CCSBT-ESC/2308/BGD 04
(*Previously* CCSBT-OMMP/2306/09)
(ESC Agenda item 7.2)

Further examination of CPUE abundance index using GAM for southern bluefin tuna based on predicted values

予測値に基づくミナミマグロについての GAM を用いた CPUE 資源豊度指数のさらなる検討

Tomoyuki ITOH and Norio TAKAHASHI
伊藤智幸・高橋紀夫
Fisheries Resources Institute, Japan Fisheries Research Agency
水産研究･教育機構 水産資源研究所

要約

CCSBT で用いている CPUE 資源量指数では、2018 年に異常に高い値となった問題があった。非操業時空間での CPUE 予測値の影響を具体的に示し、コア船 CPUE における 2018 年値の問題の原因をより明確にした。具体的な操業データが無かった時空間に対して異常に高い推定値となったことで、コア船 CPUE の 2018 年と 2019 年の値が高くなったことが分かった。一方、GAM_CPUE での 2022 年値が高いことについては、非操業時空間に対する異常に高い予測値の影響ではないことが確認された。

Summary

The CPUE abundance index used by CCSBT had a problem with anomalously high value in 2018. The influence of the CPUE predicted value in the non-operating space-time was specifically shown, and the cause of the problem of the core vessel CPUE value in 2018 was clarified. It was found that the values of the core vessel CPUE in 2018 and 2019 were high due to the anomalously high estimates for the spatiotemporal data for which there was no actual data of fishing operation. On the other hand, it was confirmed that the high value for 2022 in GAM_CPUE is not due to the anomalously high predicted value for non-operating space-time.

1. Introduction

The CPUE abundance Index is an important component in CCSBT southern bluefin tuna (*Thunnus maccoyii*; SBT) management that is directly linked to stock management through TAC determination via MP, as well as stock assessment. Data from the Japanese longline fishery have been used for the CPUE abundance Index. In 2019, a problem was found that resulted in anomalously high values in the index at year of 2018 (CCSBT 2019). At this time, a part of the data set for the core vessel is extracted from the dataset from all vessels and standardized by GLM (called core vessel CPUE). After that, the CCSBT ESC spent two years developing a new standardized method, and in 2022 agreed on a method for creating a new abundance index. In this method, all data from Japanese longline fishing for SBT (Statistical Areas 4-9, April to September) were used to determine abundance indices using the delta two-step approach using GAM (called GAM_CPUE).

In 2023, using the data up to 2022 to calculate the biomass index using the newly agreed method yielded significantly higher values than in the previous year (Itoh and Takahashi 2023). Although the rise was not as anomalously high as the core vessel CPUE in 2018, it did raise concerns about the increase. In addition, while the anomalously high core vessel CPUE value in 2018 was said to be due to the anomalously high predicted value for the time and space in which no fishing operations were conducted (CCSBT 2022a, 2022b), this was shown in the CPUE Working Group Working Group, etc., but it has not been specifically shown in a document.

Therefore, in this document, we specifically show the impact of non-operating time and space by analyzing the predicted values by standardization separately depending on whether they are supported by actual data, and thus, clarify the cause of the problem of the 2018 values in the core vessel CPUE. In addition, we discuss the 2022 rise in GAM_CPUE.

2. Materials and Methods

The data used for standardization in this document was the dataset used for the 2023 work of GAM_CPUE. The dataset contains 797,416 operations from 1969 to 2022. Details are described in Itoh and Takahashi (2023).

Standardization was performed with the GLM model (Itoh and Takahashi 2022) based on the base model for core vessels and the model for GAM (Itoh and Takahashi 2023). The GLM model is as follows.

$$log(CPUE+0.2) = Intercept + Year + Month + Area + Lat5 + BET\_CPUE + YFT\_CPUE + (Month*Area) + (Year*Lat5) + (Year*Area) + Error,$$

where year, month, area, lat5 were treated as factors. glm function of R was used.

A dummy dataset was then created. After creating data with all combinations of year / month / latitude / longitude (using R's expand.grid function), we made a dummy dataset limited to the month / latitude / longitude where the fishing was operated in the past. Consists of 139,104 records be done. The number of operations in actual fishery data was attached to each record. Predicted values were calculated for each combination of variables by GLM and GAM, respectively (both area weighted). The dataset was classified into four groups based on the number of operations actually given. Group 0 has 0 operations, Group 1 has 1 to less than 5 operations, Group 2 has 5 to less than 10 operations, and Group 3 has 10 or more operations.

A higher CPUE is expected in a space-time with a higher number of operations. This is because there would be a high probability that a vessel does not stay in a space-time with a low CPUE, and it is expected that operations are not conducted in a space-time with a low CPUE through the accumulation of historical knowledge. Boxplot is used for visualization.

Note that the GLM results obtained here are different from the core vessel CPUE results. Although the GLM model is the same, the dataset and dummy dataset are different from those used for the core vessel CPUE.

3. Results

Fig. 1 shows the proportion of each group in the dummy dataset by year. The value for 2022 is provisional and may increase as data input work progresses. The proportion has decreased since 1969, indicating a decrease in the proportion of time-space in which operations took place. The decline has continued since 2000. However, it did not decrease significantly around 2018 when it became a problem.

Fig. 2 shows the predicted CPUE values by group in data all years combined, by GLM and GAM. As expected, the time and space with higher operation numbers had higher CPUE. The same figure is shown in Fig. 3, including the boxplot outlier. It can be seen that there are anomalously high predicted outliers in GLM and not any extreme at all in GAM.

A similar figure is shown in Fig. 4 by year. From 1969 to 2007, the box part is wide and the CPUE increases according to the increase in the number of operations. From around 2008, outliers became higher as the box area was compressed in the figure. From 2018, the outliers were particularly high in GLM, and it was significantly different from the figure by GAM. Fig. 5 shows the change in outliers over time in the space-time without operations. Extremely large outliers are seen in 2018 and 2019 in GLM. These came from Area 8 between June to September (Table 1). Outliers in GAM are not extremely high.

4. Discussion

Our analysis reveals that the anomalously high value for core vessel CPUE in 2018 is due to an anomalously high predicted value for the space-time without operations. High predictive values occurred in 2018 and 2019. On the other hand, it was confirmed that the high value for 2022 in GAM_CPUE is not due to the anomalously high predicted value for non-operating space-time. There is no reason not to accept the high 2022 value of GAM_CPUE at this stage.

In the present analysis, we did not explore whether the differences are between GLM and GAM or between the relatively simple GLM model and the complex GAM model including up to 3 interactions. It has already been decided to use GAM_CPUE in CCSBT, and since no problems were observed in the current year's work, it is not essential to investigate the cause. Interestingly, when the core vessel CPUE data was added after 2021, the extremely high value in 2018 disappeared (see Fig. 50 of Itoh and Takahashi 2023). This may have been due to data extraction by core vessel, rather than later data additions, eliminating the problem. The boxplot of predicted values according to the number of operations used in this analysis would be useful for diagnosing the results of GAM_CPUE and other standardized CPUE.

5. References

CCSBT (2019) Report of the Twenty Fourth Meeting of the Scientific Committee. 7 September 2019. Cape Town, South Africa. 121pp.

CCSBT (2022a) Report of the Twelfth Operating Model and Management Procedure Technical Meeting.

20-24 June 2022. Hobart, Australia and Online. 33pp.

CCSBT (2022b) Report of the Twenty Seventh Meeting of the Scientific Committee. 29 August-5 September 2022. Online. 114pp.

Itoh, T. and N. Takahashi (2022) Update work of the core vessel data and CPUE for southern bluefin tuna in 2022. CCSBT-OMMP/2206/07.

Itoh, T. and N. Takahashi (2023) Update of CPUE abundance index using GAM for southern bluefin tuna in CCSBT up to the 2022 data. CCSBT-OMMP/2306/05.

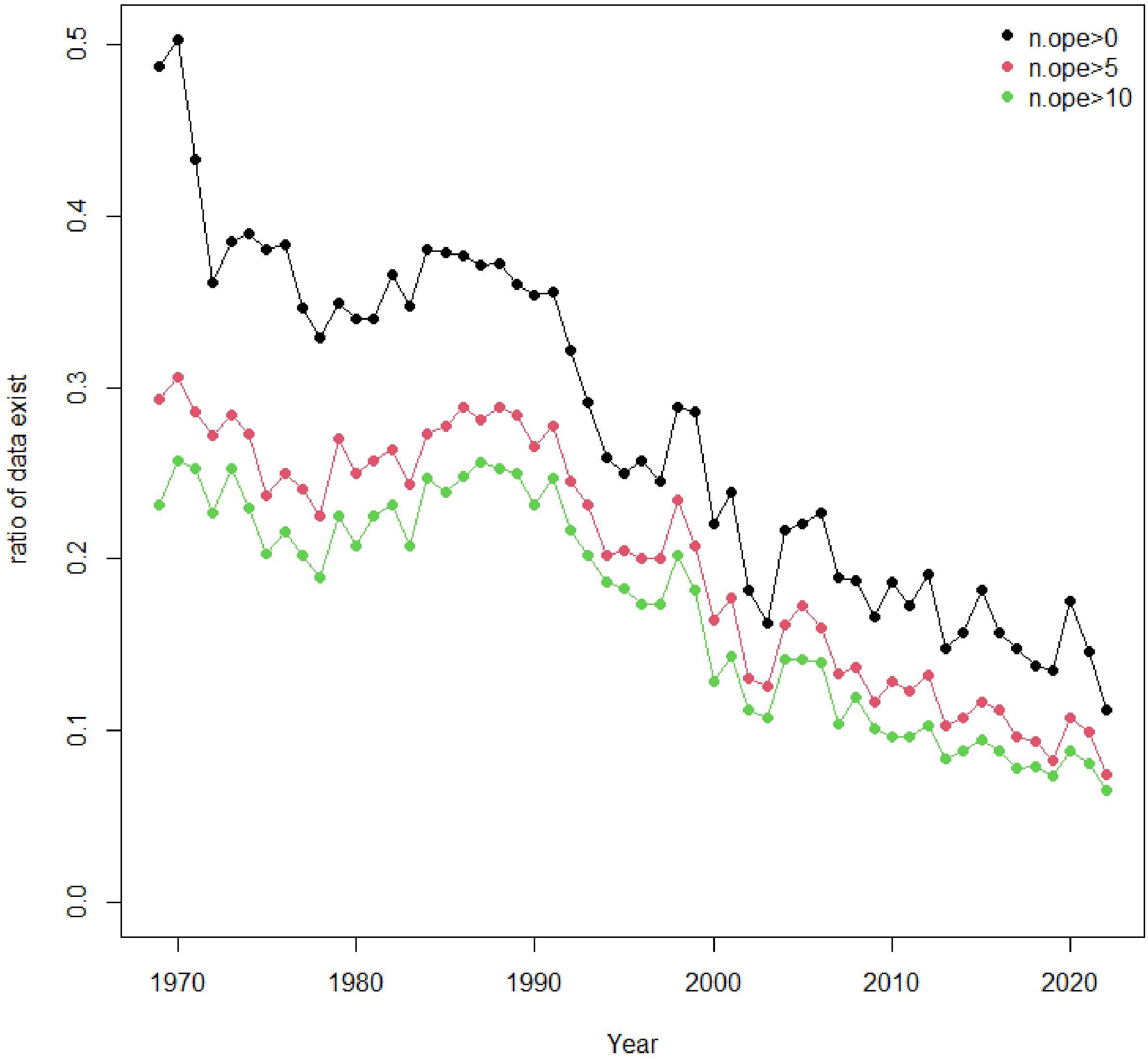


Fig. 1. Proportion of actual data existed in the dummy data used for glm/gam prediction by year.
The number of operations >0, >5 and >10 are shown in black, red and green, respectively.

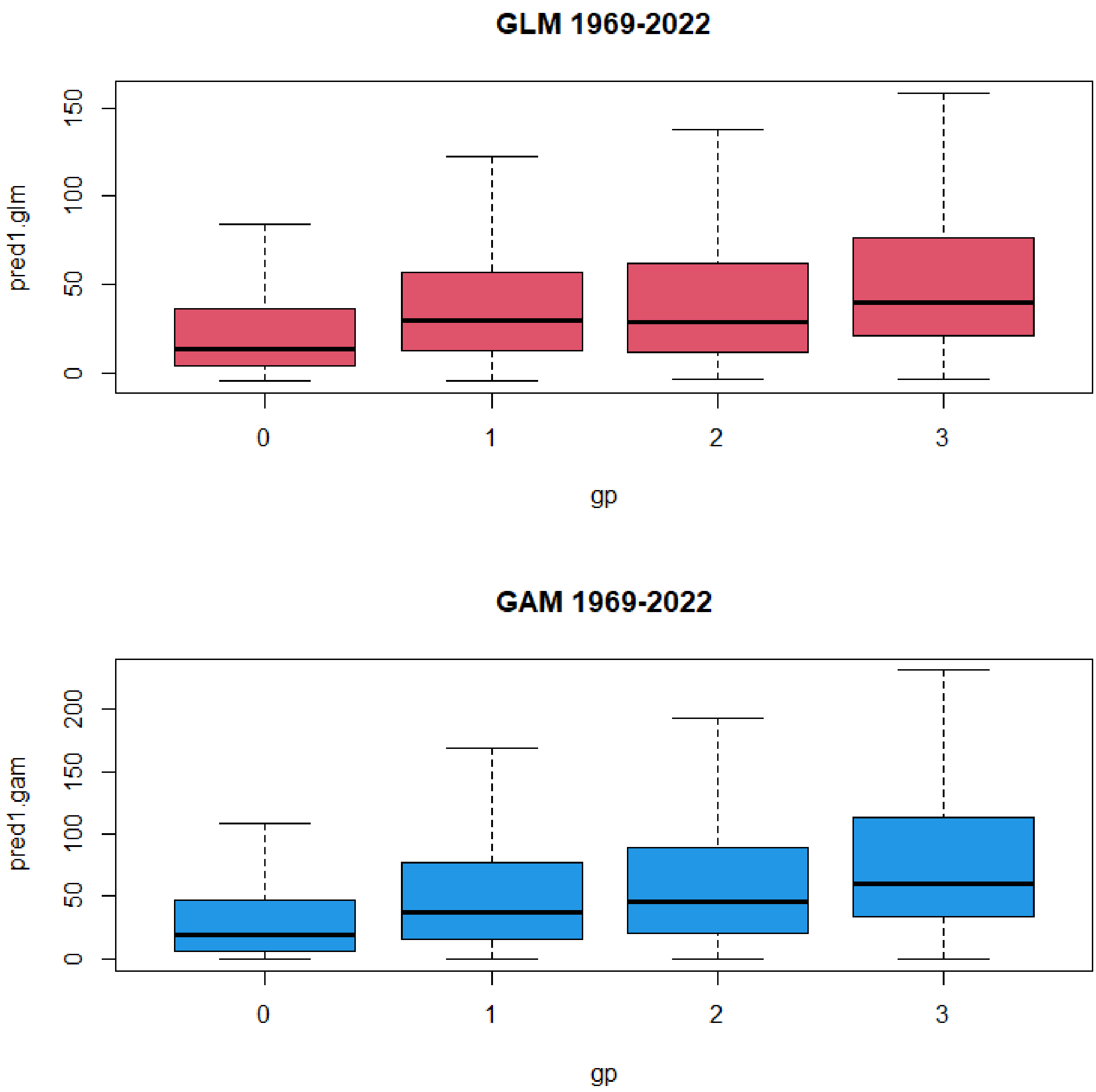


Fig. 2. Boxplot of CPUE predicted values by category group without outliers.
Group 0 is the number of operation in actual data corresponded was 0. Group 1 is >=1 and < 5 operations. Group 2 is >=5 and < 10 operations and Group 3 is > 10 operations. Data in all years were combined.

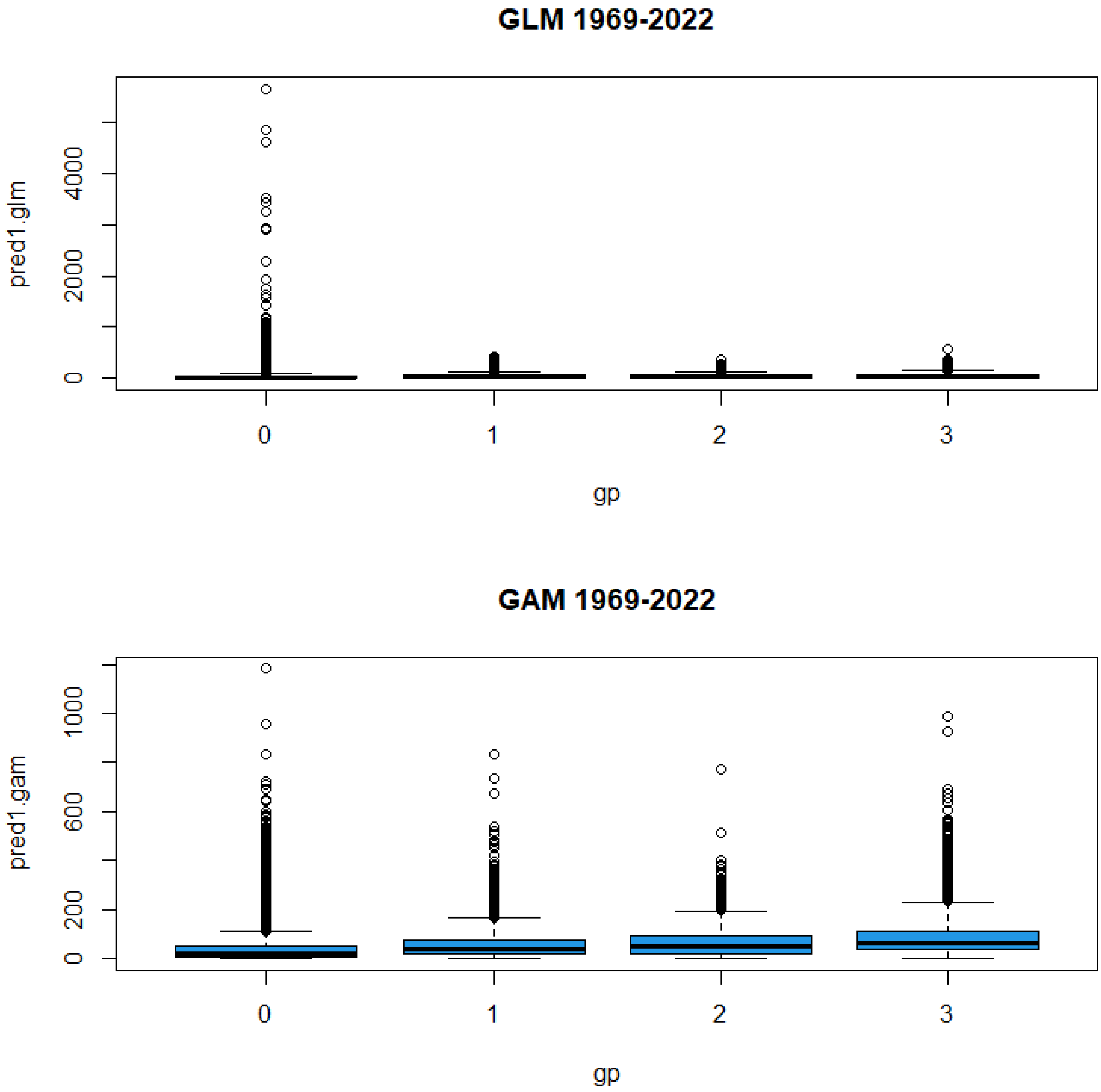


Fig. 3. Boxplot of CPUE predicted values by category group with outliers.
Group 0 is the number of operation in actual data corresponded was 0. Group 1 is >=1 and < 5 operations. Group 2 is >=5 and < 10 operations and Group 3 is > 10 operations. Data in all years were combined.

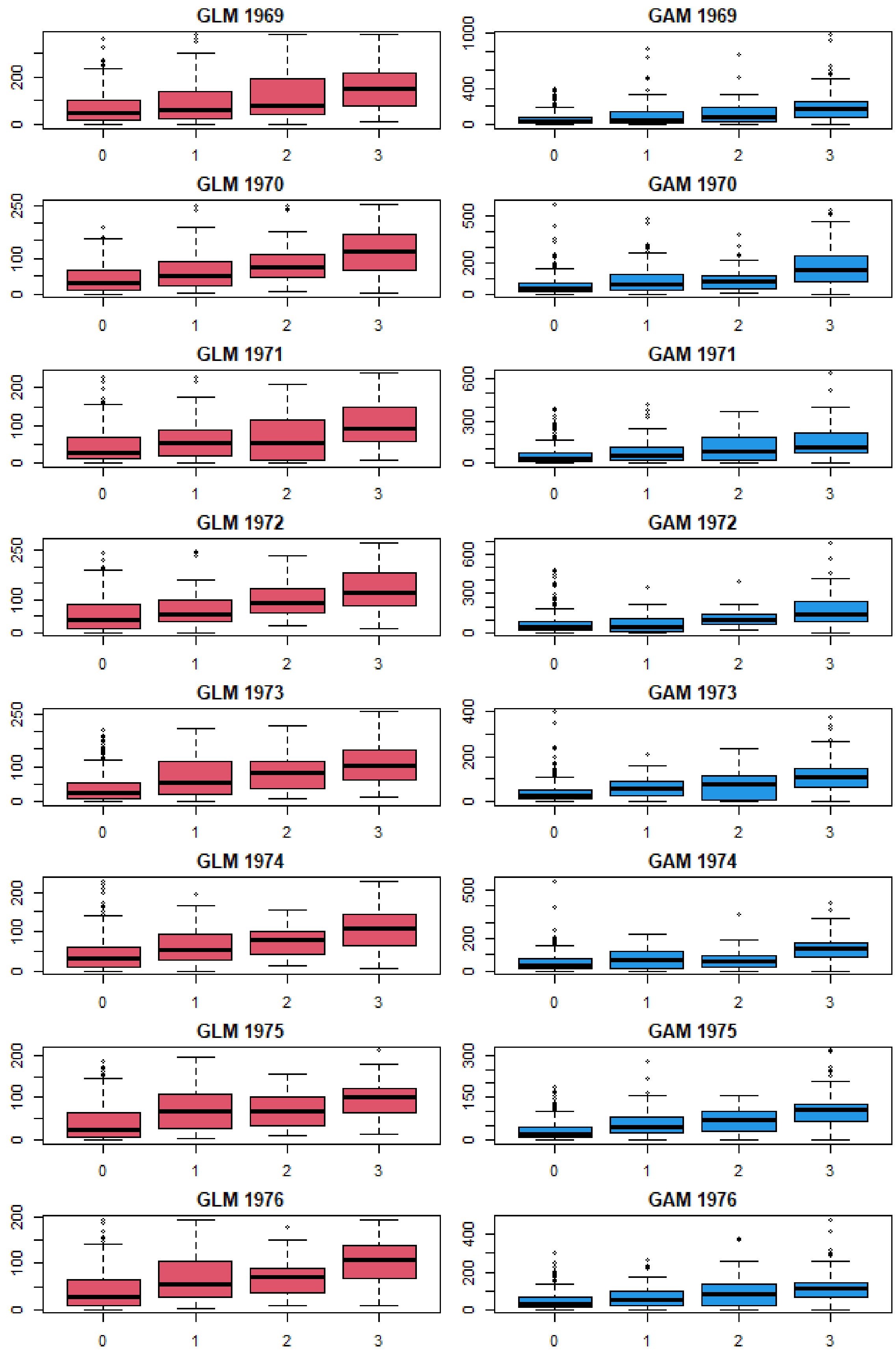


Fig. 4. Boxplot of CPUE predicted values by category group and year with outliers.

Group 0 is the number of operation in actual data corresponded was 0. Group 1 is >=1 and < 5 operations. Group 2 is >=5 and < 10 operations and Group 3 is > 10 operations.

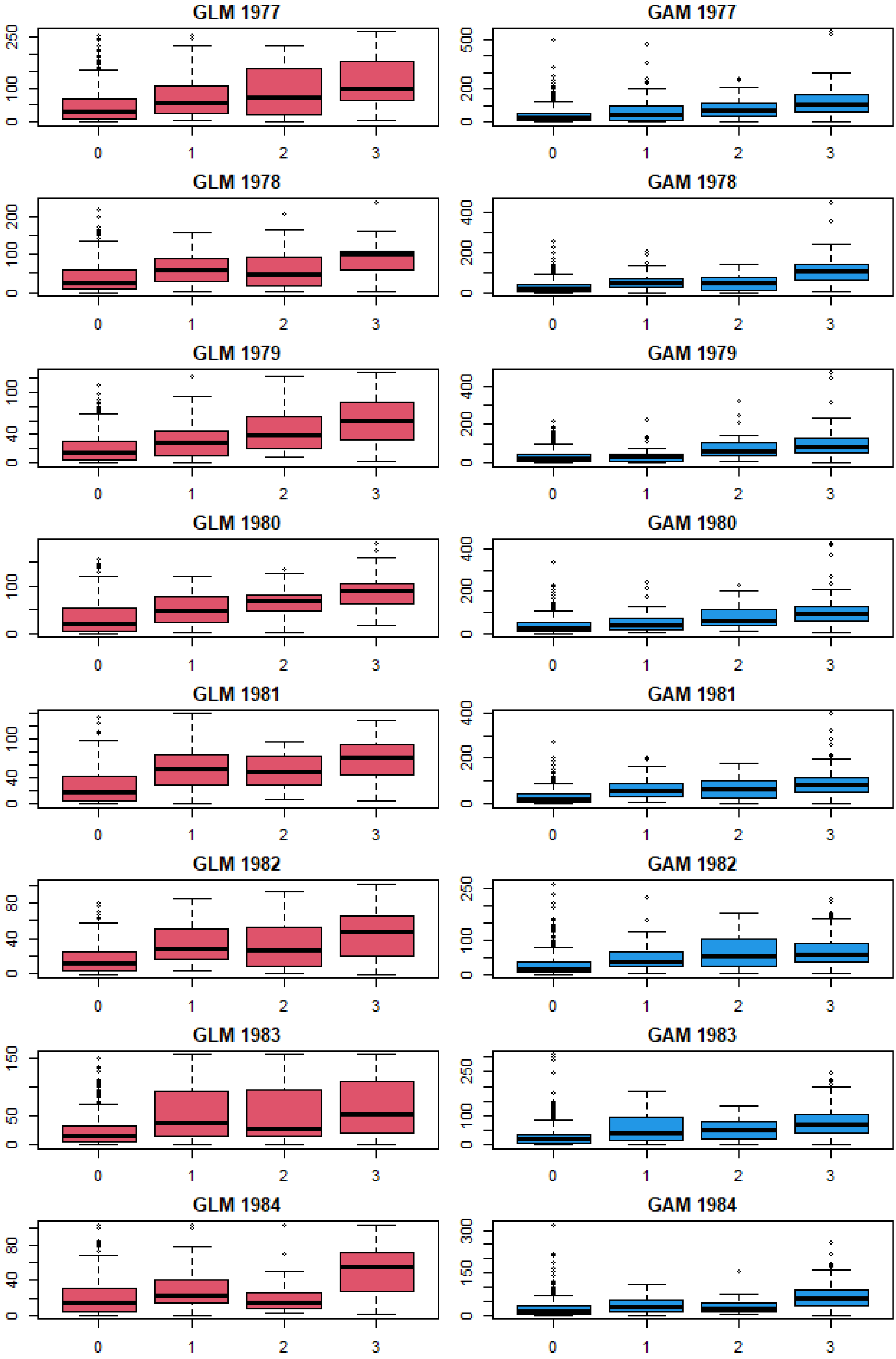

GLM 1977
GAM 1977
GLM 1978
GAM 1978
GLM 1979
GAM 1979
GLM 1980
GAM 1980
GLM 1981
GAM 1981
GLM 1982
GAM 1982
GLM 1983
GAM 1983
GLM 1984
GAM 1984
0
1
2
3

Fig. 4. (Cont'd)

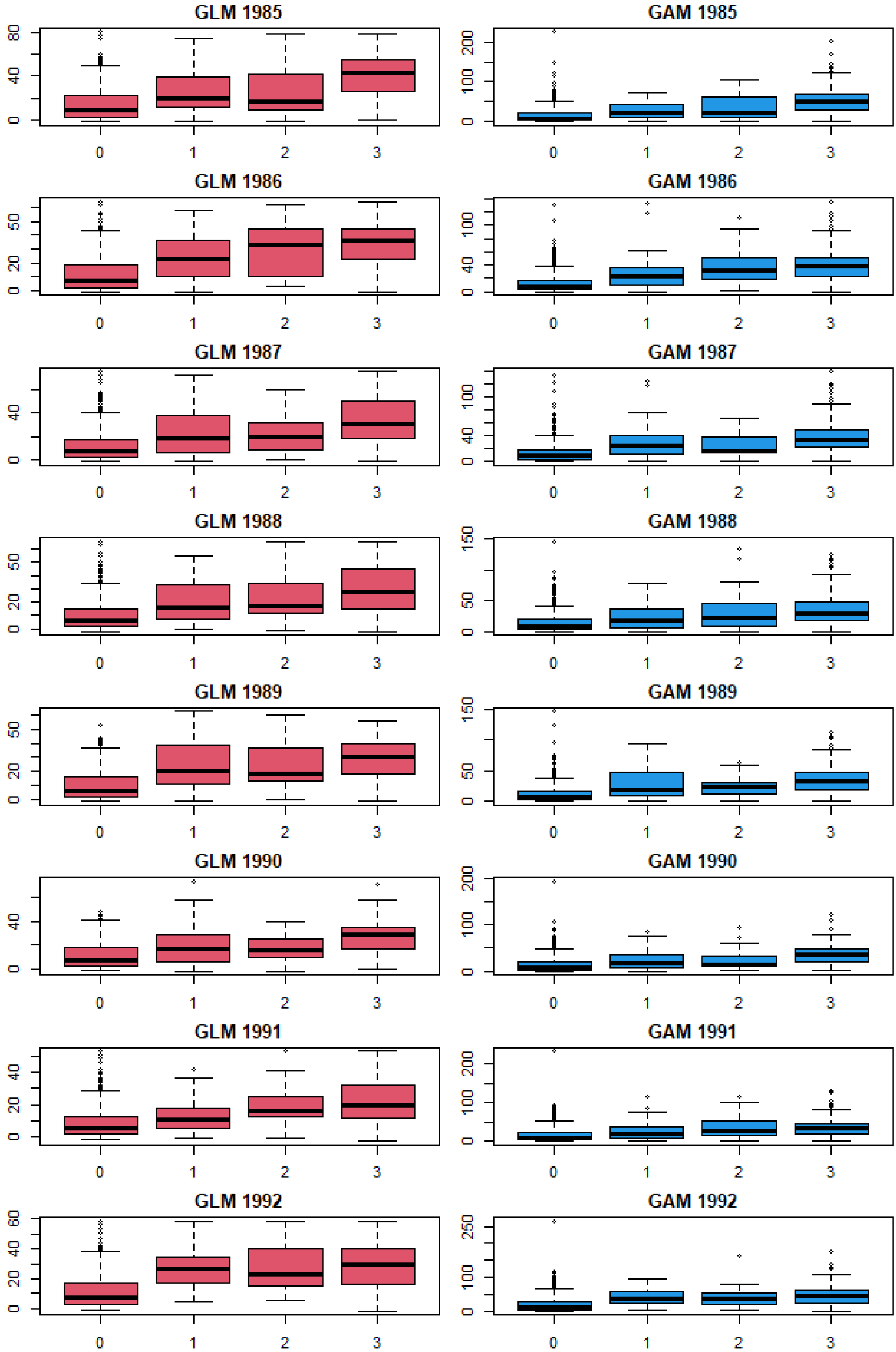

GLM 1985
GAM 1985
GLM 1986
GAM 1986
GLM 1987
GAM 1987
GLM 1988
GAM 1988
GLM 1989
GAM 1989
GLM 1990
GAM 1990
GLM 1991
GAM 1991
GLM 1992
GAM 1992

Fig. 4. (Cont'd)

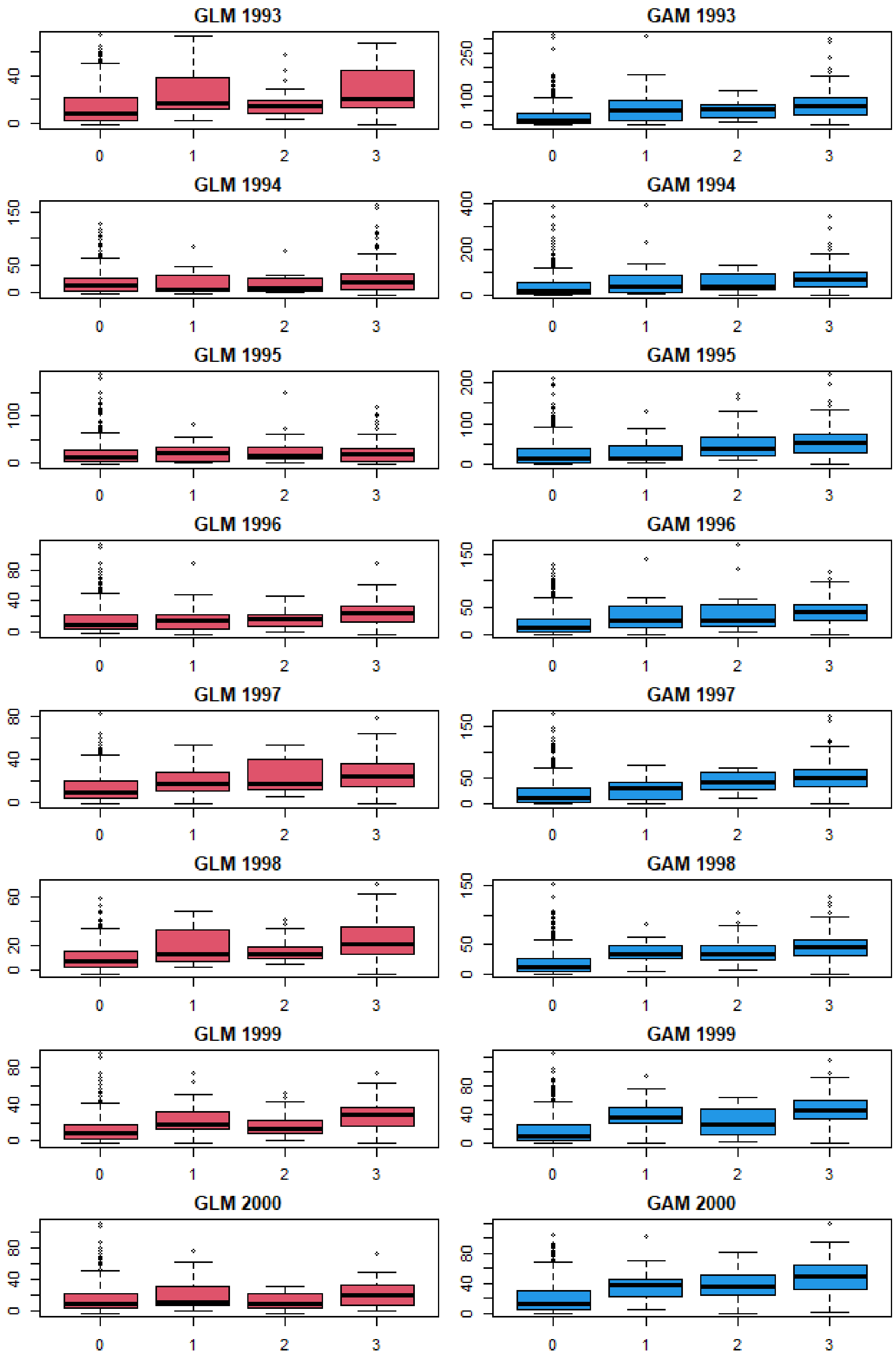

GLM 1993
GAM 1993
GLM 1994
GAM 1994
GLM 1995
GAM 1995
GLM 1996
GAM 1996
GLM 1997
GAM 1997
GLM 1998
GAM 1998
GLM 1999
GAM 1999
GLM 2000
GAM 2000
0 1 2 3

Fig. 4. (Cont'd)

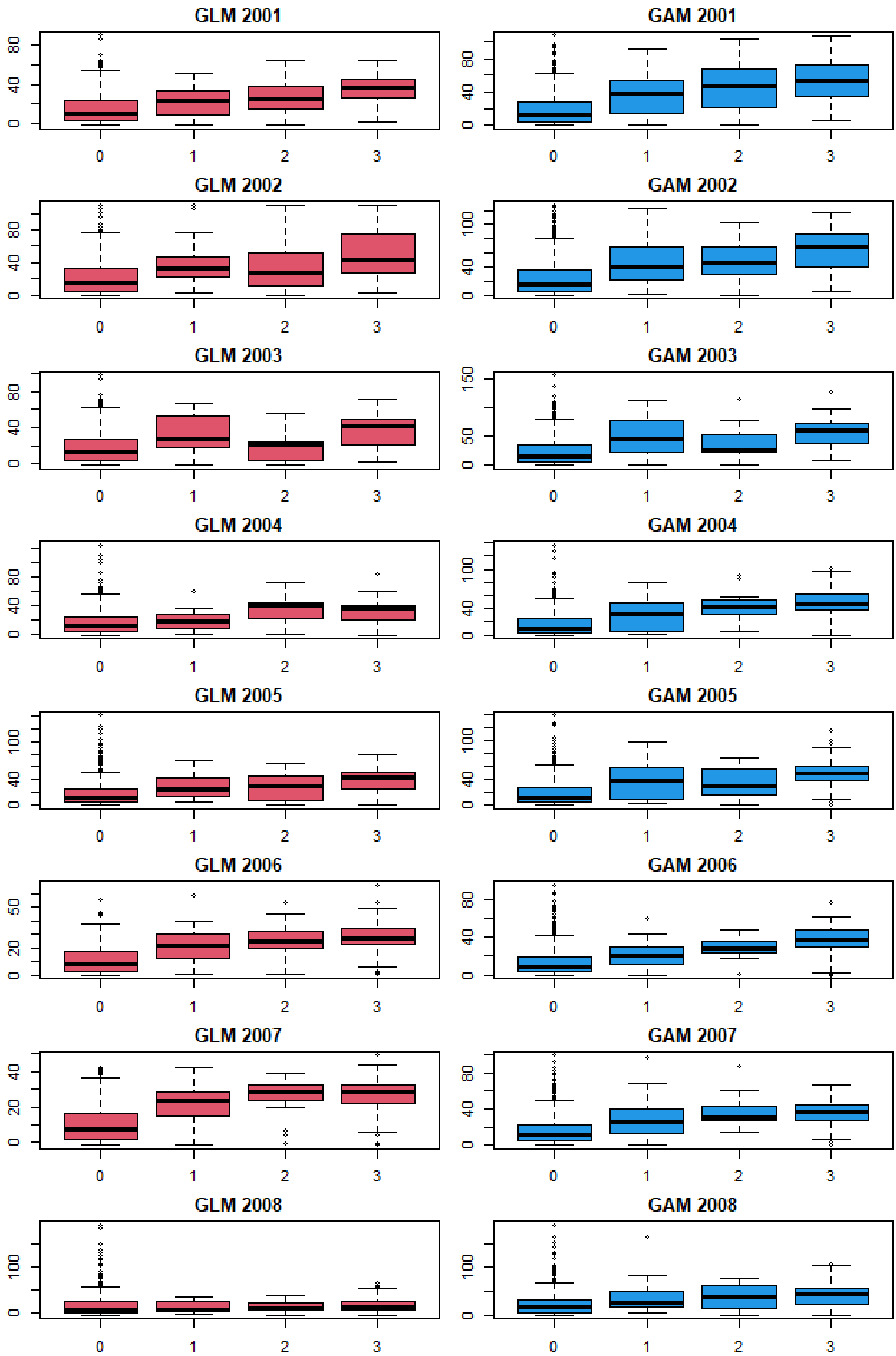

GLM 2001
GAM 2001
GLM 2002
GAM 2002
GLM 2003
GAM 2003
GLM 2004
GAM 2004
GLM 2005
GAM 2005
GLM 2006
GAM 2006
GLM 2007
GAM 2007
GLM 2008
GAM 2008
0
1
2
3

Fig. 4. (Cont'd)

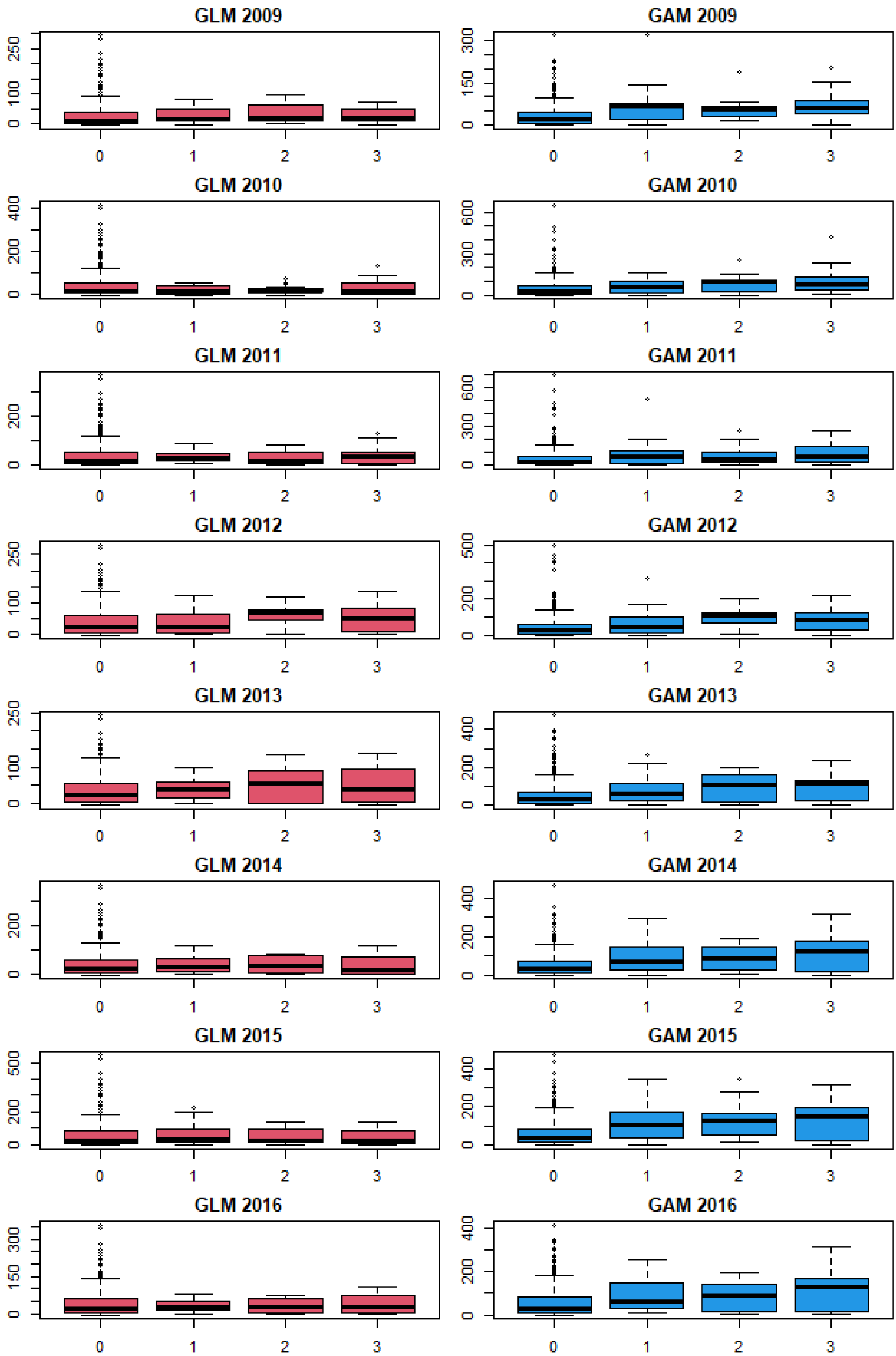

GLM 2009
GAM 2009
GLM 2010
GAM 2010
GLM 2011
GAM 2011
GLM 2012
GAM 2012
GLM 2013
GAM 2013
GLM 2014
GAM 2014
GLM 2015
GAM 2015
GLM 2016
GAM 2016

Fig. 4. (Cont'd)

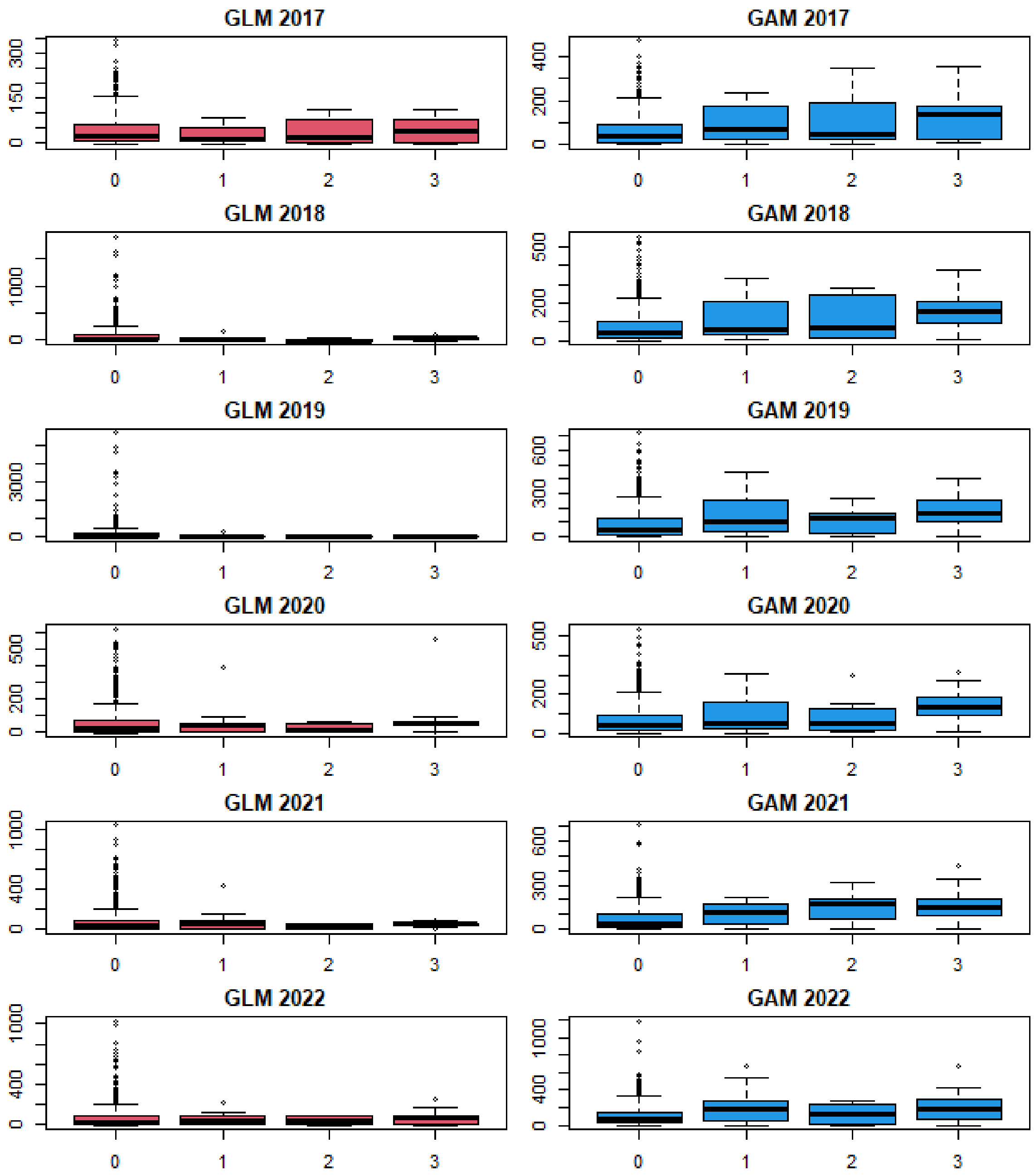
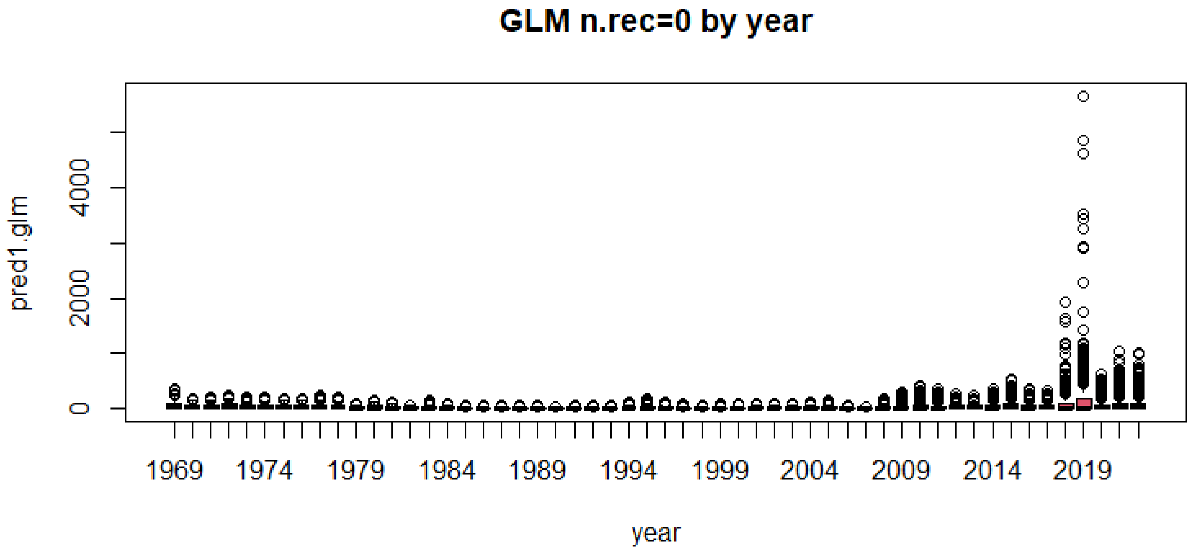

GLM 2017
GAM 2017
GLM 2018
GAM 2018
GLM 2019
GAM 2019
GLM 2020
GAM 2020
GLM 2021
GAM 2021
GLM 2022
GAM 2022
0
1
2
3

Fig. 4. (Cont'd)



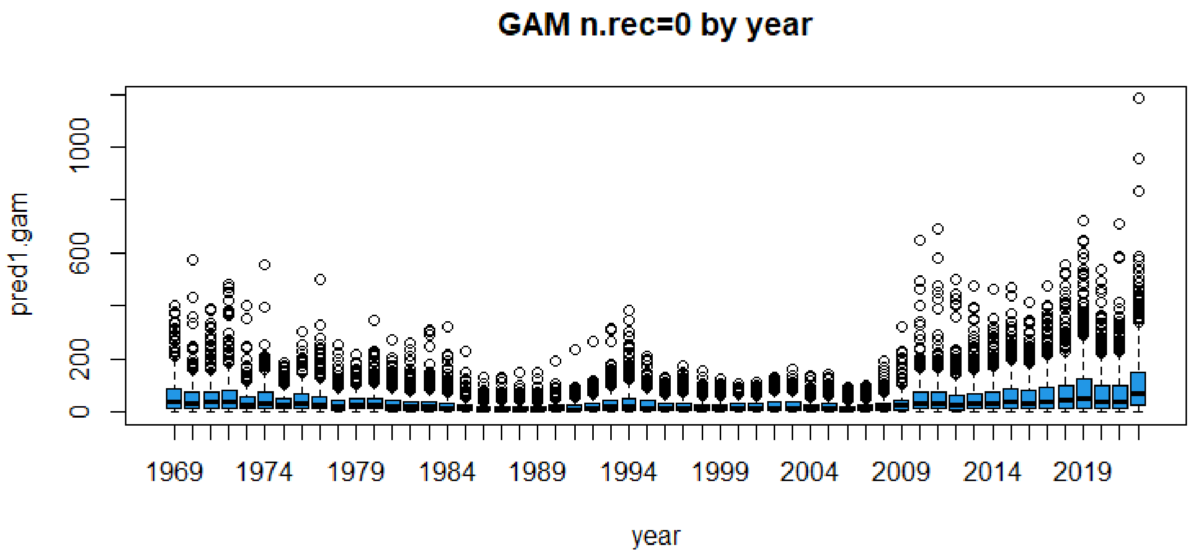


Fig. 5. Boxplot of CPUE predicted values by year in group 0, which no actual data corresponded.

Table 1. Summary statistics of the estimates where CPUE predicted value > 1500, in group 0.

Year	Area	Month	N	Mean	Max	Sum
2018	8	6	4	1,564	1,564	6,255
2018	8	7	4	1,646	1,646	6,583
2018	8	8	4	1,917	1,917	7,668
2019	8	6	20	2,899	4,622	57,977
2019	8	7	12	3,803	4,864	45,634
2019	8	8	12	4,218	5,665	50,617
2019	8	9	8	1,739	1,739	13,915