

ミナミマグロのコア船 CPUE データにおける 2018 年の極めて高い  
値の検討

Examination of an anomalously high value of the core vessel  
CPUE in 2018 for southern bluefin tuna

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

コア船の CPUE から求めたミナミマグロの資源量指数において 2018 年の値が大きく増加したことについて、その原因を探索した。原因は 8 海区と 9 海区の緯度別データ数がアンバランスであることによって年\*Area の交互作用の推定値が異常になったためであることが分かった。データセットのわずかな変更によって原因を証明した。YearArea 交互作用項を除いた GLM モデル、YearArea 交互作用項をランダム項にした GLMM によって異常な高値は無くなった。

Summary

The cause of the anomalously large increase in the value of 2018 in the abundance index of SBT calculated from the CPUE of the core vessels was searched. It was found that the cause was an abnormal estimate of the year\*Area interaction due to an unbalanced number of data by latitude in areas 8 and 9. The cause was proven by a slight modification of the dataset. The GLM model without the year\*Area interaction term and the GLMM with the year\*Area interaction term as random terms eliminated the anomalously high value.

## Introduction

コア船 CPUE 資源量指数は、CCSBT におけるミナミマグロの資源評価並びに TAC を決定する MP において用いられる重要な資源量指数である。2019 年の計算において、その 2018 年の値は 2017 年以前との比較で極めて高く、大きな誤差を含んでいる可能性が伺われた (Itoh and Takahashi 2019)。2019 年の CCSBT-ESC においては、その原因を探索することが合意された (Anon. 2019)。本文書は、2018 年高値の原因を明らかにすること、解決策を提示することを目的とした。また、将来に類似の異常値が発生したときにも検出できるよう、チェックの方法を検討した。

The core vessel CPUE abundance index is an important index used in the stock assessment and MP that determine the TAC of southern bluefin tuna (SBT) in CCSBT TAC. In the calculation of 2019, the value for 2018 was extremely high compared to that before 2017, and it was possible that it contained a large error (Itoh and Takahashi 2019). In the 2019 CCSBT-ESC, it was agreed to investigate the cause (Anon. 2019). The purpose of this document is to identify the causes of the 2018 anomalously high value and to provide solutions. In addition, we examined the check method so that it can be detected even when similar abnormal values occur in the future.

## 原因

資源量指数は以下のプロセスで作成される。まず、コア船データセットを作成する (Itoh and Takahashi 2020 を参照) (ステップ 1)。次いで、年、月、緯度 5 度、経度 5 度別にまとめ、GLM で標準化をする (ステップ 2)。GLM に用いた変数の全ての組み合わせによるテストデータを作成し、GLM 予測値を階層別に求める (ステップ 3)。これを年別に平均値を求めて LS-means とする。次いで、テストデータを面積指数データと対応させる。この時、面積指数が存在しない階層 (例えば 30-40S に位置する 4 海区においての 45S) については除かれることになる (ステップ 4)。GLM 予測値に面積指数 (年に関係ない CS または年で変動する VS) を掛け合わせて、年平均値を求め、これが資源量指数となる (ステップ 5)。本文書では、データまたは統計値がどのステップなのかを明確にした。

2019 年の作業において次のことが分かっている。GLM の年の LS-mean (ステップ 3) までは極端な値ではなく、面積指数を掛けた後 (ステップ 5) に極端な値となったこと、Area8 の 40S が関係していること、GLM のベースモデルでは極端な値になるが Reduced-model ではそうはならないこと。昨年は原因の特定には至らなかったが、Year\*Area 交互作用が関係していると疑われた。

2020 年に作成した 2019 年までのコア船データで解析した。データや方法については別文書 (Itoh 2020、Itoh and Takahashi 2020) を参照。データを 2019 年までアップデートしても、2018 年の値が異常に高いことは変わらなかった (Fig. 1)。ステップ 4 のテストデータにおいて、GLM の予測値には 2,278 件中に 3 件の極端に値の大きなレコードがあった。実験としてこの 3 データを除いてステップ 5 の資源量指数を求めると 2018 年のアノマリーが消失し、この 3 件が大きな影響を持つことが特定された (図 2)。3 件は全て

2018年、8海区、40Sのもので、7月、8月、9月のものであった。

詳細な検討の結果、原因は8海区と9海区の緯度別データ数がアンバランスであることによって年\*Areaの交互作用の推定値が異常になったためであることが分かった。2018年には9海区における35S(35-39S)のデータがわずかであった(表1)。そして、この9海区の35Sでのデータは主に4月と9月のもので主漁期から外れており、ミナミマグロのCPUEが低かった。一方、9海区の40Sのデータは豊富に存在した。一方、8海区においては35Sのデータは多くあるが40S(40S-44S)のデータは存在しなかった。GLMのYear\*Area項は単に2018年のArea8とArea9にデータが存在するだけでなく、同じ緯度で対応するデータが存在することを要求していた。

実験として、データセット(ステップ1)においてArea9の40S及び41Sのデータを緯度5度階級の35Sとして変更した。この簡単な操作によって2018年のアノマリーが消失したことから、原因を特定できた(図3)。

### Cause

The abundance index is created by the following process. First, create a core vessel dataset (see Itoh and Takahashi 2020) (step 1). Then, it is summarized by year, month, latitude 5 degrees, and longitude 5 degrees, and standardized by GLM (step 2). Create a test dataset for all combinations of strata from variables used for GLM and obtain GLM predicted values for each record (step 3). The value is averaged by year to obtain the least-square means. Next, the test data is limited to correspond to the area index data. At this time, the strata where the area index does not exist (for example, 45S in Area 4 where located at 30-40S) is excluded (step 4). The GLM predicted value is multiplied by the area index (CS that does not depend on the year or VS that depends on year), and then averaged by year, which is the abundance index (step 5). In this document, it is clarified which steps the data or statistics are.

The following are known in our work in 2019: It is not an extreme value until LS-mean (step 3) of GLM, but it becomes an extreme value after multiplying the area index (step 5), that 40S of Area 8 is related, that the value become extreme in Base model, but not in Reduced Base model. Although the cause was not identified last year, it was suspected that the Year\*Area interaction was involved.

Analysis was performed using core vessel data up to 2019 created in 2020. See other documents (Itoh 2020, Itoh and Takahashi 2020) for data and methods. Even when the data was updated to 2019, the value for 2018 remained anomalously high (Fig. 1). In the test data of step 4, GLM predicted values had three extremely large records out of 2,278. As an experiment, when the resource abundance index in step 5 was calculated by excluding these three data, the 2018 anomaly disappeared, and it was identified that these three cases had a large impact (Fig. 2). All three were from 2018, Area 8, 40S, July, August, and September.

As a result of detailed examination, it was found that the cause was that the estimated value of the year\*Area interaction became abnormal due to the unbalanced number of

data by latitude in Area 8 and Area 9. In 2018, there were few 35S (35-39S) data in Area 9 (Table 1). Those data of 35S in Area 9 were mainly from April and September, which were out of the main fishing season, and the CPUE of SBT was low. On the other hand, there are abundant data on 40S in Area 9. In Area 8, there are many 35S data but no 40S (40S-44S) data. In the calculation of GLM including year\*Area interaction, it was required that not only data for Area 8 and Area 9 in 2018 to exist, but corresponding data at the same latitude.

As an experiment, we changed the data of 40S and 41S of Area 9 in the dataset (step 1) to 35S in 5 degree strata latitude. With this simple operation, the 2018 anomaly disappeared, and the cause could be identified (Fig. 3).

## 対策

3つの解決候補を示す。第1方法は前述のようにデータに修正を加える方法である (BaseA と称す)。第2方法は、Year\*Area 項が問題となったことを考慮して GLM の Base モデルから Year\*Area 項を除いた。Reduced-base (RedB) よりも Year\*Lat 項が多い。第3方法は、GLMM を用いて Year\*Area 項をランダム項として扱う方法である。この方法では欠損値が存在しても推定値が得られる。

3方法によって2018年の指数値は2017年と2019年との中間的な値となり、増加は穏やかなものとなった (図4、表2)。3方法間の違いは小さかった。

AIC は Base モデルに対して、第1方法はほぼ同じ (データ数と自由度は同じだがデータが異なるので厳密には比較は不適當かもしれない)、第2方法は高く、第3方法はさらに高かった (表3)。BIC は Base モデルに対して、第1方法は少し高く、第2方法は低く、第3方法はわずかに低かった。

第1方法の長所は、今後も Base モデルで Year\*Area 項を考慮できることにある。Nominal CPUE の関係は年代によって海区によって異なっており (図6 of Itoh and Takahashi 2020 参照)、Year\*Area を考慮することには合理性がある。短所はデータの修正が恣意的であること、今後と同様の問題が生じた場合の対処方法が未定であることである。第2方法の長所は、AIC、BIC がほどほどに良いこと、過去から将来まで一貫した方法になることである。短所は、Year\*Area 項を無視することである。第3方法の長所も過去から未来までの一貫性にある。短所は AIC、BIC の当てはまりが悪いことである。

## Solutions

I present three possible solutions here. The first method is to modify the data as described above (called BaseA). The second method removes the year\*Area term from the GLM Base model, considering that the year\*Area term has become a problem. Note that there is more year\*Lat terms than Reduced-base model (RedB). The third method is to treat the year\*Area term as a random effect term using GLMM. With this method, an estimated value can be obtained even if there are missing values.

The calculated abundance index value for 2018 was midway between 2017 and 2019, and the increase moderated by the three methods (Figure 4, Table 2). The difference

between the three methods was small.

AIC is almost the same as the Base model in the first method (the number of data and the degree of freedom are the same, but the data may be different, so the comparison may be inappropriate), the second method is higher, and the third method is further higher (Table 3). The BIC was slightly higher in the first method, lower in the second method, and slightly lower in the third method than the Base model.

The advantage of the first method is that the year\*Area term can be taken account in the Base model in the future. The relationship of nominal CPUE differs depending on the area depending on year (see Figure 6 of Itoh and Takahashi 2020), and it is reasonable to consider the year\*Area interaction. The disadvantage is that the modification of the data is arbitrary, and what to do if similar problems occur in the future is undecided. The advantage of the second method is that AIC and BIC are reasonably good, and the method is consistent from past to future. The disadvantage is that it ignores the year\*Area interaction term. The advantage of the third method is that it is consistent from past to future. The disadvantage is that AIC and BIC do not fit well.

## 今後のチェック

図 5 に年・緯度、図 6 に年・海区別の推定値を示す。ステップ 3, 4, 5 のそれぞれで示した。BaseA は CPUE 小会合において使用しないことが合意されているが、参考のため含めた。また、CPUE 小会合で検討された GAM シリーズはトレンドがどれも類似していたが、そのうち gamx9 を参考のために含めた。2018 年の異常値はステップ 3 の段階から 40S や 8 海区で見られている。こうしたチェックを行っておけば異常値に気づくことができ、原因を早期に究明できたかもしれない。今後は、資源量指数の最終結果だけでなく、こうした途中経過や推定値の最初の段階についても注意深く検討をしていくべきであろう。

ただし異常値が見られる場合にも、面積重みづけがされるので最終的な資源量指数への影響が小さい場合もある。例えば、4 海区や縁辺の緯度（30S や 45S）では比較的大きな変動が見られるものの、面積指数が小さいために資源量指数への影響は小さい。影響の度合いが集計方法や重みづけの扱いによって異なるので、どのステップかを明確にすることが必要であろう。

## Future check

Figure 5 shows the statistics by year and latitude, and Figure 6 shows that by year and Area. Those are shown in steps 3, 4, and 5, respectively. BaseA has been agreed not to be used in the CPUE small group, but included for reference. In addition, gamx9 was included for reference that several GAM series examined at the CPUE small group had similar trends. Outliers in 2018 have been seen in 40S and Area 8 since step 3. If such a check is performed, the anomalous value may be noticed and the cause may have been clarified early. In the future, it is necessary to carefully consider not only the final result of the abundance index, but also the progress and the initial stage of the estimation.

However, even if an abnormal value is seen in earlier steps, the area weighting is applied, so that the final effect on the abundance index may be small. For example, although there are relatively large fluctuations in Area 4 and marginal latitudes (30S and 45S), the area weighting factor is small, so the effect on the abundance index is small. Since the degree of influence depends on the data making method and the treatment of weighting, it will be necessary to clarify which step.

## **Reference**

- Anonymous (2019) Report of the twenty fourth meeting of the Scientific Committee, CCSBT. September 2019. Cape Town, South Africa.
- Itoh, T. (2020) Change in operation pattern of Japanese southern bluefin tuna longliners in the 2019 fishing season. CCSBT-OMMP/2006/10.
- Itoh, T., and Takahashi, N. (2019) Update of the core vessel data and CPUE for southern bluefin tuna in 2019. CCSBT-OMMP/1906/09.
- Itoh, T., and Takahashi, N. (2020) Update of the core vessel data and CPUE for southern bluefin tuna in 2020. CCSBT-OMMP/2006/11.

Table 1. Number of operations in the original dataset in Area 8 and 9 in 2018 by one degree latitude. Records in enclosed in Area 9 was assigned to be 35S in five degrees strata in an experiment.

Latitude	Area8	Area9
45S		14
44S		232
43S		712
42S		767
41S		174
40S		42
39S		4
38S	85	3
37S	175	1
36S	56	18
35S	36	11

Table 2. Estimated abundance index from the core vessel CPUE

1) W0.8

Year	Nominal	Base	Reduced Base	Base with SxS	Base A	Base_ noYA	GLMM
1969		2.2841	2.2841	2.2841	2.2841	2.28411	2.28411
1970		2.2268	2.2268	2.2268	2.2268	2.22684	2.22684
1971		2.0654	2.0654	2.0654	2.0654	2.06542	2.06542
1972		2.1669	2.1669	2.1669	2.1669	2.16691	2.16691
1973		1.8263	1.8263	1.8263	1.8263	1.82627	1.82627
1974		1.8989	1.8989	1.8989	1.8989	1.89893	1.89893
1975		1.4556	1.4556	1.4556	1.4556	1.45555	1.45555
1976		1.8715	1.8715	1.8715	1.8715	1.87152	1.87152
1977		1.6556	1.6556	1.6556	1.6556	1.65562	1.65562
1978		1.4300	1.4300	1.4300	1.4300	1.43003	1.43003
1979		1.1472	1.1472	1.1472	1.1472	1.14724	1.14724
1980		1.3862	1.3862	1.3862	1.3862	1.3862	1.3862
1981		1.3103	1.3103	1.3103	1.3103	1.31025	1.31025
1982		1.0285	1.0285	1.0285	1.0285	1.02855	1.02855
1983		1.0103	1.0103	1.0103	1.0103	1.01026	1.01026
1984		1.0261	1.0261	1.0261	1.0261	1.0261	1.0261
1985		0.8578	0.8578	0.8578	0.8578	0.85784	0.85784
1986	2.089	0.6311	0.6463	0.6390	0.6315	0.62907	0.63351

CCSBT-OMMP/2006/12

Year	Nominal	Base	Reduced Base	Base with SxS	Base A	Base_ noYA	GLMM
1987	2.132	0.6433	0.6640	0.6340	0.6439	0.65552	0.65133
1988	2.059	0.5402	0.5216	0.5816	0.5409	0.51757	0.53373
1989	1.983	0.5047	0.5064	0.5433	0.5055	0.51105	0.51229
1990	1.963	0.5341	0.5843	0.4731	0.5341	0.56587	0.54881
1991	1.895	0.4385	0.5037	0.4260	0.4387	0.45911	0.44894
1992	2.251	0.5426	0.6068	0.5109	0.5430	0.56639	0.55353
1993	2.970	0.7279	0.6979	0.7048	0.7272	0.7168	0.72097
1994	2.667	0.6890	0.5832	0.6992	0.6889	0.65119	0.68339
1995	2.715	0.7298	0.7376	0.7890	0.7297	0.76133	0.74033
1996	2.426	0.5882	0.5578	0.6171	0.5882	0.56046	0.57914
1997	2.083	0.5143	0.5478	0.4943	0.5145	0.53114	0.52287
1998	2.204	0.5574	0.5790	0.5315	0.5575	0.57612	0.5663
1999	2.321	0.5653	0.5785	0.5407	0.5653	0.59189	0.57685
2000	2.413	0.5337	0.5233	0.5181	0.5338	0.52554	0.52977
2001	2.644	0.6009	0.6122	0.5908	0.6008	0.62133	0.60999
2002	3.441	0.9388	0.8047	0.8752	0.9372	0.83485	0.85656
2003	2.564	0.6715	0.6935	0.6310	0.6708	0.71679	0.68801
2004	1.962	0.6386	0.6751	0.6493	0.6386	0.65726	0.64359
2005	1.851	0.5287	0.5352	0.6795	0.5284	0.53562	0.52404
2006	1.350	0.3747	0.3524	0.3739	0.3748	0.36126	0.37428
2007	1.407	0.2796	0.3264	0.3021	0.2797	0.28616	0.28505
2008	1.879	0.5793	0.5147	0.5478	0.5794	0.52002	0.56906
2009	2.644	0.7507	0.7040	0.6531	0.7510	0.74965	0.74021
2010	3.676	0.9866	0.6925	0.9654	0.9858	0.94277	0.97359
2011	3.510	0.9013	0.7559	0.9153	0.9013	0.8743	0.88188
2012	3.360	1.0718	0.7447	1.0080	1.0718	1.02282	0.99859
2013	3.890	1.0466	0.8822	1.0504	1.0473	1.05189	1.02441
2014	4.185	1.2125	0.9254	1.0424	1.2121	1.24432	1.20784
2015	5.651	1.3216	1.0575	1.3802	1.3213	1.41746	1.34381
2016	4.497	1.2336	1.0639	1.0872	1.2339	1.25188	1.21903
2017	5.523	1.2996	0.8875	1.4757	1.2952	1.35782	1.31097
2018	6.054	2.2809	1.3388	2.4186	1.5980	1.59363	1.61264
2019	7.780	1.7429	1.3438	2.2209	1.7414	1.6482	1.65602



## 2) W0.5

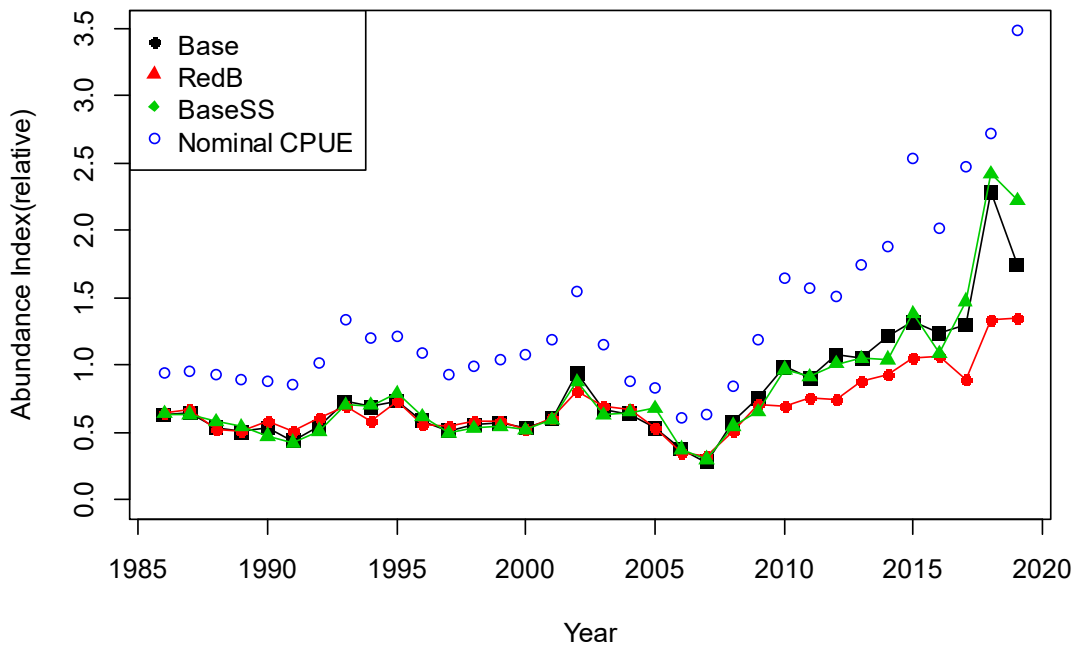
Year	Nominal	Base	Reduced Base	Base with SxS	Base A	Base_ noYA	GLMM
1969		2.4934	2.4934	2.4934	2.4934	2.49337	2.49337
1970		2.4169	2.4169	2.4169	2.4169	2.41687	2.41687
1971		2.2054	2.2054	2.2054	2.2054	2.20542	2.20542
1972		2.2273	2.2273	2.2273	2.2273	2.22735	2.22735
1973		1.9271	1.9271	1.9271	1.9271	1.92713	1.92713
1974		1.9710	1.9710	1.9710	1.9710	1.97095	1.97095
1975		1.4974	1.4974	1.4974	1.4974	1.49745	1.49745
1976		1.9279	1.9279	1.9279	1.9279	1.92793	1.92793
1977		1.6850	1.6850	1.6850	1.6850	1.685	1.685
1978		1.3820	1.3820	1.3820	1.3820	1.38202	1.38202
1979		1.2558	1.2558	1.2558	1.2558	1.25578	1.25578
1980		1.3852	1.3852	1.3852	1.3852	1.38524	1.38524
1981		1.2917	1.2917	1.2917	1.2917	1.29166	1.29166
1982		1.0220	1.0220	1.0220	1.0220	1.02197	1.02197
1983		1.0228	1.0228	1.0228	1.0228	1.02279	1.02279
1984		1.0603	1.0603	1.0603	1.0603	1.06032	1.06032
1985		0.8861	0.8861	0.8861	0.8861	0.88613	0.88613
1986	2.089	0.6663	0.6841	0.6718	0.6664	0.6665	0.66836
1987	2.132	0.6692	0.6848	0.6605	0.6696	0.67843	0.67492
1988	2.059	0.5561	0.5264	0.5928	0.5567	0.52366	0.54648
1989	1.983	0.5330	0.5305	0.5653	0.5336	0.53473	0.53871
1990	1.963	0.5294	0.5700	0.4745	0.5294	0.55263	0.54059
1991	1.895	0.4492	0.5036	0.4396	0.4494	0.46439	0.45746
1992	2.251	0.5361	0.5865	0.5094	0.5363	0.55182	0.54407
1993	2.970	0.6664	0.6363	0.6643	0.6659	0.6587	0.66078
1994	2.667	0.5803	0.4926	0.5927	0.5802	0.54877	0.57485
1995	2.715	0.6569	0.6593	0.6918	0.6569	0.68439	0.66504
1996	2.426	0.5318	0.5166	0.5608	0.5318	0.51714	0.52668
1997	2.083	0.4679	0.4966	0.4557	0.4681	0.48153	0.47387
1998	2.204	0.5421	0.5551	0.5117	0.5422	0.55228	0.54855
1999	2.321	0.5449	0.5551	0.5206	0.5449	0.56726	0.55393
2000	2.413	0.4760	0.4649	0.4693	0.4760	0.47303	0.47367
2001	2.644	0.5596	0.5629	0.5495	0.5596	0.57243	0.566
2002	3.441	0.7660	0.6676	0.7185	0.7650	0.6963	0.71173

Year	Nominal	Base	Reduced Base	Base with SxS	Base A	Base_ noYA	GLMM
2003	2.564	0.5596	0.5734	0.5353	0.5591	0.58674	0.569
2004	1.962	0.5774	0.6011	0.5786	0.5774	0.58764	0.57953
2005	1.851	0.4823	0.4842	0.5961	0.4821	0.47934	0.47691
2006	1.350	0.3284	0.3203	0.3285	0.3285	0.32367	0.32885
2007	1.407	0.2327	0.2616	0.2501	0.2327	0.23708	0.23662
2008	1.879	0.4309	0.4093	0.4055	0.4309	0.40415	0.42605
2009	2.644	0.5543	0.5334	0.4799	0.5544	0.56431	0.54927
2010	3.676	0.6920	0.5073	0.6748	0.6915	0.67127	0.68545
2011	3.510	0.6544	0.5572	0.6689	0.6543	0.64056	0.64285
2012	3.360	0.7624	0.5377	0.7201	0.7624	0.71844	0.71173
2013	3.890	0.7317	0.6192	0.7455	0.7321	0.73418	0.7178
2014	4.185	0.8654	0.6625	0.7387	0.8650	0.88469	0.86276
2015	5.651	0.9527	0.7518	0.9877	0.9526	1.00794	0.96546
2016	4.497	0.8977	0.7678	0.7895	0.8979	0.90959	0.88917
2017	5.523	0.9429	0.6330	1.0373	0.9403	0.97909	0.94982
2018	6.054	1.5246	0.9373	1.5814	1.1244	1.12889	1.13831
2019	7.780	1.2017	0.9322	1.4904	1.2008	1.14806	1.15172

Table 3. AIC and BIC of GLM/GLMM models for the core vessel CPUE series.

Model	AIC	BIC	Dataset
Base	7,685	9,292	original
RedB	8,108	8,496	original
BaseA	7,697	9,304	modified
Base_noYA	7,826	8,823	original
GLMM	8,202	9,205	original

W0.8



W0.5

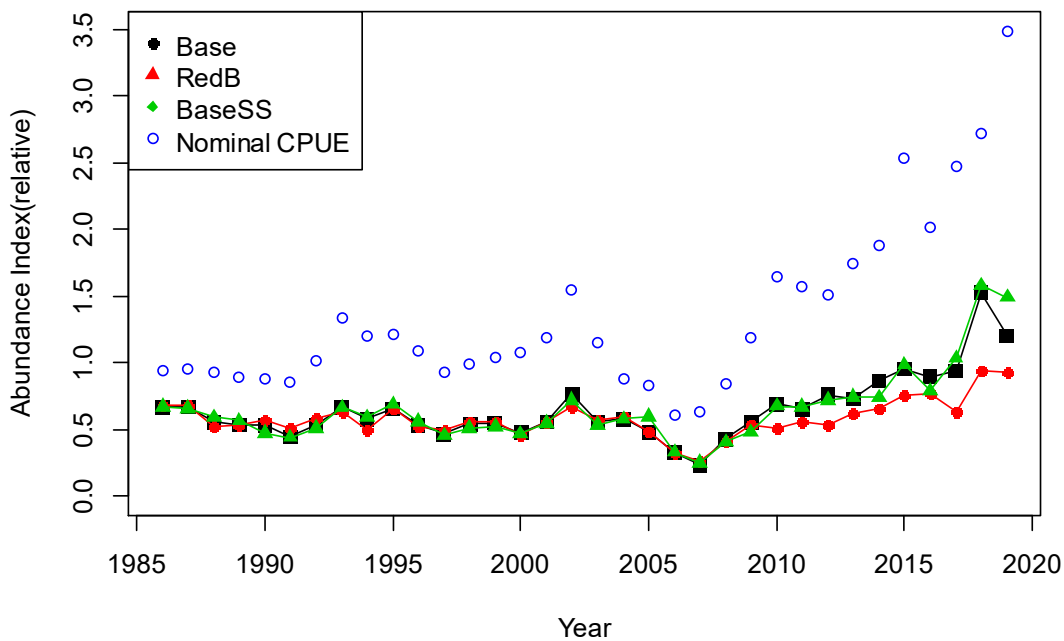


Fig. 1. Area weighed standardized CPUEs.  
Nominal CPUE of the core vessels is also shown.

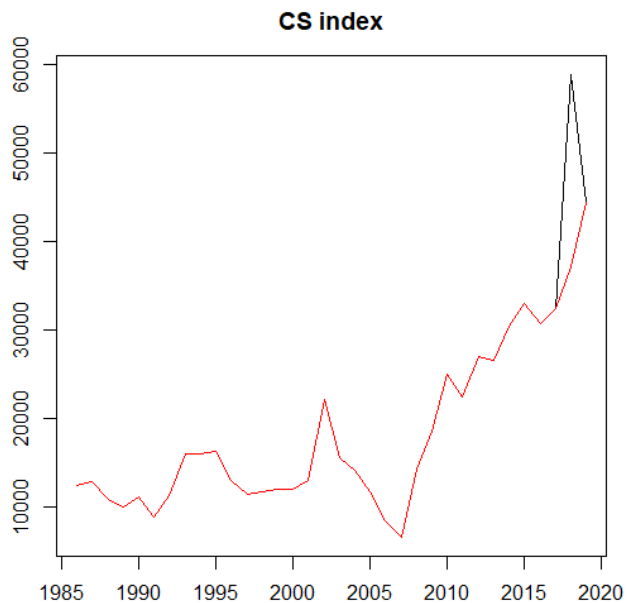


Fig. 2. CS abundance index by Base model. Black line use the original test data. Red line use the test data that eliminated three records in 40S, Area 8, year 2018 in July, August and September whose prediction values were extremely large.

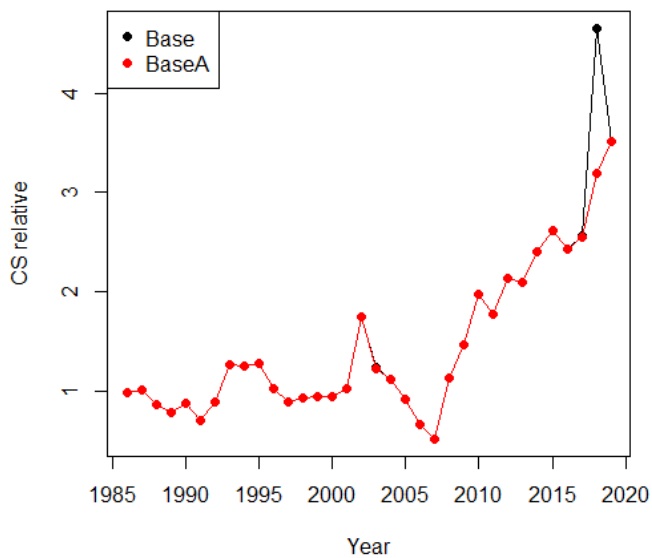


Fig. 3. Experiment of data manipulation of latitude in Area 9 for effect on the 2018 anomalous value (BaseA).

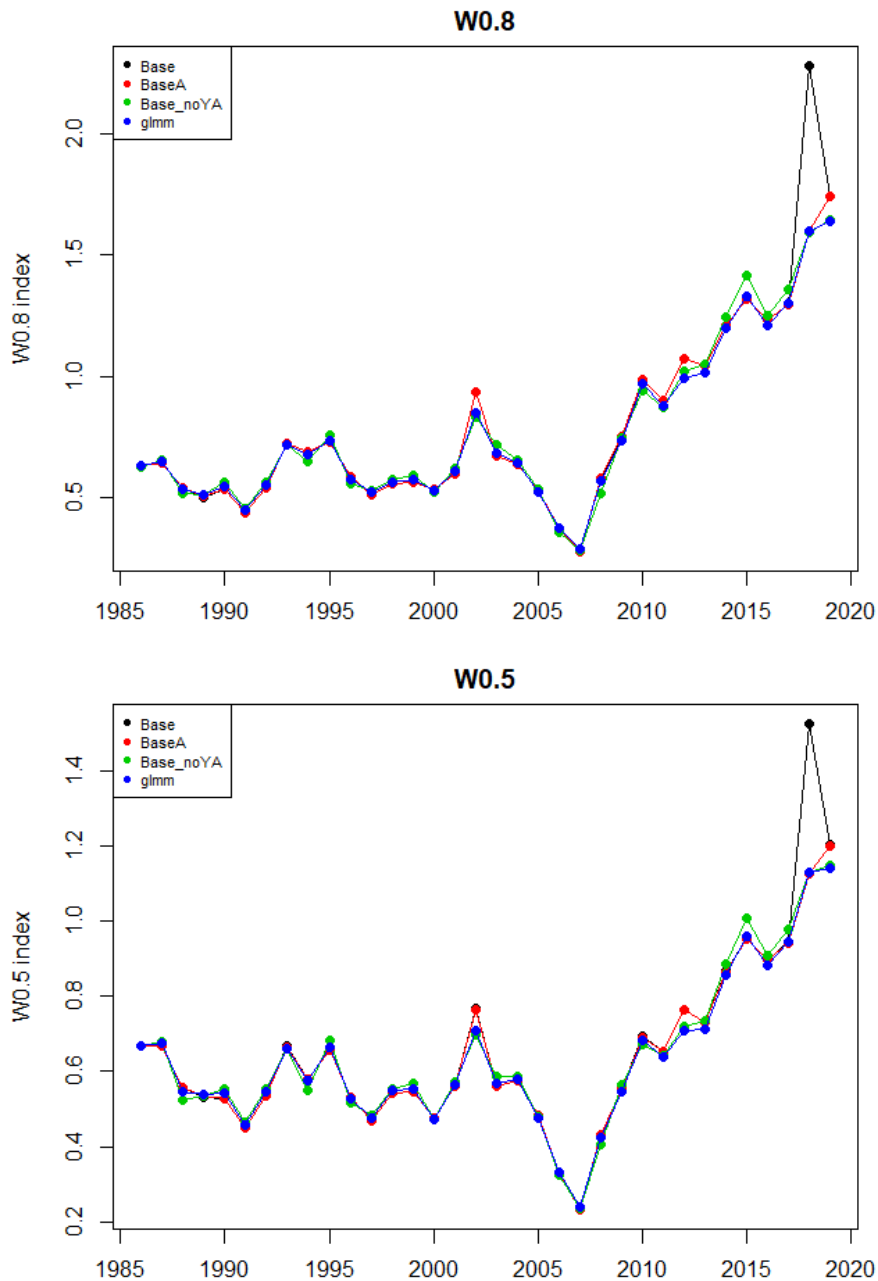


Fig. 4. Area weighed standardized CPUEs by various GLM/GLMM models.

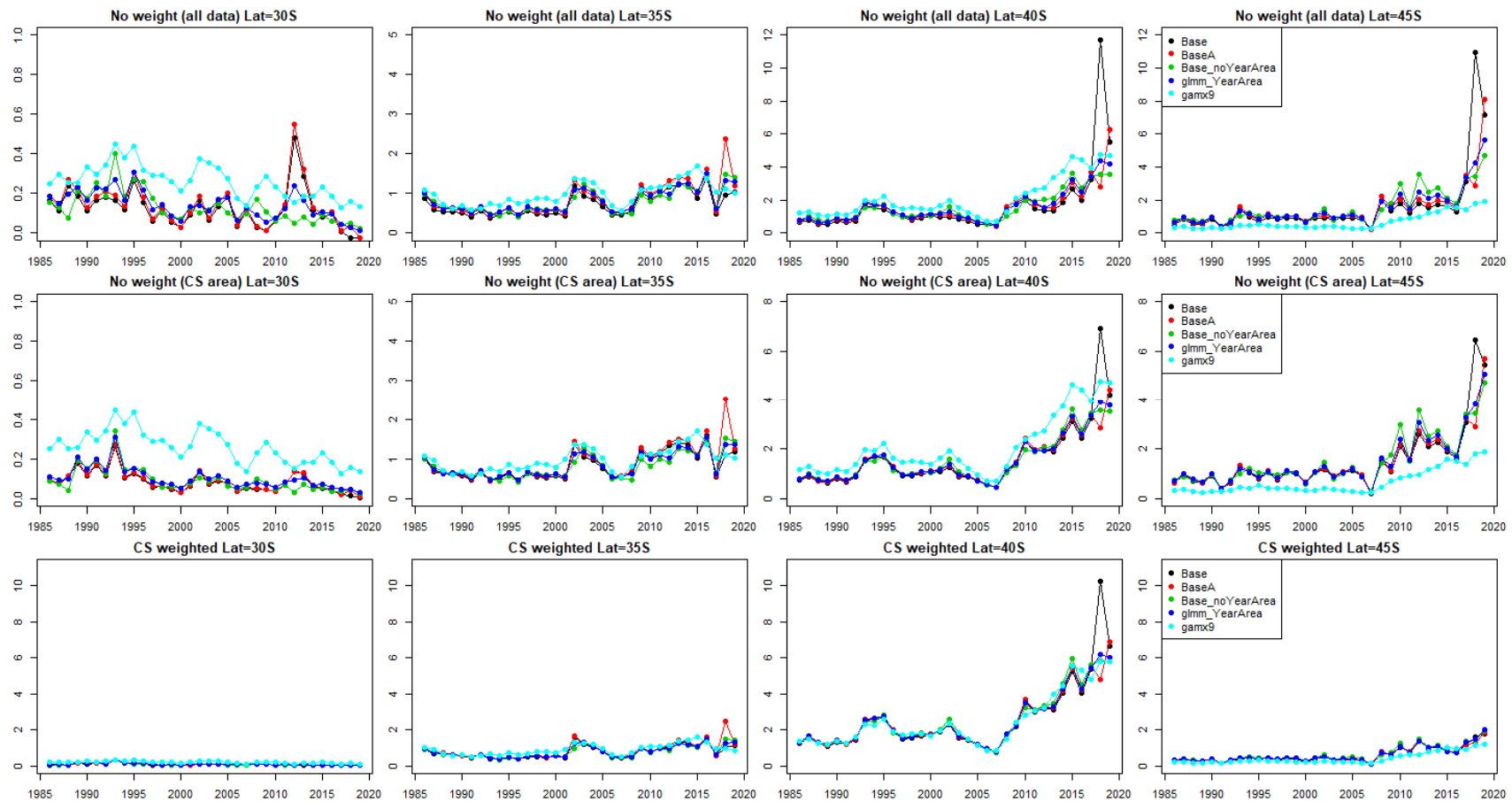


Fig. 5. Statistical value from GLM/GLMM/GAM by latitude.

Upper panels use the test data (Step 3), middle panels use the test data that eliminated strata where CS area weighting data absent (Step 4), and lower panels use the test data that CS area weighted (Step 5). The values are relative to the mean of predicted value of all the test data.

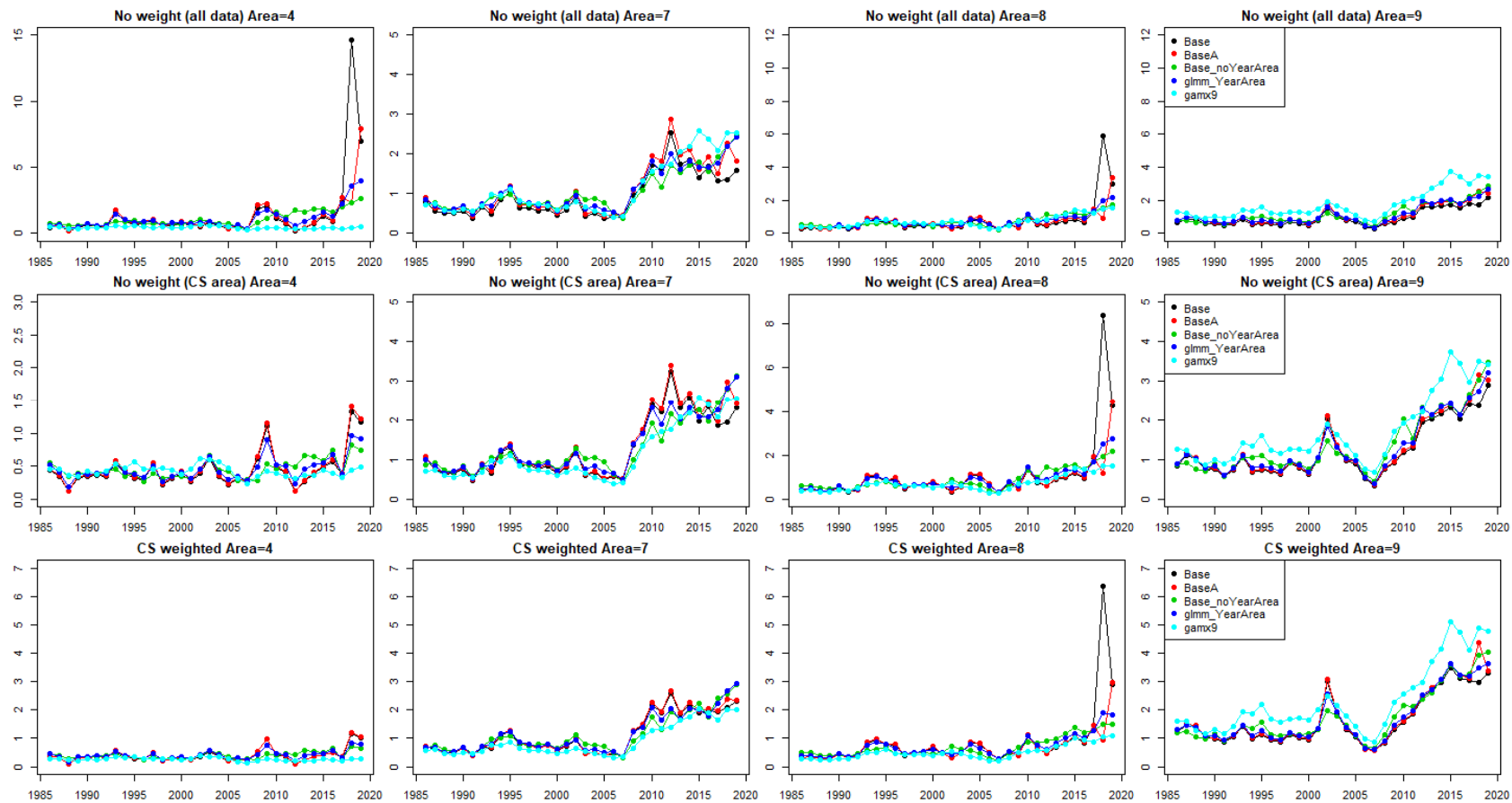


Fig. 6. Statistical value from GLM/GLMM/GAM by Area.

Upper panels use the test data (Step 3), middle panels use the test data and eliminated strata where CS area weighting data absent (Step 4), and lower panels use the test data and CS area weighted (Step 5). The values are relative to the mean of predicted value of all the test data.