

## ミナミマグロの CPUE 標準化への GAM 利用についての検討

### Examination of the abundance index for southern bluefin tuna calculated through GAM CPUE standardization

伊藤智幸

Tomoyuki ITOH

国立研究開発法人水産研究・教育機構 国際水産資源研究所

National Research Institute of Far Seas Fisheries,

Japan Fisheries Research and Education Agency

#### 要旨

CCSBT でミナミマグロの資源評価に用いる資源量指数を、GLM によるものから GAM によるものに 2020 年に変更することとなった。新たな資源量指数の特性を探索した。Nominal CPUE が 2019 年に高いこととの乖離について、Nominal CPUE は一つの時空間（2019 年 5 月、7 海区、40S、150E）に操業が集中したことで引き上げられており、標準化した資源量指数で低下したことは妥当と考えられた。新たな指数はモデルに多くの説明変数を含めたことから残差絶対値は小さくなりフィッティングは良くなった。しかし変数別には当てはまりの悪い部分もあった。2013-2016 年の資源量指数が高かったことについては、6 海区の 45S のデータに起因していた。ここでのデータは 2015 年以前にも少なく、2016 年以降は存在しないことから、推定結果の解釈には注意が必要である。

#### Summary

In 2020, the abundance index used for the stock assessment for southern bluefin tuna in CCSBT will be changed from that of GLM to that of GAM. The characteristics of the new abundance index were searched. Regarding the divergence between Nominal CPUE being high in 2019, Nominal CPUE has been raised due to the concentration of operations in one spatiotemporal space (May 2019, 7 areas, 40S, 150E), and the standardized abundance index was considered appropriate. Since the new index includes many explanatory variables in the model, the absolute value of the residual is small and the fitting is good. However, there was also a poor fit in some part of variables. The high 2013-2016 abundance index was attributed to the 45S data for Area 6. The data here is

small before 2015 and does not exist after 2016, so caution should be exercised in interpreting the estimation results.

## 緒言

CCSBTにおいてミナミマグロ *Thunnus maccoyii* の資源量指数については、2007年のCPUEワークショップによってコア船データからGLMで標準化したものを作成してきた(Anon. 2007)。2019年のESC24では、2018年の資源量指数値が極めて高かったことから原因の探求が指示された(Anon. 2019)。2020年CPUEグループ会合では、原因が8海区と9海区との35S及び40Sでのデータの非対応にあることが特定された。同会議ならびにその後のOMMP11ではGAMを用いた資源量指数とGLMMを用いた資源量指数が提案され、GAMのものが採択された(Anon. 2020)。

ただしGAMによる資源量指数の挙動にも懸念点が指摘され、ESC25までに追及することとなった。懸念点は(1) Nominal CPUEは継続して増加しており2019年は特に高いが、資源量指数は2019年に2015年や2018年よりも低いのはなぜか?(2) GAM資源量指数では2015年ごろ(2013-2016年)が従来のもの(GLMのBase)よりも高くなっているのはなぜか?本文書ではそれらの検討結果を示す。また、資源量指数の推定の評価を行う。

## Introduction

In CCSBT, abundance index of southern bluefin tuna *Thunnus maccoyii* in CCSBT have been prepared by generalized linear model (GLM) standardization from the data of core vessels since the 2007 CPUE Workshop (Anon. 2007). At ESC24 in 2019, the exploration of the cause was instructed because the abundance index value in 2018 was extremely high (Anon. 2019). The 2020 CPUE Group Meeting identified that the cause was the non-correspondence of data at 35S and 40S in Areas 8 and 9. At the same meeting and thereafter at OMMP11, an abundance index using generalized additive model (GAM) and an abundance index using generalized linear mixed model (GLMM) were proposed, and GAM was adopted (Anon. 2020).

However, there were concerns about the behavior of the abundance index by GAM, and it was decided to pursue it by ESC25. Concerns are (1) Nominal CPUE continues to increase and is particularly high in 2019, but why is the abundance index lower in 2019 than in 2015 or 2018? (2) Why is the GAM abundance index around 2015 (2013-2016) higher than the previous one (Base of GLM)? This document shows the results of examination of these points. In addition, the estimation of the abundance index is evaluated.

## 材料と方法

元データ：コア船データ。年・月・緯度 5 度・経度 5 度別（セルと称す）。詳細は CCSBT-OMMP/2006/11 を参照。なお、5 海区は 4 海区に統合され、6 海区は 7 海区に統合されている。

CPUE の標準化は GLM、GLMM、GAM を使用した場合で実施した。

GLM は以下のモデルで実施した。これは Base series とよばれるものである。

$$\log(\text{CPUE}+0.2) = \text{Intercept} + \text{Year} + \text{Month} + \text{Area} + \text{Lat5} + \text{BET\_CPUE} + \text{YFT\_CPUE} + (\text{Month}*\text{Area}) + (\text{Year}*\text{Lat5}) + (\text{Year}*\text{Area}) + \text{Error};$$

GLMM は GLM の Base モデルを変形した以下のモデルで実施した。

$$\log(\text{CPUE}+0.2) = \text{Intercept} + \text{Year} + \text{Month} + \text{Area} + \text{Lat5} + \text{BET\_CPUE} + \text{YFT\_CPUE} + (\text{Month}*\text{Area}) + (\text{Year}*\text{Lat5}) + (1/\text{Year}:\text{Area}) + \text{Error};$$

GAM は以下のモデルで実施した。

$$\log(\text{CPUE} + 0.2) \sim \text{Intercept Year} + \text{te}(\text{Lon}, \text{Lat}) + \text{te}(\text{Month}, \text{Lat}) + \text{te}(\text{Lon}, \text{Month}) + \text{te}(\text{Year}, \text{Lat}) + \text{te}(\text{Year}, \text{Month}) + \text{te}(\text{Lat}, \text{Lon}) + \text{te}(\text{Lat}, \text{Lon}) + \text{s}(\text{BETcpue}) + \text{s}(\text{YFTcpue}) + \text{Error};$$

詳細は、GLM については CCSBT-OMMP/2006/11 を、GLMM については CCSBT-OMMP/2006/12 を、GAM については CCSBT-OMMP/2006/15 をそれぞれ参照のこと。

CPUE 及び資源量指数では、数値とデータセットの組み合わせで以下の 7 種類の数字が生じる。

値 1: オリジナルデータを使った Nominal CPUE。(合計 SBT 数) / (合計 Hook 数)。

値 2: オリジナルデータを使った標準化していない CPUE の平均値。CPUE = (階層別の SBT 数) / (階層別の Hook 数)。CPUE の全階層での平均値。

値 3: オリジナルデータによる標準化した係数値。coefficient of variable。

値 4: オリジナルデータに当てはめた予測値。派生として残差 (予測値 - 観測値)。

値 5: newdata に当てはめた予測値で、重みづけしない。newdata は変数の階層の総組み合わせで作成される。

値 6: newdata のうち重みづけデータと対応する範囲に限定し、重みづけしないもの。例えば 4 海区の 45S は削除される。

値 7: newdata で、重みづけした値。CS、VS、それらの中間的な W0.5、W0.8 など。

この文書の解析では Nominal CPUE の検討については値 1 と値 2 を用いている。モデルのフィッティングは値 4 を用いている。資源量指数の検討 (2015 年ごろの高まり) については値 7 を用いている。

解析は R (version 3.6.1) を用いた。

## Materials and methods

Data source is the core ship data. The data were summarized by year, month, latitude 5 degrees, and longitude 5 degrees (called cells). See CCSBT-OMMP/2006/11 for details. In addition, Area 5 was integrated into Area 4, and Area 6 was integrated into Area 7. CPUE was standardized by GLM, GLMM, and GAM.

GLM was implemented with the following model. This has been called the Base series.

$$\log(\text{CPUE}+0.2) = \text{Intercept} + \text{Year} + \text{Month} + \text{Area} + \text{Lat5} + \text{BET\_CPUE} + \text{YFT\_CPUE} + (\text{Month}*\text{Area}) + (\text{Year}*\text{Lat5}) + (\text{Year}*\text{Area}) + \text{Error};$$

GLMM was implemented by the following model, which is a modified version of GLM's Base model.

$$\log(\text{CPUE}+0.2) = \text{Intercept} + \text{Year} + \text{Month} + \text{Area} + \text{Lat5} + \text{BET\_CPUE} + \text{YFT\_CPUE} + (\text{Month}*\text{Area}) + (\text{Year}*\text{Lat5}) + (1/\text{Year}:\text{Area}) + \text{Error};$$

GAMM was implemented by the following model.

$$\log(\text{CPUE} + 0.2) \sim \text{Intercept} \text{Year} + \text{te}(\text{Lon}, \text{Lat}) + \text{te}(\text{Month}, \text{Lat}) + \text{te}(\text{Lon}, \text{Month}) + \text{te}(\text{Year}, \text{Lat}) + \text{te}(\text{Year}, \text{Month}) + \text{te}(\text{Lat}, \text{Lon}) + \text{te}(\text{Lat}, \text{Lon}) + \text{s}(\text{BETcpue}) + \text{s}(\text{YFTcpue}) + \text{Error};$$

For details, see CCSBT-OMMP/2006/11 for GLM, CCSBT-OMMP/2006/12 for GLMM, and CCSBT-OMMP/2006/15 for GAM.

Note that in the CPUE and resource abundance indices, the following seven types of values occur in combination with numbers and data sets.

Value 1: Nominal CPUE using original data. (Total SBT number) / (Total hook number).

Value 2: Average of unstandardized CPUE using original data. CPUE=(number of SBT by layer)/(number of hooks by layer). Average value of all levels of CPUE.

Value 3: Standardized coefficient value from the original data. coefficient of variable.

Value 4: The predicted value fitted to the original data. Residuals as a derivative (predicted value-observed value).

Value 5: The predicted value applied to newdata, without weighting. newdata is created by the total combination of variable hierarchies.

Value 6: New data is limited to the range corresponding to the area weighted data and is not weighted. For example, 45S in Area 4 will be deleted.

Value 7: newdata weighted value. Constant squares (CS), variable squares (VS), their intermediate W0.5, W0.8, etc.

The analysis in this document uses value-1 and value-2 for Nominal CPUE considerations. The model fitting uses value-4. Value-7 is used for the examination of the abundance index (increased around 2015).

R (version 3.6.1) was used for the analysis.

## 結果と考察

### 1. 2019 年の Nominal CPUE との違いの原因の特定

Fig. 1 に Nominal CPUE と平均 CPUE を示す。両者には違いがあり、ほとんど全ての年で Nominal CPUE が上回っていたが、2009 年のようにわずかに逆転している年もあった。2019 年には特に違いが大きかった。2017 年にも大きく、2018 年には違いが小さかった。違いは操業回数が海域によって異なることで引き起こされる。2019 年に違いが大きかった原因を探索した。

CPUE が高い海域で操業回数が多いかと思われ、セルを単位として CPUE と Hook 数の関係をプロットした (Fig. 2; 全ての年については Attachment 1 に示す)。両者間の相関が有意かをピアソンの相関係数の ANOVA の F 検定で検討した。データでは多くの年で正の相関がみられ、高い CPUE 海域への努力量の集中が示唆された。しかし相関が有意の年は 34 年中の 9 年にすぎず、強い相関は見られなかった。また、Hook が多いが CPUE が低いというデータ点も多くみられた。よって船は高 CPUE 海域に素早く集まっているわけではないと考えられた。

あるセルが Nominal CPUE にどれだけ強い影響を与えているかを以下のスコアを求めて調べた。

①年別 Nominal CPUE (N1) を求める。

②ジャックナイフ (データを 1 件だけ削除) して年別 Nominal CPUE (N2) を求める。

③各セルでスコアを計算する。

$$\text{スコア} = (N1 - N2) / N1$$

スコアは正の値 (そのセルが Nominal CPUE を引き上げた) と負の値 (そのセルが Nominal CPUE を引き下げた) の場合がある。

Fig. 3 にスコアを示す。この図ではスコアを順に並べて示している。2019 年には一つのセルが大きな影響を与えていた。このセルは Nominal CPUE を 9% (差のうちの 58%) 増加させていた。

この影響の強いセルは、5 月の 40S、150E (7 海区) であった (Fig. 4; 全ての月については Attachment 2 に示す)。7 海区には多くの年で最も正の影響の強いセルが存在し、また、多くの年で最も負の影響の強いセルが存在していた。CPUE が高いが操業が集中していない区画もあった (e.g. 2011 年 5 月の 7 海区)。

ここまでの解析から以下にまとめられる。

- ① Nominal CPUE とセル別 CPUE の平均値とは異なる場合がある。
- ② 多くの年では Nominal CPUE の方が高いが、一部の年ではほぼ同値であったり、1 年間ではわずかに逆転した場合があった。
- ③ 差は、セル (年・月・緯度 5 度・経度 5 度) 毎の hook 数が異なることに起因する。
- ④ 2019 年に Nominal CPUE が高かったのは 1 セル (7 海区 5 月 40S、150E) が強く影響しており、高い CPUE 海域に操業が集中していたため。
- ⑤ 操業は常に高い CPUE 海域に集中するわけではなかった。

よって、Nominal CPUE は一つの時空間 (2019 年 5 月、7 海区、40S、150E) に操業が集中したことで引き上げられており、標準化した資源量指数で低下したことは妥当と考えられた。

## 2. 残差によるモデル結果の評価

GLM、GLMM、GAM 間で CS 資源量指数は、1986 年から 2012 年まではほとんど一致した年トレンドを示している (Fig. 5)。しかし 2013 年から 2016 年には GLM および GLMM に比較して、GAM の値が高くなっている。まずモデル間の残差による評価を行い、その後、この 2015 年ごろの資源量指数の高まりの確実性を評価した。

QQ プロットによる全体的な当てはまりは、GAM と GLMM が同等で、GLM は劣った (Fig. 6)。newdata におけるレコード別の CS 値をモデル間で比較した (Fig. 7)。GAM または GLMM に対して外れ値は見られていない。GLM では 8 海区 40S での外れ値が 2018 年と 2019 年に存在した。この問題は既に Itoh (2020; CCSBT-OMMP/2006/11) で記述してある。

残差 (値 4) について要約した。残差の 2 乗によるモデル間の比較を Fig. 8 に示す。残差 (2 乗しない) によるモデル間の比較を、一変数について Fig. 9 に、二変数について Fig. 10 (年・緯度)、Fig. 11 (年・月)、Fig. 12 (年・海区) に示す。経度を含めたすべての結果は Attachment 3 で利用可能である。以下の特徴がみられた。

- 残差の 2 乗において、GAM は優れていた。
- それぞれのモデルで 2019 年は最も残差 2 乗値が大きかった。
- 残差 (2 乗しない) の検討において、GAM の当てはまりは、GLM/GLMM に対して、海区、緯度、経度で優れていた。ただし GAM はこれらをモデル項に含んでいるので当然の結果である。
- GAM の年への当てはまりは、GLMM に劣った。
- 年別緯度 (Fig. 10) について 45S (グラフで-47.5 と表記) と 40S において、GLM と GLMM では過大推定で最近 10 年間について経年的な増加が見られた。GAM では経年的な振幅の増加がみられた。35S については 3 モデルとも最近 10 年で過小推定が経年的に増加していた。
- 年別月 (Fig. 11) について、2019 年に当てはまりが悪かったがその誤差 (5 月の過小推定、7 月の過大推定、9 月の過小推定) の方向と程度は 3 モデルで共通して

いた。

- 年別海区 (Fig. 12) について、2019 年に当てはまりが悪かったがその誤差 (4 海区の過小推定、7 海区の過大推定、8 海区の過小推定) の方向と程度は 3 モデルで共通していた。
- 年別海区について、7 海区では 3 モデルで共通して 2012 年から過大推定が続いている。

以上から、GAM のフィッティングは悪くはないが、他のモデルよりも圧倒的に優れるものではなく、変数のある部分に対しては当てはまりが悪い場合があった。

### 3. GAM で 2015 年ごろに値が高いことの信頼性

GLM、GLMM、GAM 間で CS 資源量指数は、1986 年から 2012 年まではほとんど一致した年トレンドを示している (Fig. 5)。しかし 2013 年から 2016 年には GLM および GLMM に比較して、GAM の値が高くなっている。この 2015 年ごろの資源量指数の高まりの確実性を評価した。

プロGRESSプロットで検討したところ GAM において 2015 年ごろに上昇したのは、主に年・緯度項が影響していた (Fig. 13)。次いで一部を年・月項が影響していた。合意された GAM (GAM11) から年・緯度項および年・月項を削除した GAM

(GAM\_YMTLat) を実施し、CS 値を求め、2012 年に両者が一致するようにし、グラフ上での GAM11 との違いの領域の広さで原因を探索した (Fig. 14)。その結果、大きな影響を与えていたのは 7 海区の 4 月、5 月、6 月であった。これらは 7 海区の中心的な操業時期であったことから、2015 年ごろの高まりが少数の外れ値によって引き起こされているわけではないことを示している。

緯度は 45S が大きな影響を与えており、40S もそれより小さな程度で影響を与えていた。コア船データに含まれる 7 海区 (前述のように 6 海区との統合) のデータは、45S では少なく操業回数も少ない (Table 1)。2016 年からは無くなっている。また 45S での操業は 2002 年から 2015 年までは、160E より東で、6 海区のものであった (Table 2)。

GAM は 45S (グラフの-47.5) での 2015 年のフィットが悪かった (Fig. 10)。7 海区に限定したグラフにおいても 45S のフィットが悪く、2015 年に過大推定であった (Fig. 15)。40S(-42.5)では、GAM だけでなく GLM 及び GLMM も、2013-2019 年は過大推定となっていた (Fig. 15)。

以上から、資源量指数の 2013 年から 2016 年の高まりは 6/7 海区の 45S に起因し、それはモデルの当てはまりが悪いこと、データが少ないことに起因することを認識する必要がある。2015 年ごろ (2013-2016 年) の高まりについての信頼度が低いと考えられる。

## Results and discussion

### 1. Identifying the causes of the difference from the 2019 Nominal CPUE

Figure 1 shows Nominal CPUE and average CPUE. There was a difference between

the two, with Nominal CPUE surpassing almost every year, but in some years, like 2009, it was slightly reversed. The difference was particularly large in 2019. It was also large in 2017 and small in 2018. The difference is caused by the number of operations depending on Area. The causes of the large difference in 2019 is further searched.

It seems that the number of operations is high in the sea where CPUE is high, and the relationship between CPUE and hook number used is plotted in units of cells (Fig. 2; Attachment 1 for all years). We examined whether the correlation between the two was significant by the ANOVA F test of Pearson's correlation coefficient. The data showed a positive correlation for many years, suggesting a concentration of effort in the high CPUE area. However, the correlation was significant only in 9 out of 34 years, and no strong correlation was found. In addition, there were many data points that there were many hooks used but low CPUE. Therefore, it was considered that the vessels were not quickly gathering in the high CPUE area.

We investigated how much a certain cell had effect on Nominal CPUE by obtaining the following scores.

- 1 Calculate the Nominal CPUE (N1) by year.
- 2 Calculate the Nominal CPUE (N2) for each year by jackknife method (delete only one data).
- 3 Calculate the score in each cell.

$$\text{Score} = (N1-N2)/N1 ,$$

Scores can be positive (the cell raised Nominal CPUE) and negative (the cell lowered Nominal CPUE).

Figure 3 shows the score. In this figure, the scores are arranged in order. One cell had a large positive impact in 2019. This cell increased Nominal CPUE by 9% (58% of the difference).

The cells with strong influences were 40S and 150E (Area 7) in May (Fig. 4; all months are shown in Attachment 2). Area 7 had cells with the most positive impacts for many years, and cells with the most negative impacts for many years. There were also some cells with high CPUE but not concentrated operations (e.g. Area 7 in May 2011).

The analysis up to this point is summarized below.

- (1) The average value of Nominal CPUE and cell-specific CPUE may differ.
- (2) Nominal CPUE was higher in many years, but it was almost the same in some years or slightly reversed in one year.
- (3) The difference is due to the different number of hooks for each cell (year/month/latitude 5°/longitude 5°).
- (4) Nominal CPUE was high in 2019 because 1 cell (May 40S, 150E in Area 7) had a strong influence, and operations were concentrated in the high CPUE area.
- (5) Operations were not always concentrated in high CPUE area.

Therefore, the Nominal CPUE has been raised due to the concentration of operations in one spatiotemporal space in 2019 (May, Area 7, 40S, 150E), and it is considered appropriate that the standardized abundance index has decreased.

## 2. Evaluate model results with residuals

The CS abundance indices among GLM, GLMM, and GAM show almost the same annual trend from 1986 to 2012 (Fig. 5). However, from 2013 to 2016, GAM has a higher value than GLM and GLMM. First, we evaluated the residuals between the models, and then evaluated the certainty of the increase in the abundance index around 2015.

The overall fit from the QQ plot was that GAM and GLMM were equivalent and GLM was inferior (Fig. 6). The CS values for each record in newdata were compared between the models (Fig. 7). No outliers have been found for GAM or GLMM. GLM had outliers in the 40s in Area 8 in 2018 and 2019. This problem has already been described in Itoh (2020; CCSBT-OMMP/2006/11).

The residuals (value-4) were summarized. Figure 8 shows a comparison between the models based on the square of the residual. Comparison between models by residuals (not squared) is shown in Fig. 9 for one variable, Fig. 10 (year/latitude), Fig. 11 (year/month), and Fig. 12 (year/sea area) for two variables. All results, including longitude, are available in Attachment 3. The following features were observed.

- GAM was excellent in the residual squared.
- Each model had the largest residual squared value in 2019.
- In the examination of residuals (not squared), the GAM fit was superior to GLM/GLMM in terms of Area, latitude, and longitude. However, since GAM includes these in the model term, this is an expected result.
- GLMM fits the year better than GAM.
- GLM and GLMM overestimate over the last 10 years at 45S (expressed as -47.5 in the graph) and 40S for by year and latitudes (Fig. 10), and have shown an increase with year. In GAM, the amplitude increased with year. For 35S, underestimation of all three models has increased over the last 10 years.
- Regarding year-by-month (Fig. 11), the fit was poor in 2019, but the direction and degree of the error (underestimation in May, underestimation in July, underestimation in September) were common to all three models.
- Regarding Area by year (Fig. 12), the fit was poor in 2019, but the direction and degree of the error (underestimation of Area 4, overestimation of Area 7, underestimation of Area 8) are common to all 3 models.
- Regarding Area by year, overestimation has continued from 2012 in common for the three models in Area 7.

From the above, the fitting of GAM is not bad, but it is not overwhelmingly superior

to other models, and there are cases where it does not fit well to parts with variables.

### 3. Reliability of high value of GAM around 2015

The CS abundance indices among GLM, GLMM, and GAM show almost the same annual trend from 1986 to 2012 (Fig. 5). However, from 2013 to 2016, GAM has a higher value than GLM and GLMM. We evaluated the certainty of this rise in the abundance index around 2015.

As a result of examination using a progress plot, the increase in GAM around 2015 was mainly due to the year and latitude terms (Fig. 13). Next, partly affected by the year and month term. Perform GAM (GAM\_YMTLat) that removes the year/latitude term and year/month term from the agreed GAM (GAM11), calculate the CS value, and adjust to make them match in 2012, and compare them with GAM11 on the graph. We searched for the cause by the extent of the difference (Fig. 14 as yellow area). As a result, it was in April, May, and June of Area 7 that had a great influence. These months were the main operating seasons in Area 7, indicating that the rise around 2015 was not caused by a few outliers.

For latitude, 45S had a large impact, and 40S had a lesser impact. The data for Area 7 (integrated with Area 6 as described above) included in the core vessel data is low for 45S and the number of operations is also small (Table 1). It has disappeared since 2016. From 2002 to 2015, the operation at 45S was to the east of 160E in Area 6 (Table 2).

The GAM had a poor fit in 2015 at 45S (-47.5 on the graph) (Fig. 10). The 45S fit was poor even in the graph limited to Area 7, and was overestimated in 2015 (Fig. 15). At 40S (-42.5), not only GAM but also GLM and GLMM were overestimated during 2013-2019 (Fig. 15).

From the above, it is necessary to recognize that the increase in the abundance index from 2013 to 2016 is due to 45S in Area 6/7, which is due to poor model fit and limited data. It is considered that the degree of confidence in the rise around 2015 (2013-2016) is low.

### Acknowledgement

Simon Hoyle に、プログレスプロットのコードの提供、解析上の様々な協力に感謝する。

Simon Hoyle is appreciated for providing the progress plot code and for all of the analysis help.

## Reference

- Anonymous (2007) Report of the Second CPUE Modelling Workshop. CCSBT. 21-25 May 2007. Shimizu, Japan.
- Anonymous (2019) Report of the Extended Scientific Committee for the Twenty Fourth Meeting of the Scientific Committee. CCSBT. 2-7 Sep. 2019. Cape Town, South Africa.
- Anonymous (2020) Report of the Eleventh Operating Model and Management Procedure Technical Meeting. CCSBT. 15-19, 22 and 24 June. Web meeting.
- Itoh, T., and Takahashi, N. (2020) Update of the core vessel data and CPUE for southern bluefin tuna in 2020. CCSBT-OMMP/2006/11.
- Itoh, T. (2020) Examination of an anomalously high value of the core vessel CPUE in 2018 for southern bluefin tuna. CCSBT-OMMP/2006/12.
- Hoyle, S., (2020) Exploratory analyses for primary CCSBT CPUE index. CCSBT-OMMP/2006/15.

Table 1. Number of longline operation in Area 7 by year and latitude in the core vessel data.  
Assuming 3000 hooks per operation.

Lat		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
-47.5	45S-49S	247	301	281	273	472	483	222	198	126	186	51	417	254	253	88	112	109	86	107	47
-42.5	40S-44S	156	293	528	757	1172	1001	1122	803	1175	910	609	550	756	1613	1191	1837	1287	995	736	581
-37.5	35S-39S	3		5			7	6			10		11		5	37				4	

Lat		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
-47.5	45S-49S	55	47	88	35	54	45	28	14	9	15				
-42.5	40S-44S	317	646	443	461	448	611	750	1106	1037	1135	1197	1468	1427	1143
-37.5	35S-39S														

Table 2. Number of operation in 45S in Area 7 by year and latitude in the core vessel data.  
Assuming 3000 hooks per operation.

Lon	Area	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
132.5	7		6	7	8	43			13	4		18	37								6
137.5	7		77	29	61	81	142	18	12	15	7		17	169	137	74	13				4
142.5	7		11	113	47	19	46		8				3	28	42						
147.5	7		4	8				9		3	44	5	11								
152.5	7																				
157.5	7																				3
162.5	6						5	11							6	5	7				12
167.5	6	110	78	63	71	128	262	190	166	69	181		86	74	137	48	73	109	86	95	47
172.5	6	44	65	100	66	96	148						91		35	20	17				
177.5	6				26	14	38						3								

Lon	Area	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
132.5	7														
137.5	7														
142.5	7														
147.5	7														
152.5	7							9							
157.5	7														
162.5	6		8	7		6									
167.5	6	47	40	88	28	54	45	18	14	9	15				
172.5	6														
177.5	6														

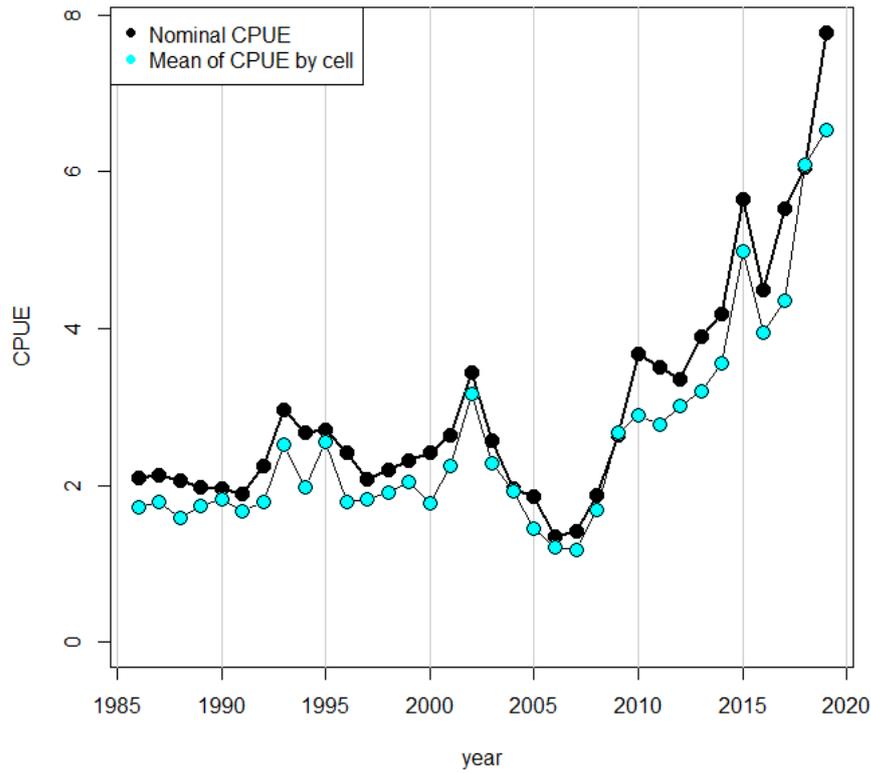


Fig. 1. Comparison of Nominal CPUE and mean CPUE in the core vessel data.

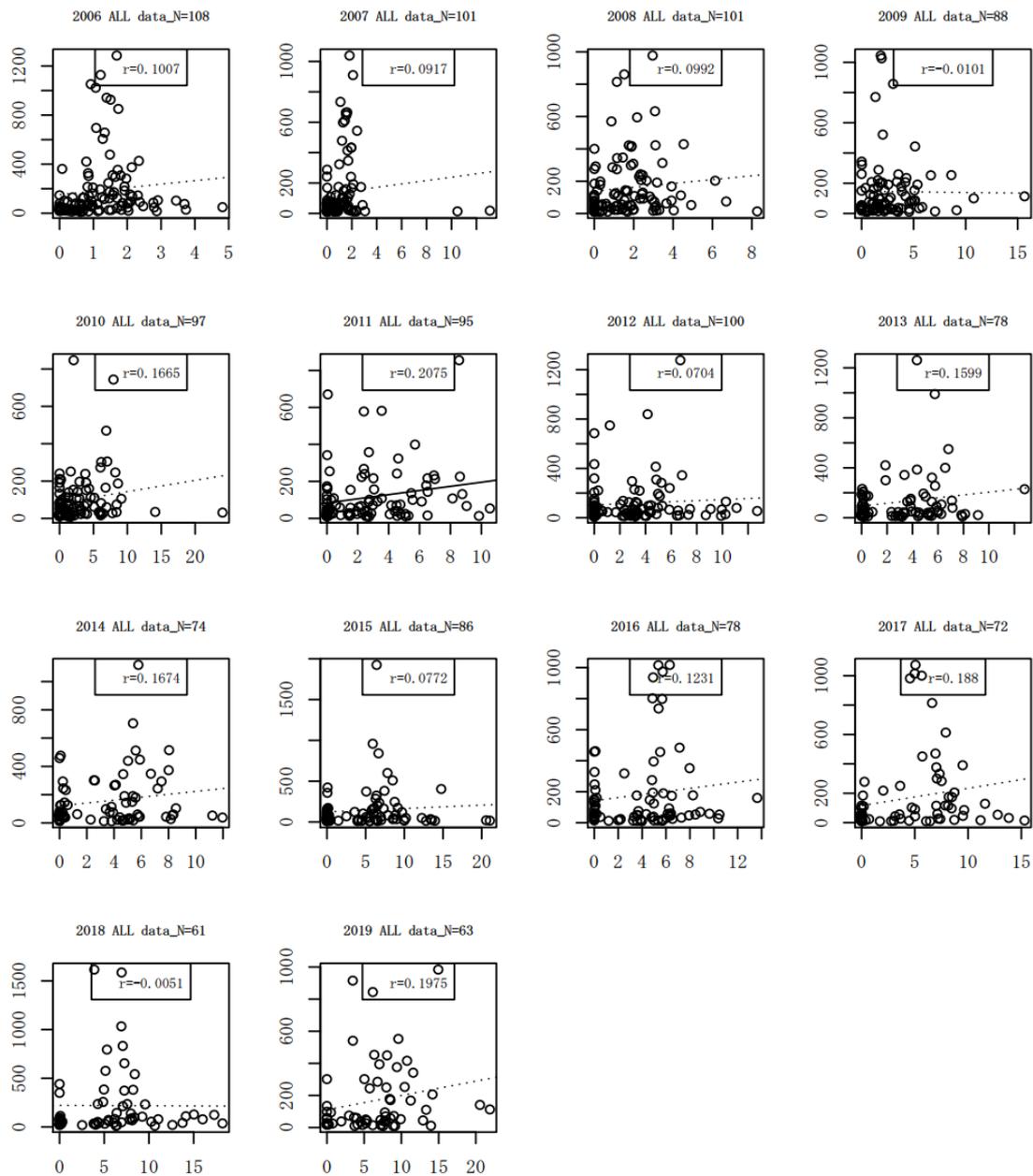


Fig. 2. Plot of hooks (Y axis in 1000 hooks unit) to CPUE (X axis) by cell.

A cell is by year, month, latitude in 5 degree and longitude in 5 degree. Regression line is solid if slope was significantly different from 0. r is the Pearson's correlation coefficient.

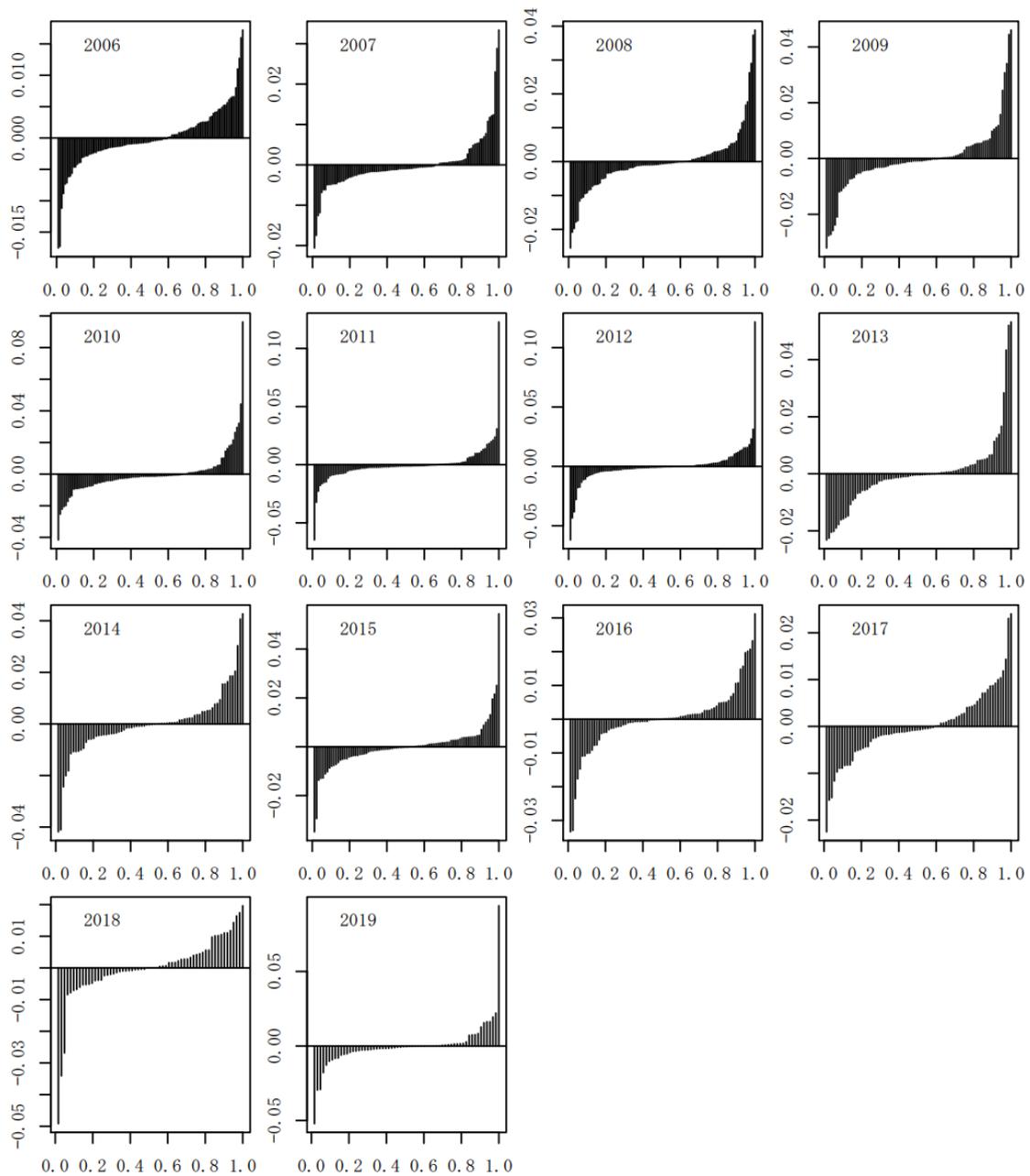


Fig. 3. Proportion of change in Nominal CPUE by one cell record

The scores are arranged in order. X axis is the order of record. Y axis is score which is the influence that changes Nominal CPUE, see text.

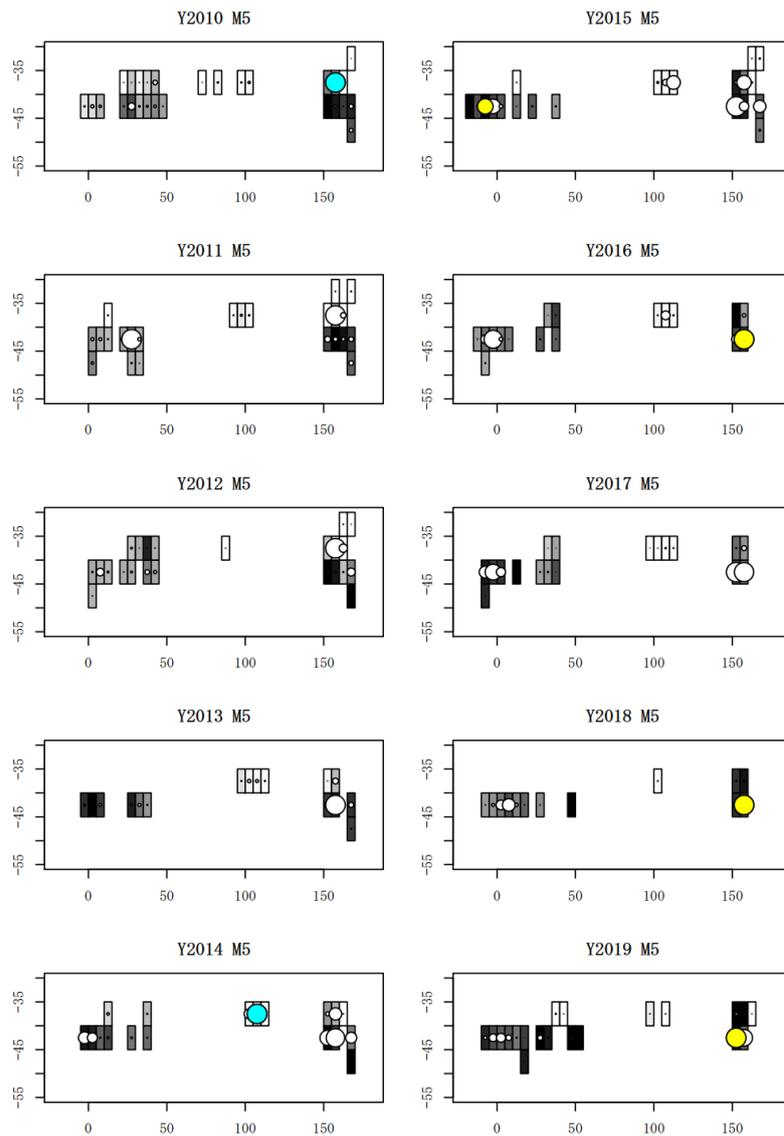


Fig. 4. Mapping of CPUE and hooks

CPUE and hook in May are plotted by year, latitude in 5 degree, and longitude in 5 degree. Longitude is on the X-axis and latitude is on the Y-axis. The color of the section is CPUE (white = 0, black = maximum CPUE for the year/month). Hook is a relative value to the maximum hook of the year and month. Hook in yellow indicates maximum score, light blue indicates minimum score, and white indicates intermediate score. Score is the influence that changes Nominal CPUE, see text.

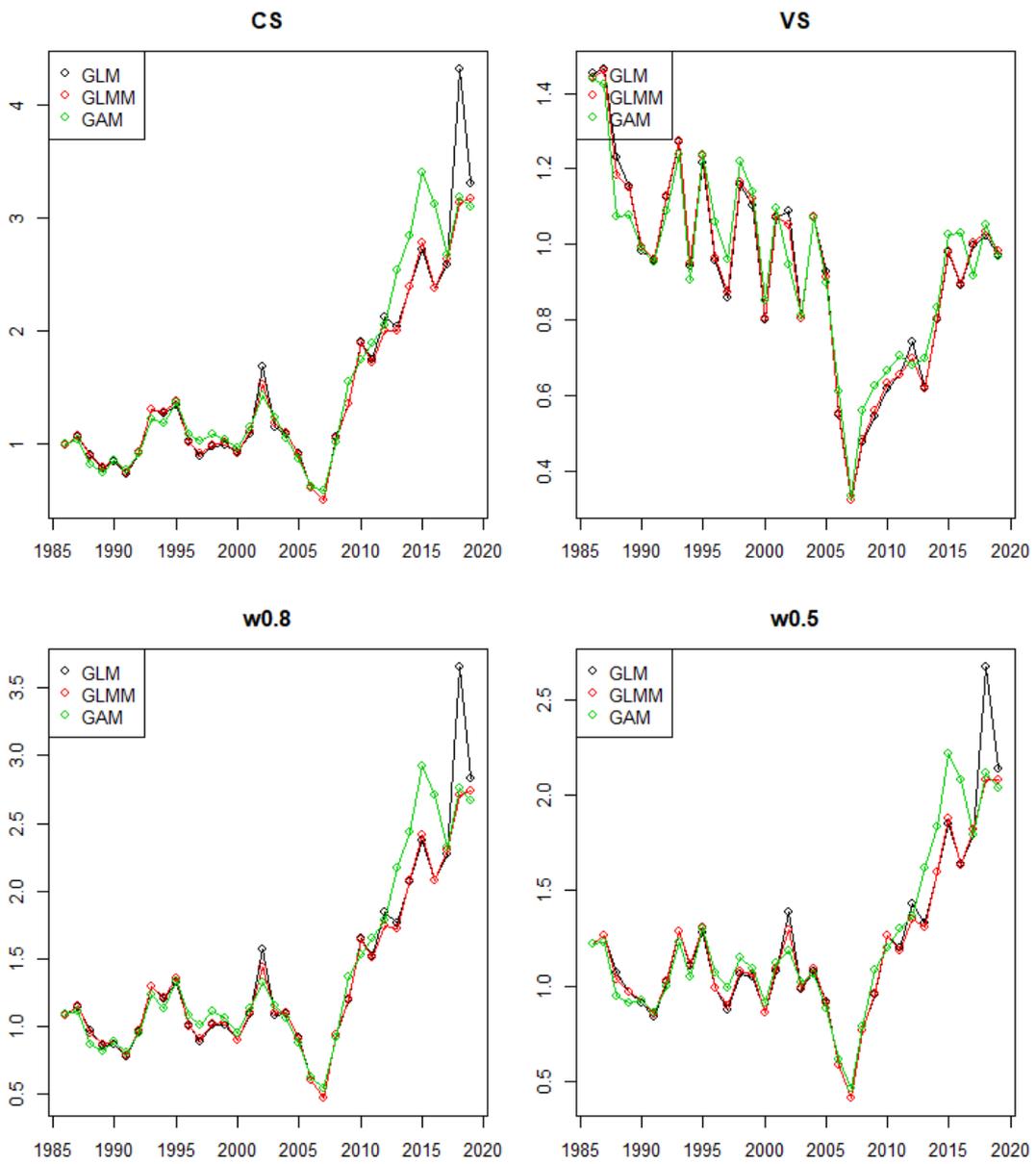


Fig. 5. Abundance indices by 3 models

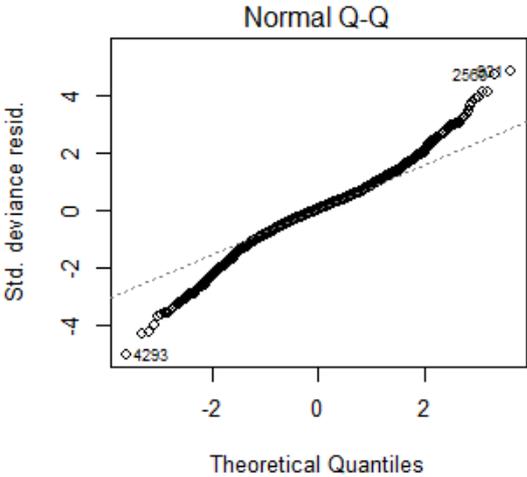


Fig. 6. QQ plot (GLM)

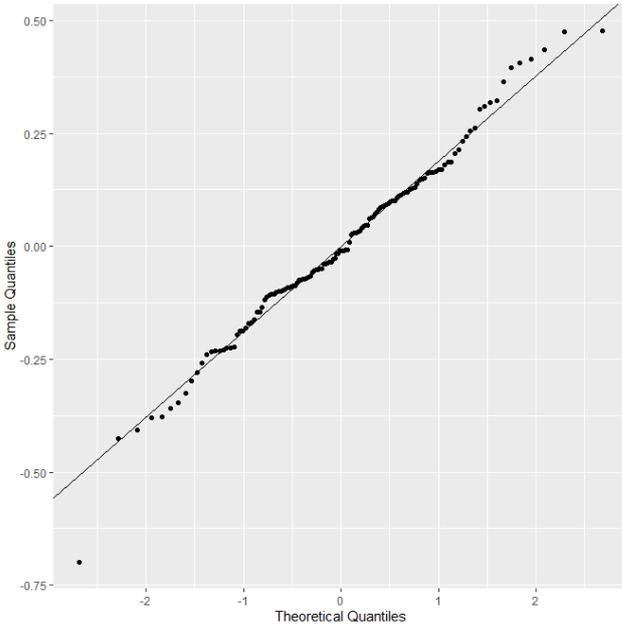


Fig. 6. (cont'd) (GLMM)

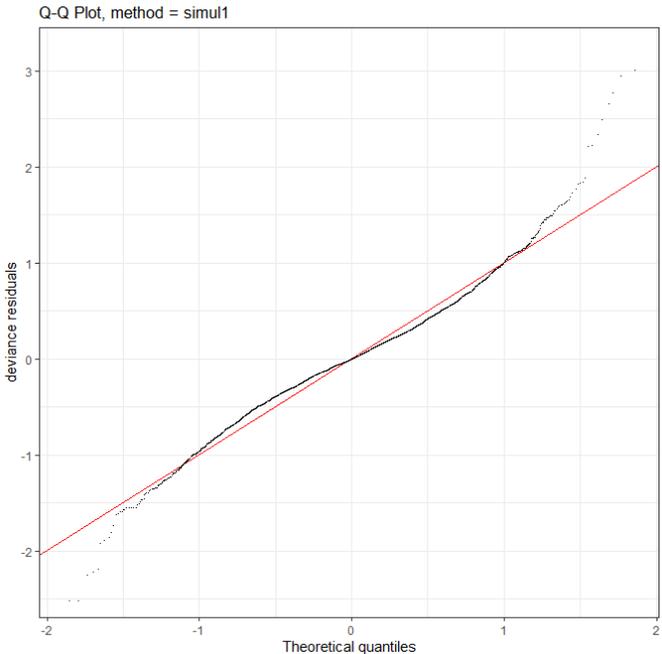


Fig. 6. (cont'd) (GAM)

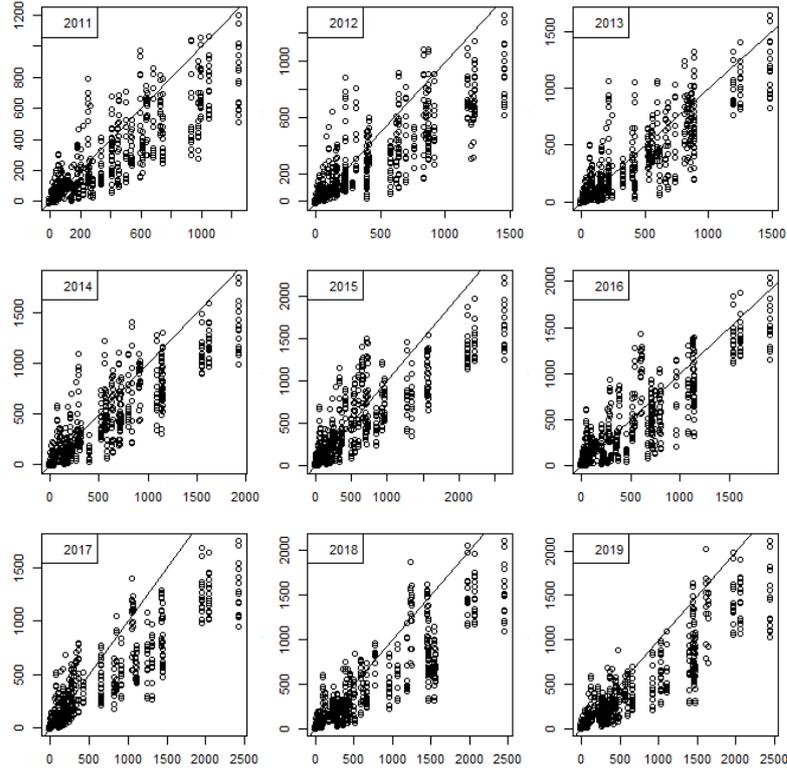


Fig. 7. Comparison of predicted value among models  
Plot CS in newdata. X axis is GLMM, and Y axis is GAM. Line denotes X=Y.

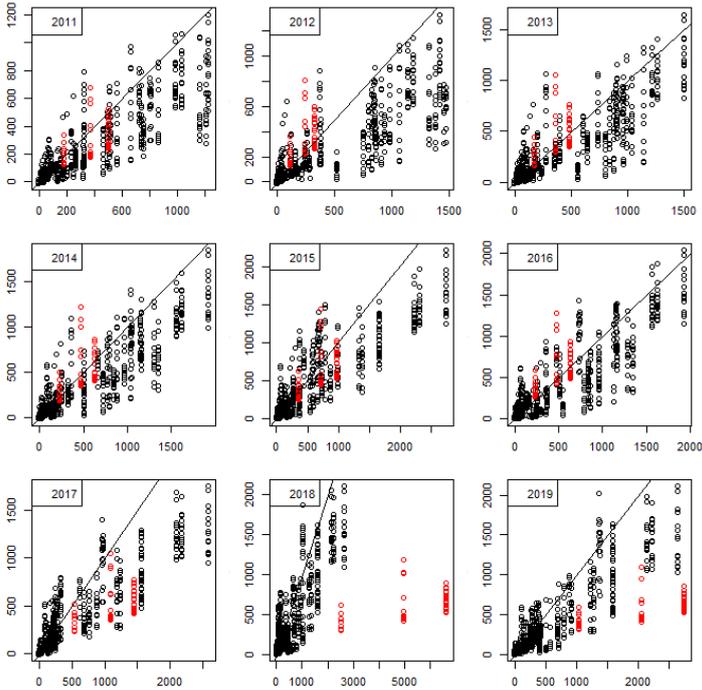


Fig. 7. (cont'd)

X axis is GLM, and Y axis is GAM. Red denotes 40S in Area 8 between June and September.

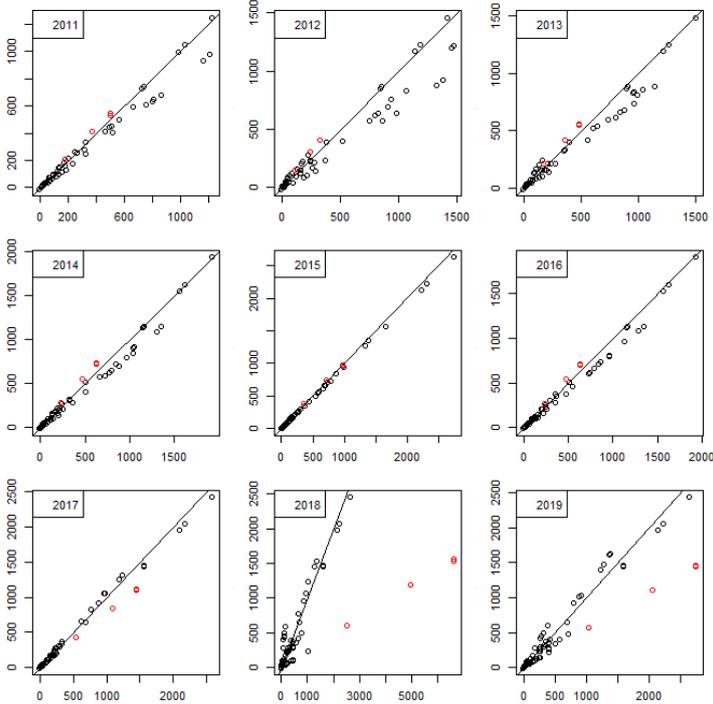


Fig. 7. (cont'd)

X axis is GLM, and Y axis is GLMM. Red denotes 40S in Area 8 between June and September.

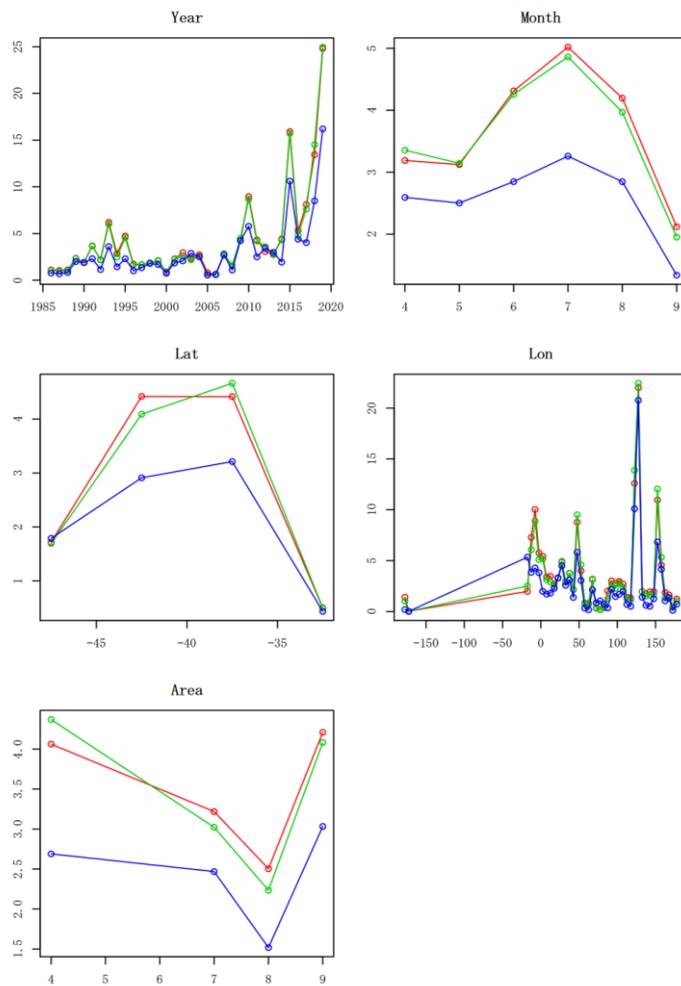


Fig. 8. Squared residuals by variables

Red is GLM, green is GLMM, and blue is GAM.

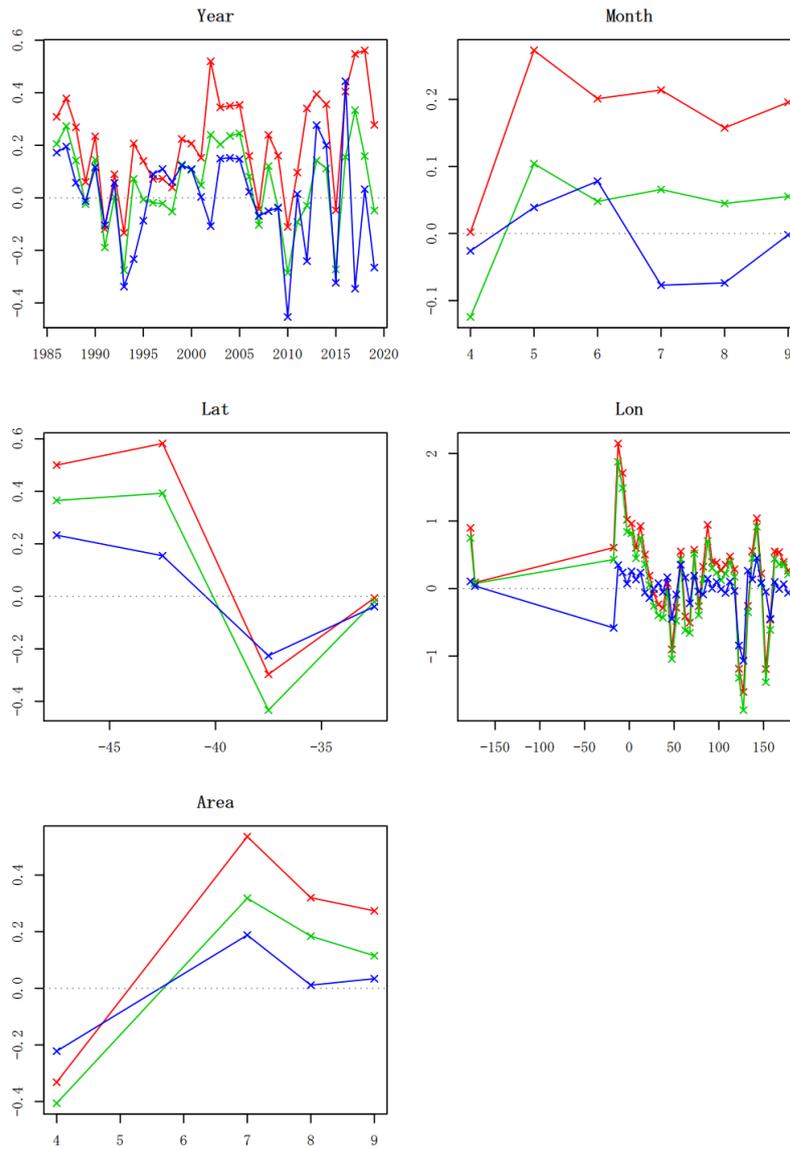


Fig. 9. Residuals by one variable

Red is GLM, green is GLMM, and blue is GAM.

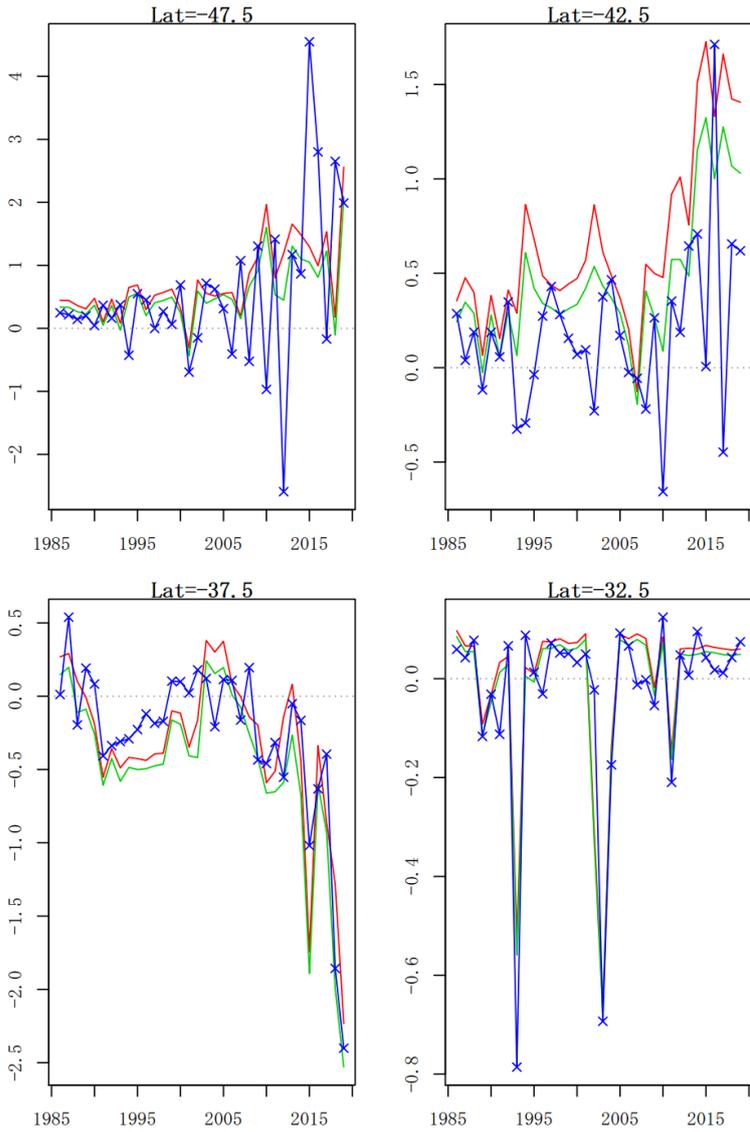


Fig. 10. Residuals by two variables (Year, Latitude)  
Red is GLM, green is GLMM, and blue is GAM.

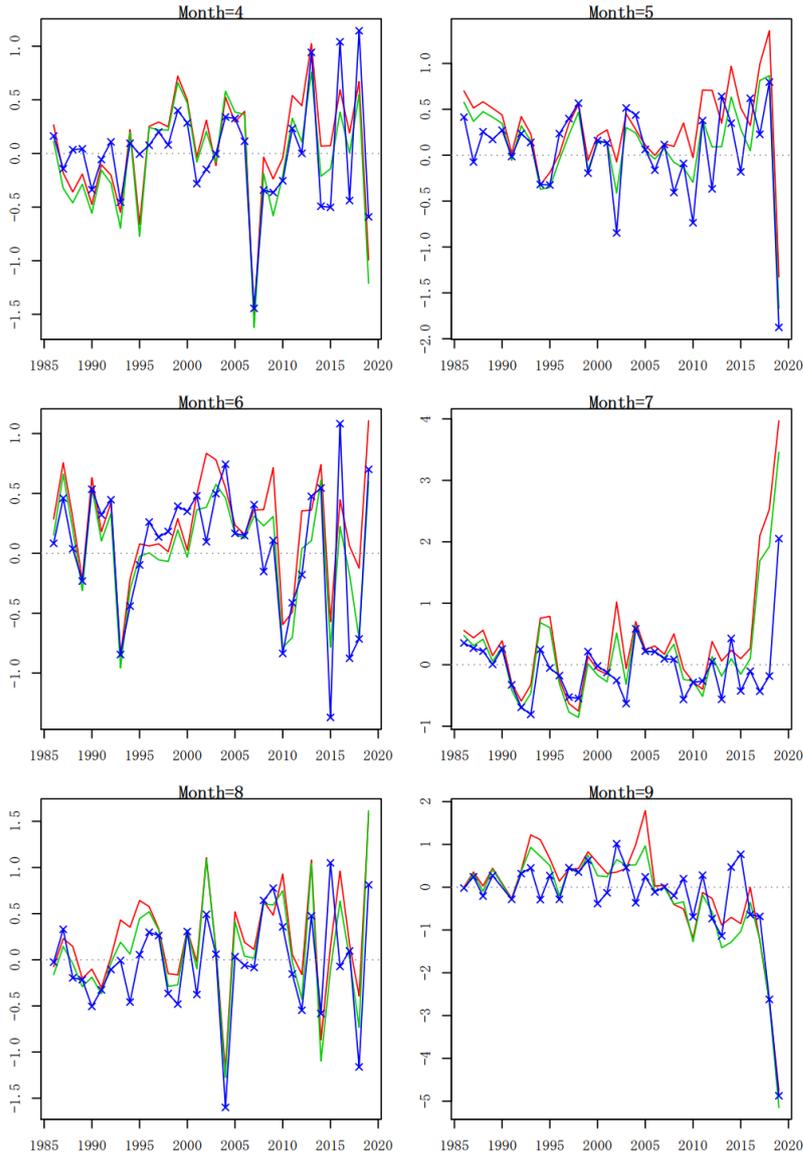


Fig. 11. Residuals by two variables (Year, Month)  
Red is GLM, green is GLMM, and blue is GAM.

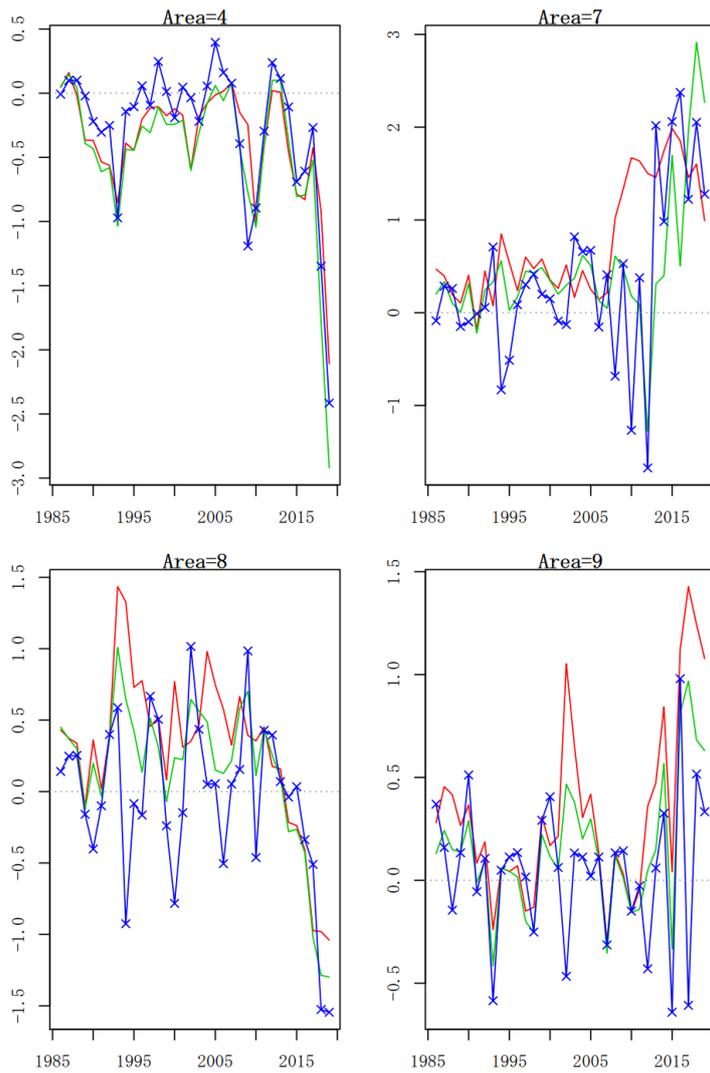


Fig. 12. Residuals by two variables (Year, Area)  
Red is GLM, green is GLMM, and blue is GAM.

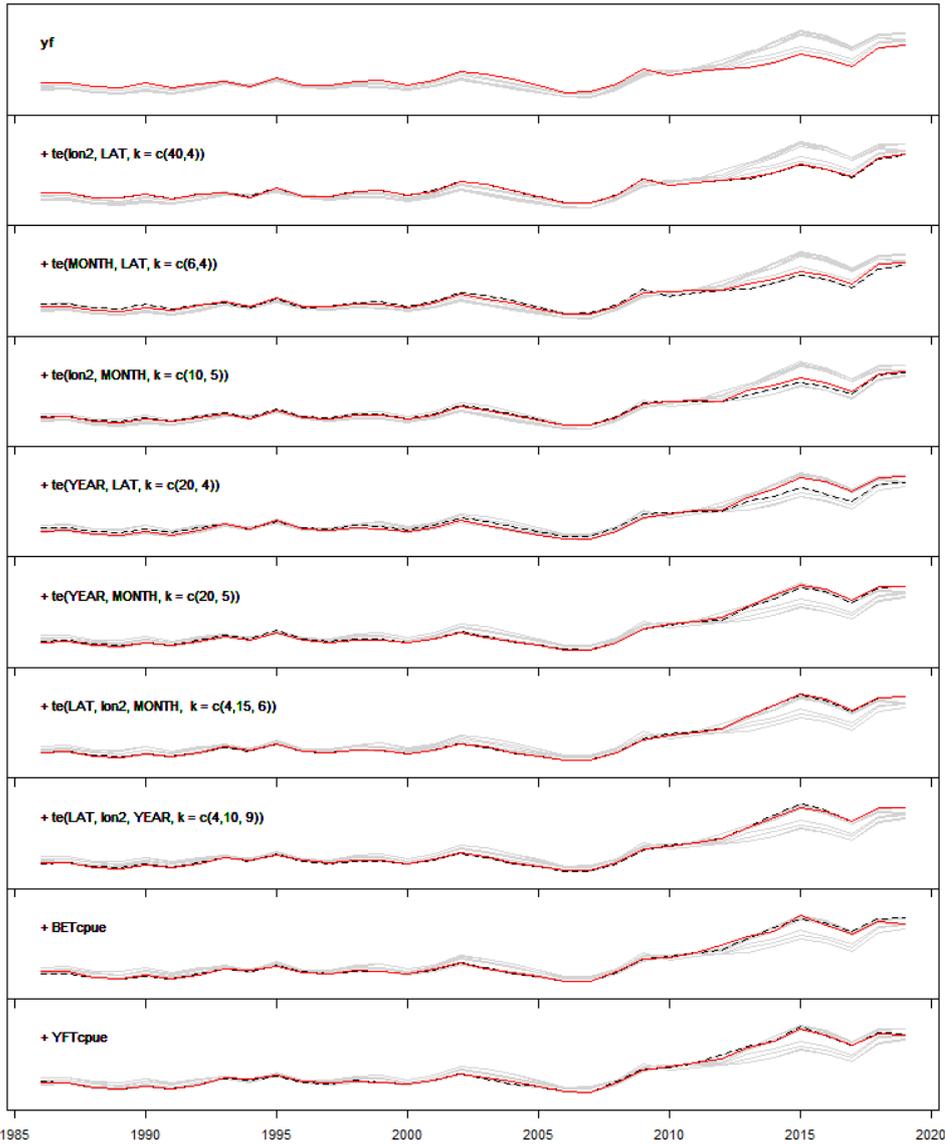


Fig. 13. Progress plot of GAM

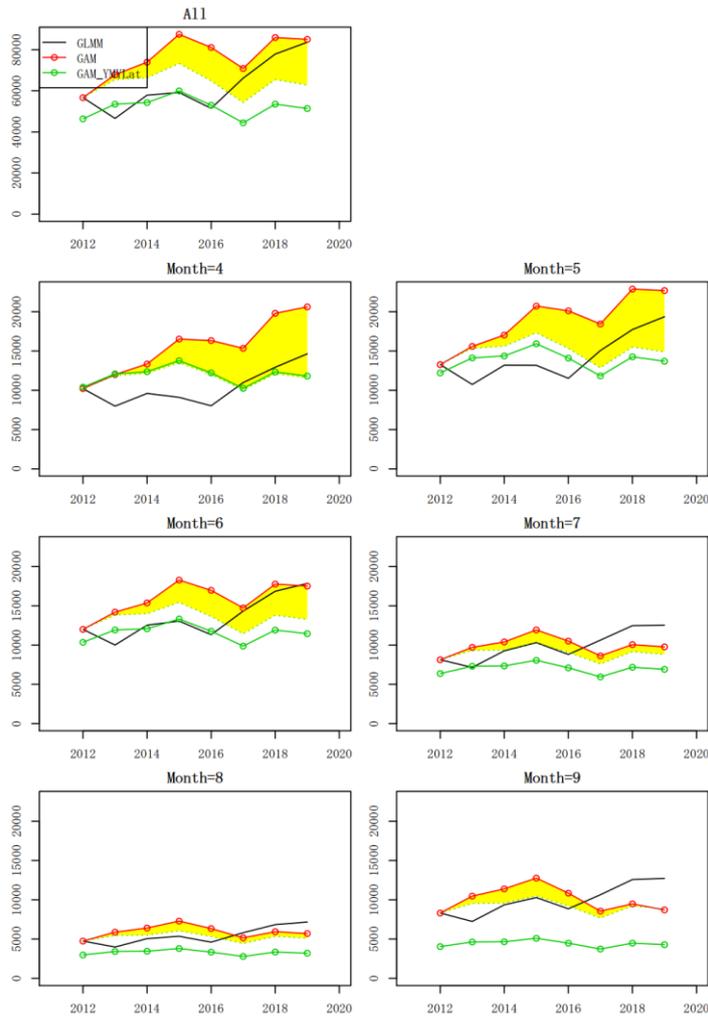


Fig. 14. Comparison of CS abundance index among models by month in Area 7.

X is year, and Y is CS abundance value. The top panel is all months, and the others are by month. In the monthly chart, the Y-axis range is unified. Red is GAM11. Black is for GLMM and adjusted with the value of 2012 GAM11. Green (GAM\_YMYLat) is based on GAM11 except for the year/month and year/latitude terms, which had a strong effect on raising around 2015 in the step plot. The solid and dotted lines in green are the nominal value and adjusted with the 2012 GAM11 value, respectively. The yellow polygon is the area between the red line and the green dotted line, and indicates that the CS value of GAM11 is increased by the year/month term and the year/latitude term.

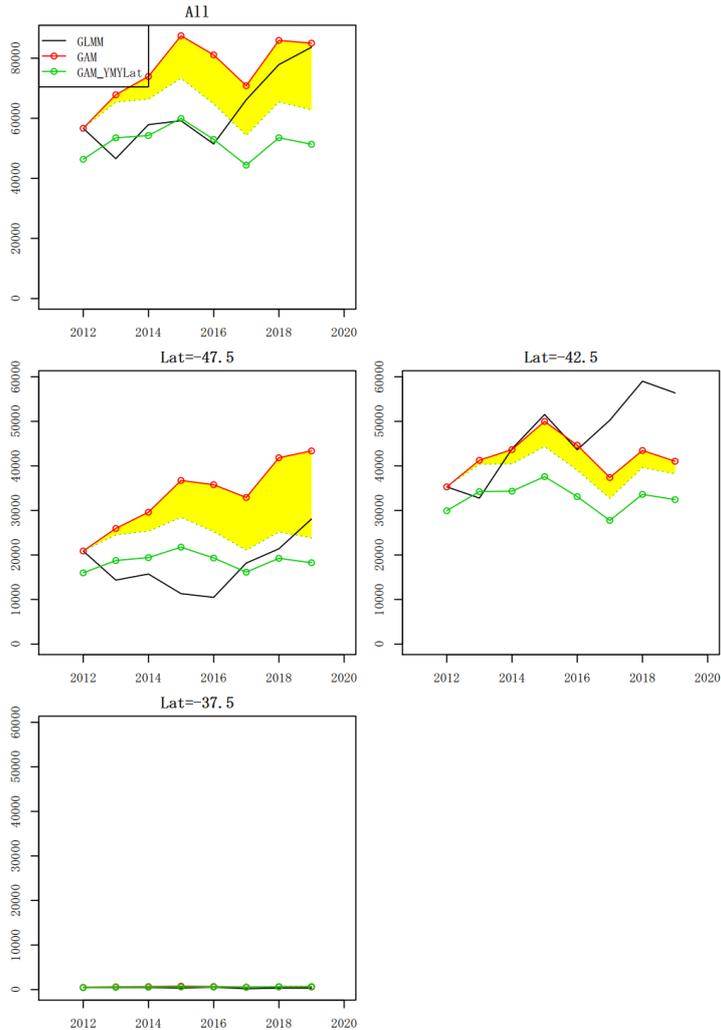


Fig. 14. (cont'd) (by Latitude in Area 7)

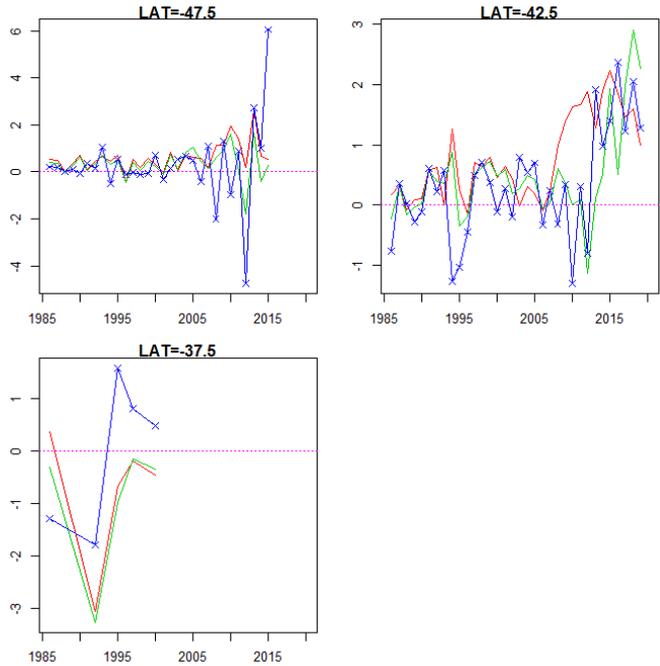


Fig. 15. Fitting by year and latitude in the data between April and June in Area 7.

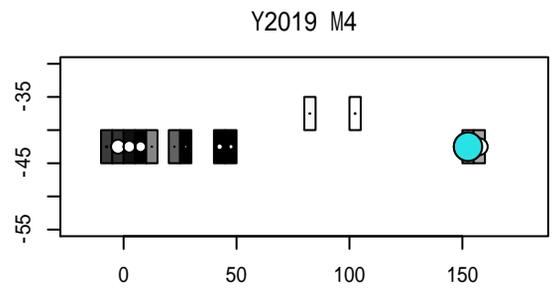
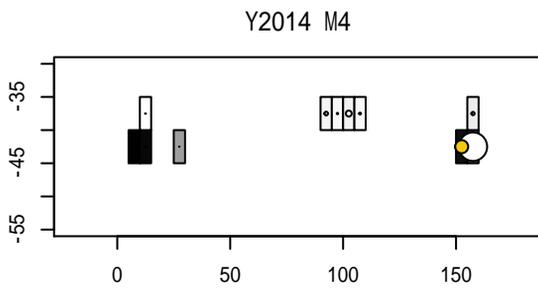
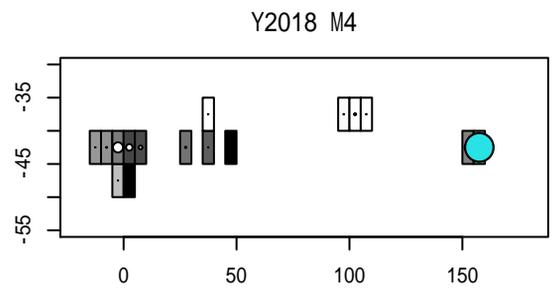
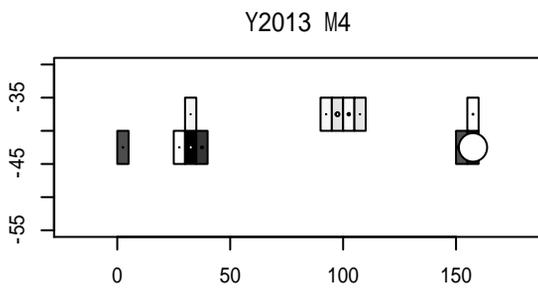
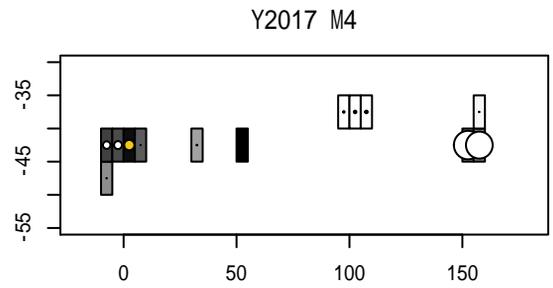
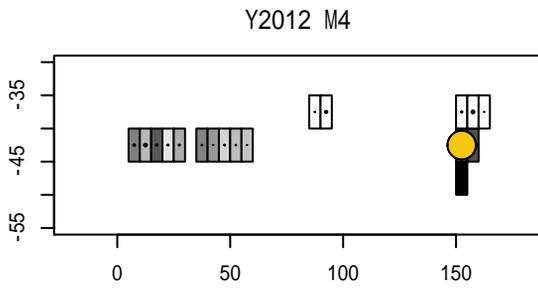
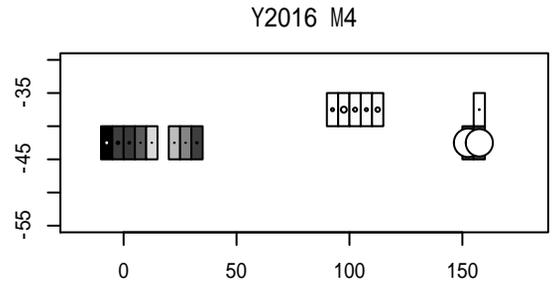
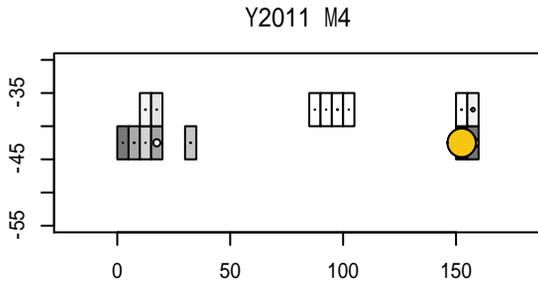
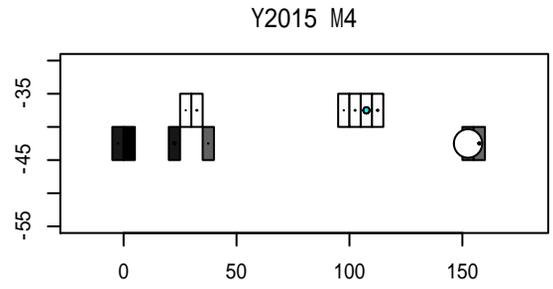
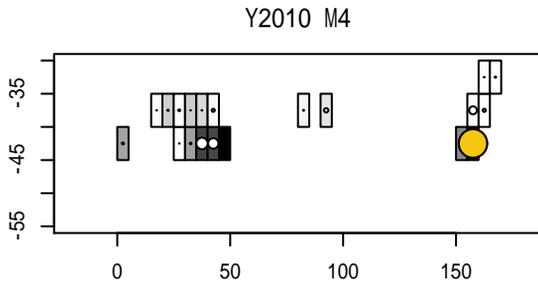
## **Attachments**

### **Examination of the abundance index for southern bluefin tuna calculated through GAM CPUE standardization**

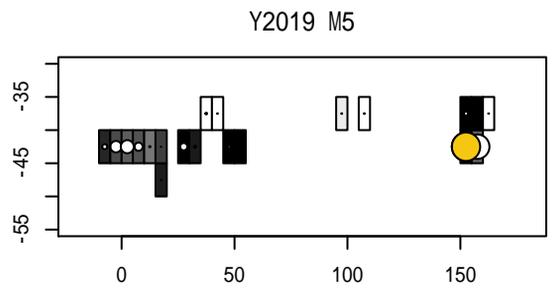
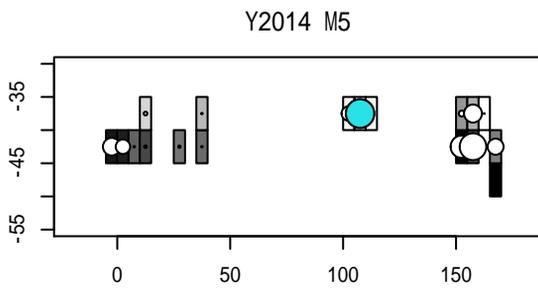
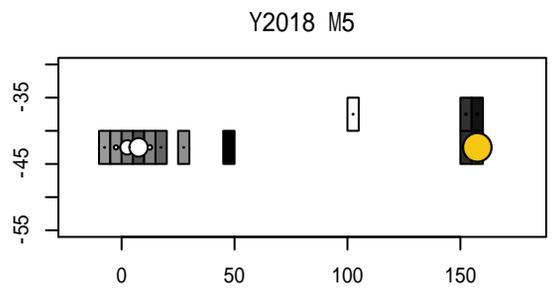
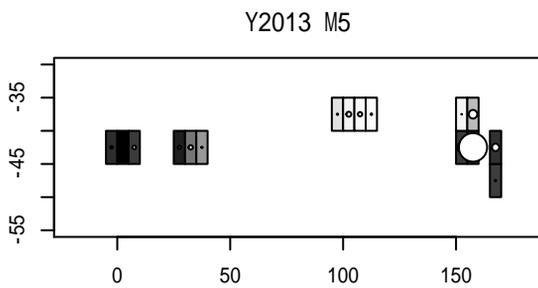
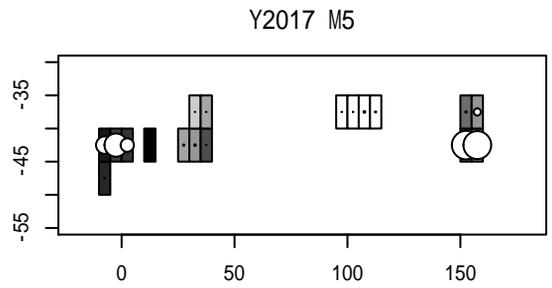
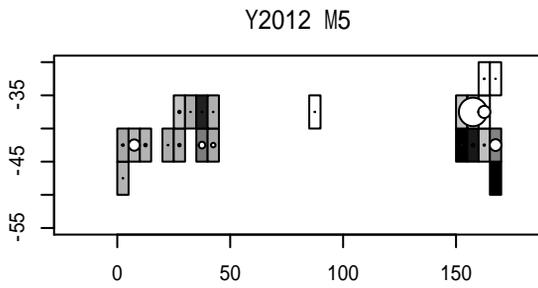
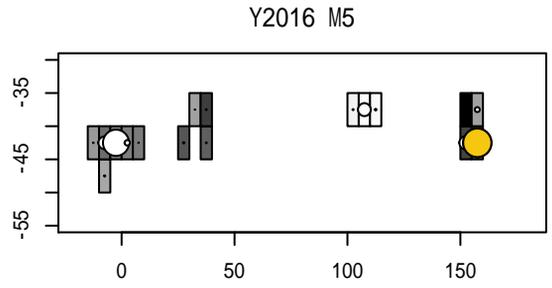
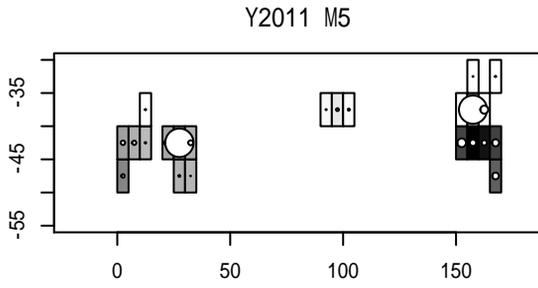
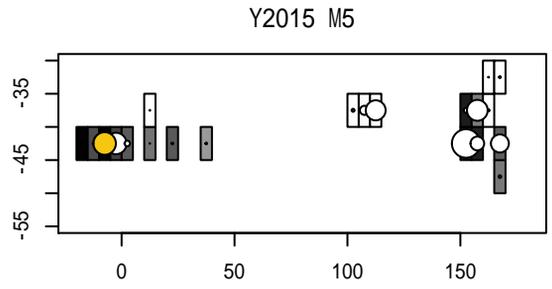
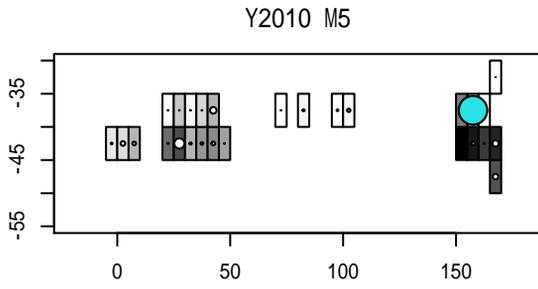
Tomoyuki ITOH

#### **Attachment 1. Mapping of CPUE and hooks**

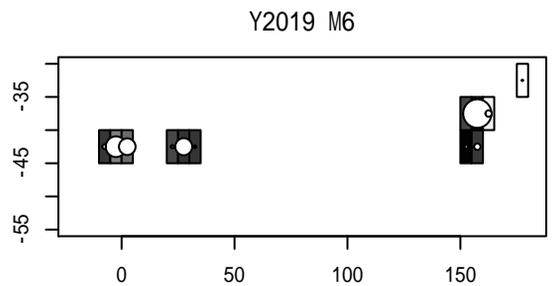
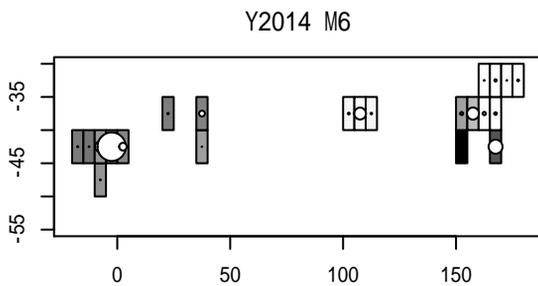
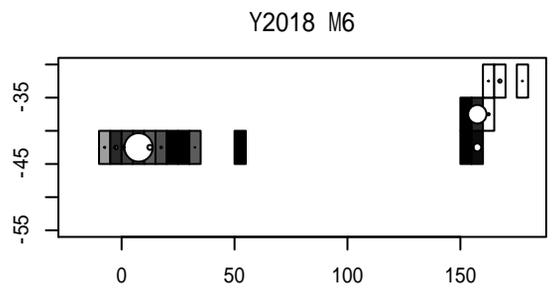
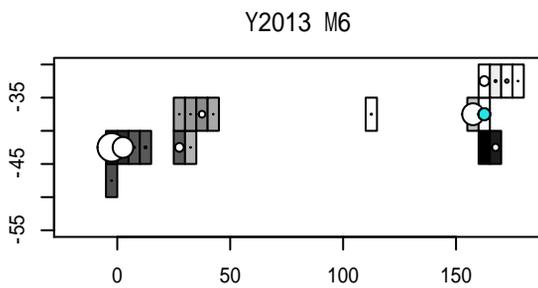
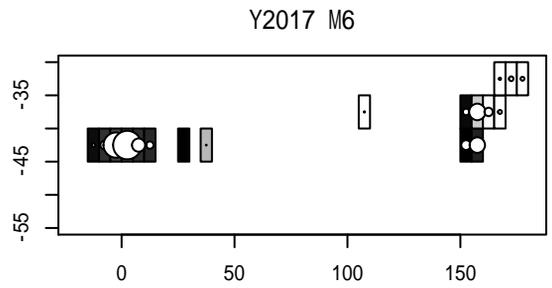
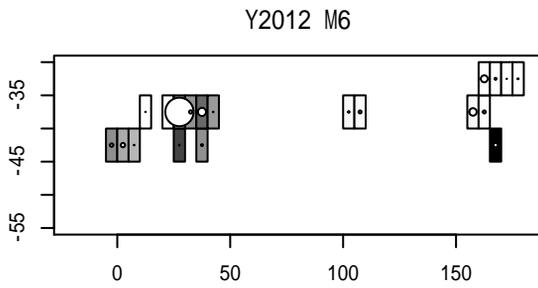
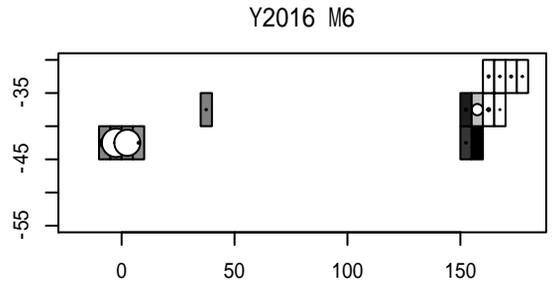
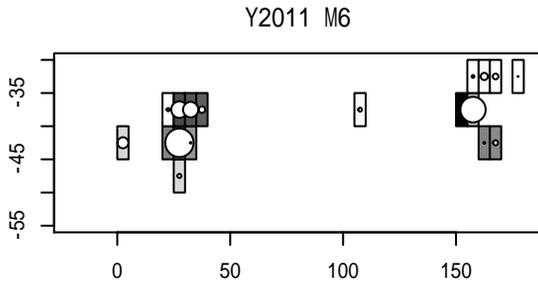
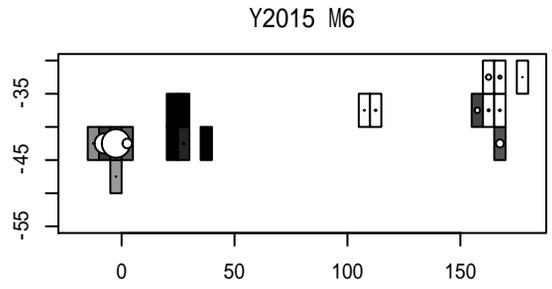
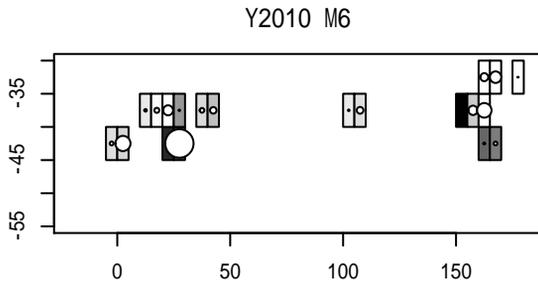
CPUE and hook in May are plotted by year, month, latitude in 5 degree, and longitude in 5 degree. Longitude is on the X-axis and latitude is on the Y-axis. The color of the section is CPUE (white = 0, black = maximum CPUE for the year/month). Hook is a relative value to the maximum hook of the year and month. Hook in yellow indicates maximum score, light blue indicates minimum score, and white indicates intermediate score. Score is the influence that changes Nominal CPUE, see text.



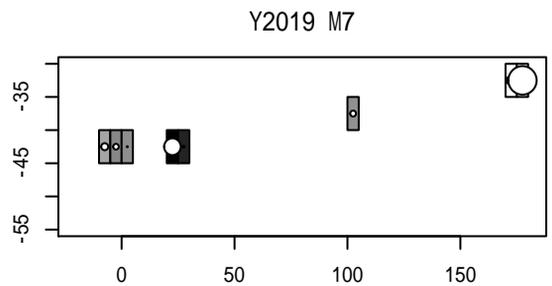
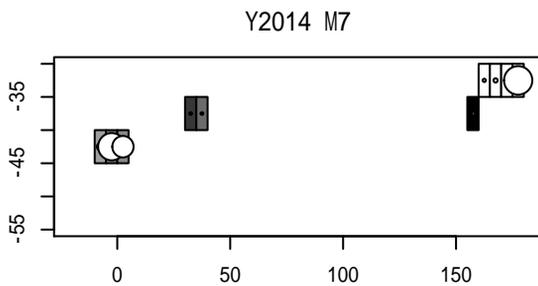
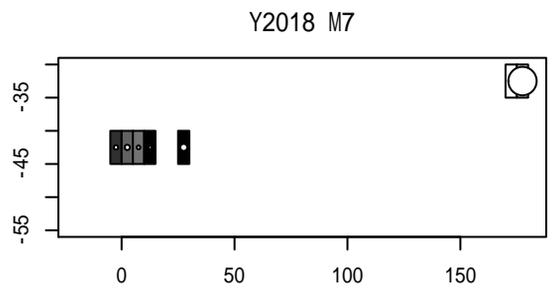
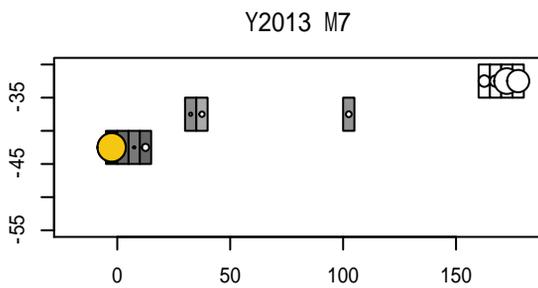
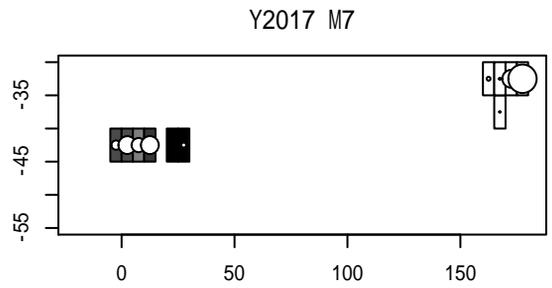
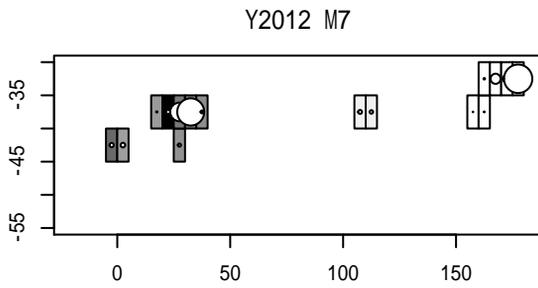
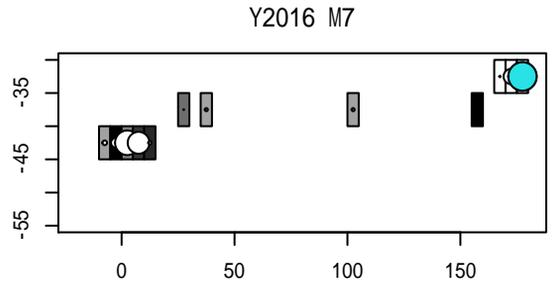
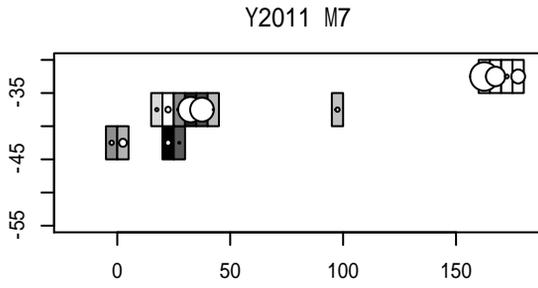
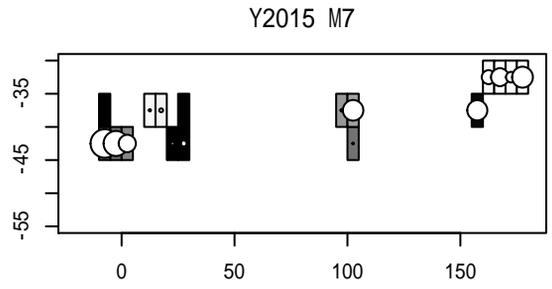
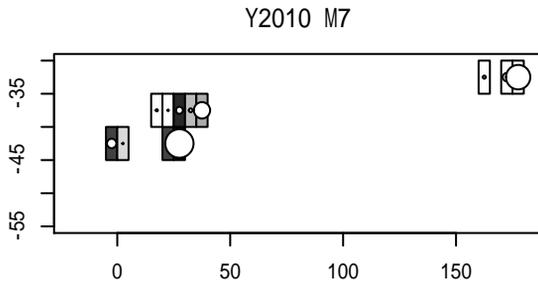
Back ground is CPUE. Black is the highest CPUE of the year.  
 Circle denotes the effect to change Nominal CPUE.  
 Yellow and lightblue is the highest and lowest of the year, respectively.



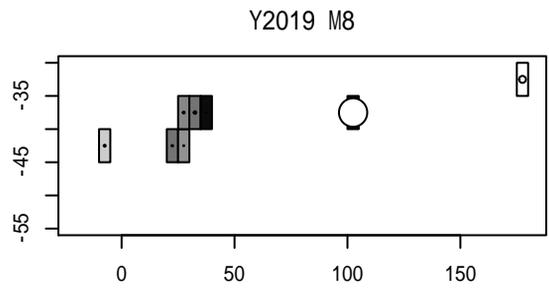
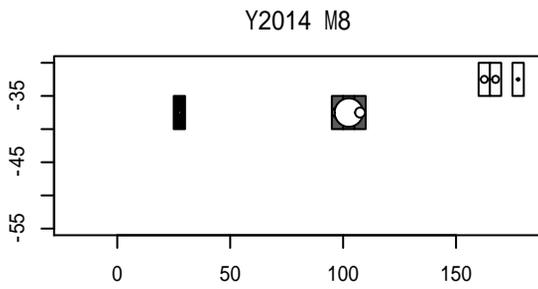
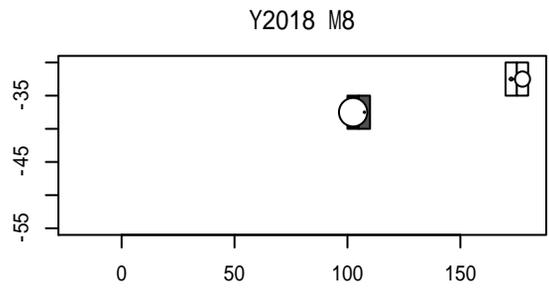
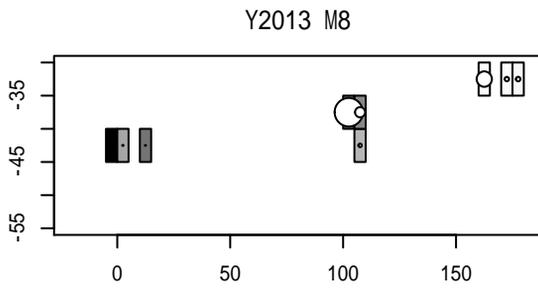
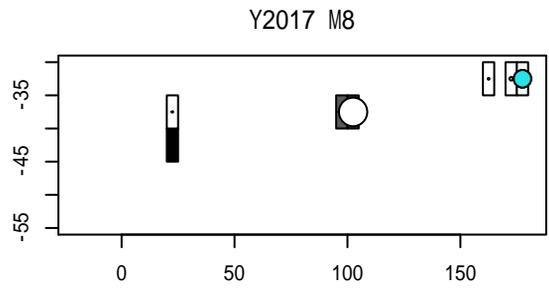
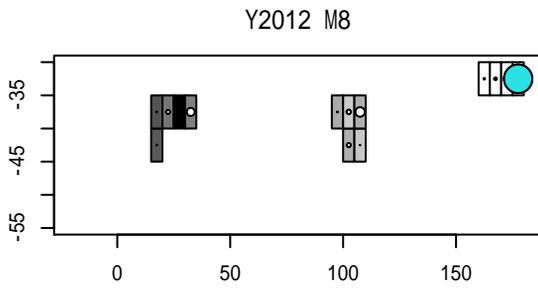
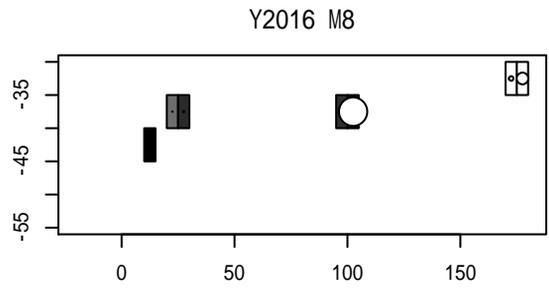
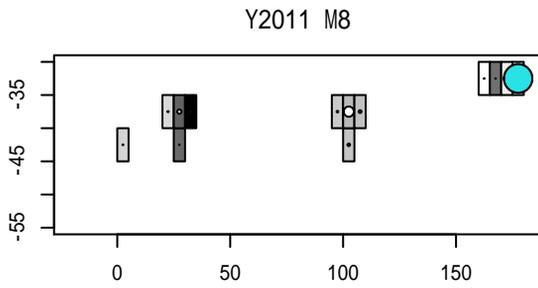
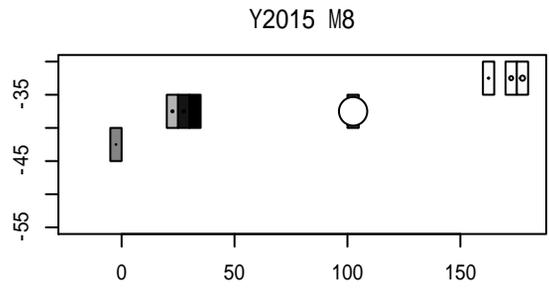
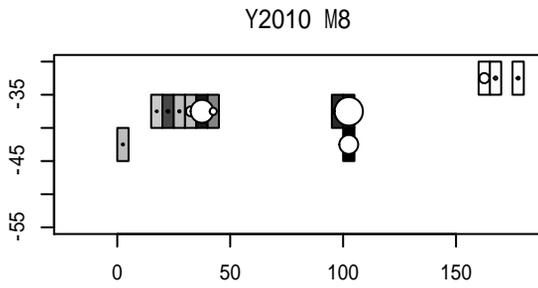
Back ground is CPUE. Black is the highest CPUE of the year.  
 Circle denotes the effect to change Nominal CPUE.  
 Yellow and lightblue is the highest and lowest of the year, respectively.



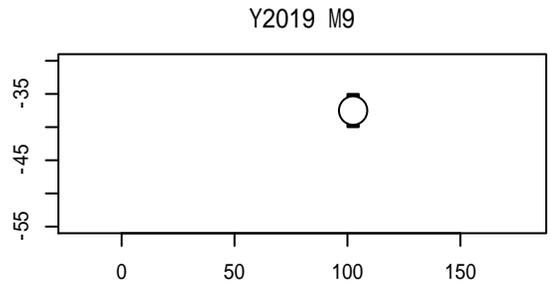
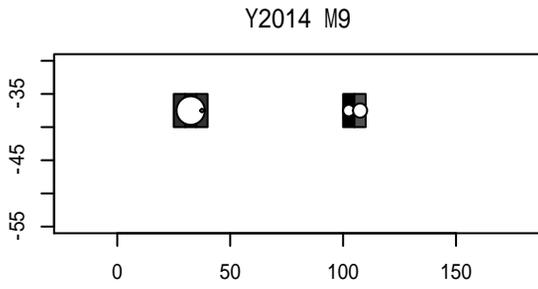
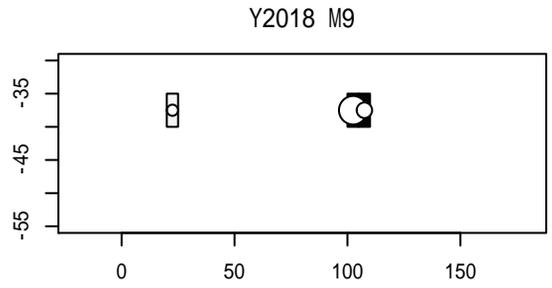
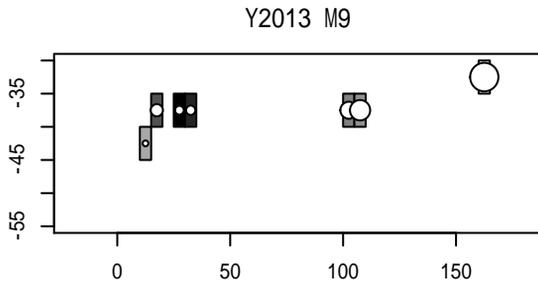
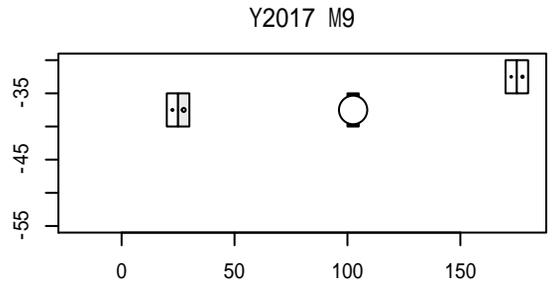
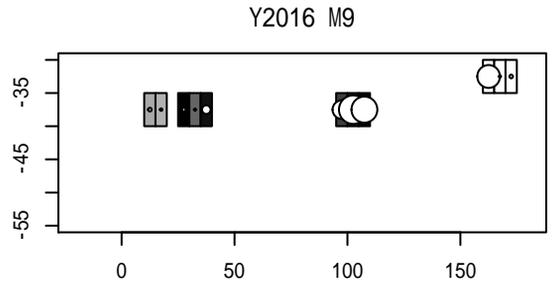
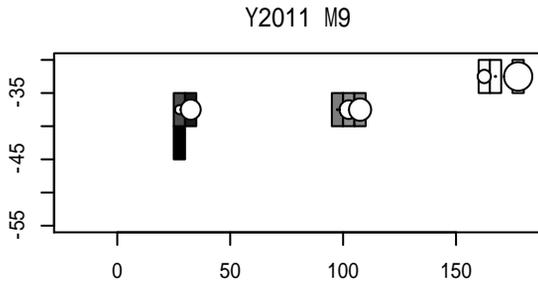
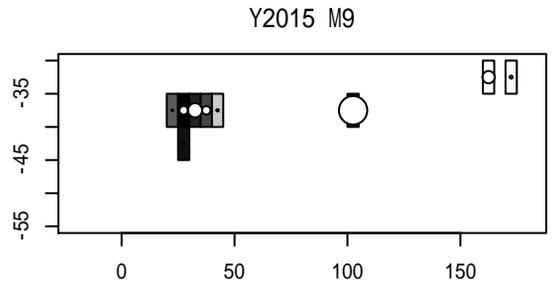
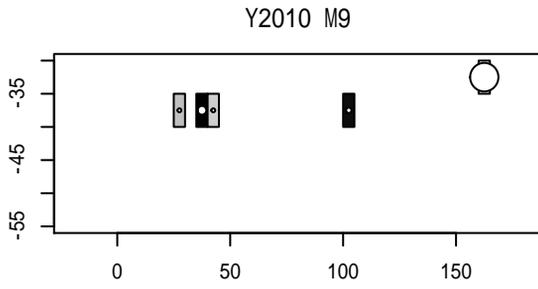
Back ground is CPUE. Black is the highest CPUE of the year.  
 Circle denotes the effect to change Nominal CPUE.  
 Yellow and lightblue is the highest and lowest of the year, respectively.



Back ground is CPUE. Black is the highest CPUE of the year.  
 Circle denotes the effect to change Nominal CPUE.  
 Yellow and lightblue is the highest and lowest of the year, respectively.



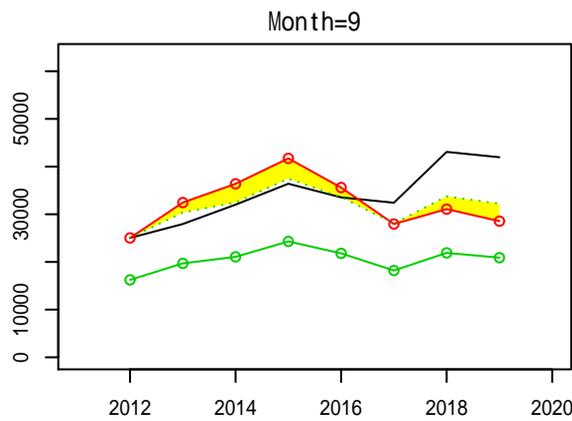
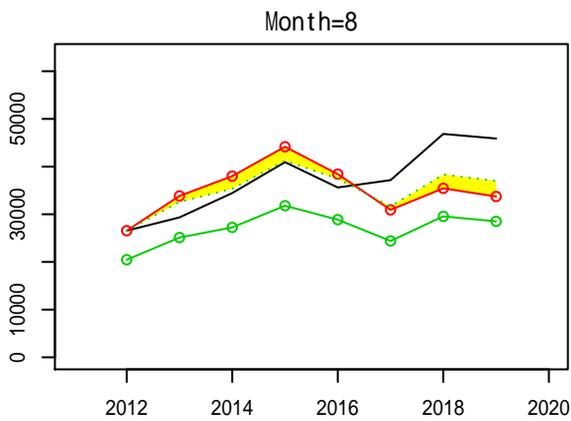
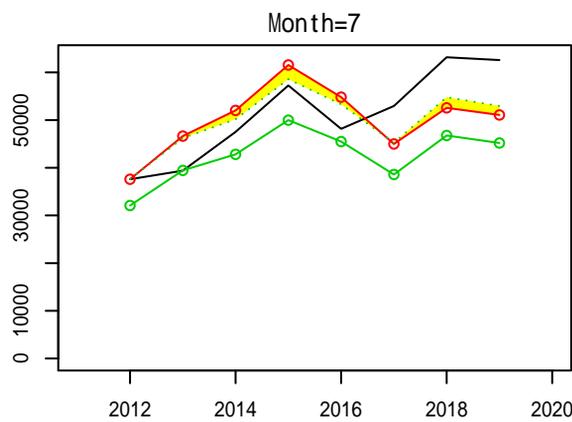
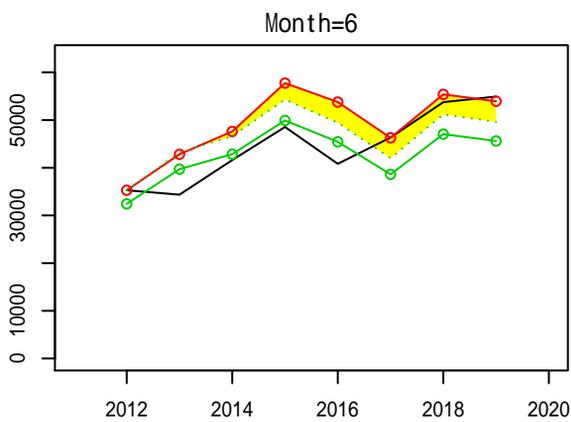
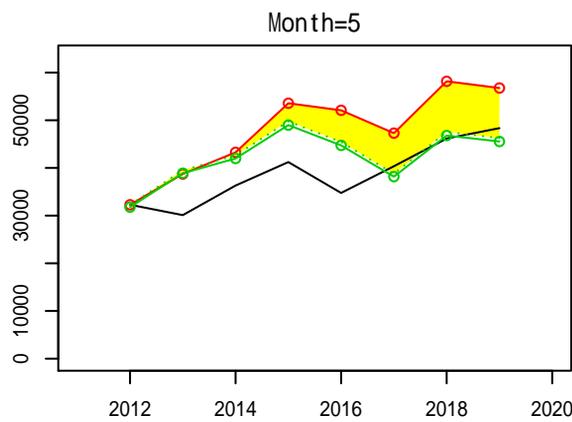
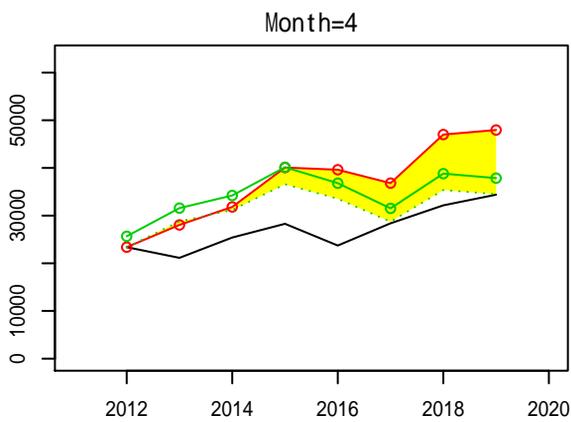
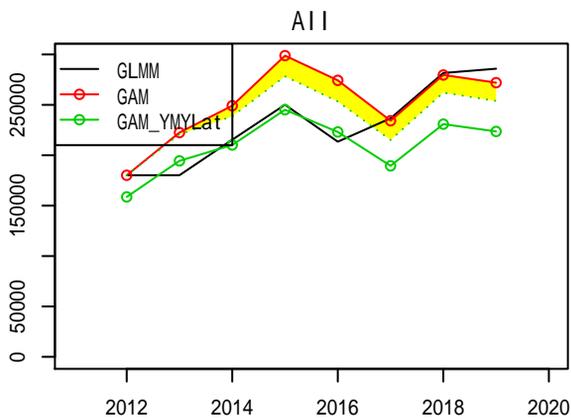
Back ground is CPUE. Black is the highest CPUE of the year.  
 Circle denotes the effect to change Nominal CPUE.  
 Yellow and lightblue is the highest and lowest of the year, respectively.

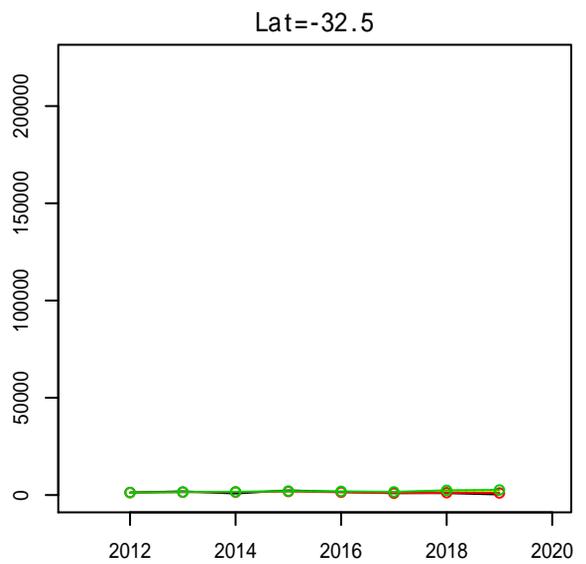
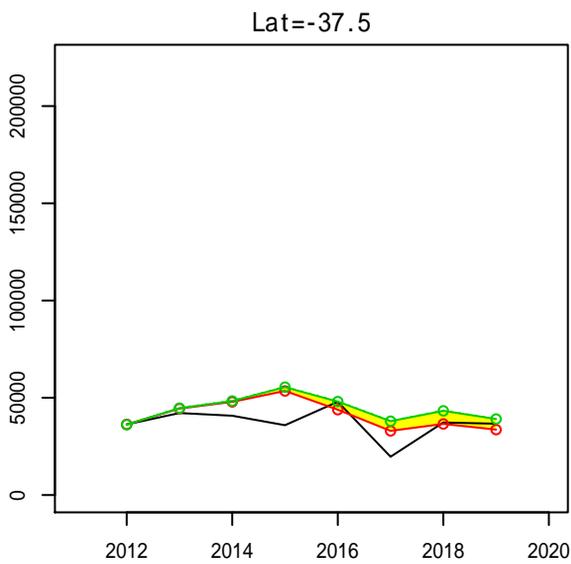
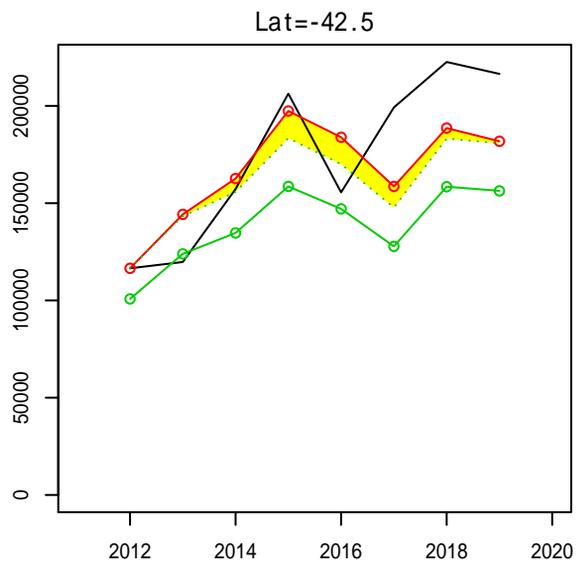
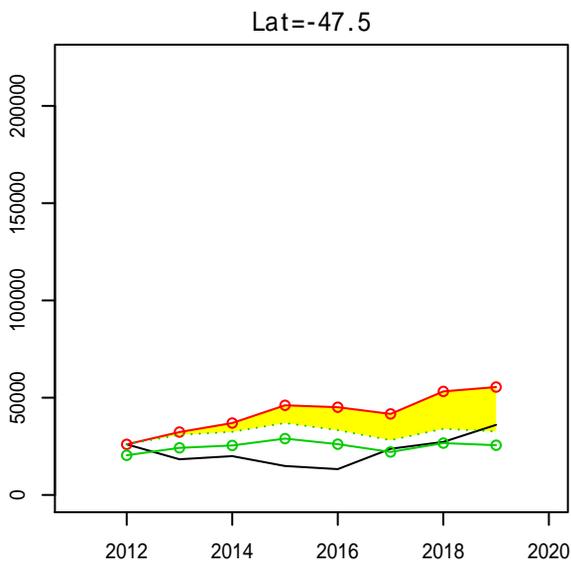
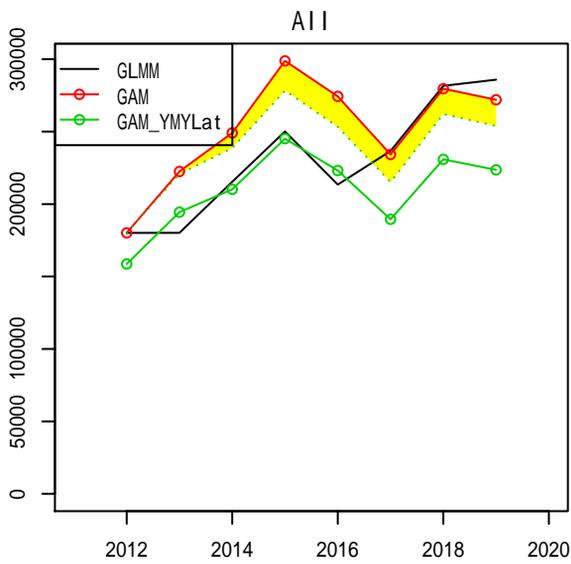


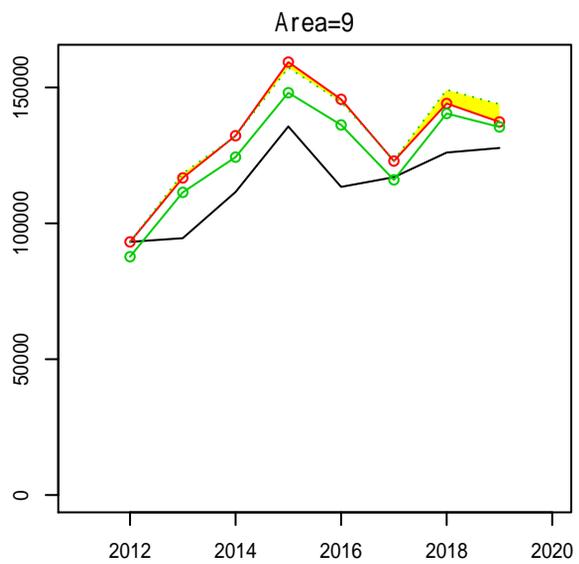
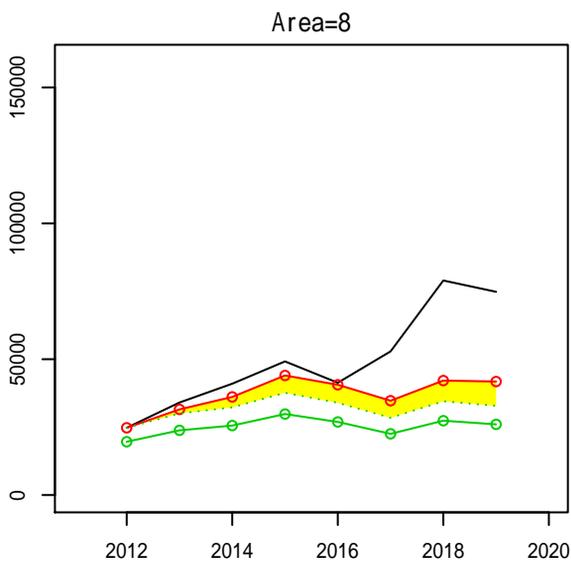
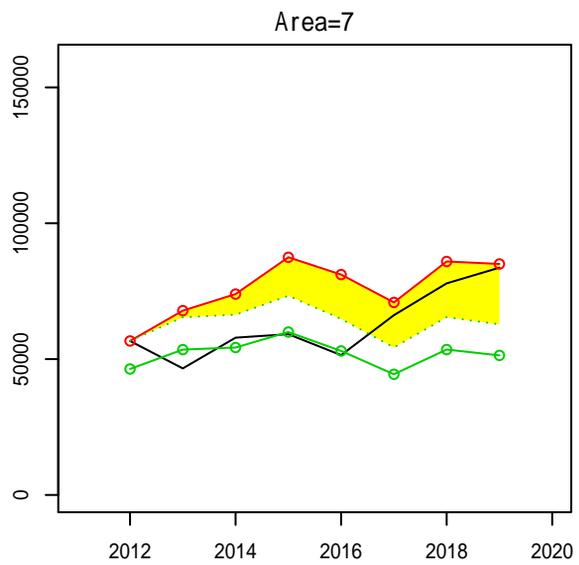
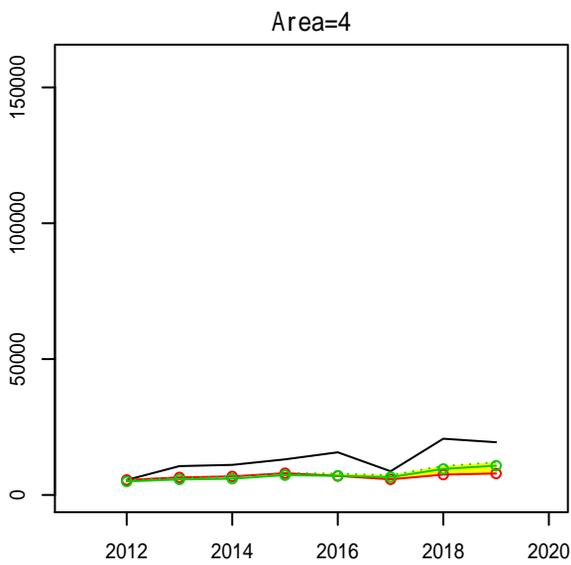
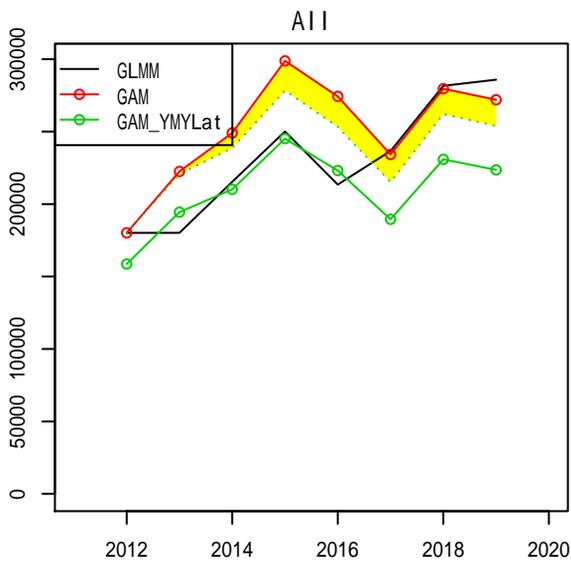
Back ground is CPUE. Black is the highest CPUE of the year.  
 Circle denotes the effect to change Nominal CPUE.  
 Yellow and lightblue is the highest and lowest of the year, respectively.

**Attachment 2. Comparison of CS abundance index among models by month**

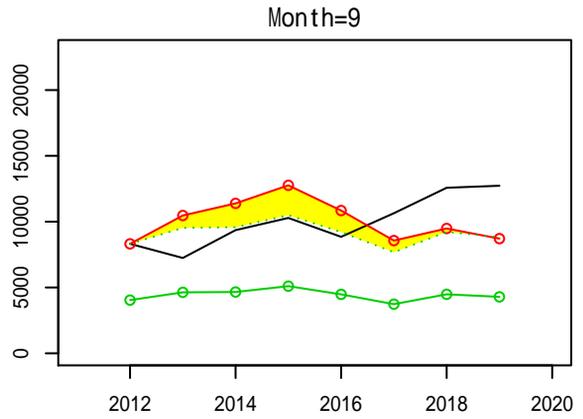
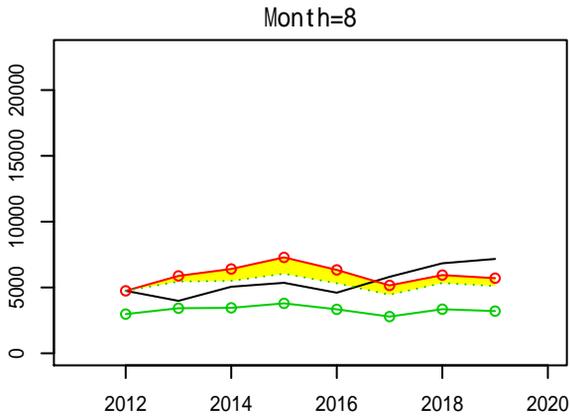
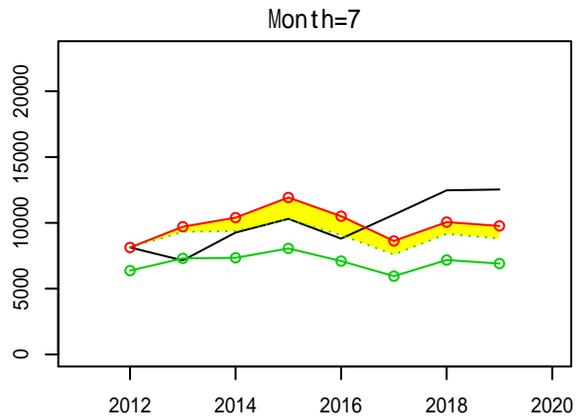
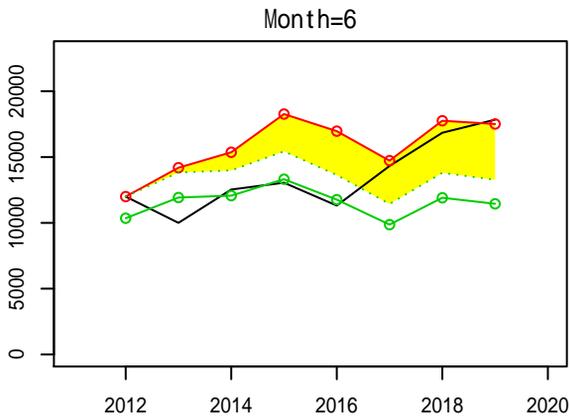
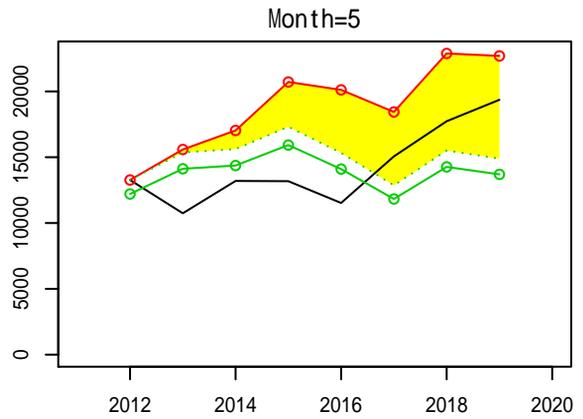
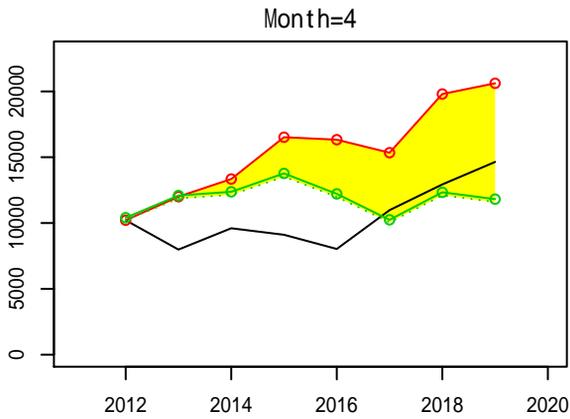
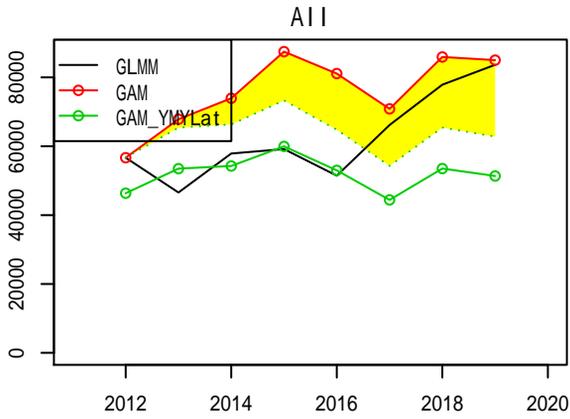
X is year, and Y is CS abundance value. The top panel is all months, and the others are by month. In the monthly chart, the Y-axis range is unified. Red is GAM11. Black is for GLMM and adjusted with the value of 2012 GAM11. Green (GAM\_YMYLat) is based on GAM11 except for the year/month and year/latitude terms, which had a strong effect on raising around 2015 in the step plot. The solid and dotted lines in green are the nominal value and adjusted with the 2012 GAM11 value, respectively. The yellow polygon is the area between the red line and the green dotted line, and indicates that the CS value of GAM11 is increased by the year/month term and the year/latitude term.



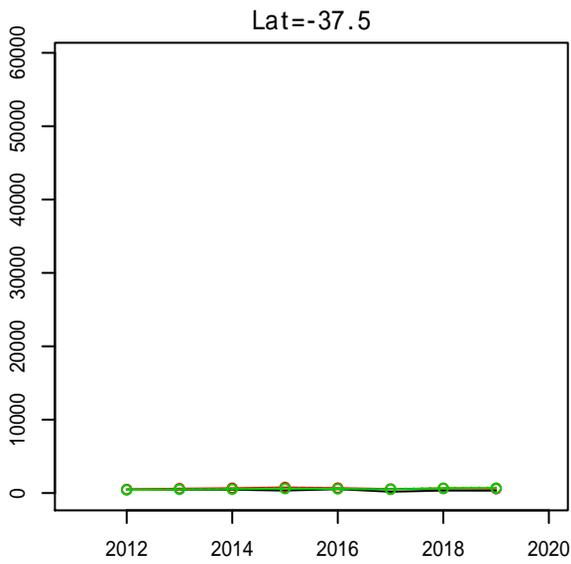
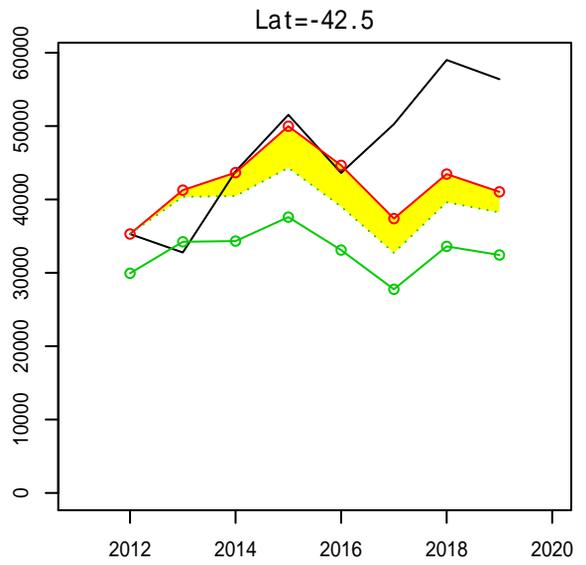
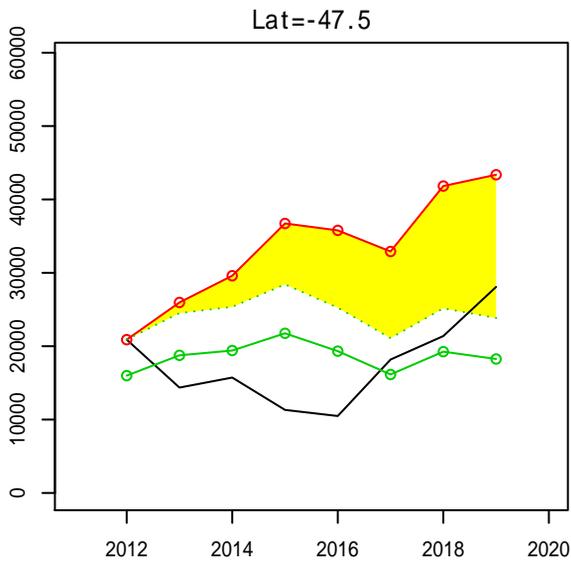
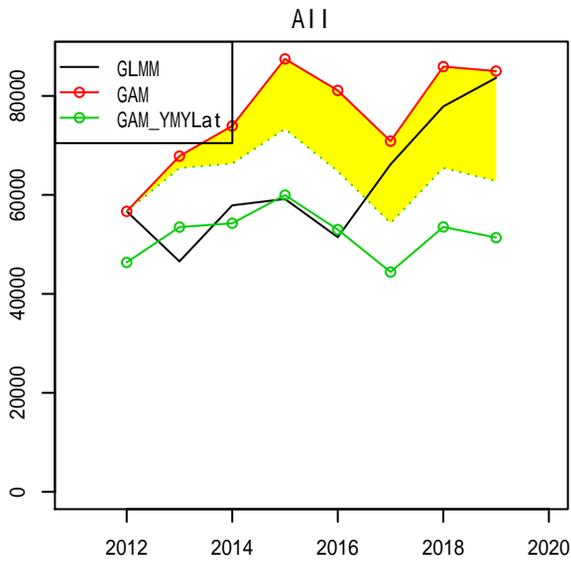




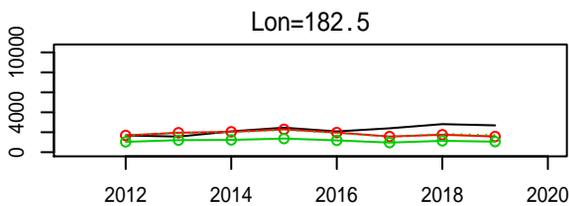
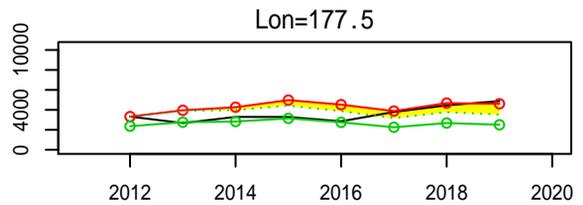
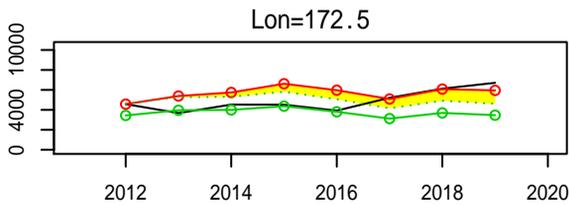
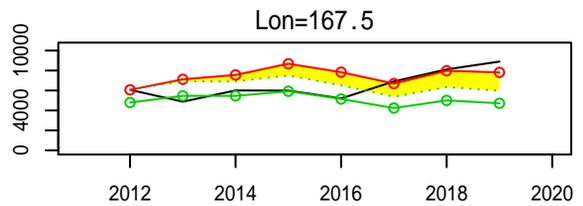
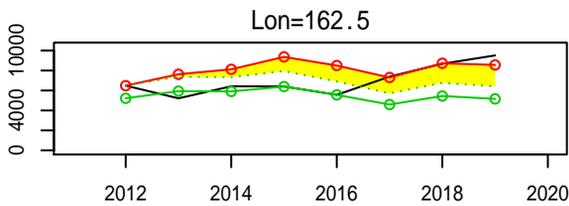
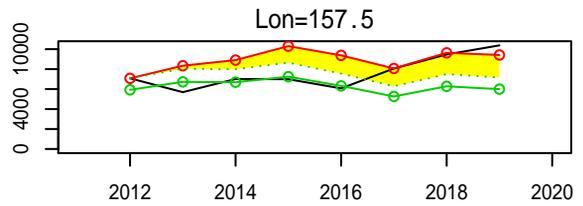
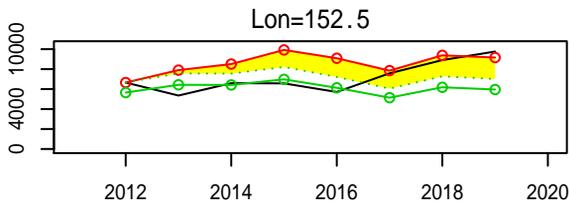
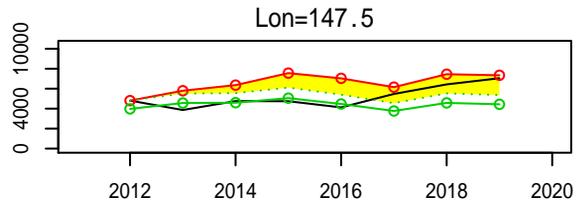
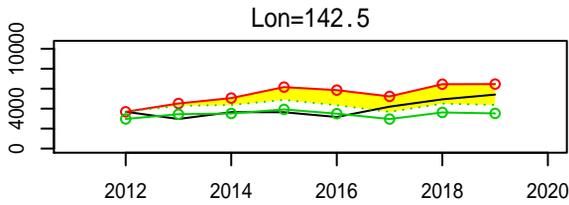
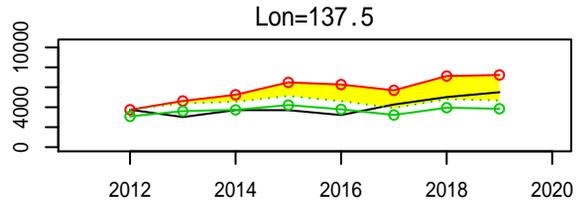
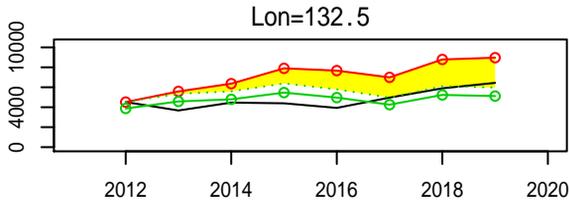
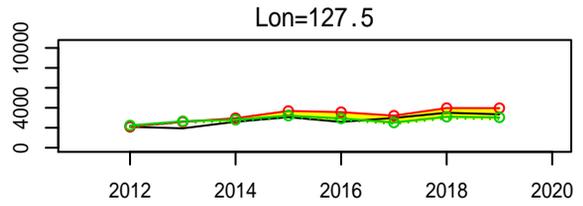
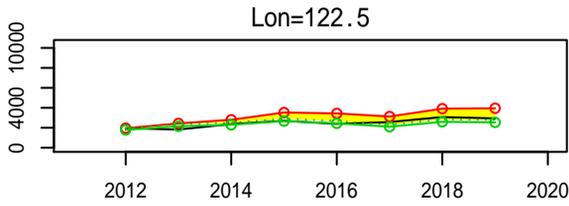
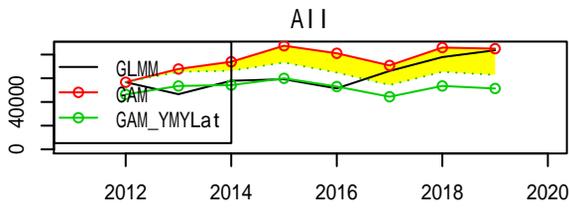
Area=7  
Month



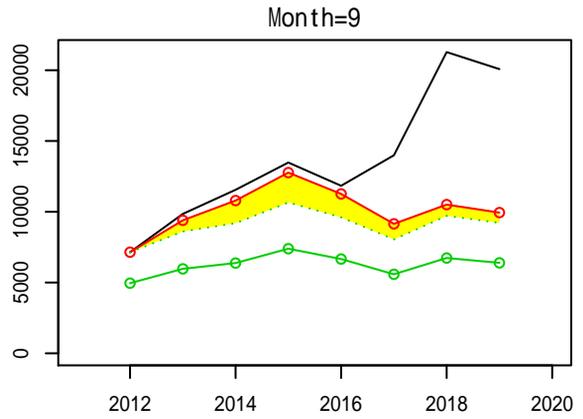
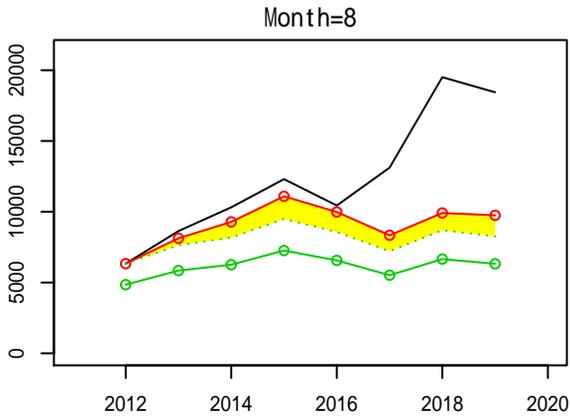
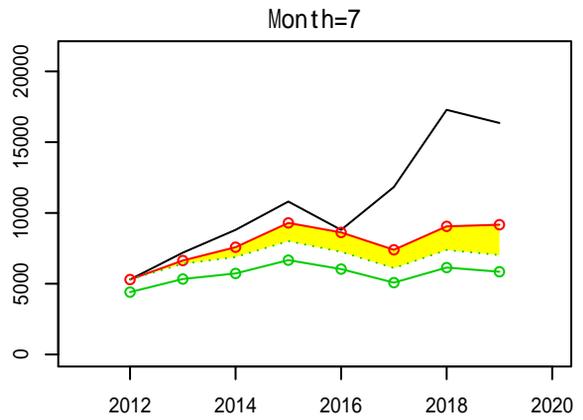
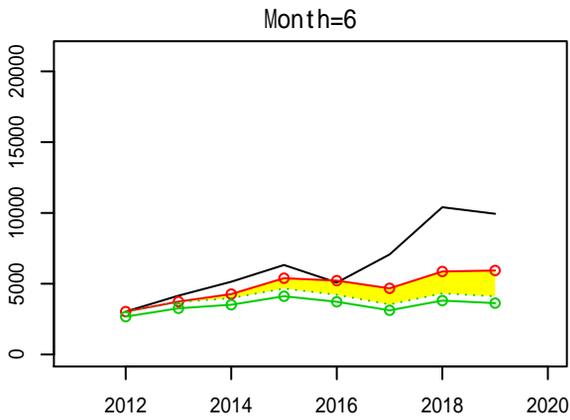
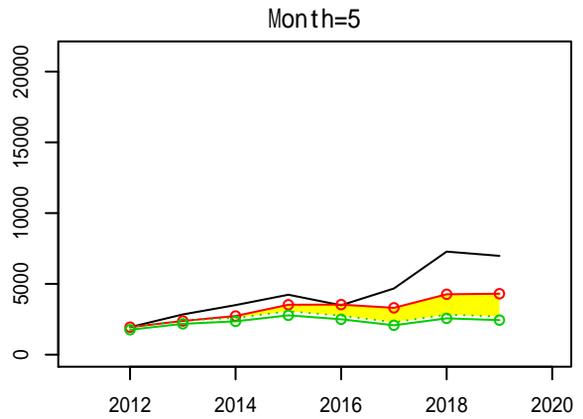
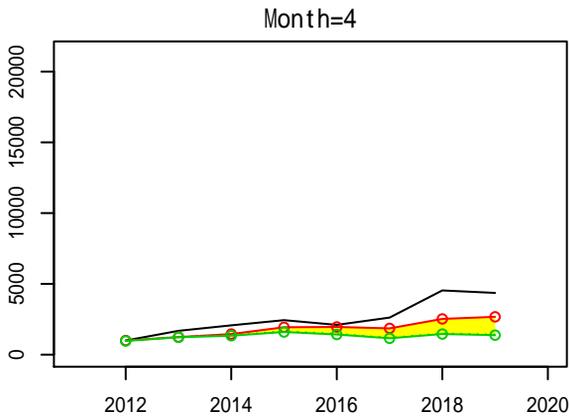
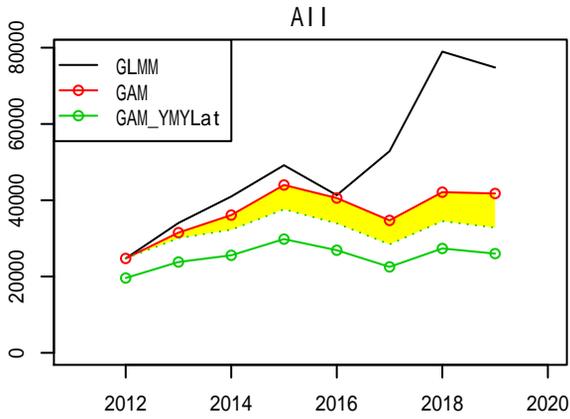
Area=7  
Lat



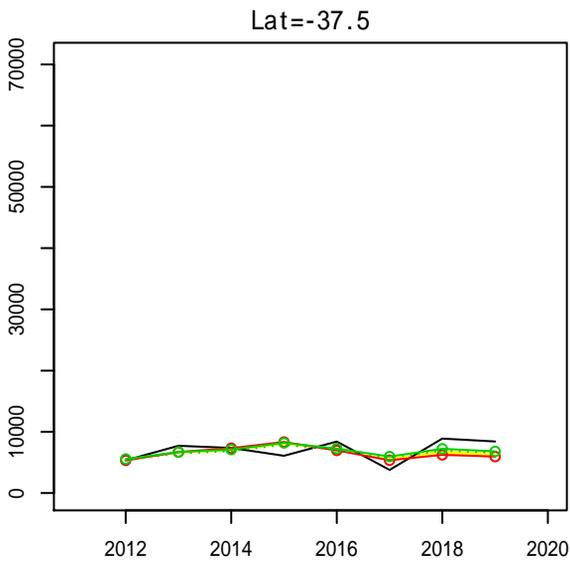
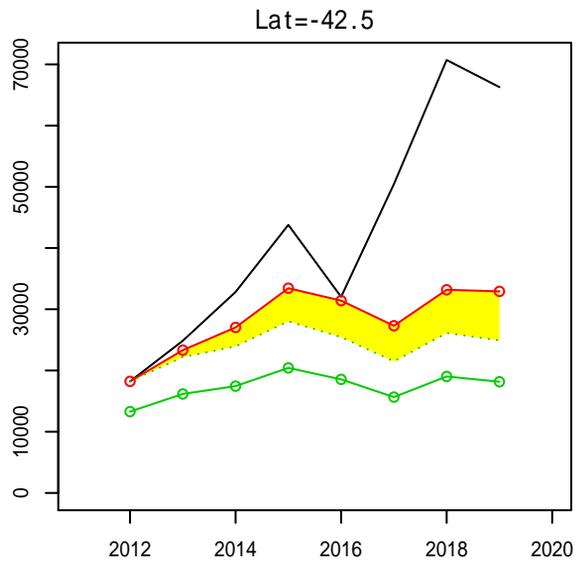
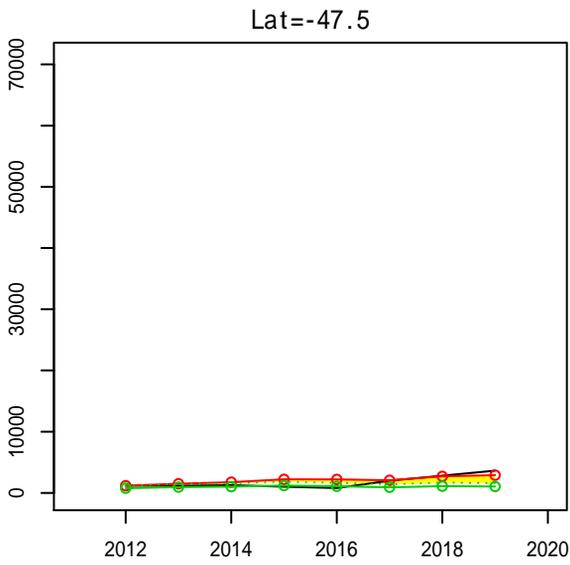
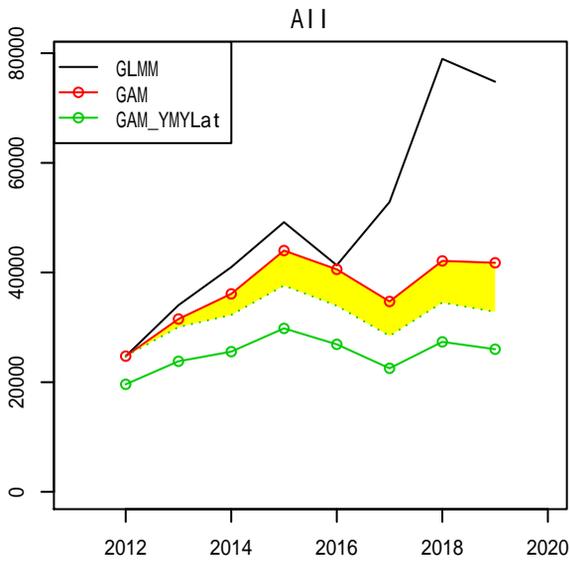
Area=7  
Lon



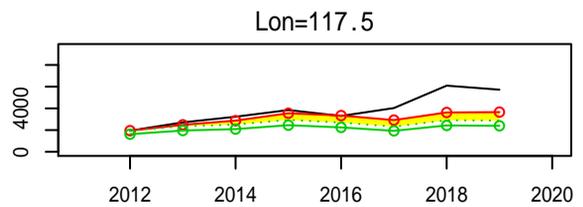
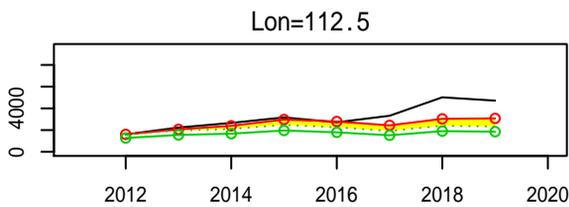
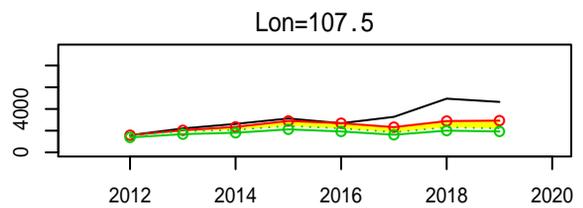
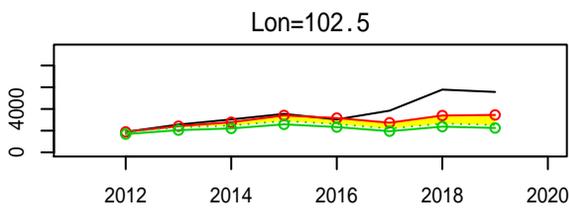
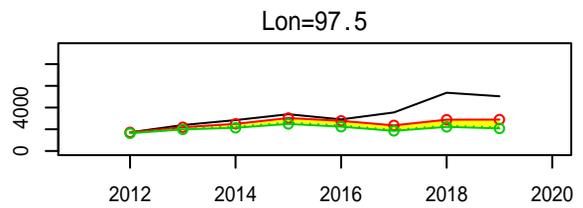
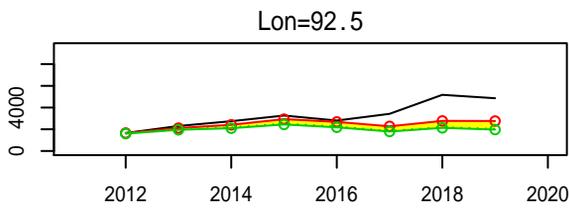
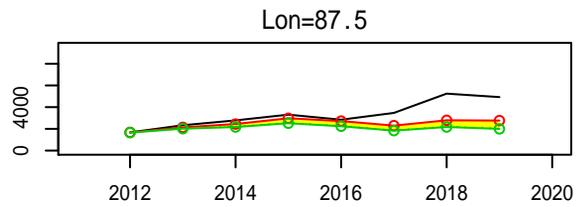
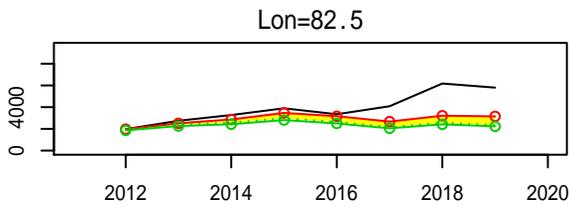
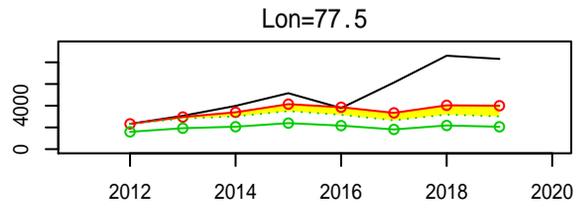
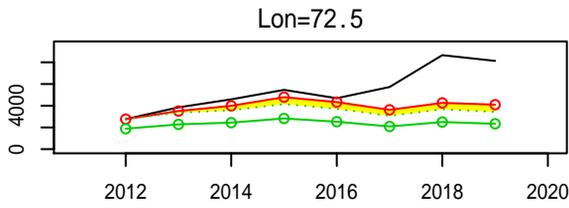
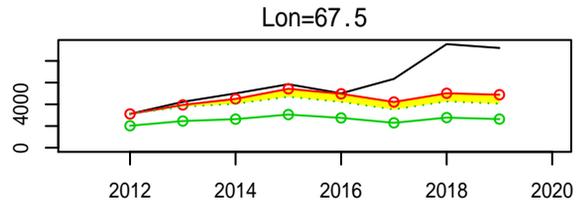
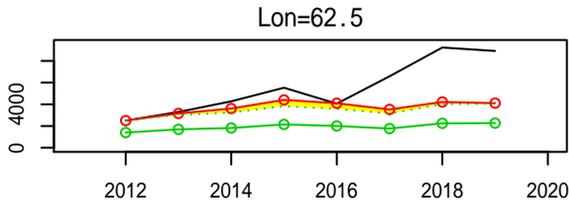
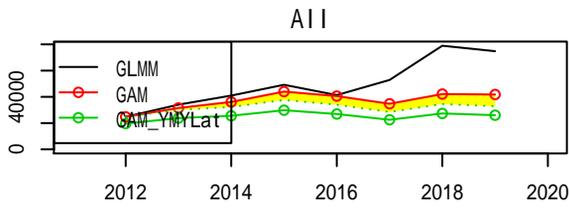
Area=8  
Month



Area=8  
Lat



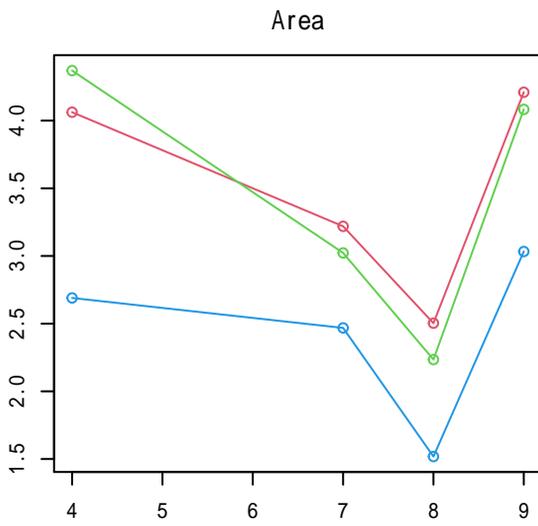
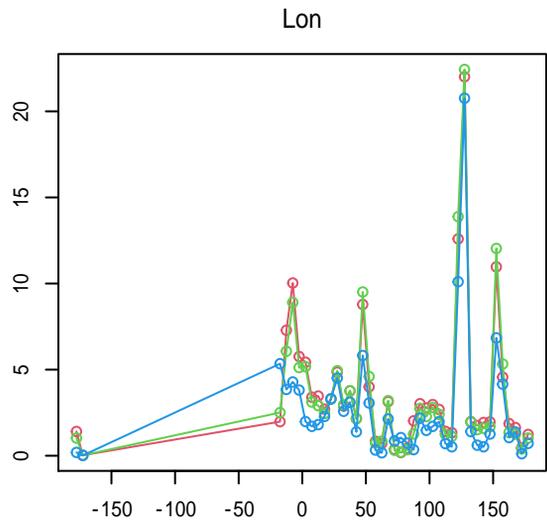
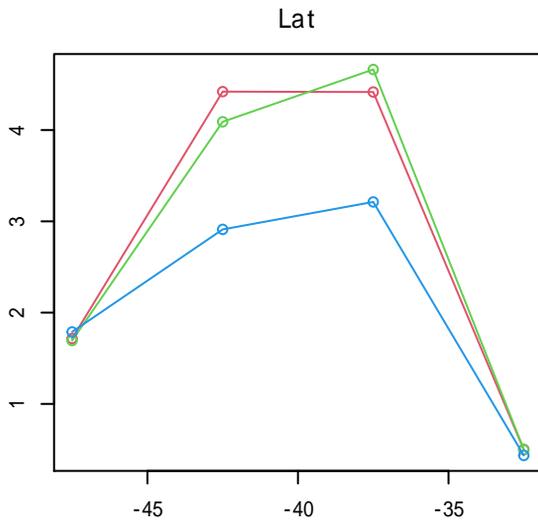
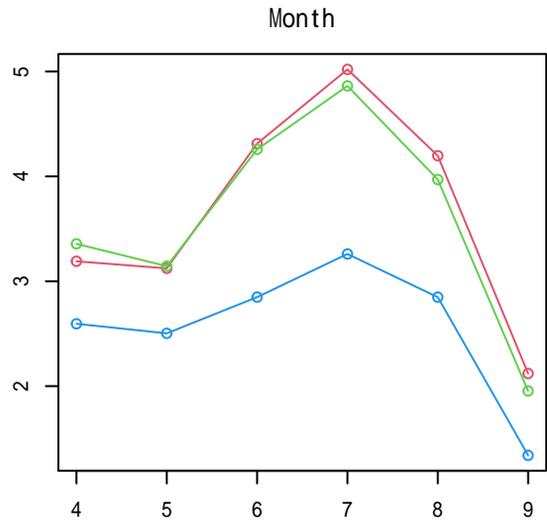
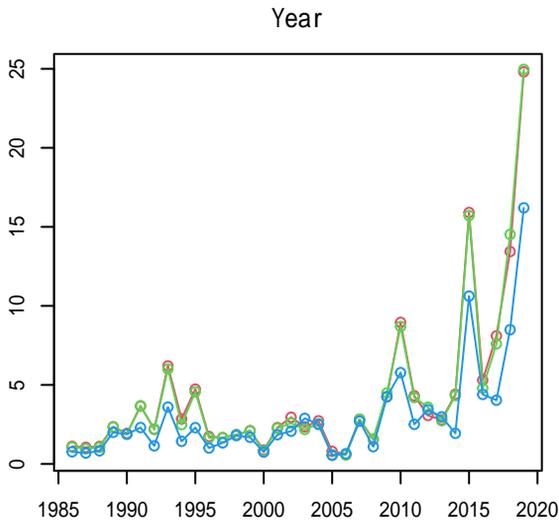
Area=8  
Lon



Attachment 3. Residuals by one or two variables

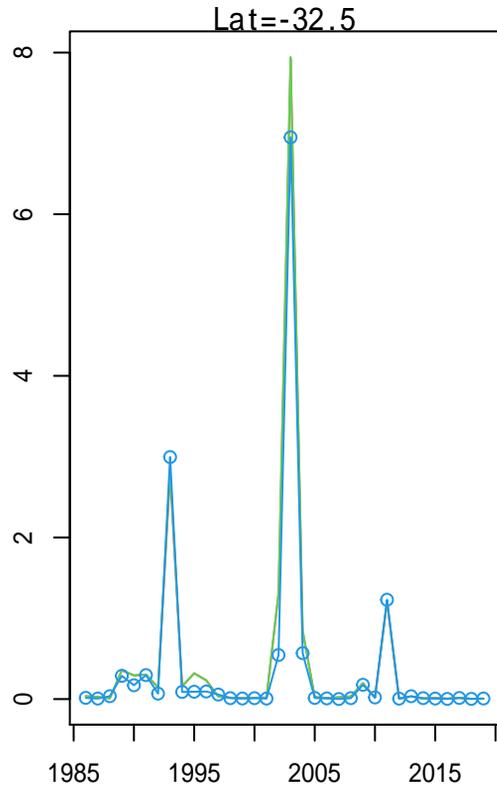
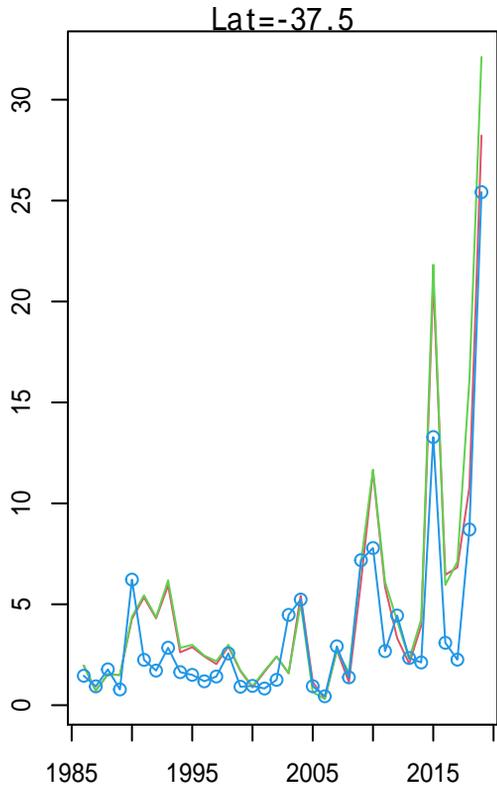
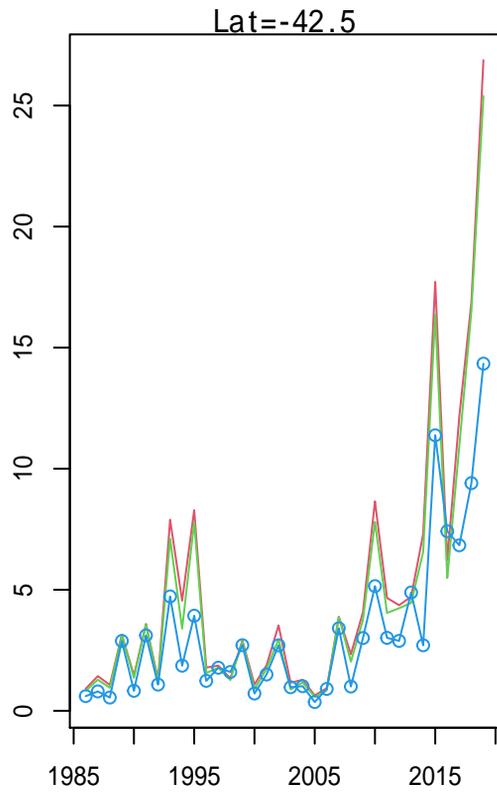
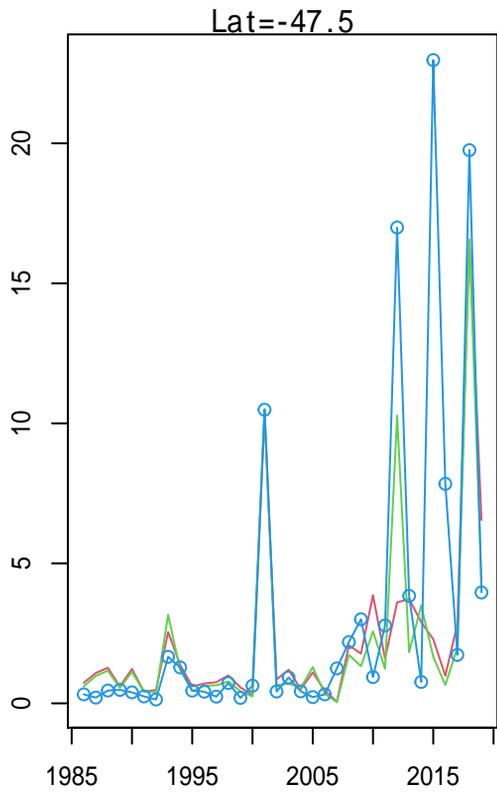
Red is GLM, green is GLMM, and blue is GAM.

# 1 variable. SQRT(residuals)



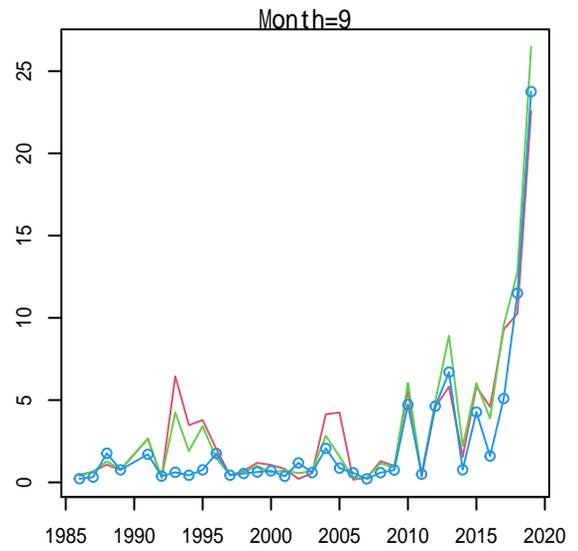
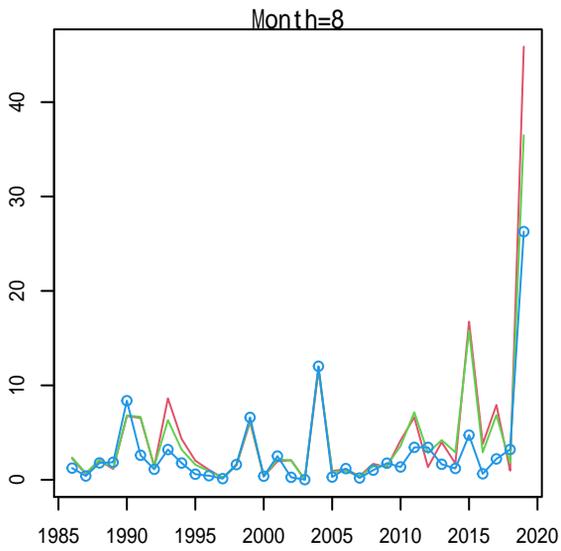
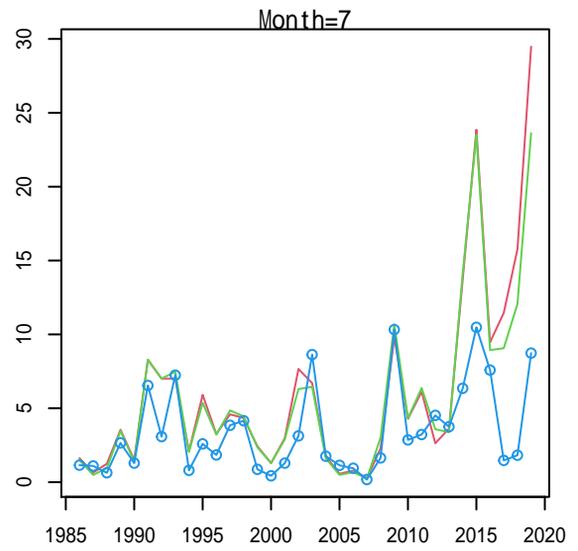
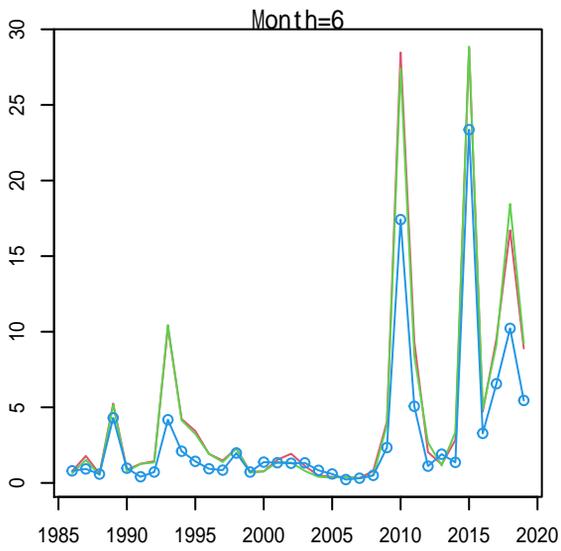
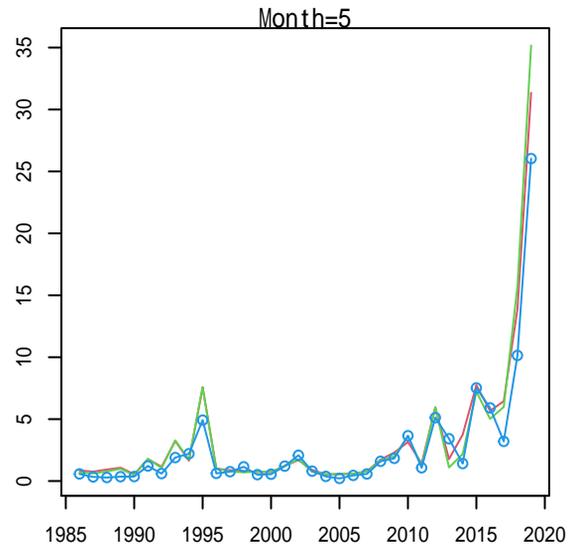
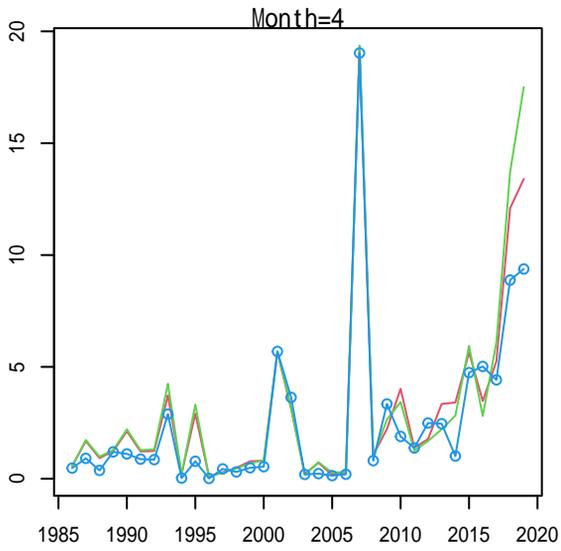
Red:GLM, Green:GLMM, Blue:GAM

2 variables. SQRT(residuals)



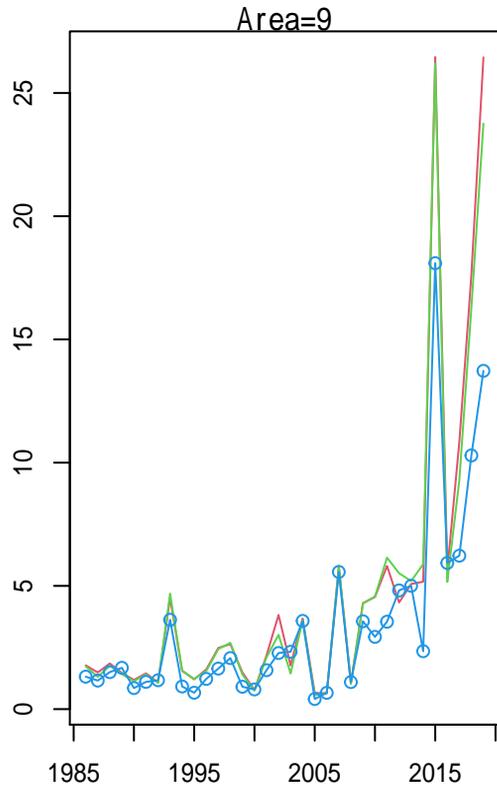
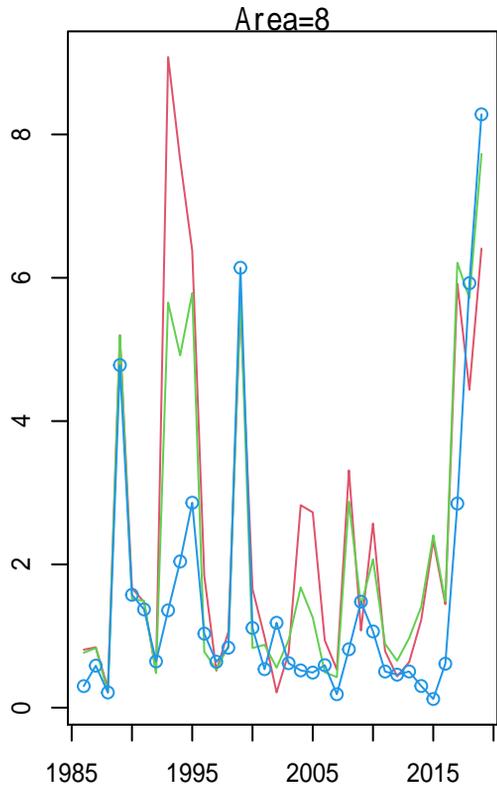
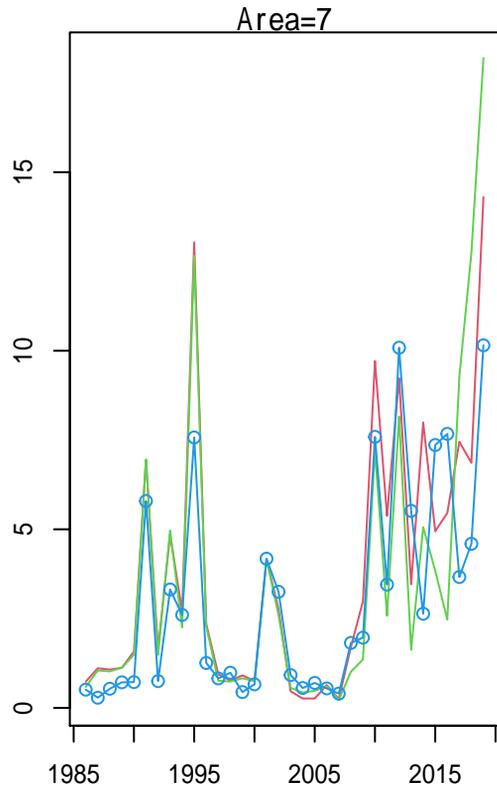
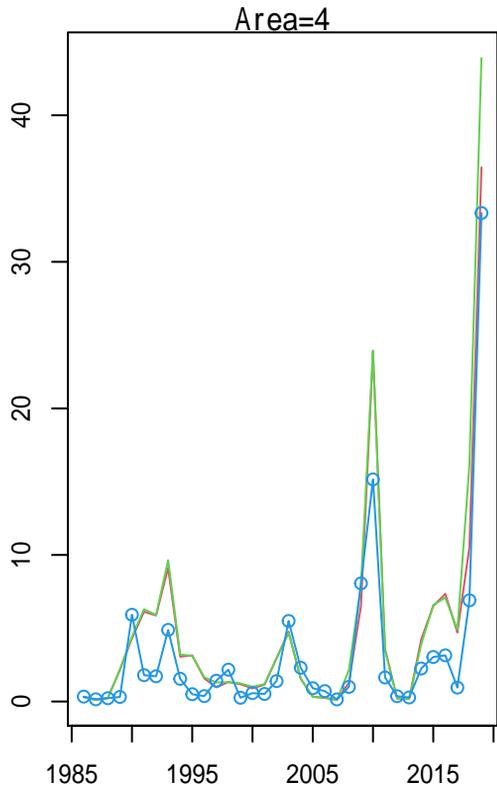
Red:GLM, Green:GLMM, Blue:GAM

# 2 variables. SQRT(residuals)

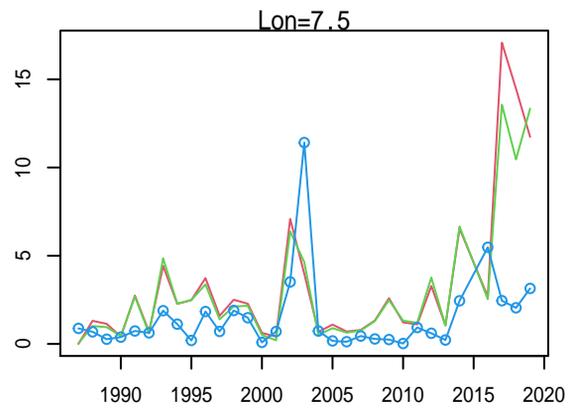
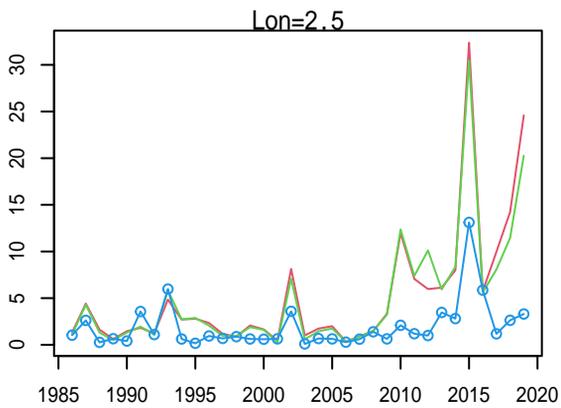
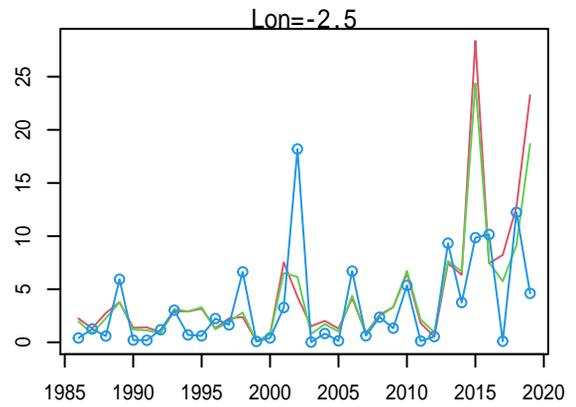
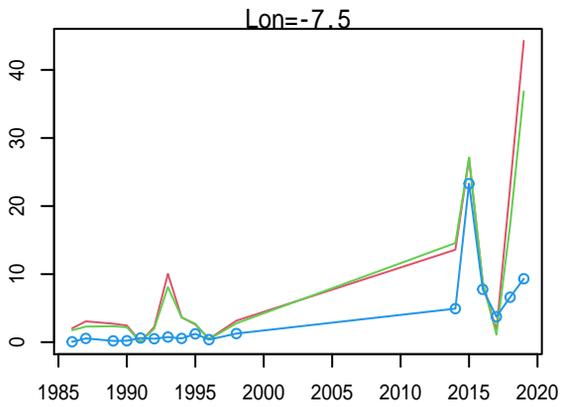
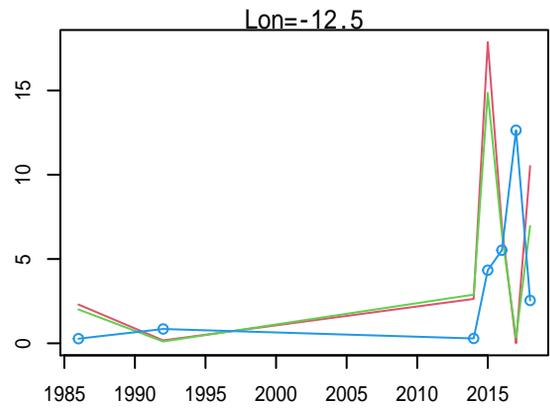
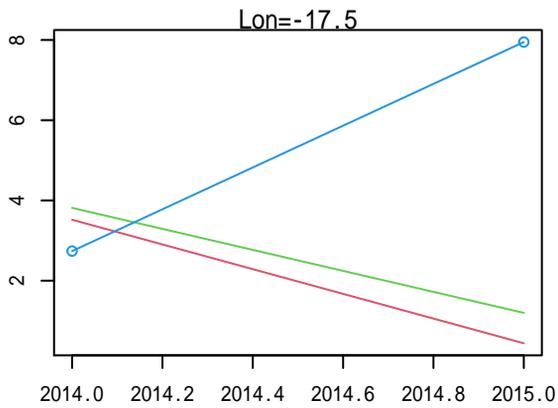
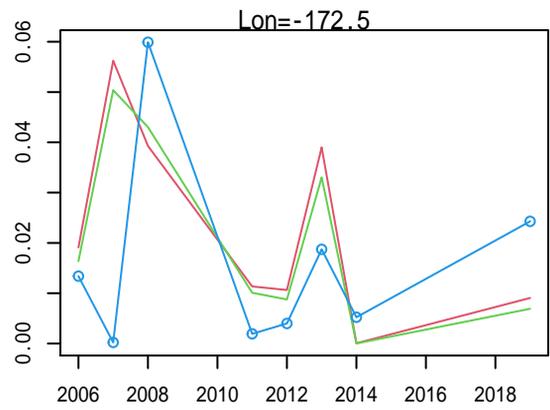
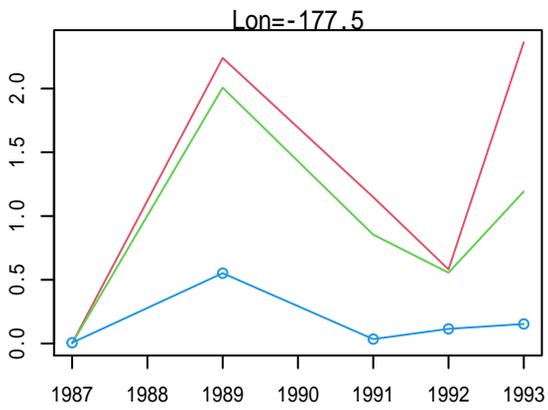


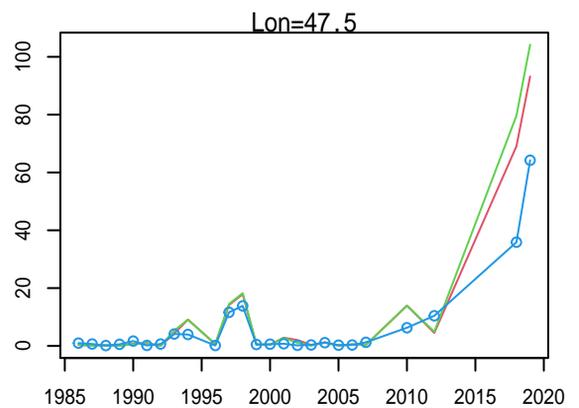
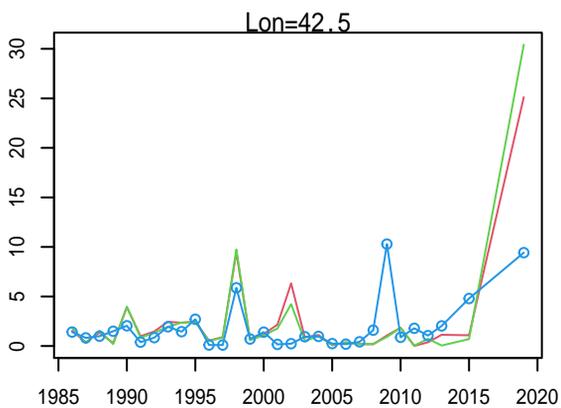
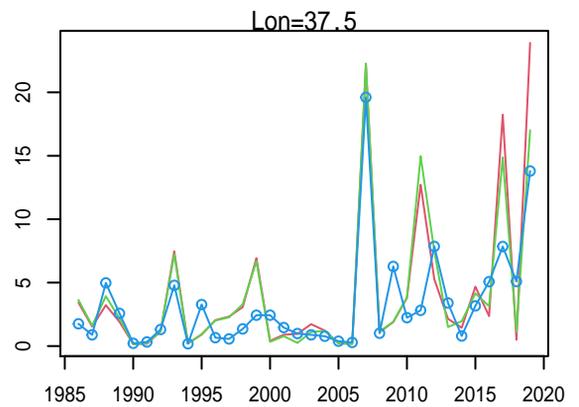
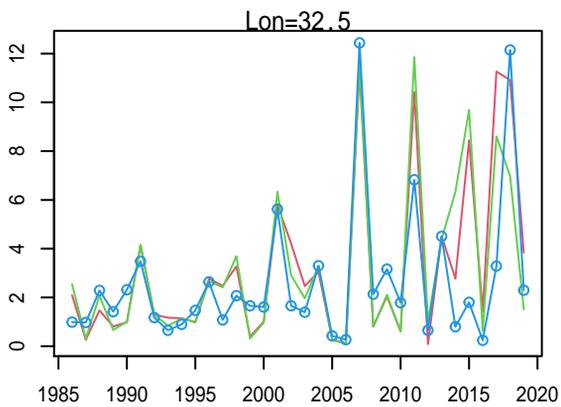
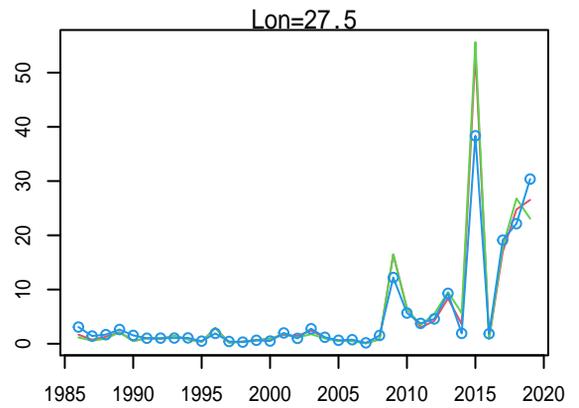
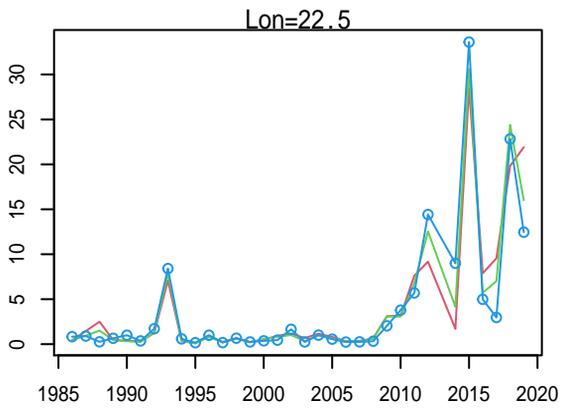
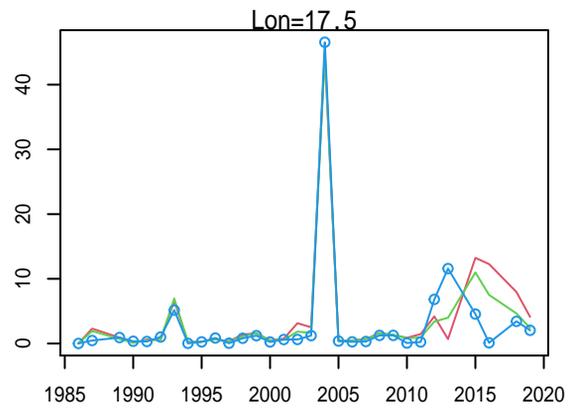
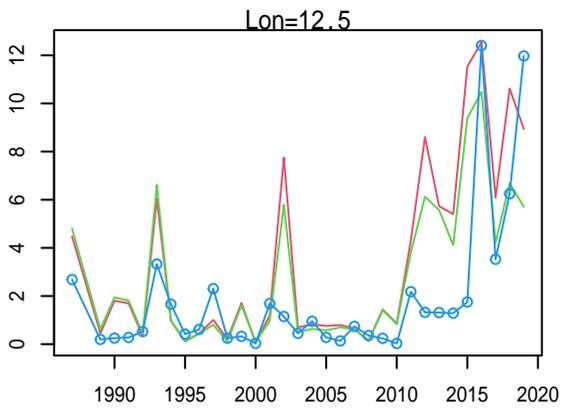
Red:GLM, Green:GLMM, Blue:GAM

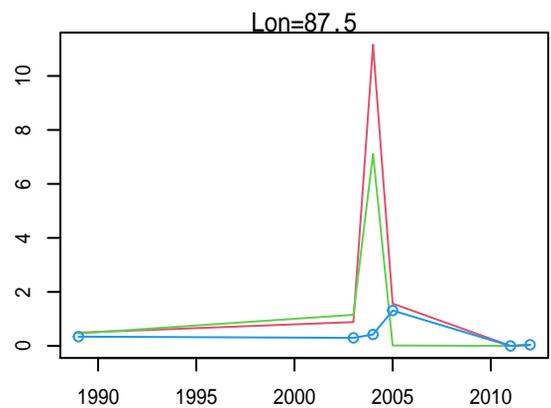
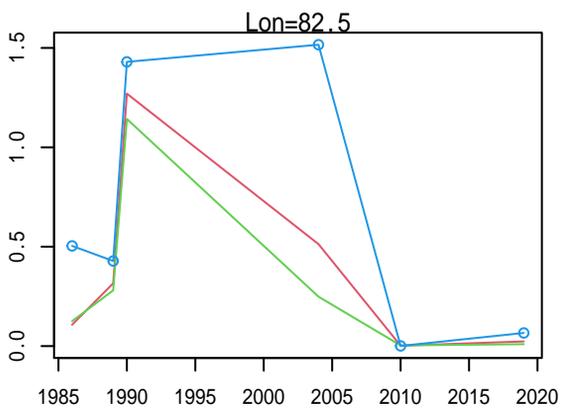
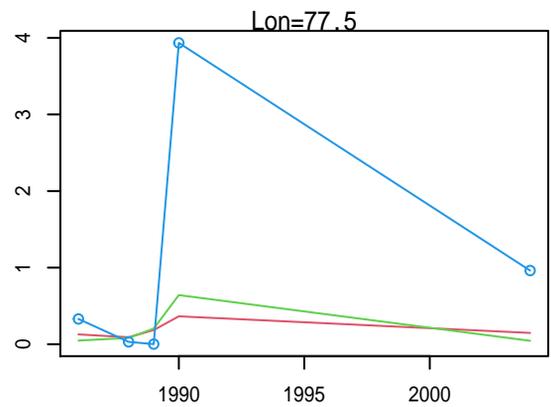
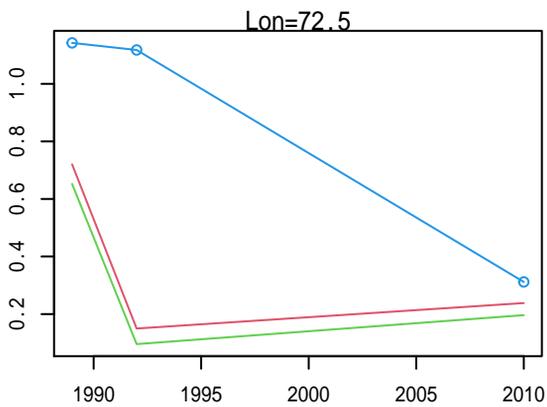
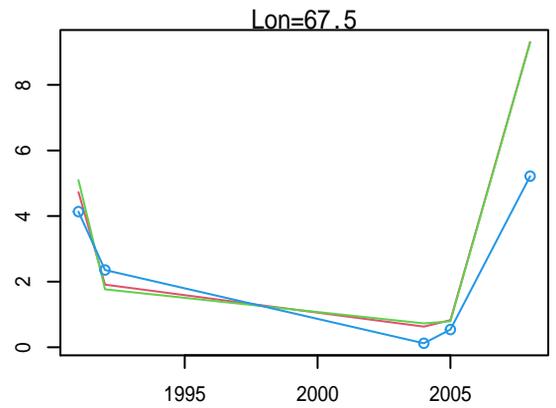
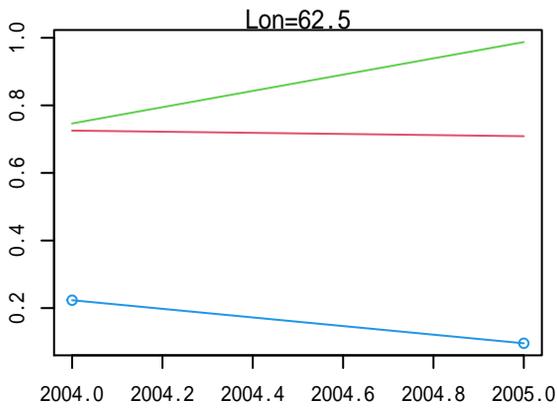
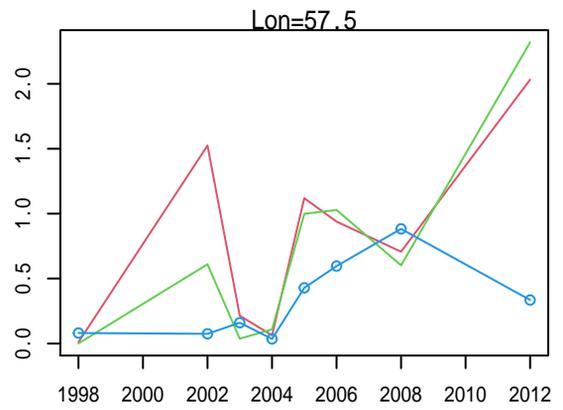
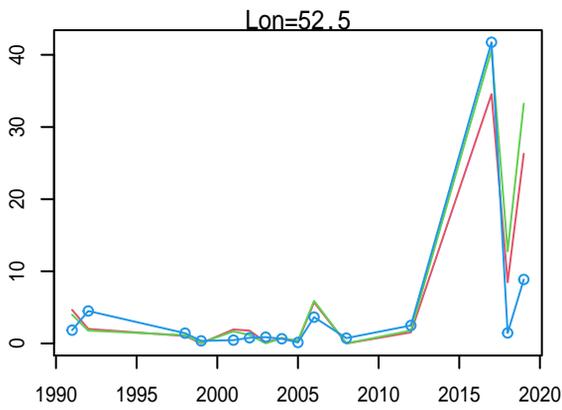
2 variables. SQRT(residuals)

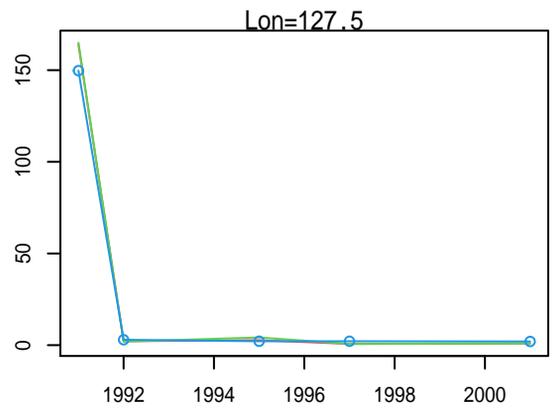
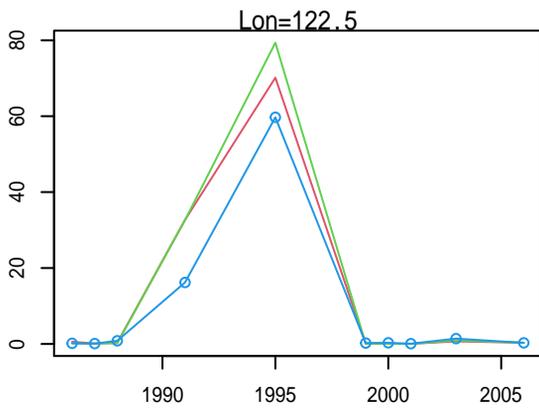
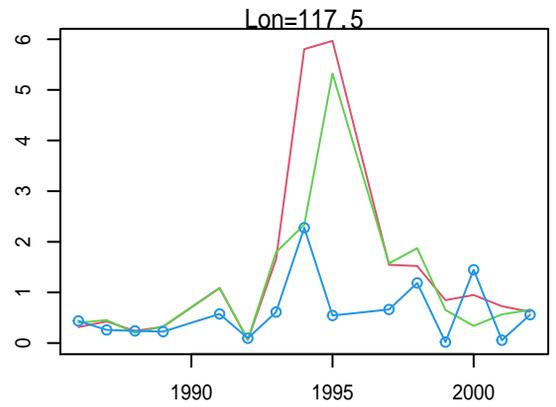
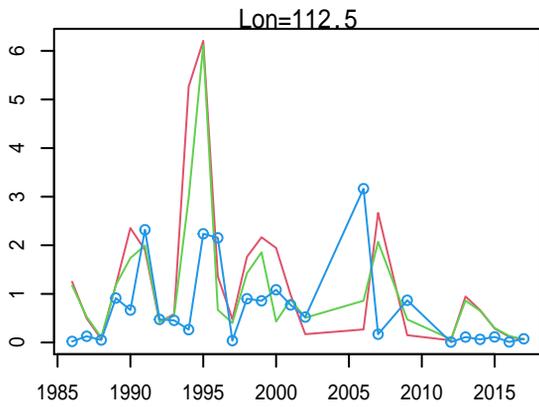
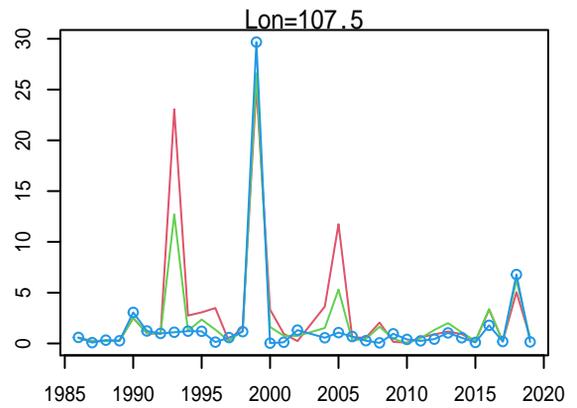
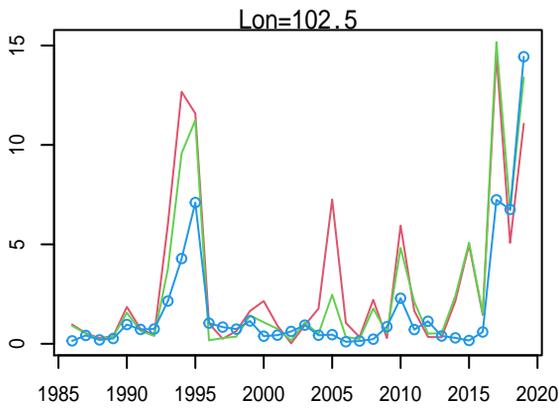
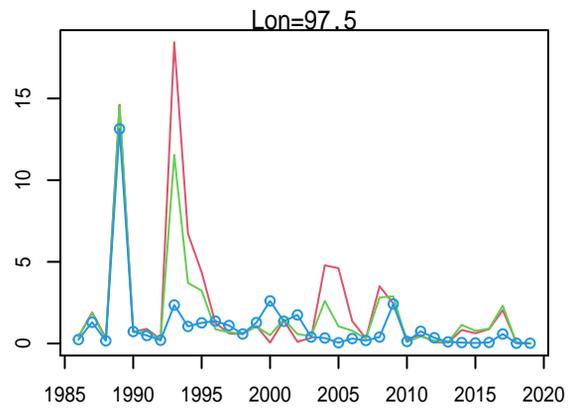
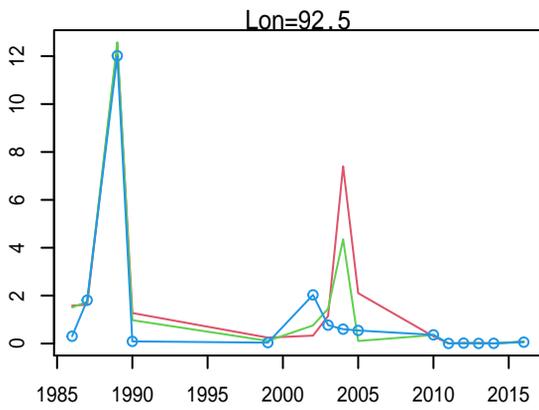


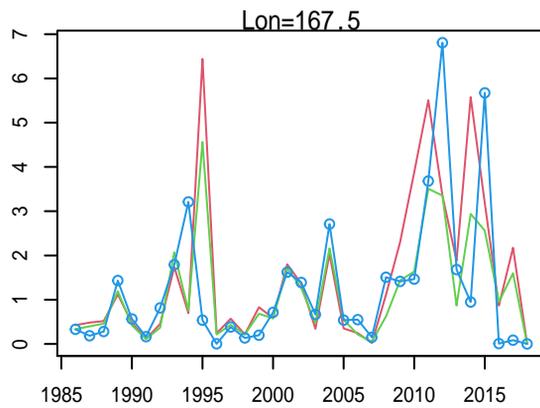
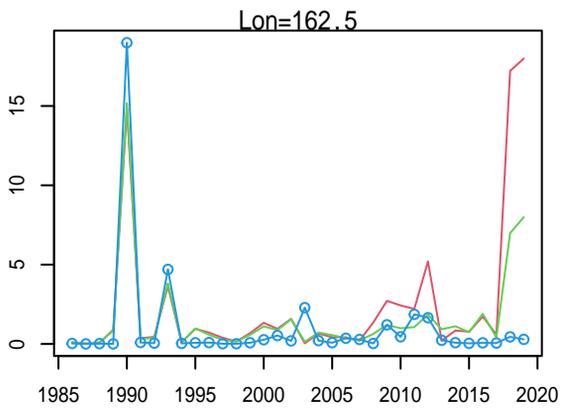
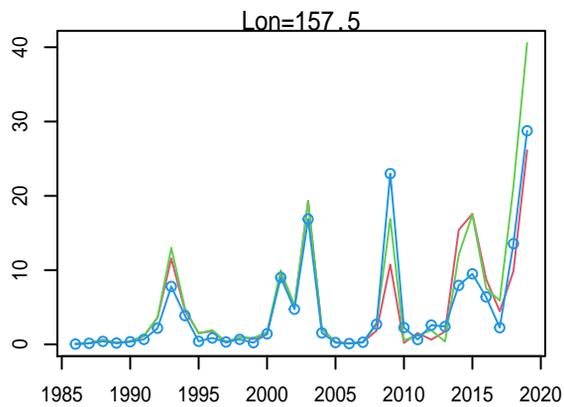
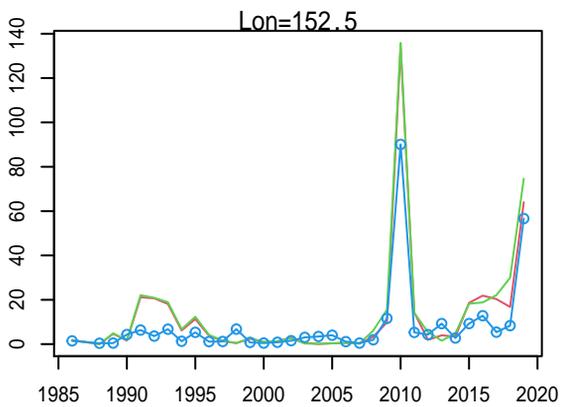
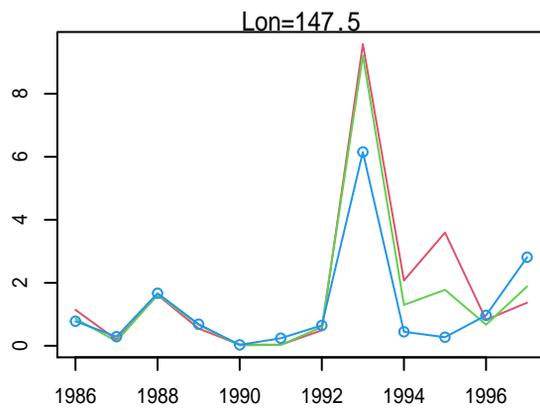
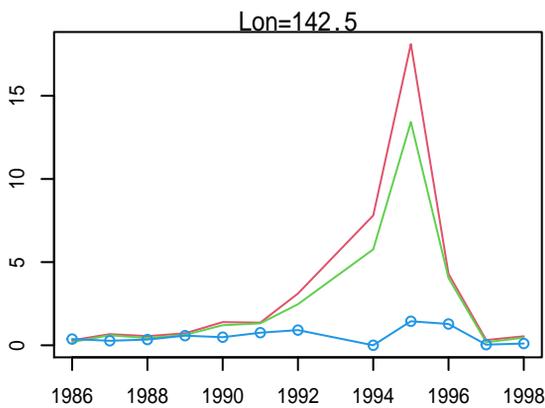
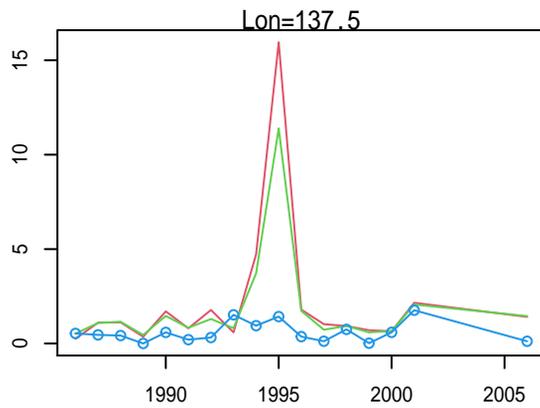
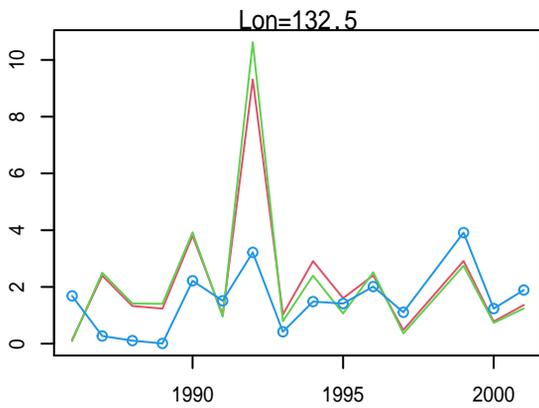
Red:GLM, Green:GLMM, Blue:GAM



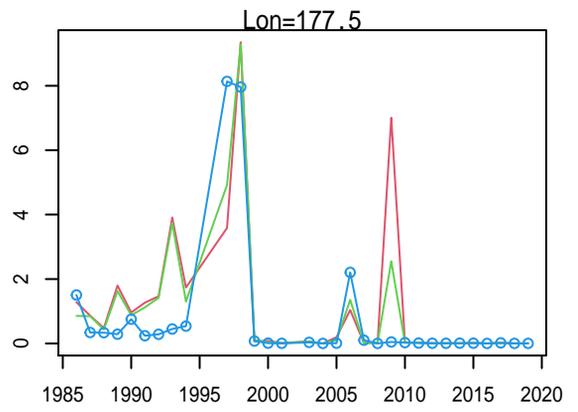
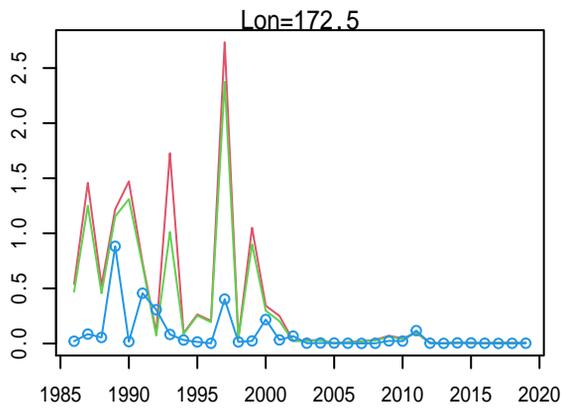






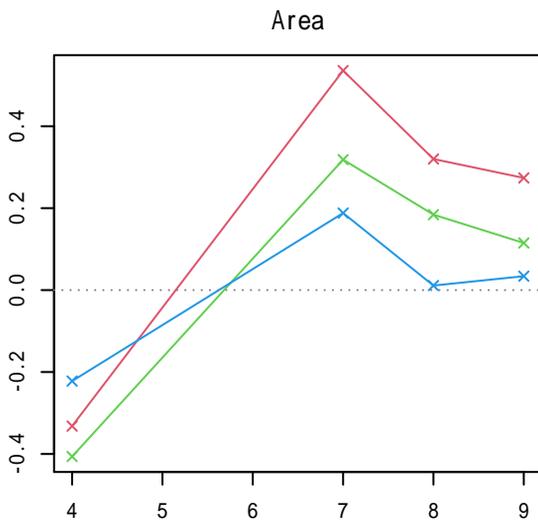
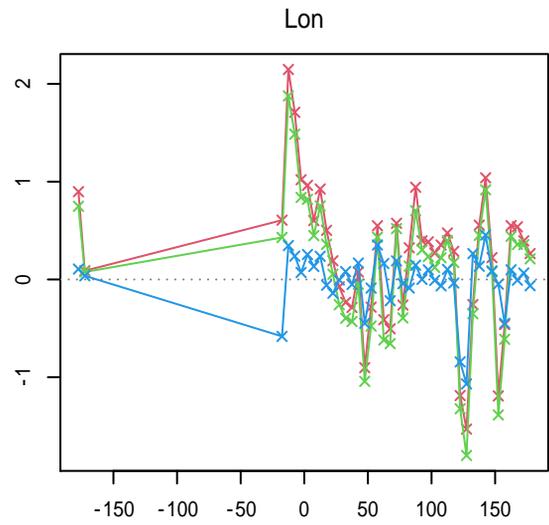
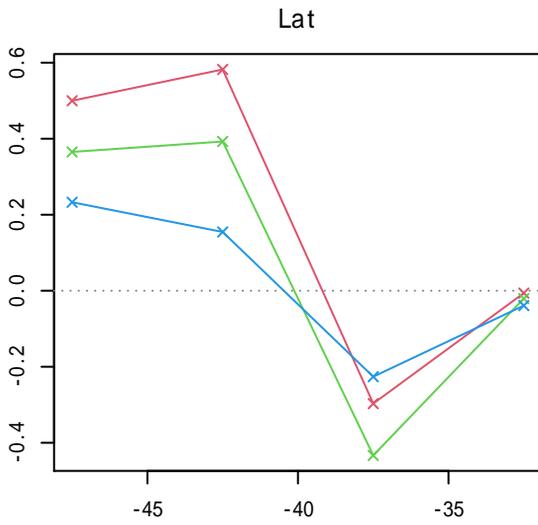
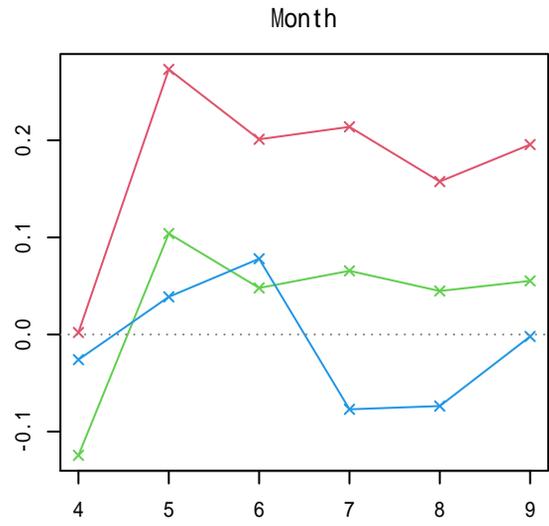
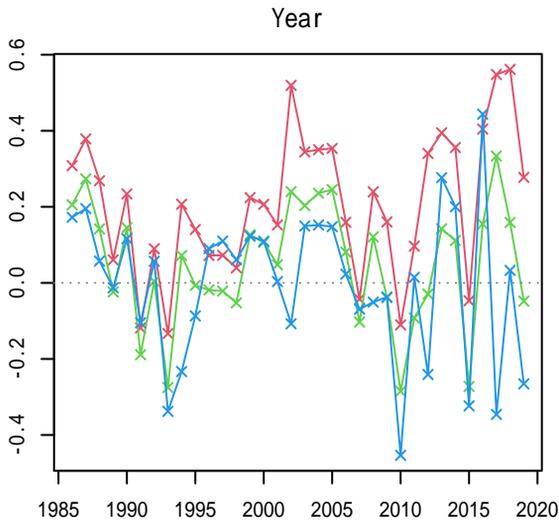


2 variables. SQRT(residuals)



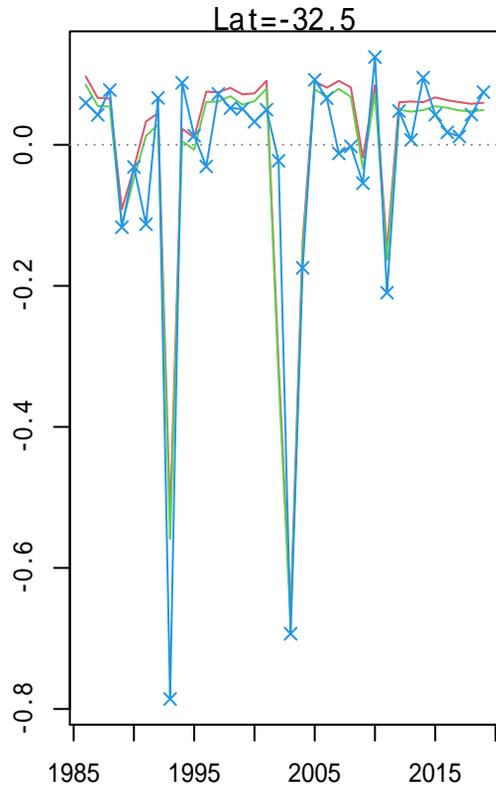
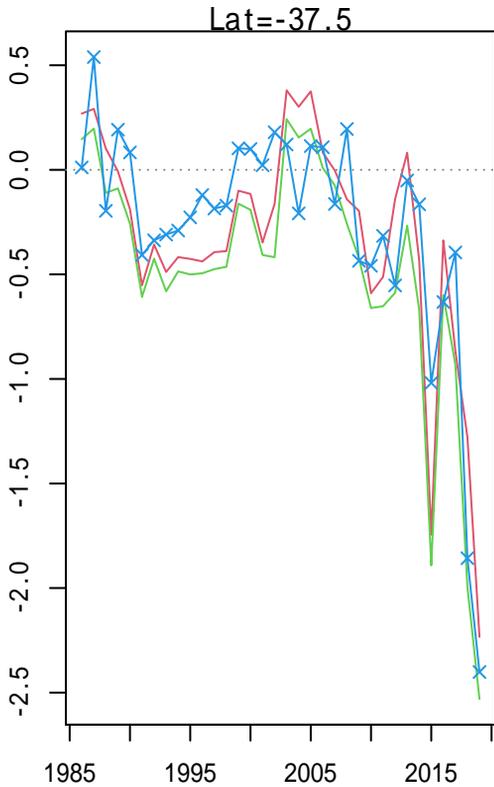
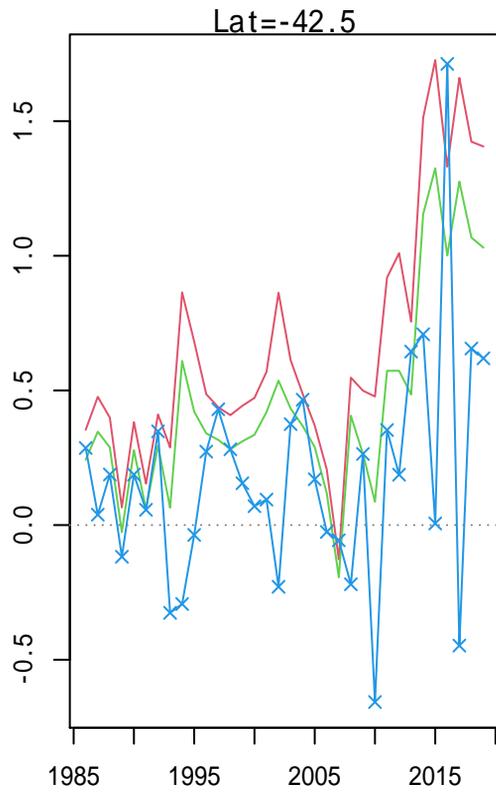
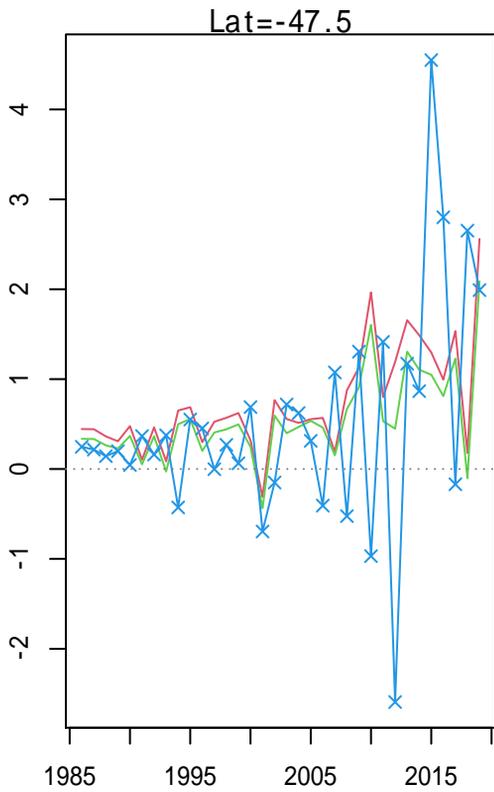
Red:GLM, Green:GLMM, Blue:GAM

# 1 variable. Residuals



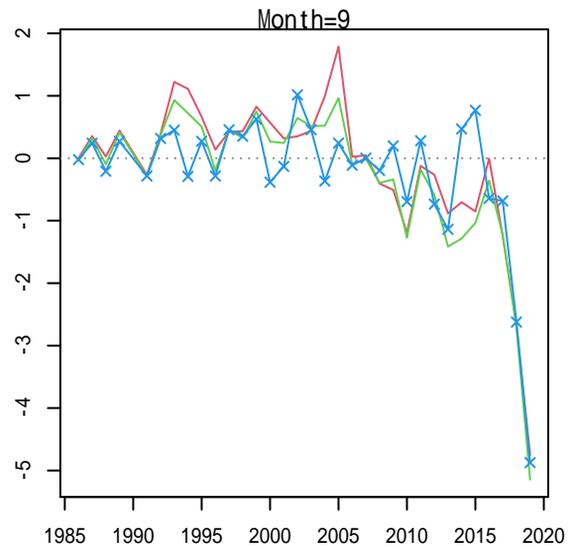
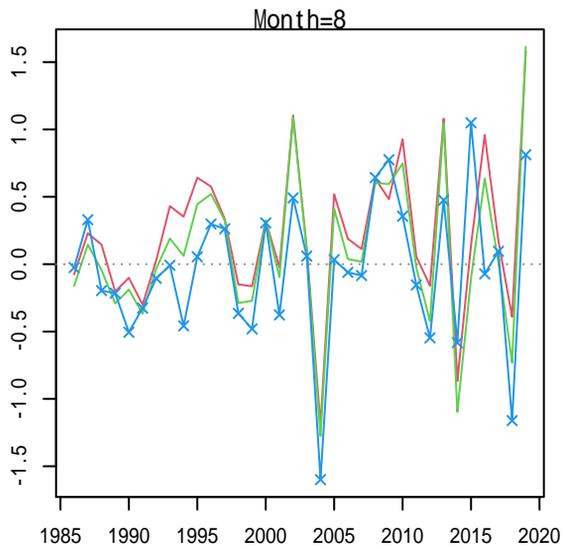
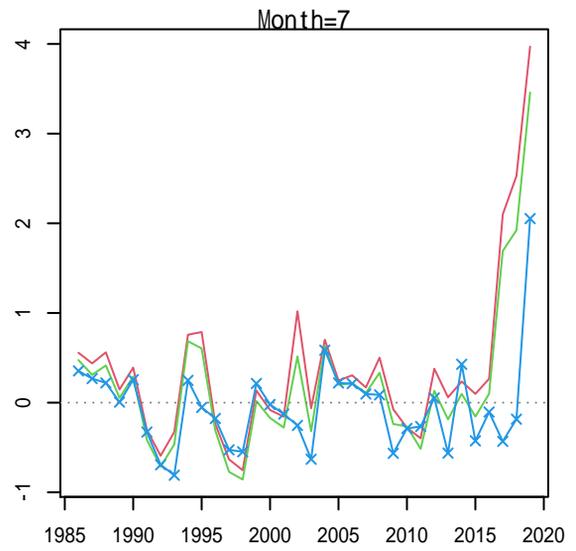
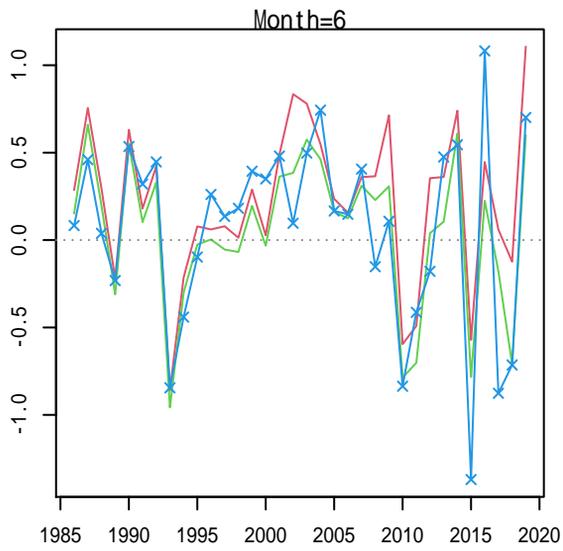
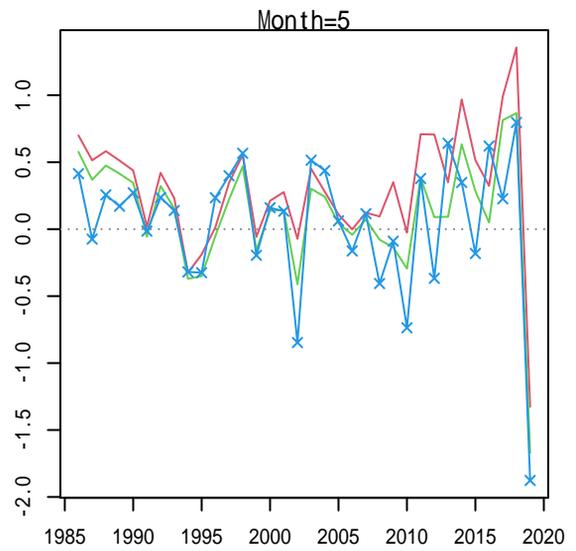
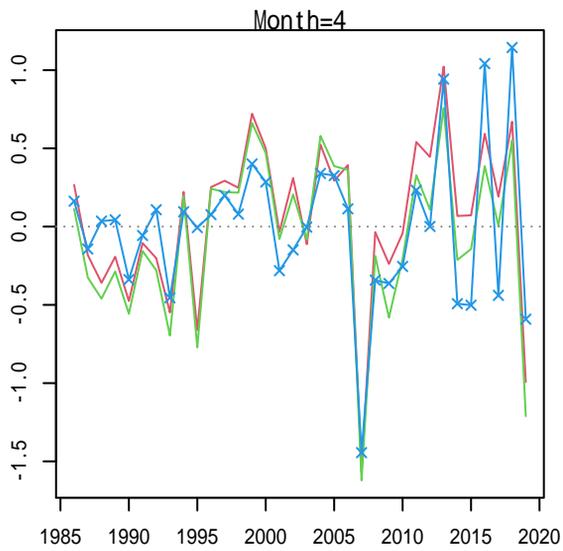
Red:GLM, Green:GLMM, Blue:GAM

2 variable. Residuals



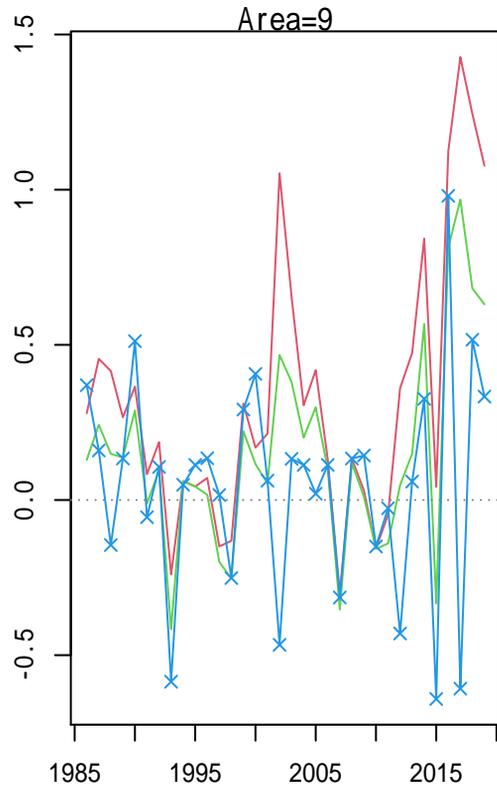
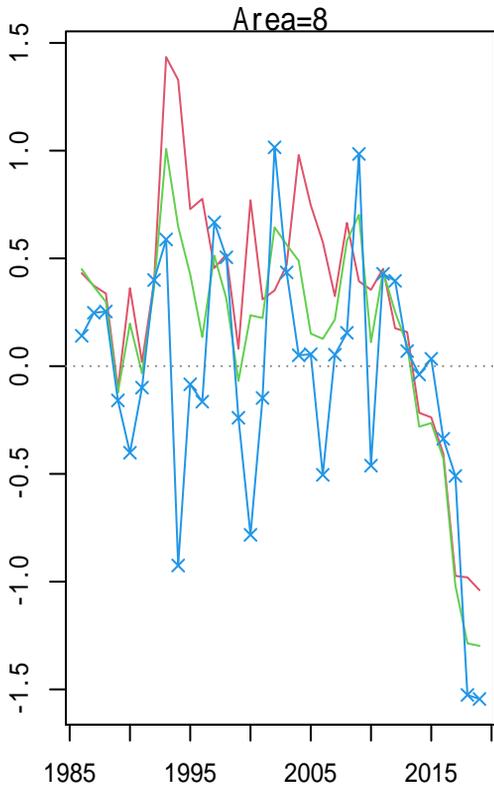
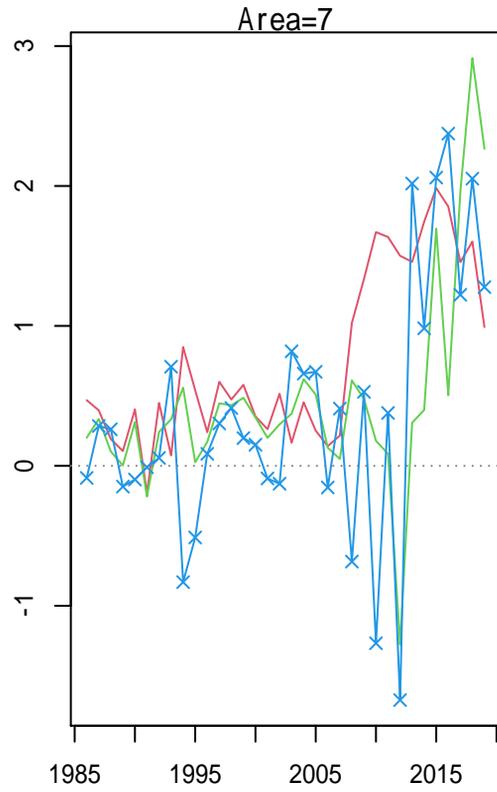
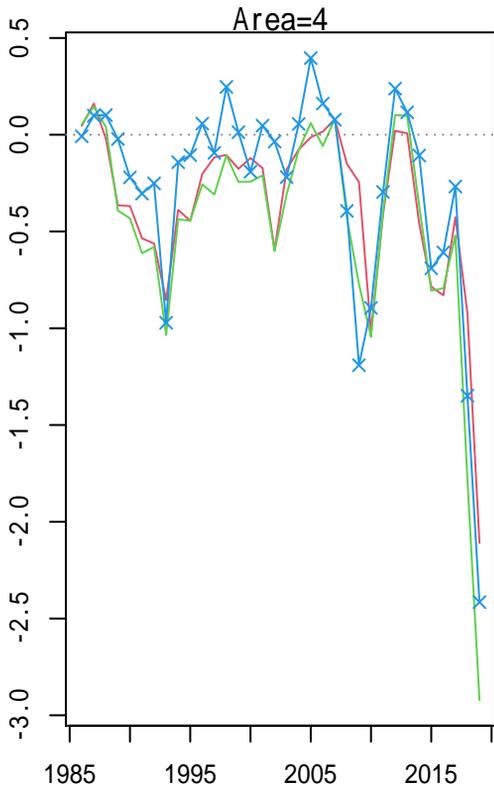
Red:GLM, Green:GLMM, Blue:GAM

## 2 variable. Residuals

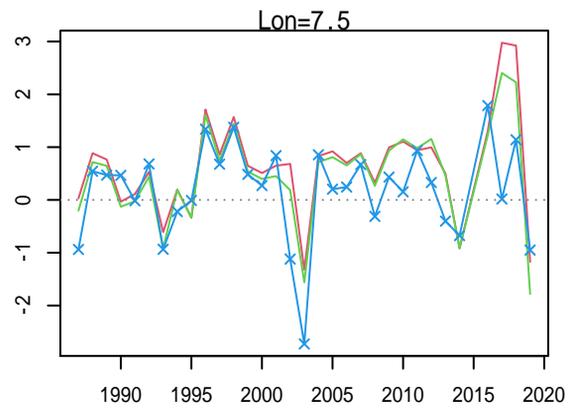
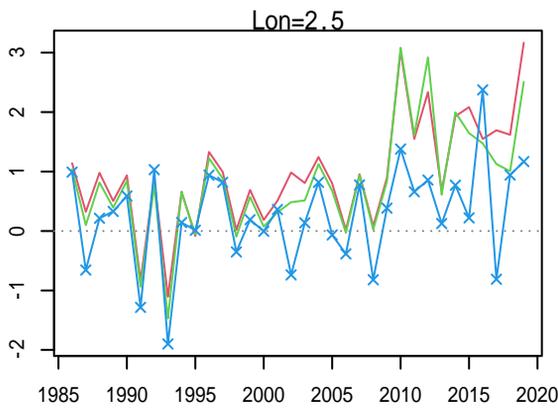
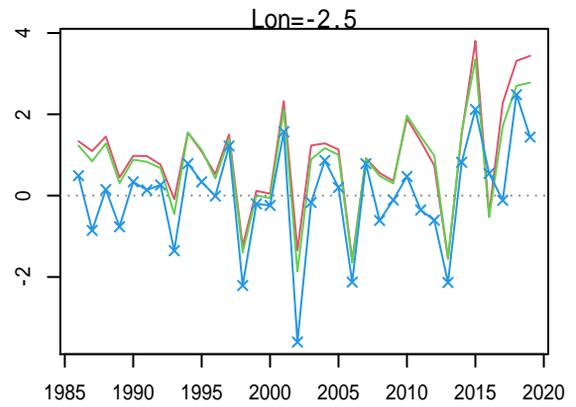
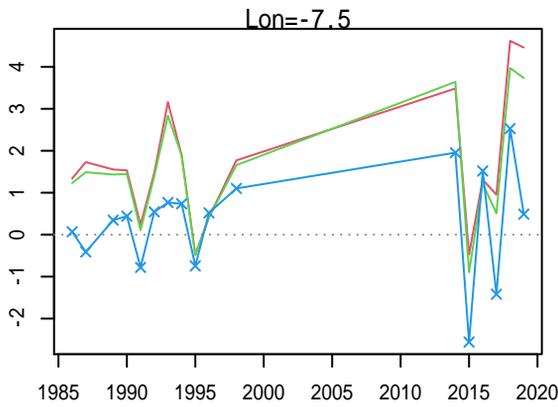
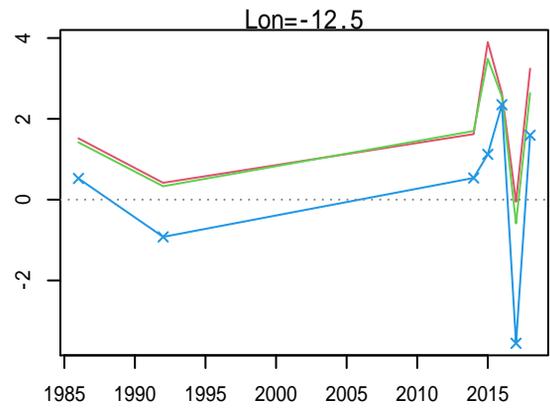
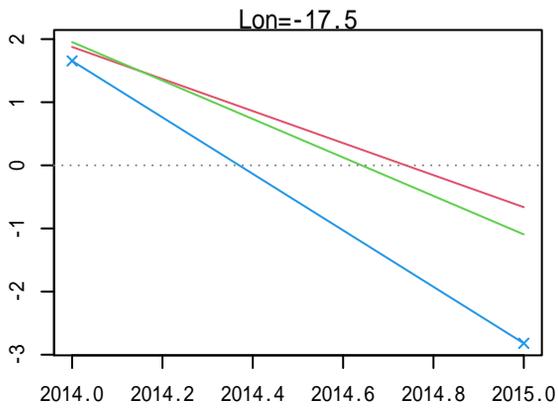
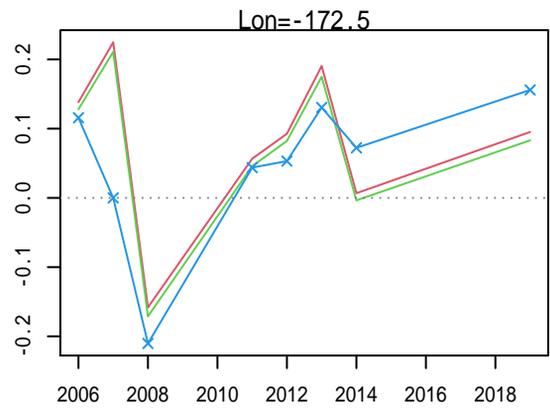
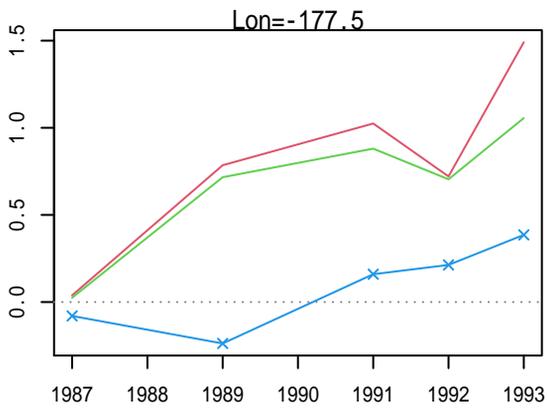


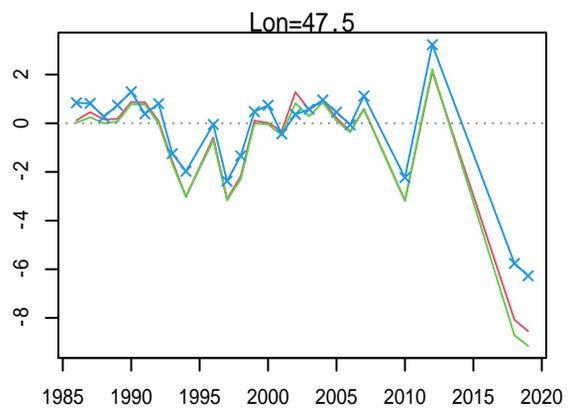
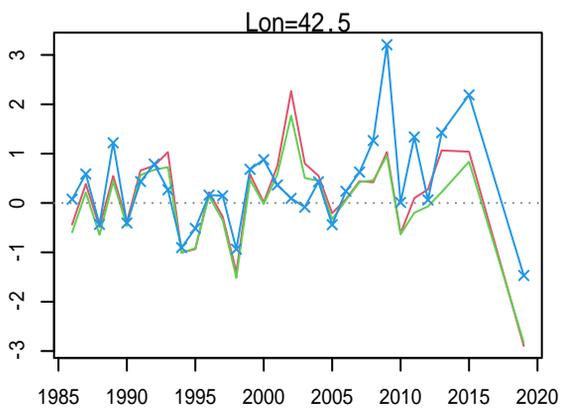
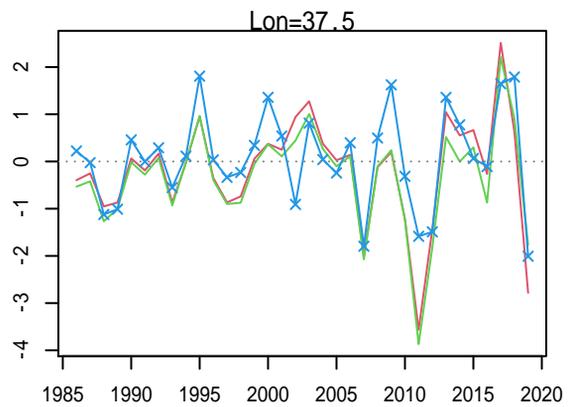
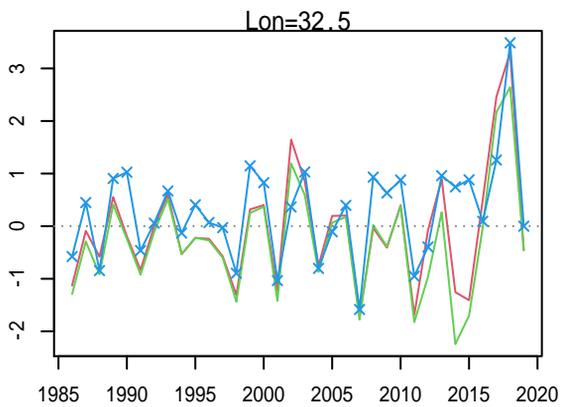
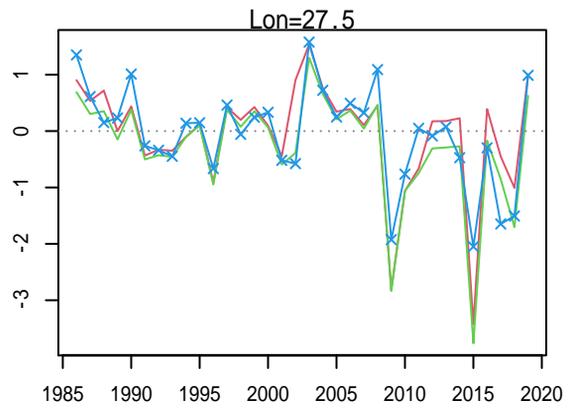
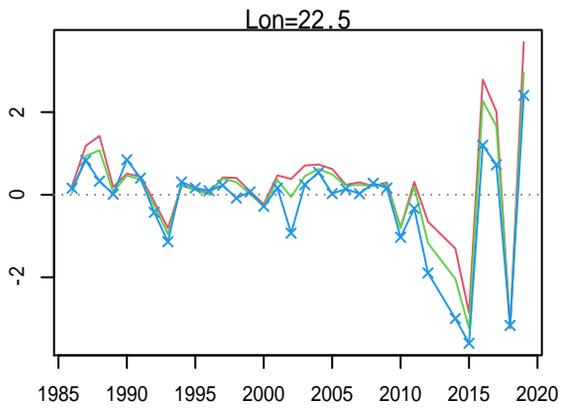
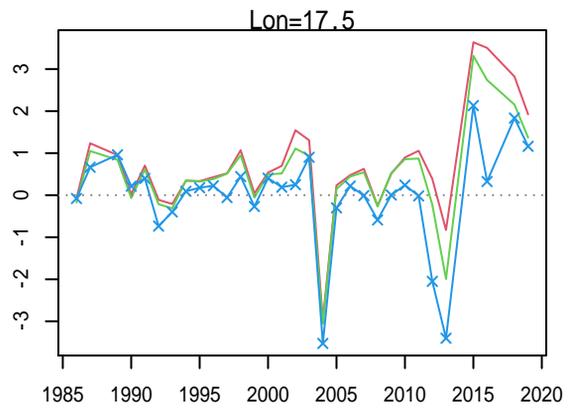
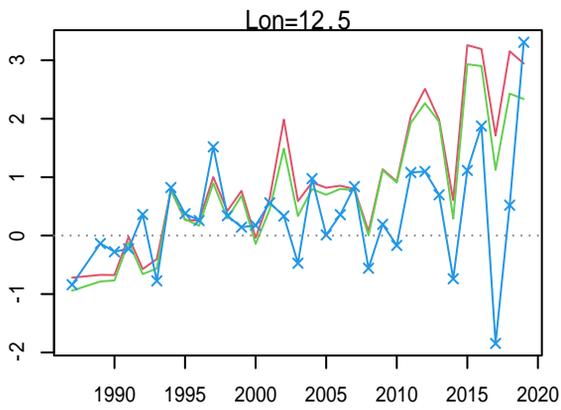
Red:GLM, Green:GLMM, Blue:GAM

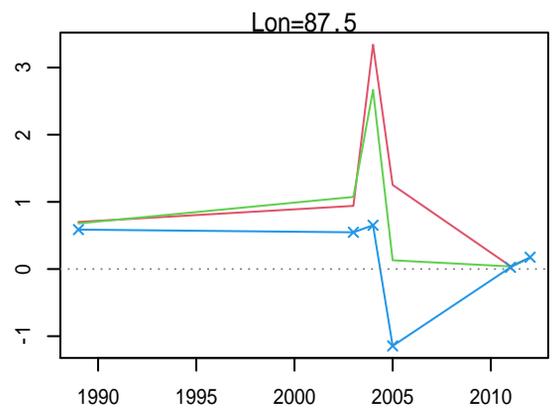
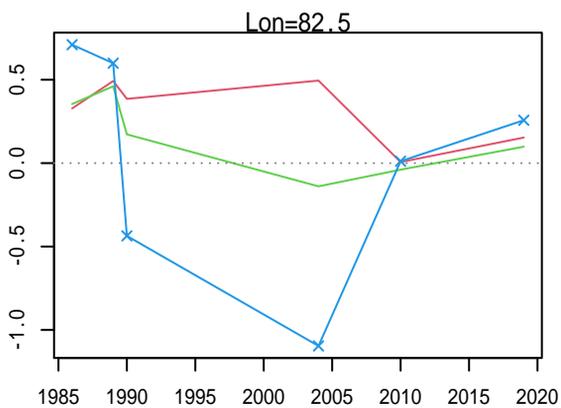
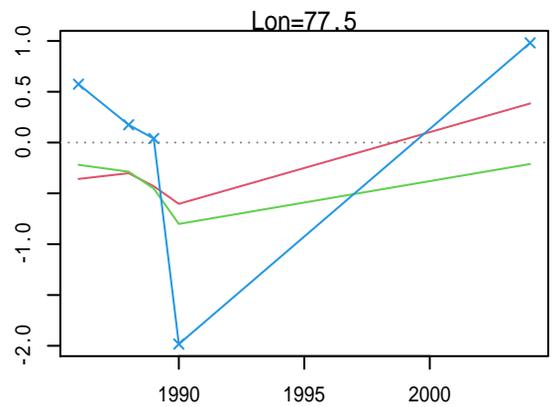
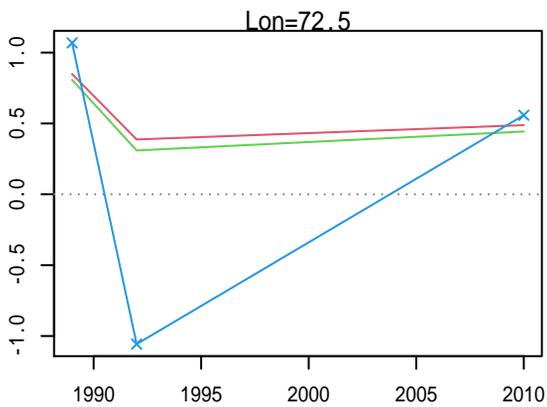
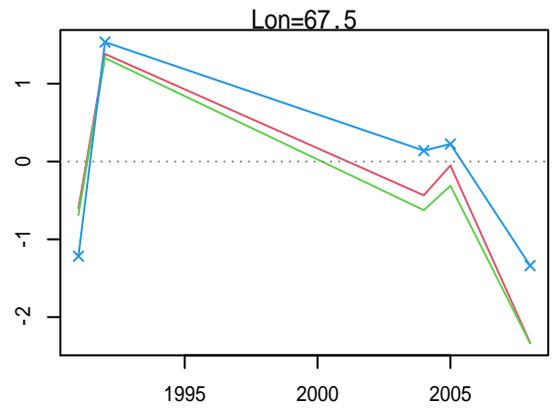
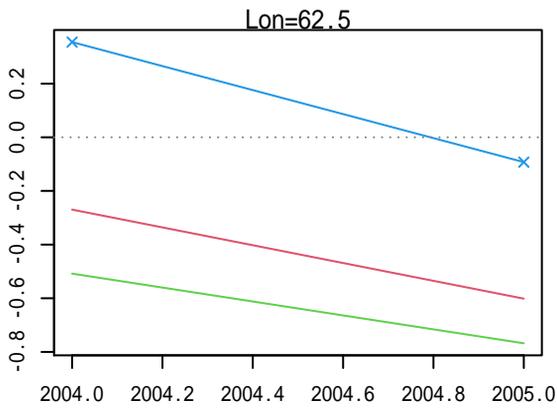
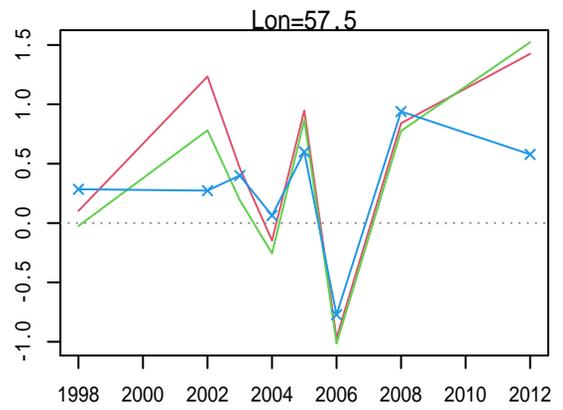
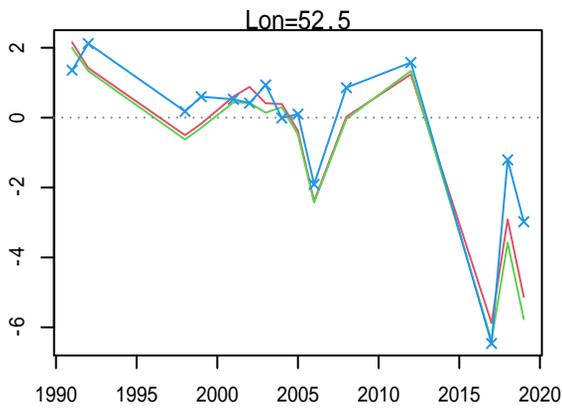
## 2 variable. Residuals

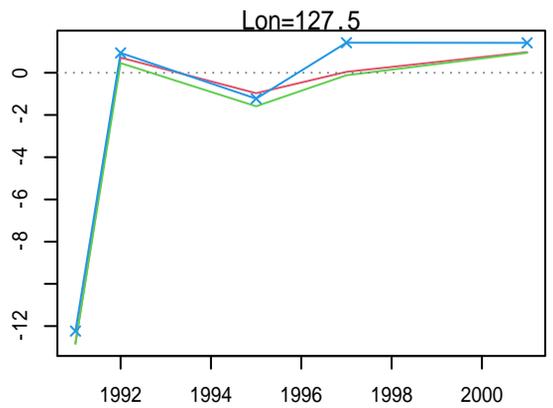
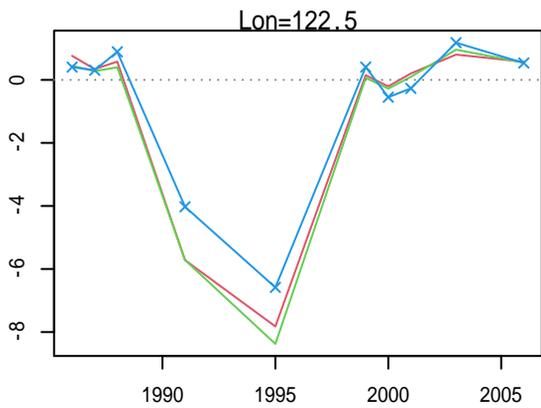
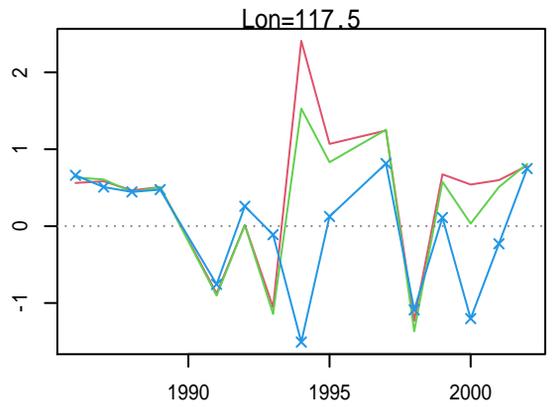
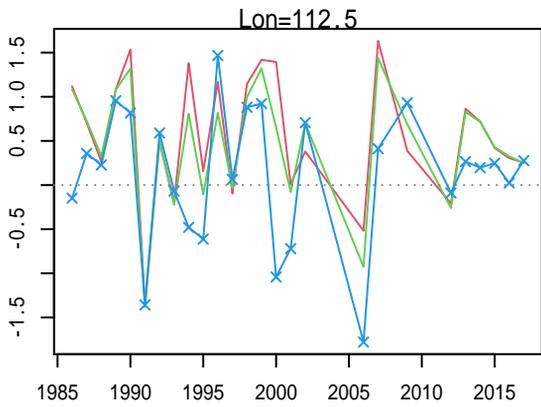
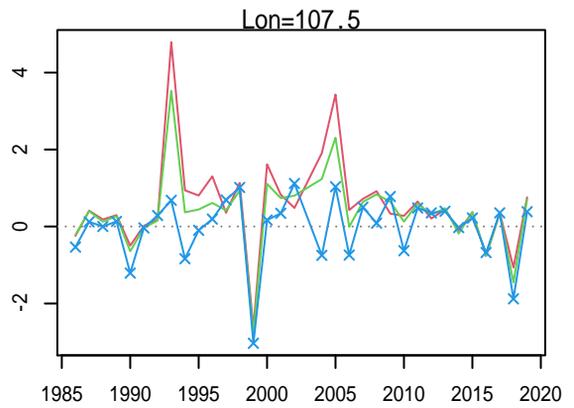
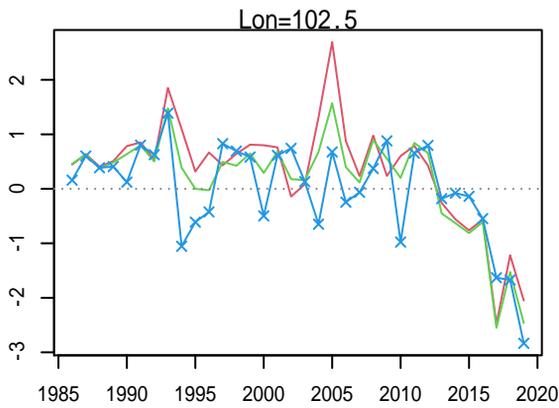
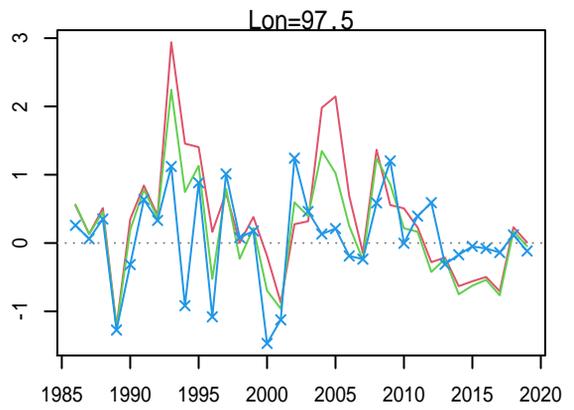
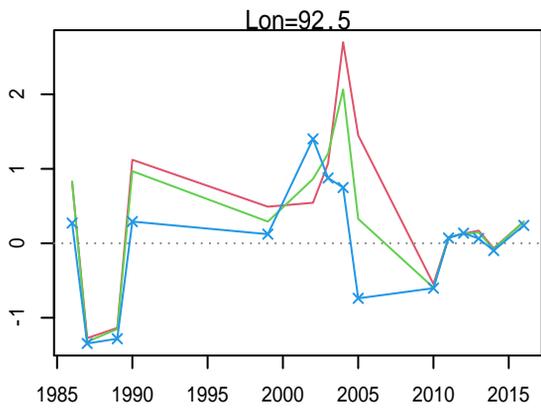


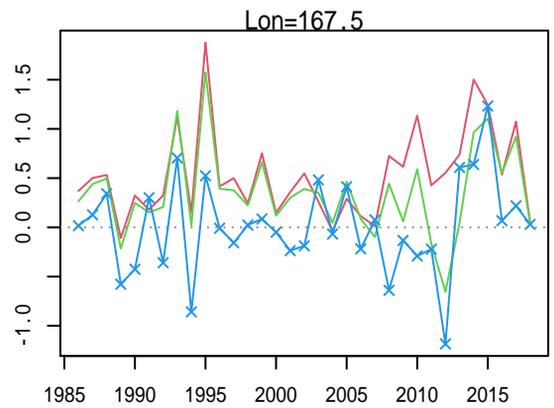
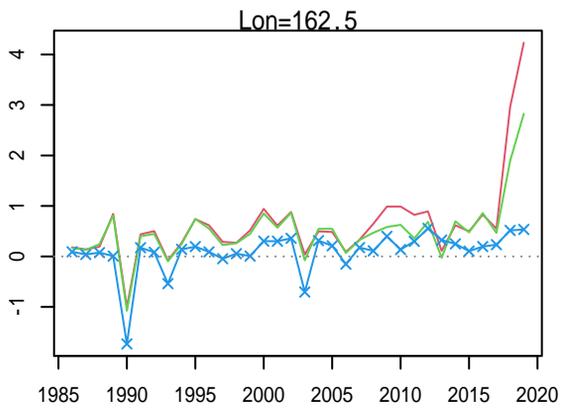
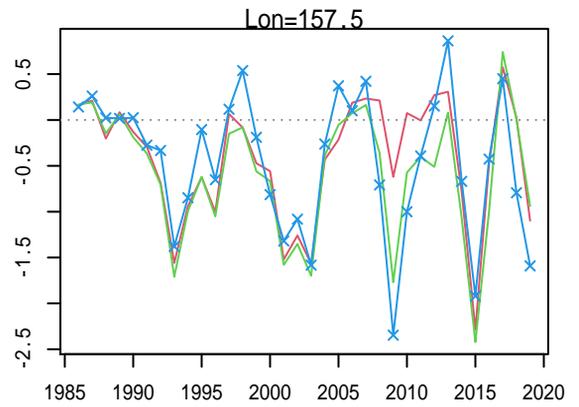
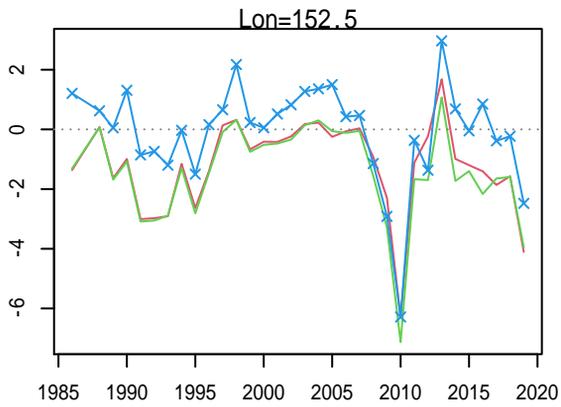
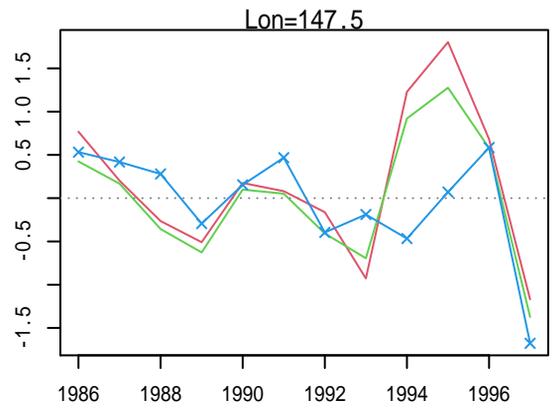
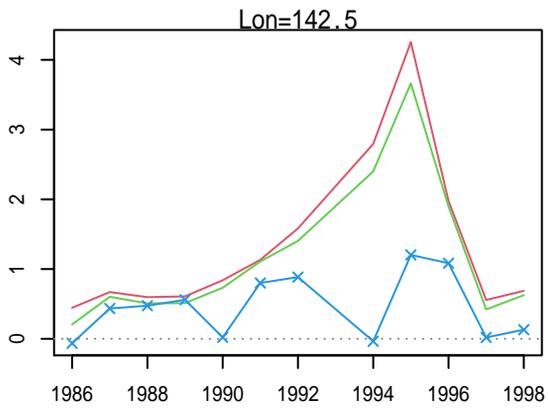
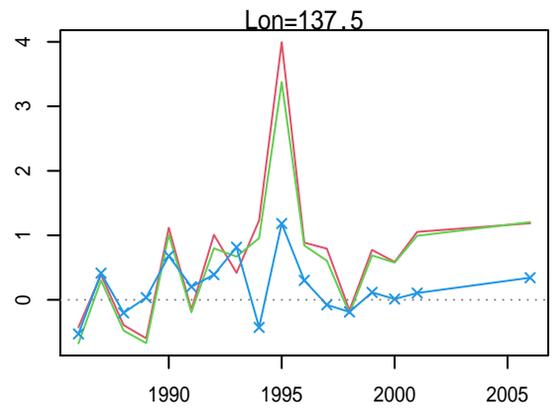
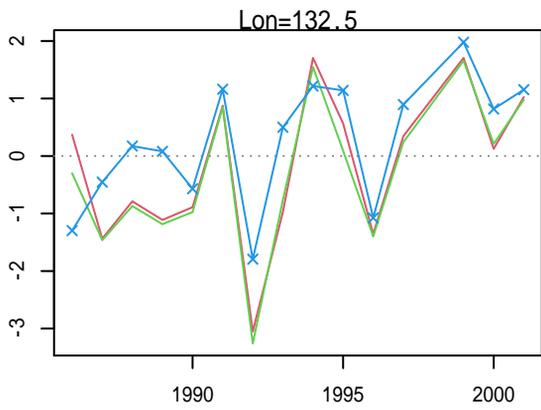
Red:GLM, Green:GLMM, Blue:GAM



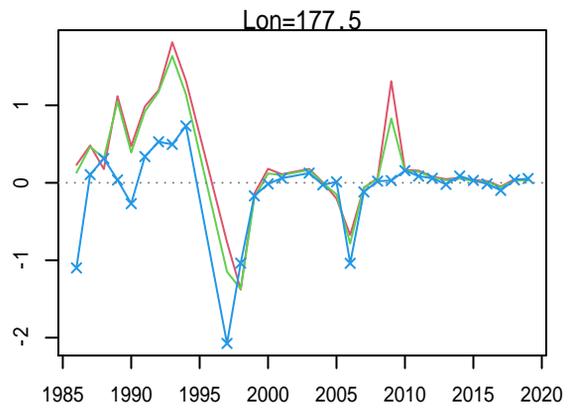
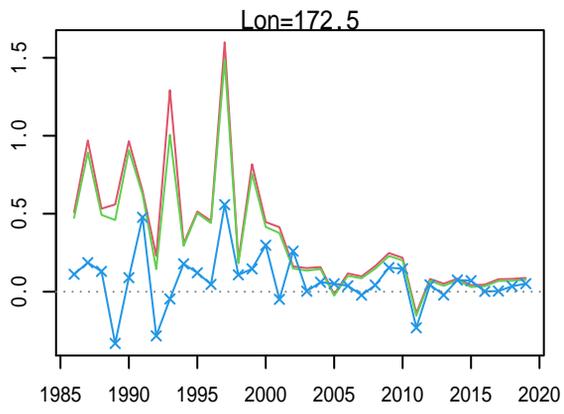








## 2 variable. Residuals



Red:GLM, Green:GLMM, Blue:GAM