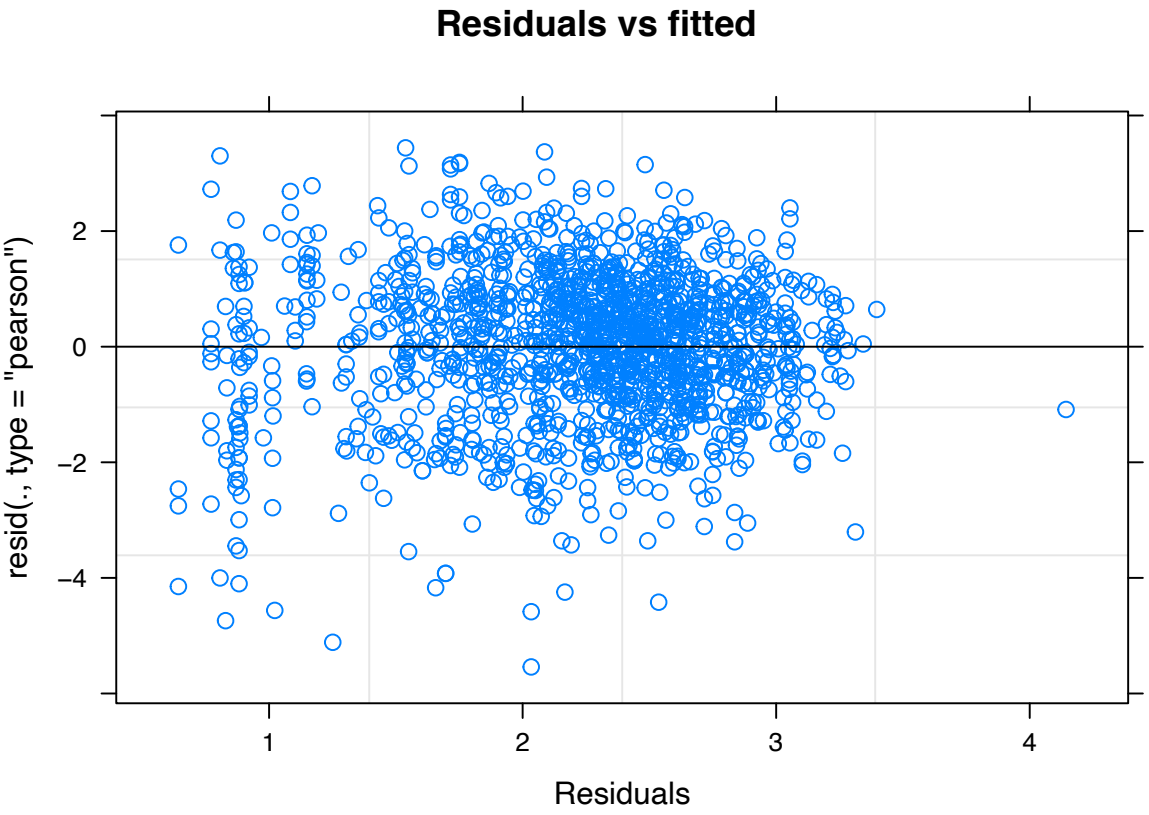

Figure 30: Plot of residuals versus fitted values for component 3 (Catch per positive free school set of SKJ) for the 2010-2018 period. Model including Gulland index.

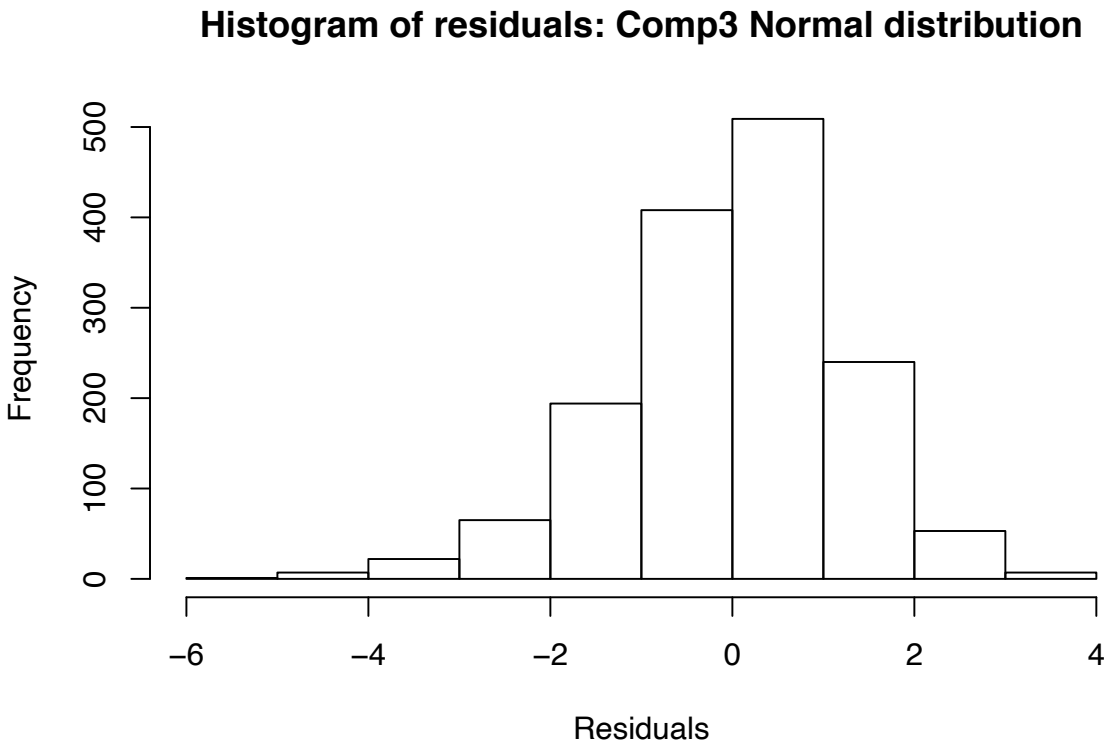

Figure 31: Histogram of residuals for component 3 (Catch per positive free school set of SKJ) for the 2010-2018 period. Model including a spatial variable (cwp55).

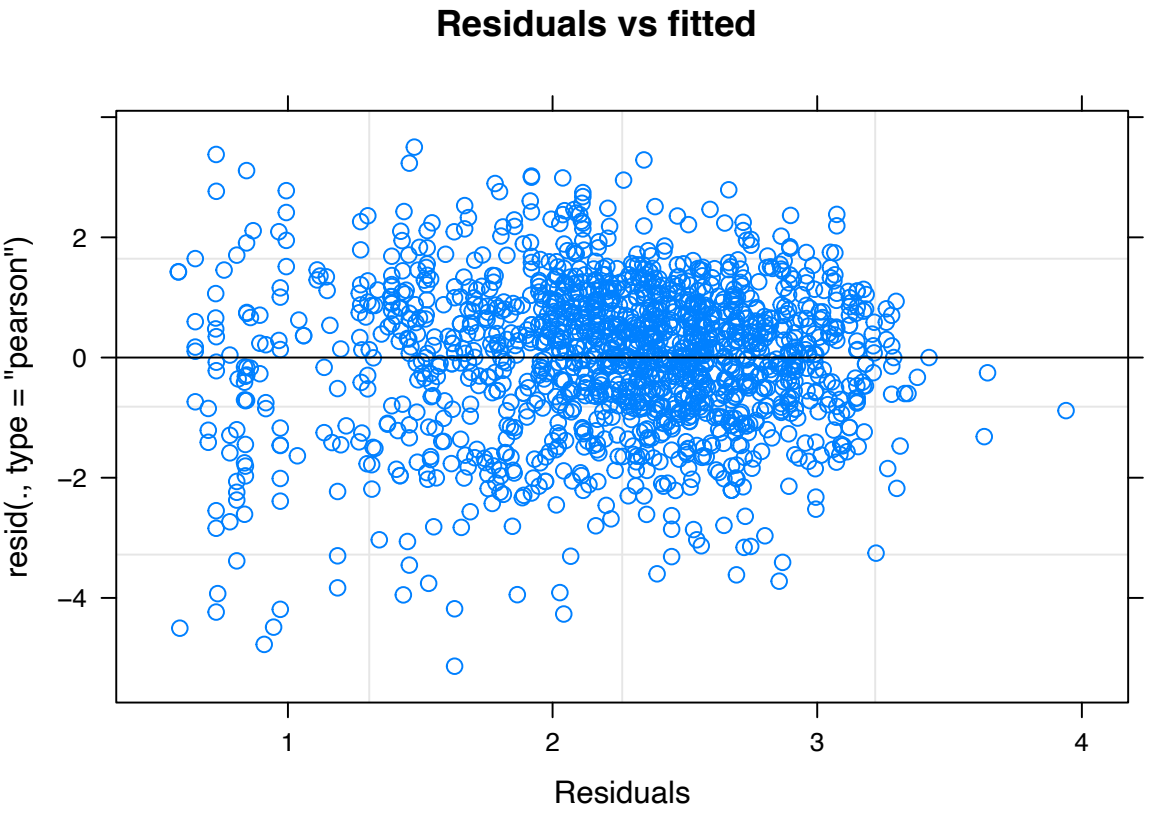

Figure 32: Plot of residuals versus fitted values for component 3 (Catch per positive free school set of SKJ) for the 2010-2018 period. Model including a spatial variable (cwp55)

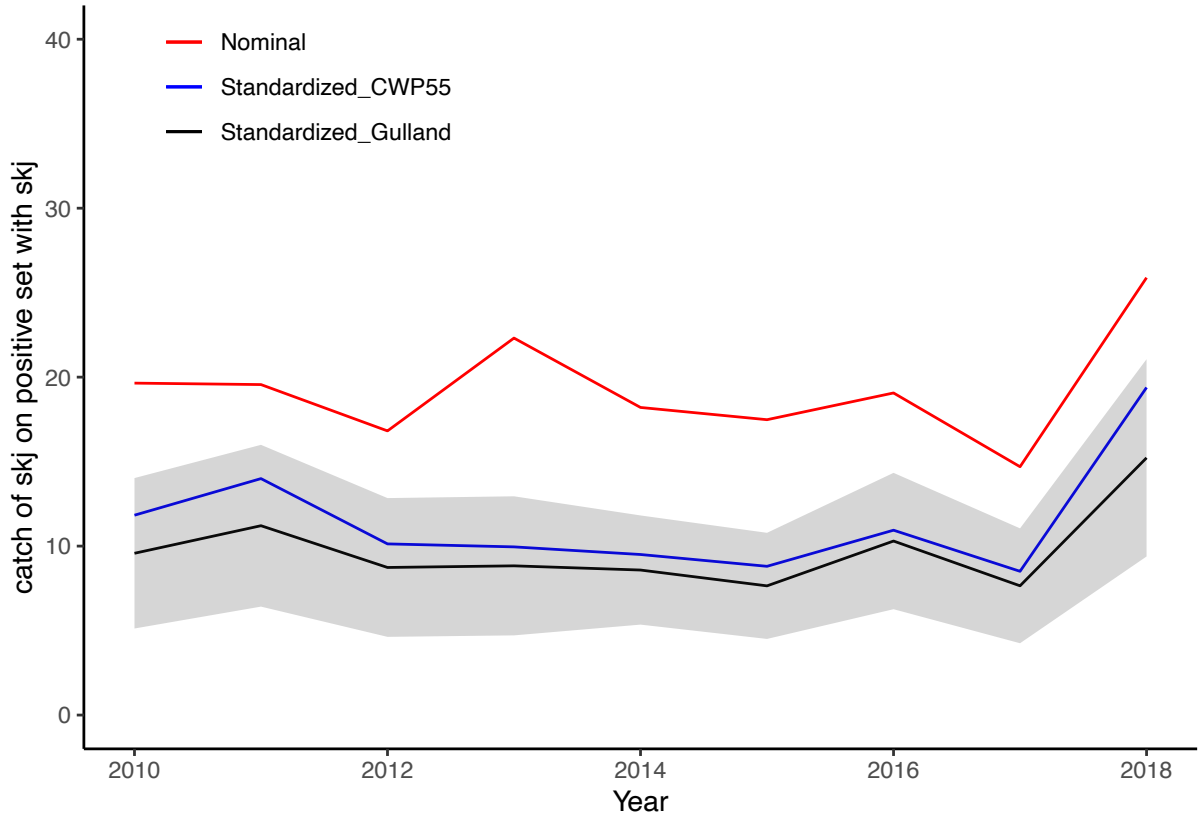

Figure 33: Standardised time series (2010-2018) by year of the catch of skj on positive FSC set (component 3) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

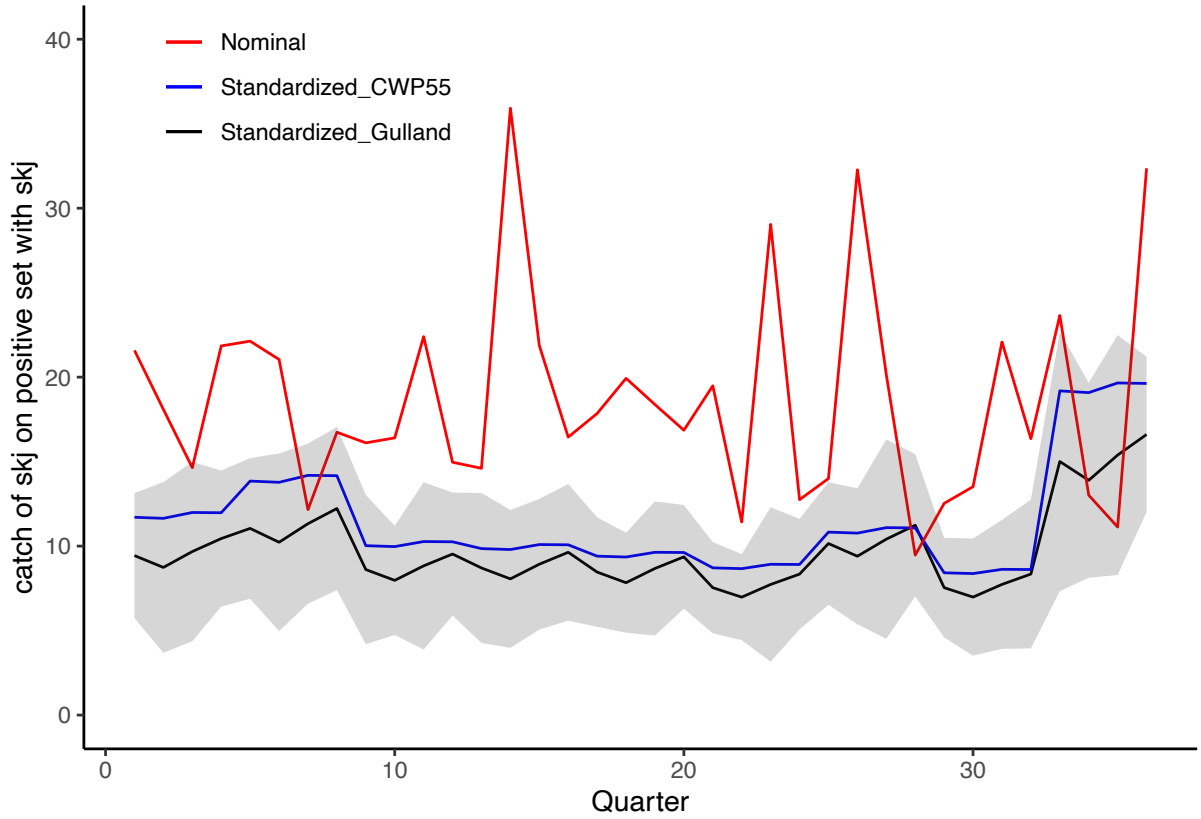

Figure 34: Standardised time series (2010-2018) by quarter of the catch of skj on positive FSC set (component 3) with model including either Gulland Index (black) or CWP55 position (blue) and a 97.5% confidence intervals (grey) to be compared to the nominal (red)

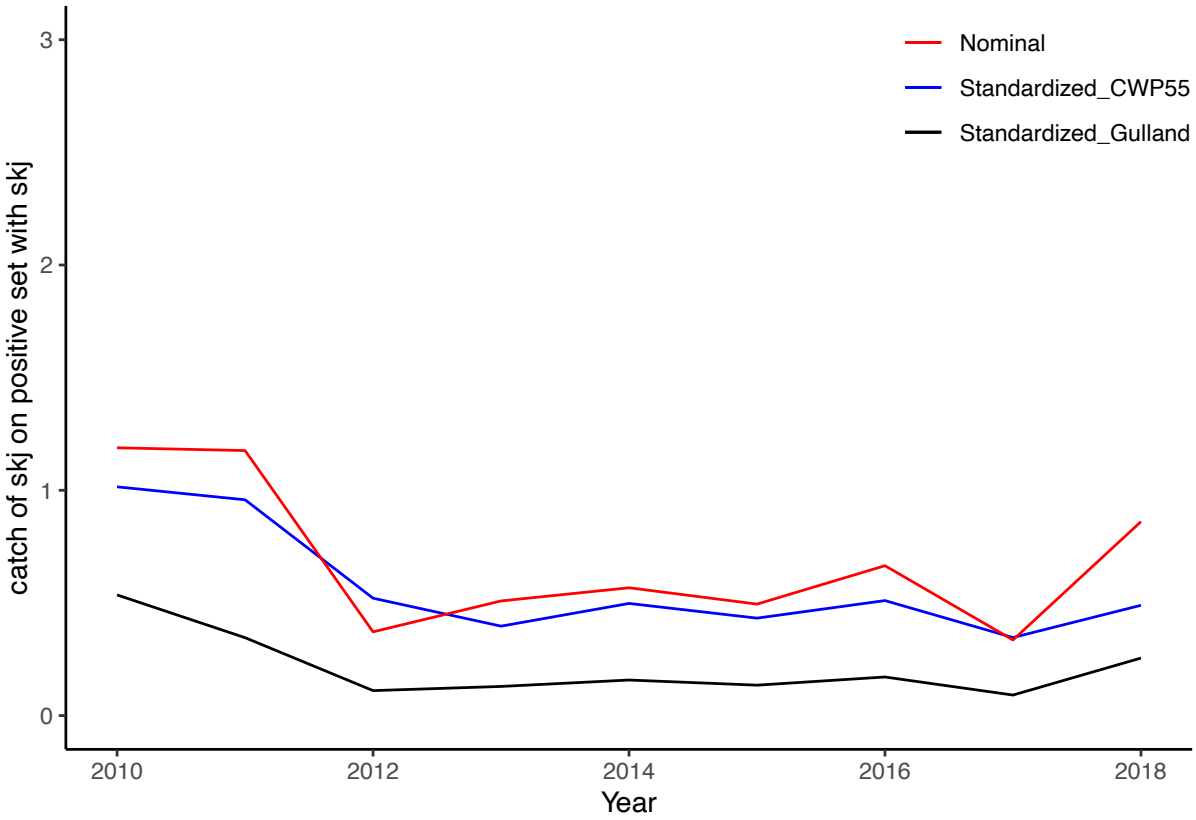

Figure 35: Standardised SKJ CPUE on FSC sets by year for the 2010-2018 period with Gulland Index (black), CWP55 position (blue) compared to nominal (red)

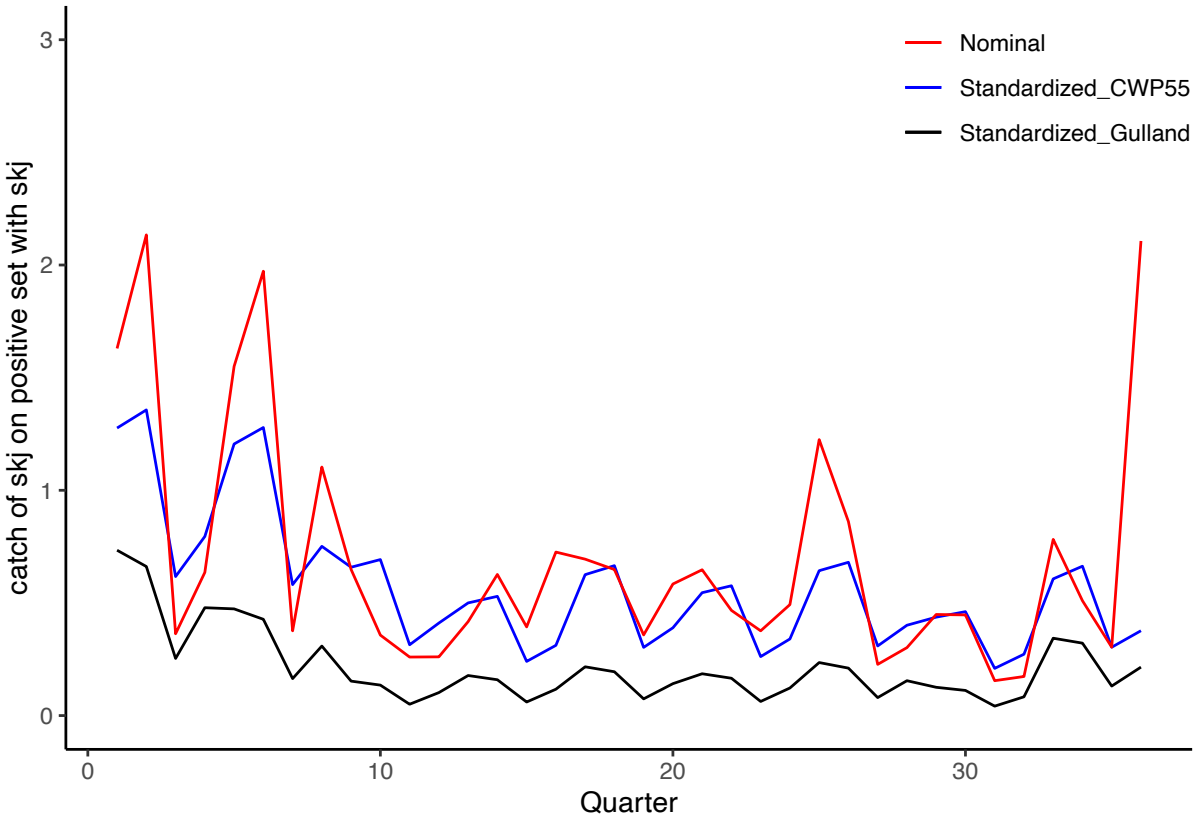

Figure 36: Standardised SKJ CPUE on FSC sets by quarter for the 2010-2018 period with Gulland Index (black), CWP55 position (blue) compared to nominal (red)

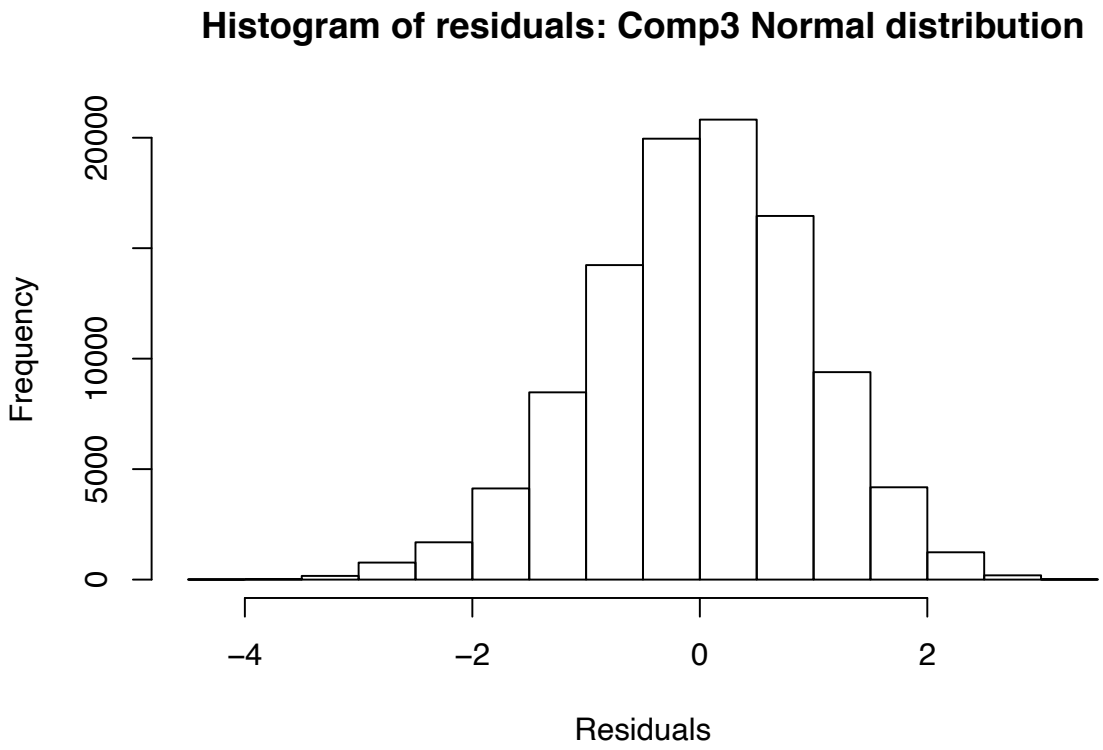

Figure 37: Histogram of residuals for component 3 (Catch per positive FAD set of SKJ) for the 1991-2018 period. Model including Gulland index.

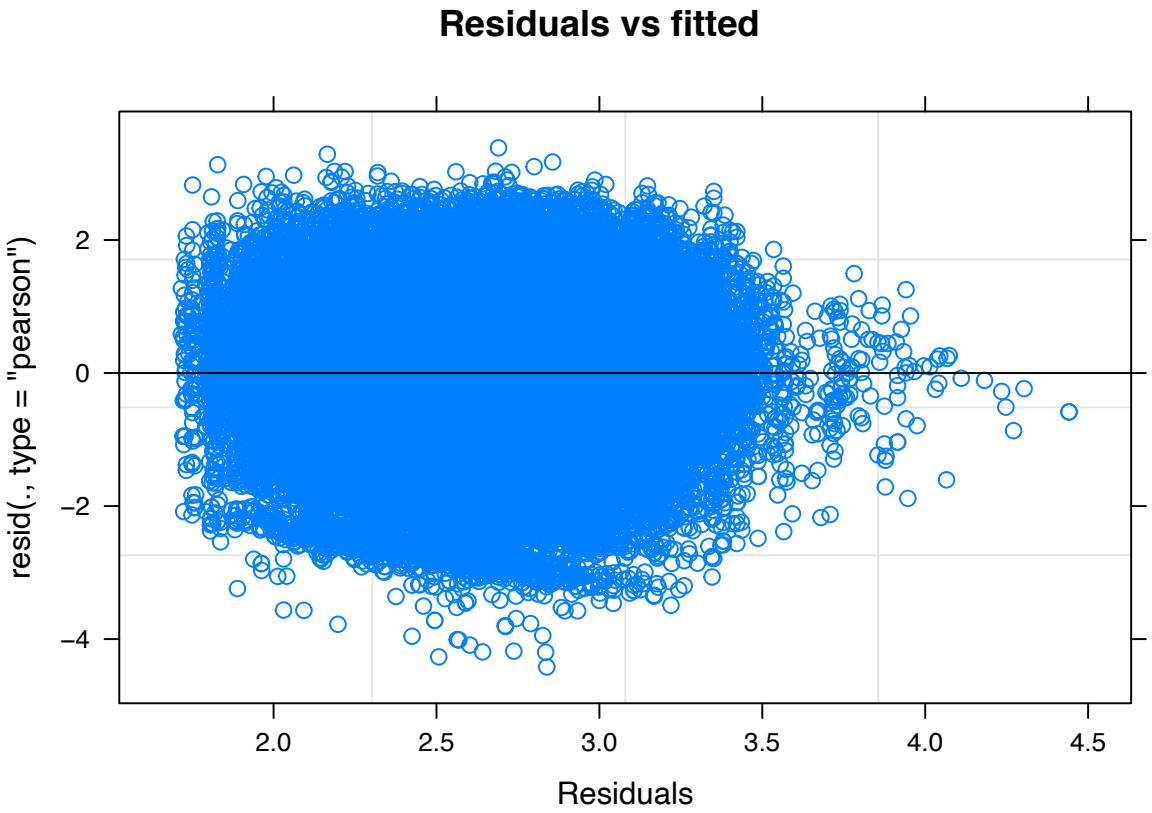

Figure 38: Plot of residuals versus fitted values for component 3 (Catch per positive FAD set of SKJ) for the 1991-2018 period. Model including Gulland index.

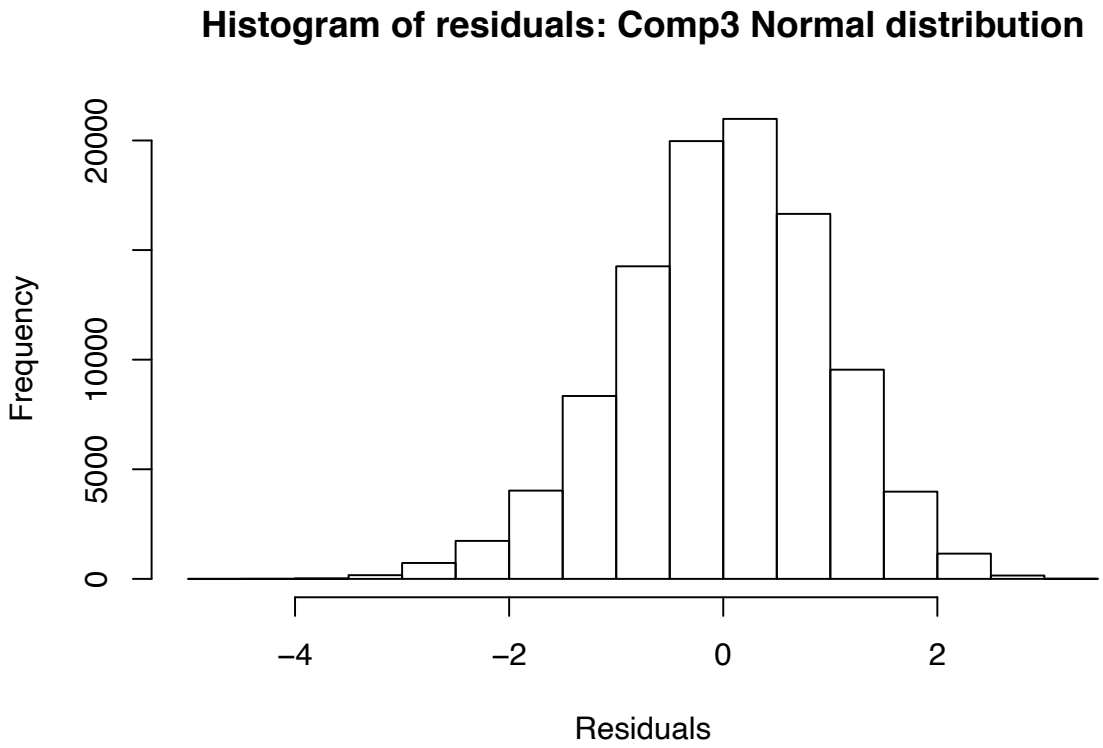

Figure 39: Histogram of residuals for component 3 (Catch per positive FAD set of SKJ) for the 1991-2018 period. Model including a spatial variable (cwp55).

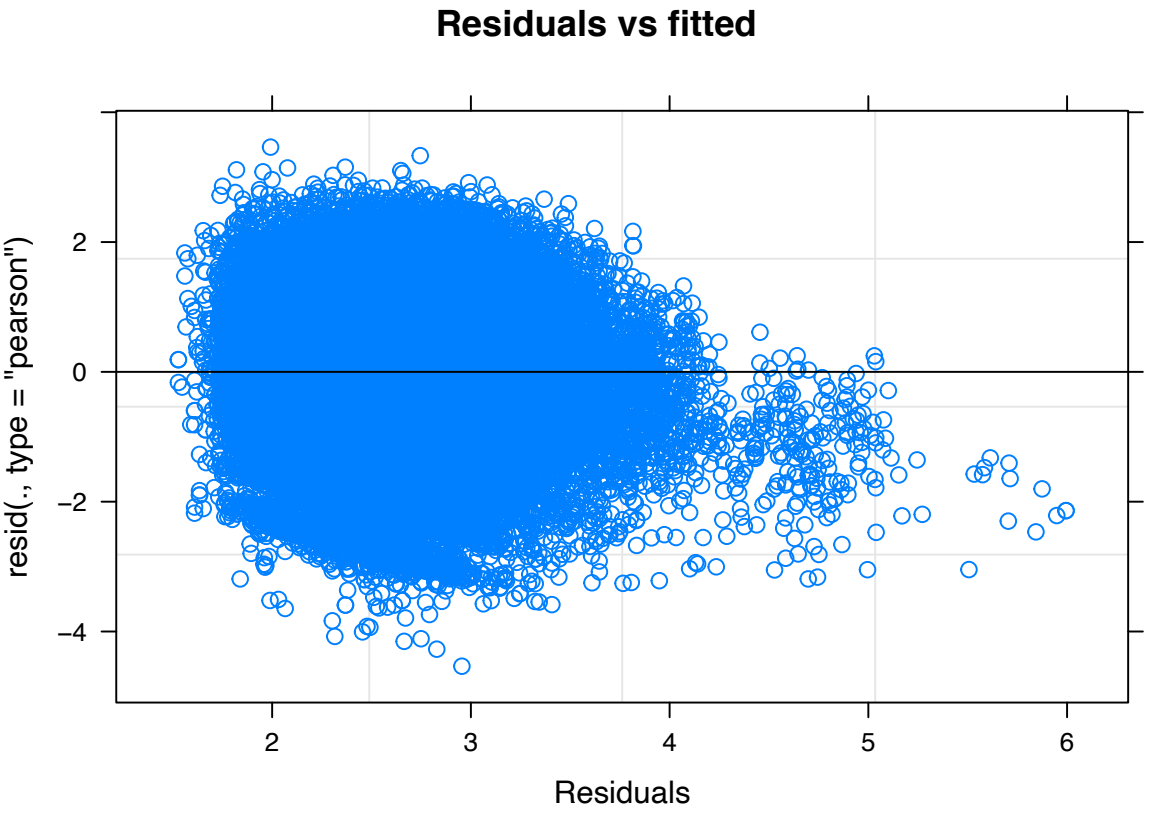

Figure 40: Plot of residuals versus fitted values for component 3 (Catch per positive FAD set of SKJ) for the 1991-2018 period. Model including a spatial variable (cwp55).

```
## Warning: Removed 5 row(s) containing missing values (geom_path).
```

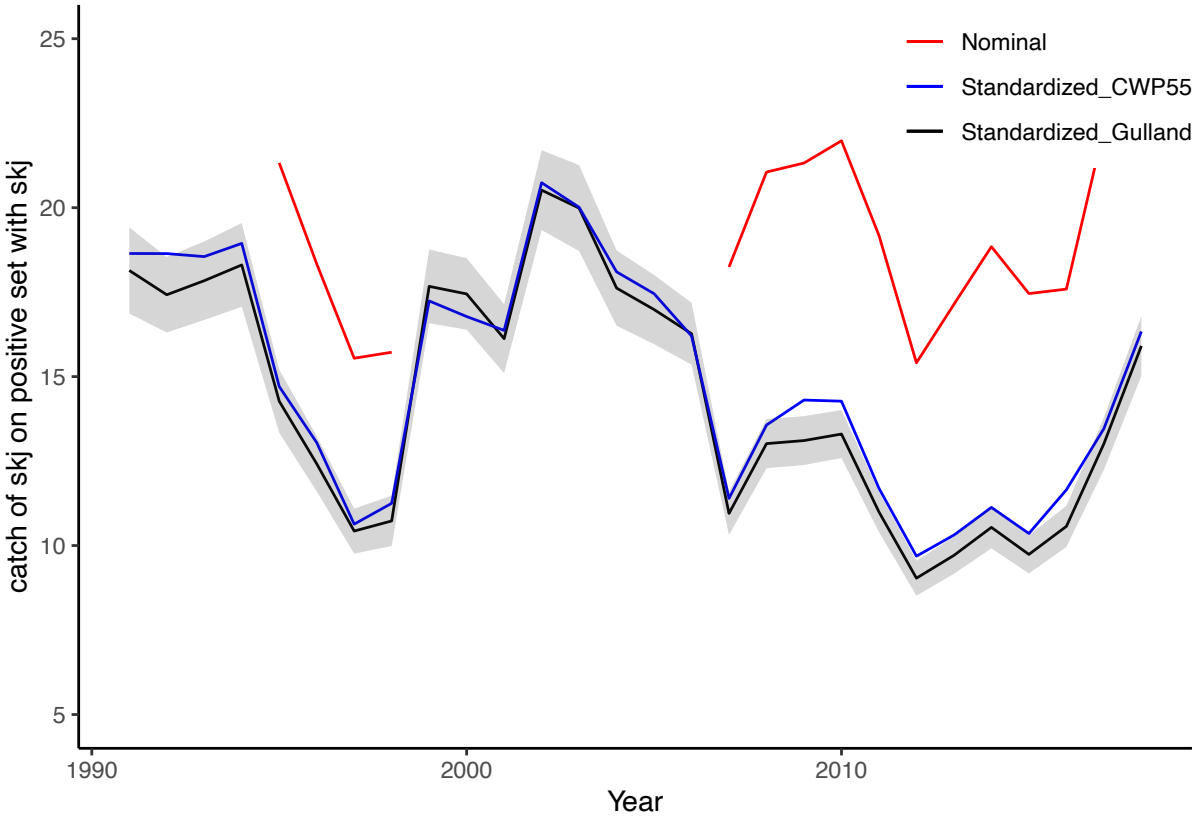

Figure 42: FAD sets - 1991-2018 - catch of skj on positive set with skj; standardised time series by year with Gulland Index (black), CWP55 position (blue) with 97.5% confidence intervals (grey) compared to nominal (red)

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```

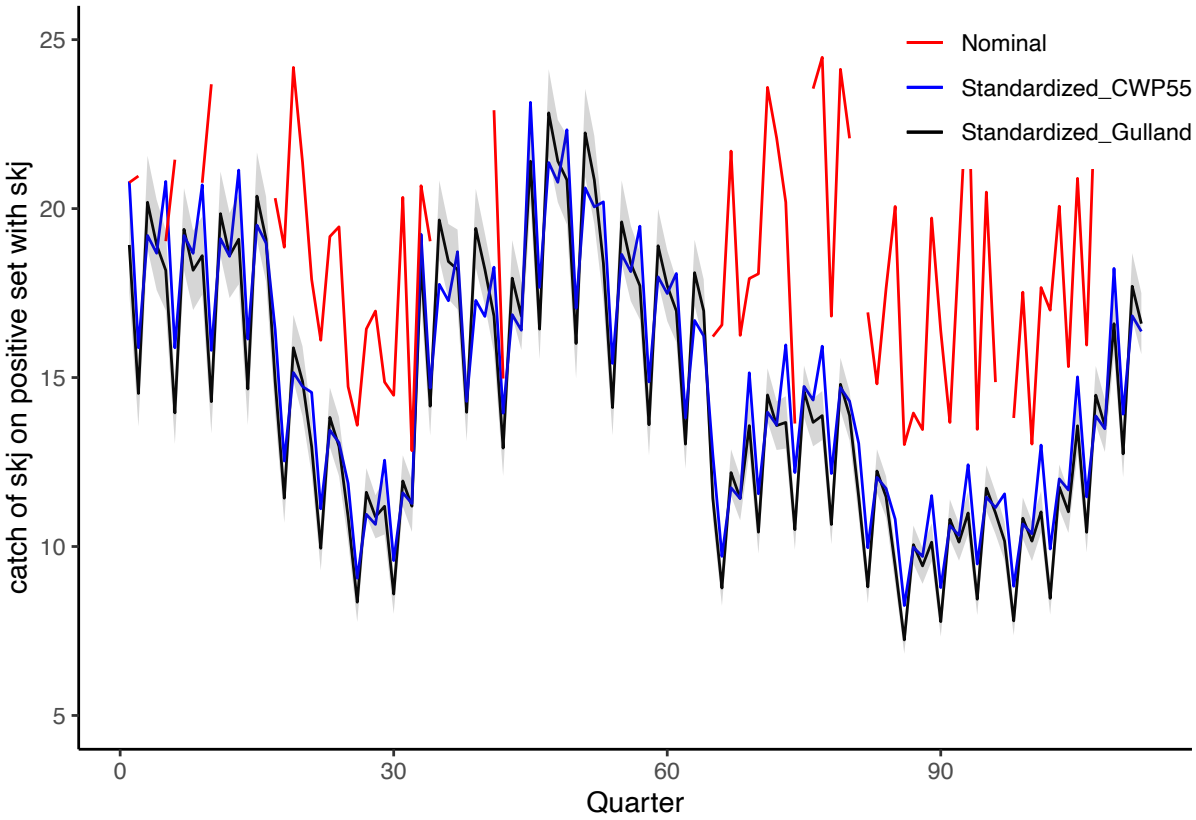

Figure 43: FAD sets - 1991-2018 - catch of skj on positive set with skj; standardised time series by quarter with Gulland Index (black), CWP55 position (blue) with 97.5% confidence intervals (grey) compared to nominal (red)

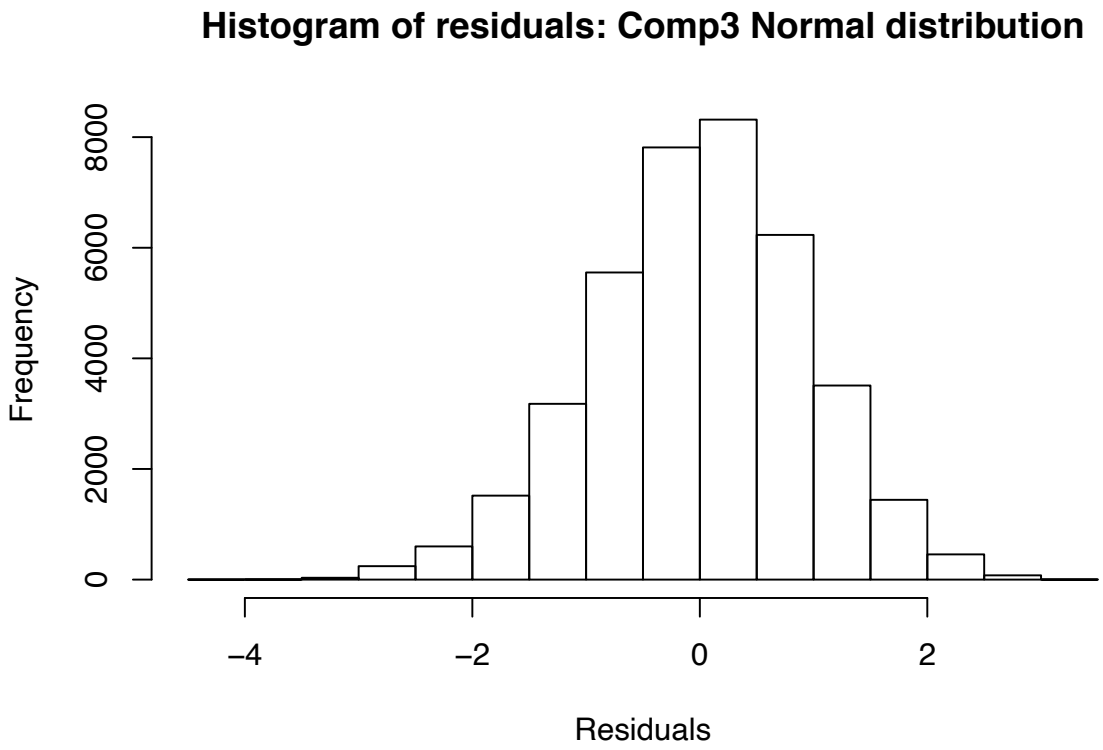

Figure 43: Histogram of residuals for component 3 (Catch per positive FAD set of SKJ) for the 2010-2018 period. Model including Gulland index.

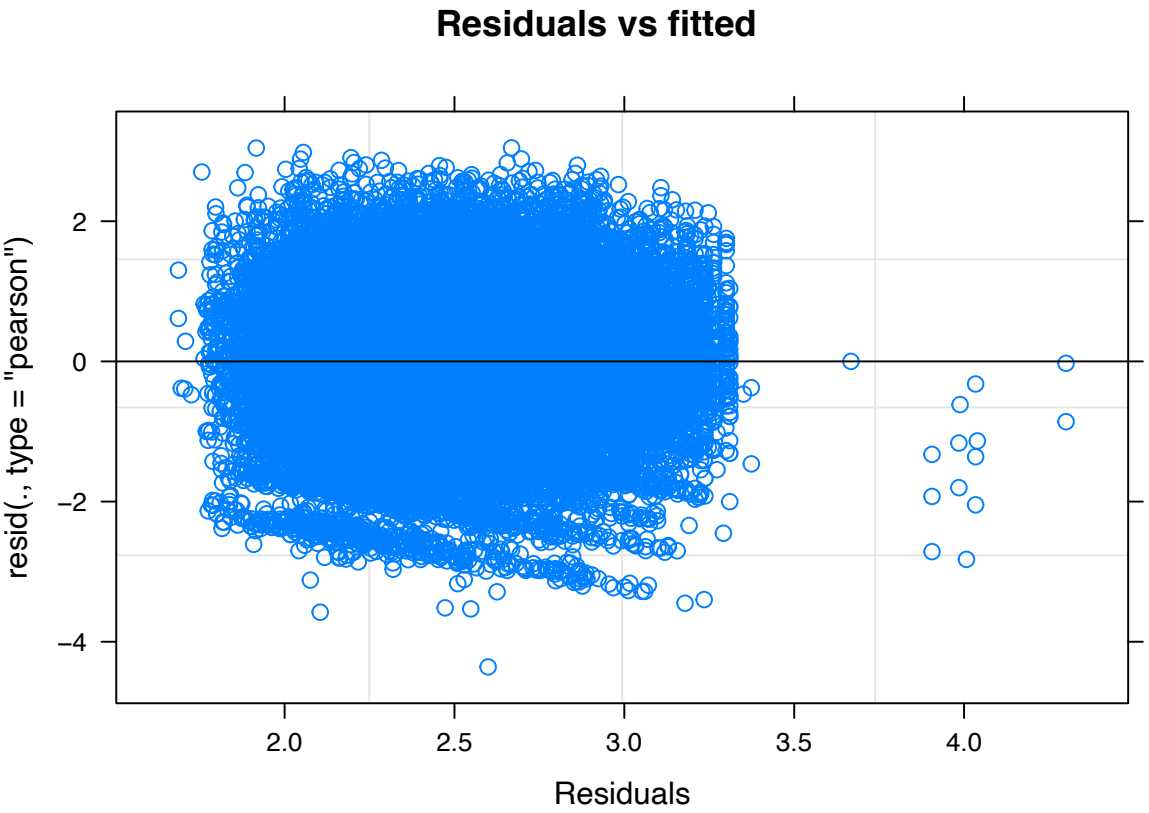

Figure 44: Plot of residuals versus fitted values for component 3 (Catch per positive FAD set of SKJ) for the 2010-2018 period. Model including Gulland index.

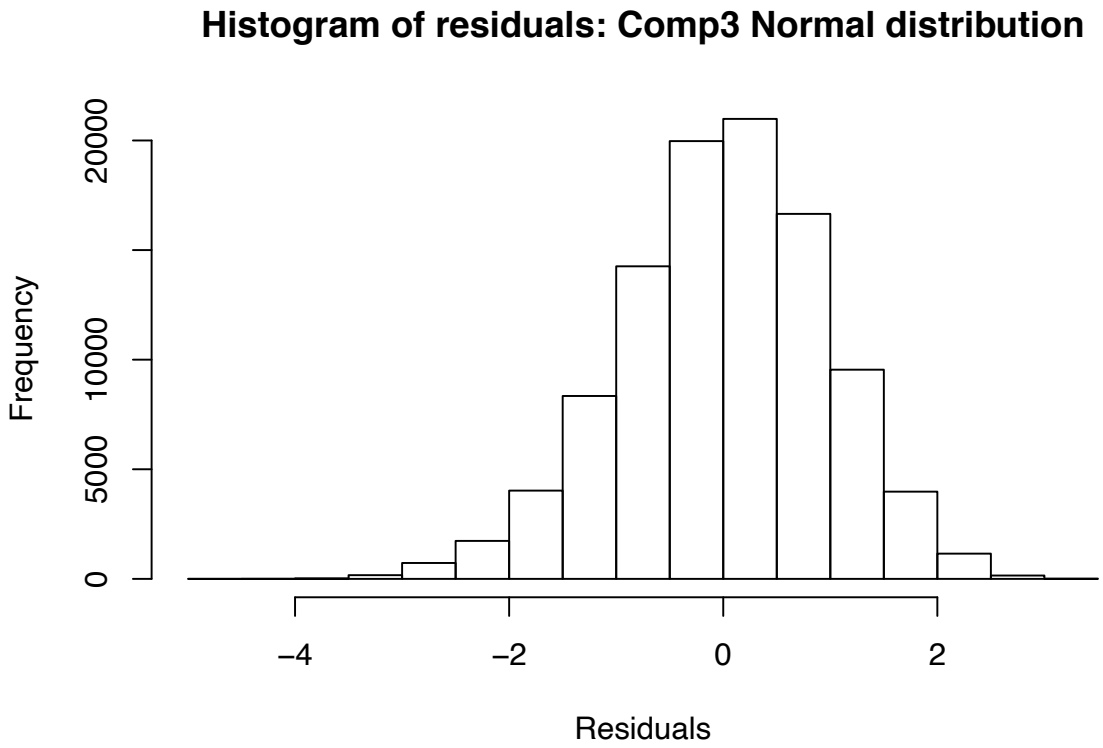

Figure 45: Histogram of residuals for component 3 (Catch per positive FAD set of SKJ) for the 1991-2018 period. Model including a spatial variable (cwp55).

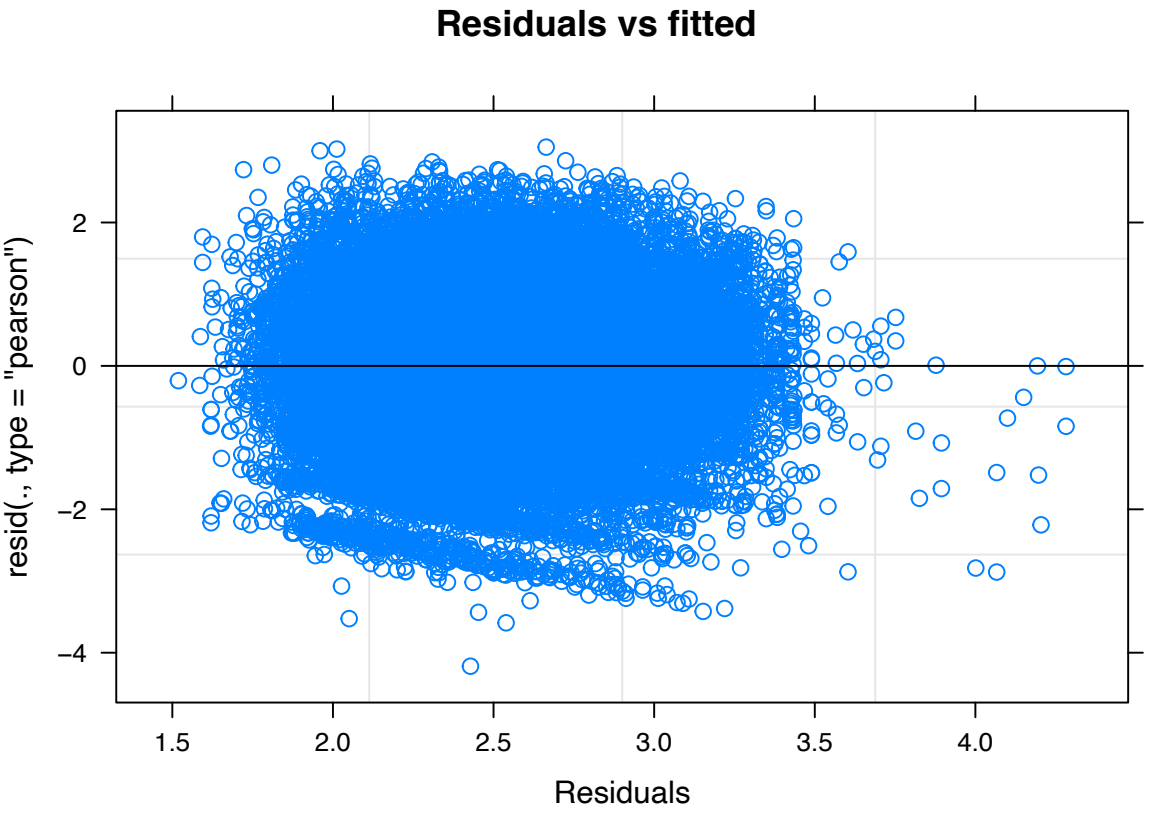

Figure 46: Plot of residuals versus fitted values for component 3 (Catch per positive FAD set of SKJ) for the 2010-2018 period. Model including a spatial variable (cwp55).

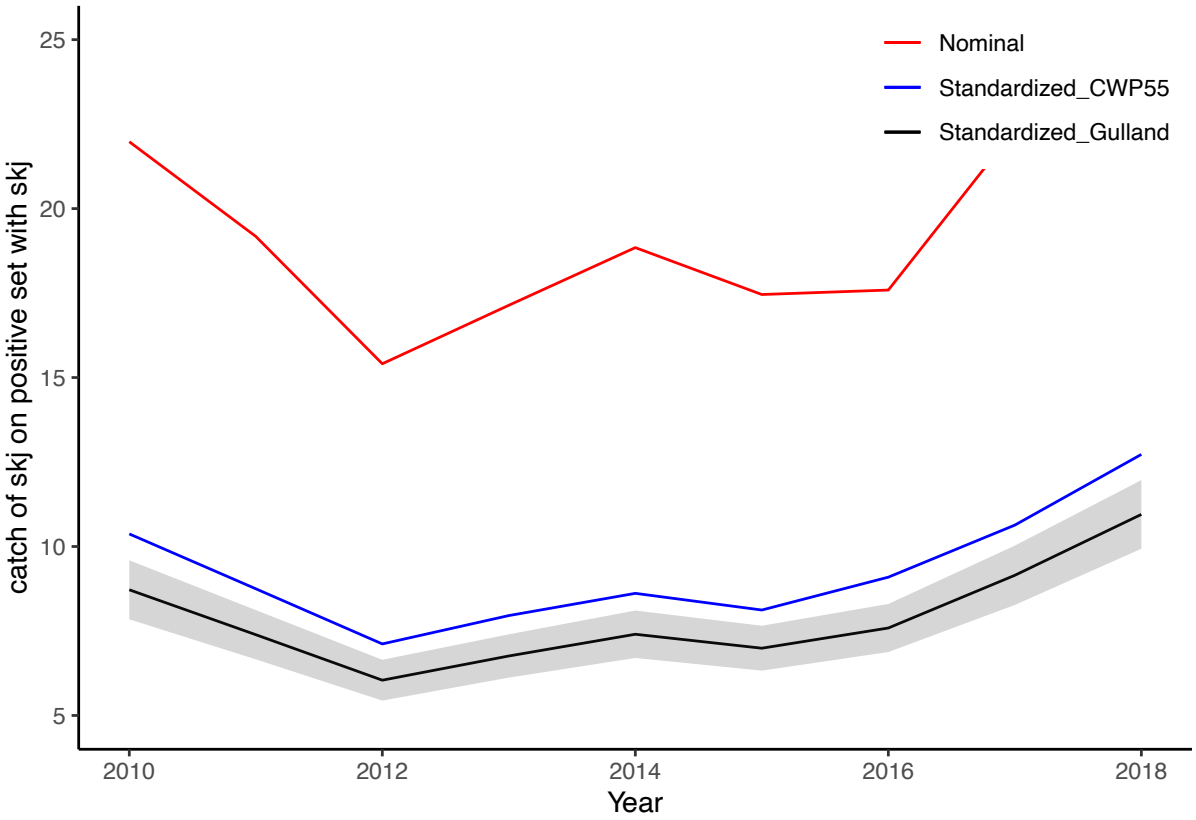

Figure 48: FAD sets - 2010-2018 - catch of skj on positive set with skj; standardised time series by year with Gulland Index (black), CWP55 position (blue) with 97.5% confidence intervals (grey) compared to nominal (red)

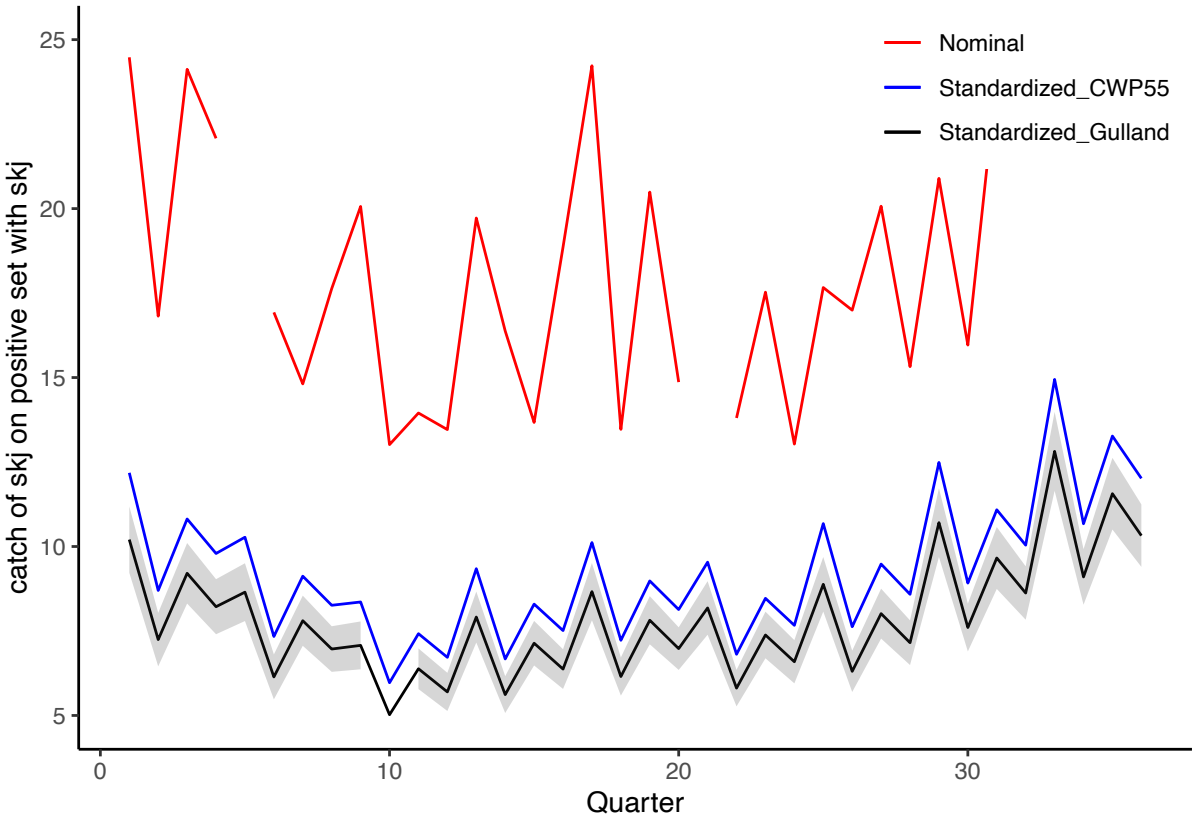

Figure 49: FAD sets - 2010-2018 - catch of skj on positive set with skj; standardised time series by quarter with Gulland Index (black), CWP55 position (blue) with 97.5% confidence intervals (grey) compared to nominal (red)