CCSBT-ESC/2608/BGD 04
(*Previously* CCSBT-OMMP/2606/07 (Rev.1))
(ESC Agenda item 7.1)

はえ縄 CPUE から求めたミナミマグロの資源量指数をチェックする手法の提案

Proposal of a method for checking abundance indices of southern bluefin tuna obtained from longline CPUE

伊藤智幸

Tomoyuki ITOH

水産研究・教育機構 水産資源研究所

Fisheries Resources Institute,

Japan Fisheries Research and Education Agency

要約

CCSBT では、日本はえ縄漁業による漁獲率（CPUE）から求めた資源量指数を、資源評価並びに管理のツールである MP での TAC 計算に用いている。はえ縄 CPUE による資源量指数の歴史は長く、標準化モデルやデータの扱いが変化してきた。近年では 2018 年の資源量指数値が急増した問題から、標準化モデルの改善が行われ、現在用いている GAM22 へと移行した。2018 年資源量指数値では、非操業時空間に対して極端に高い異常値が推定されたことが原因であった。GAM22 は現在のところ頑健な資源量指数値を提示してはいるが、今後も信頼できる資源量指数値を求めるため、非操業時空間に対する CPUE 予測値に注目した新たな診断方法を開発した。

Summary

CCSBT has been using an abundance index derived from the catch rate (CPUE) of Japanese longline fisheries for the stock assessment and MP to calculate TAC. The history of abundance indices based on longline CPUE is long, and the standardization model and data handling have changed over time. In recent years, due to the problem of a sharp increase in the abundance index value in 2018, the standardization model was improved, and the current GAM22 model was adopted subsequently. The reason for the high abundance index value in 2018 was that an extremely high outlier was estimated for non-operating time and space. Although GAM22 currently provides robust abundance index values, in order to obtain reliable abundance index values in the future, a new diagnostic method focusing on CPUE predictions for non-operating time and space was developed.

Introduction

In the CCSBT, the abundance index is calculated from the catch rate (CPUE; number of fish per 1000 hooks) of southern bluefin tuna (SBT; *Thunnus maccoyii*) caught by Japanese longline fisheries. This abundance index is used for the assessment of SBT stock and is also an important input for the Management Procedure (MP) used to determine the Total Allowable Catch (TAC), making it crucial for stock management.

The history of abundance indices using longline CPUE is long, and standardization models and data handling have evolved with improvements in computing power. Because it utilizes information obtained from fishing, an economic activity, the data does not rely on systematic sampling. It is based on sampling that is biased in time and space. Therefore, longline CPUE undergoes statistical processing called standardization. Model approaches such as GLM and GAM are commonly used. After estimating model parameters with real data, CPUE predictions are obtained from the model for hypothetical data assuming no bias in time or space. The aggregated value of this hypothetical data (generally the mean, sometimes weighted) is used as the abundance index.

The abundance index can exhibit abnormal behavior in a given year. This can be due to problems with the standardization of CPUE, or problems arising in the process of calculating the abundance index. In the case of SBT stock, in 2019, a phenomenon was detected in which the 2018 value in the abundance index was significantly higher than the 2017 value, which was considered problematic (10th OMMP Report). Subsequently, CCSBT decided to develop a new abundance index using CPUE (25th ESC Report, 2020). As a result of this development work, the stock assessment for 2020 was successfully completed by transitioning from the conventional core vessel GLM to GAM11. However, the abundance index using GAM11 also showed extreme increases afterward, which was considered problematic. Following a comprehensive review work, GAM22 was developed, and it was agreed in 2022 to transition from GAM11 to GAM22. These problems were subjectively perceived as problems because the abundance index values increased sharply.

Subsequent analysis suggested that the extremely high abundance index values for the core GLM vessel in 2018 and 2019 were due to the estimated year*Area interaction of the GLM model with respect to spatiotemporal strata without operations (Itoh 2020).

It is not impossible that the abundance index may show abnormal values in the future. In order to appropriately manage the SBT stock through MP within the CCSBT, it is necessary to have objective criteria for determining whether the abundance index for a given year is abnormal or normal.

This paper proposes a novel diagnostic method focusing on CPUE predictions for non-operating spatiotemporal strata. This paper does not examine CPUE standardization itself. Evaluative methods for standardization, such as QQ plots, residual distribution, and AIC-based model selection, are well established. Furthermore, there are options for different statistical models, such as GLM, GAM, and VAST. Of course, exploring the best CPUE

standardization is a necessary step.

However, in the context of this paper, what is important is to recognize that even with sophisticated standardization, there are significant areas where estimations are based on predictions without operational data, and this carries a risk of major errors. The objective of this paper is to optimize (or prevent major errors in) the overall goal of stock assessment and stock management.

Furthermore, this paper does not delve deeply into the causes of abnormally high estimates in the absence of operational data. Rather than investigating the causes, it prioritizes the practical aspect of reliably detecting observed phenomena and using them to inform management decision-making.

Material and method

The data used was longline CPUE data for determining GAM22, created in the 2026 work (Itoh and Takahashi 2026). This data includes Japanese longline data from 1969 to 2025. Data was limited to April to September from CCSBT statistical area 4 to 9. SBT catch numbers were limited to fish aged 4 years or older based on fork length data and the growth equation (fork length of all individuals has been available since 1995, so there is little uncertainty about size data in this dataset). To address the issue of target species, catch numbers for bigeye tuna (*T. obesus*), yellowfin tuna (*T. albacares*), albacore (*T. alalunga*), and swordfish (*Xiphias gladius*) were also included.

The standardization method used was the core vessel GLM model (Itoh and Takahashi 2020), the GAM11 model (Hoyle 2020) which used for the 2020 stock assessment, and the GAM22 model (Itoh and Takahashi 2026). Note that the core vessel GLM model and GAM11 limited data to vessels targeting SBT when calculating the abundance index, and further calculated the abundance index using intermediate values (w0.5, w0.6, w0.8, w0.9) between the Constant Square and Variable Square hypotheses. However, the analysis in this paper uses data from all vessels and calculates based solely on the Constant Square hypothesis. Also, regarding the k-parameter of GAM11, while it might be necessary to increase it with the increasing number of data years, the original values were used. The purpose of this analysis is to examine how much the results differ depending on the model, and it merely utilizes representative models adopted throughout the history of CCSBT.

The model used for core vessel GLM is as follows:

$$\log(\mathrm{CPUE}+0.2) = \mathrm{Intercept} + \mathrm{f.year} + \mathrm{f.month} + \mathrm{f.area} + \mathrm{lat} + \mathrm{BET_CPUE} + \mathrm{YFT_CPUE} + (\mathrm{f.month}*\mathrm{f.area}) + (\mathrm{f.year}*\mathrm{lat}) + (\mathrm{f.year}*\mathrm{f.area}) + \mathrm{Error}$$

The model used for GAM11 is as follows:

log(CPUE + 0.2) ~ Intercept +f.year + s(month) + s(lon) + s(lat) +
ti(lon*lat) + ti(month*lat) + ti(lon*month) + ti(year*lat) + ti(year*month) +
ti(lat*lon*month) + ti(lat*lon*year) + Error

The models used for GAM22 are as follows:

Binomial submodel
cpue > 0 ~ Intercept + f.year +s(month) + s(lon) + s(lat) +
ti(lon, lat) + ti(month, lat) + ti(lon, month) + ti(year, lat) + ti(year, lon) +
ti(year, month) + cl + s(log(hook)) + Error

Positive catch submodel
log(CPUE) ~ Intercept + f.year +s(month) + s(lon) + s(lat) +
ti(lon, lat) + ti(month, lat) + ti(lon, month) + ti(year, lat) + ti(year, lon) +
ti(year, month) + ti(lat, month, year) + ti(lat, lon, month) + ti(lat, lon, year) +
ti(year, lon, month) + cl + s(log(hook)) + Error

where,
year: Year in number
f.year: Year in factor.
month: Month in number
f.month: Month in factor.
lat: Latitude in 5 degree in number. Represented by the middle (e.g. -47.5 from 45.0S to 49.9S).
lon: Longitude in 5 degrees in number. Represented by the middle (e.g. 32.5 for 30.0E to 34.9E). Convert to 360 degree while >240 was converted by -360 so that lon ranged from -22.5 to 187.5 continuously.
f.area: CCSBT statistical area in factor.
cl: Cluster in factor. 1, 2, 3, and 4. Cluster was calculated based on the number of SBT, bigeye tuna, yellowfin tuna, albacore, and swordfish (see details in Itoh and Takahashi 2026).
hook: Number of hooks used in number.
BET_CPUE: Nominal CPUE of bigeye tuna.
YFT_CPUE: Nominal CPUE of yellowfin tuna.

The abundance index is calculated by determining the CPUE predicted value for dummy data and taking the annual average. First, the dummy data was created using all possible

combinations of the model factors (using the expand.grid function in R). For continuous variables, representative values (mean values) were fitted. Next, the dummy data was limited to the spatialotemporal strata in which SBT were historically caught (monthly, 5x5 degrees). At the same time, a weighted index was also calculated (the number of 1x1 degree sections in which SBT were caught within the monthly, 5x5 degree section (maximum 25) multiplied by the area size of 1x1 degree, taking into account that the area size changes with latitude). The abundance index is calculated by summing (dummy data predicted value × area index) for each year and dividing by the average of the annual sums.

For detailed analysis, the dummy data was assigned the number of actual operational data by year, month, latitude (5 degrees), and longitude (5 degrees). The dummy data was then divided into sub-datasets: one with actual operations and one without. The three indices (core GLM, GAM11, GAM22) were compared, and abundance indices were calculated for both the sub-datasets with and without actual operations to investigate which sub-dataset was responsible for the differences observed in the comparison of the three indices.

The annual average predicted CPUE was calculated for the sub-datasets where actual operations occurred, and the relative value of the predicted CPUE was calculated using this as the denominator. The relative predicted CPUE values for the two sub-datasets were examined using box plots.

In the sub-datasets where no actual operations occurred, a threshold (x_cut) was set. x_cut=3, 5, and 10 were calculated. The change in the abundance index was examined when data records with extremely high predicted CPUE values corresponding to x_cut were removed. The spatial and temporal distributions of the data records removed by x_cut were plotted to examine whether there was a bias towards specific spatiotemporal strata. The software R (R4.5.2) was used for analysis.

Result

In the dummy dataset used to calculate the abundance index, the proportion of cells (year, month, 5 degrees latitude, 5 degrees longitude) where actual operations took place was only about 50% even in the year with the highest value (Fig. 1). This shows that the abundance index has always relied heavily on predicted values for spatiotemporal strata where actual data is unavailable. Currently, this proportion is about 10%, meaning that predicted values are calculated for 90% of spatiotemporal strata without actual data.

The three indices for the core vessel GLM, GAM11, and GAM22 were compared (Fig. 2, top). Compared to GAM22, the core vessel GLM was higher in 2018 and 2019, and also higher in 2025. GAM11 showed extremely high values in 2024 and 2025, and also a relatively high value around 2011, showing values that differed significantly from GAM22. When the indices were

calculated using sub-datasets where actual operational data existed, GAM11 and GAM22 were very similar, and the overall trajectory of the core vessel GLM was also similar (Fig. 2, middle). In contrast, when the indices were calculated using sub-datasets where actual operational data did not exist, the differences shown for all data were observed (Fig. 2, bottom). This suggests that the differences in the three indices originate from the CPUE prediction values for sub-datasets where actual operational data did not exist.

The ratio of CPUE predictions, where the ratio of the annual predicted CPUE to the average value for sub-datasets with operational data, is shown in Figure 3. For the core vessel GLM, the CPUE predictions in sub-datasets with operational data were at most less than 12 times the average value (Fig. 3, top left). However, in sub-datasets without actual operational data, extremely high CPUEs exceeding 70 times the average CPUE were predicted for 2018 and 2019 (Fig 3, top right). For GAM11, the CPUE predictions in sub-datasets with operational data were at most less than 6 times the average value (Fig. 3, middle left). However, in sub-datasets without actual operational data, extremely high CPUEs exceeding 100 times were predicted for many years since 2000 (Fig. 3, middle right). For GAM22, the CPUE predictions in sub-datasets with operational data were at most less than 7 times the average value (Fig. 3, bottom left). In sub-datasets where actual operational data is unavailable, the maximum multiplier is less than 20, and there have been no values of 5 times or higher in the last five years. Values of 10 times or higher are from the past, specifically the 1990s and around 2010 (Fig. 3, bottom right). This reinforces the idea that the increase in the index originates from CPUE predictions for sub-datasets where actual operational data was unavailable.

Examples of setting x_cut to 3, 5, and 10 are shown (Fig. 4). It can be seen that the core vessel GLM experienced significant changes in 2018 and 2019 depending on x_cut. In GAM11, a large overall change occurred with x_cut, and a trend similar to that of GAM22 was obtained. In GAM22, there was almost no change in the abundance index even when x_cut changed.

Figure 5 plots the spatial and temporal regions where the CPUE prediction values were abnormally high for the core vessel GLM. The abnormally high CPUE prediction values in 2018 and 2019 were not due to a decrease in a limited, biased spatiotemporal region, but rather spanned a wide geographical area and a wide time period (April to September). This suggests that the problem cannot be solved by provisional measures that ignore some spatiotemporal regions. At that time, it was found that the abnormally high values were caused by an imbalance in the number of data points between Area 8 and Area 9, resulting in Year*Aera interaction (Ioth 2020).

Figure 6 plots the spatial and temporal regions where GAM11 produced abnormally high CPUE prediction values. While these values were observed across a wide range of years and months, the latitudes were limited to the 45-49S region. This suggests a problem with the estimation for this latitude.

Discussion

If a significant difference is observed between x_cut=5, for example, and the abundance index when x_cut=NA (the normal abundance index), then this alone can be used to determine that attention should be paid to the abundance index. In 2019, the ESC issued a warning because the abundance index for the core vessel GLM increased sharply, but using x_cut allows for a more objective and simpler determination of whether a warning should be issued.

The actions taken after determining a warning will vary. The timing of the actions will also differ, depending on whether they are for the meeting that year or for review in subsequent years. For example, as a response at the meeting that year, it would be good to observe the response when the model conditions are changed, investigate the cause, and make improvements. However, with complex models where a single calculation takes several tens of minutes, it may be difficult to investigate the cause through trial and error with many runs within the limited number of meeting days (and limited number of PCs). As a provisional measure, it may be a good idea to conduct a stock assessment based on the CPUE resource index with x_cut=5 as one of the sensitivity tests. This could also serve as a decision method in the meta-rule process of the MP in CCSBT. If the TAC in the MP changes based on the CPUE abundance index with x_cut=5, this should be recommended to the manager.

If a warning flag is raised and improvement is needed, comprehensive work will be carried out in subsequent years. Standardization methods (GLM or GAM, or other), models, and settings (such as the k-value in GAM) should be reviewed. However, there is no guarantee that improvement will be possible. For example, in the case of GAM11 shown in this example, the problem is suggested with the estimates for 45-49S, implying a structural problem, which gives hope that improvement is possible. However, in the core vessel GLM, problems were only seen in 2018 and 2019, the spatial bias is small, and it does not seem that structural improvement is possible. Furthermore, abnormally high prediction values have not occurred since 2023, and it is possible that in later years this can be dismissed as a past temporary anomaly. The stage of raising a warning flag and the subsequent stage of improvement should be clearly distinguished, and the improvement stage should set tasks according to the target time.

Fortunately, the GAM22 currently used in CCSBT did not produce abnormally high prediction values, and there was little change in x_cut. However, there is a risk that abnormally high prediction values may occur in a given year in the future for unknown reasons, as was the case with the core vessel GLM in 2018. Therefore, it would be beneficial for the OMMP and ESC groups to include box plots (Figure 3) showing the cases with and without operational data, and the stock abundance index at x_cut=NA, 3, 5, and 10 (Figure 4), in the annual GAM22 update papers to help them make informed decisions.

In the dummy data, actual operational data was unavailable in many spatiotemporal areas. Even in the era when fisheries covered a wide spatiotemporal strata, 50% of the estimates were based on data without actual data. While the reduction in operational spatiotemporal

area due to the decline in fisheries also has an impact, it is necessary to recognize that the stock abundance index is inherently based on many predictions.

Typically, the abundance index is based on the Least Square-Means, which calculates the annual average of dummy data created for all possible combinations. In this case, the proportion of data without actual fishing operations increases, which will further exacerbate the impact of outlier estimates. CCSBT limits the calculation to cases where fishing has historically occurred, so the dummy data is confined to a realistic spatial range, and the impact of outliers is likely mitigated.

In addition to the x_cut method presented here, it will be necessary to further verify the reliability of abundance indices and develop diagnostic methods that allow us to confidently present abundance indices to managers.

References

Anon. 2019 Report of the Tenth Operating Model and Management Procedure Technical Meeting. 24 - 28 June 2019. Seattle, USA. 27pp.

Anon. 2020 Report of the Extended Scientific Committee for the Twenty Fifth Meeting of the Scientific Committee. 31 August – 7 September 2020. Online. 133pp.

Hoyle, S. 2020. Exploratory analyses for primary CCSBT CPUE index. CCSBT-OMMP/2006/15.

Itoh, T. 2020 Examination of an anomalously high value of the core vessel CPUE in 2018 for southern bluefin tuna. CCSBT-OMMP/2006/12.

Itoh, T. and N. Takahashi 2020 Update of the core vessel data and CPUE for southern bluefin tuna in 2020. CCSBT-OMMP/2006/11.

Itoh, T. and N. Takahashi 2026 Update of CPUE abundance index using GAM for southern bluefin tuna in CCSBT (GAM22) up to the 2025 data. CCSBT-OMMP/2606/06.

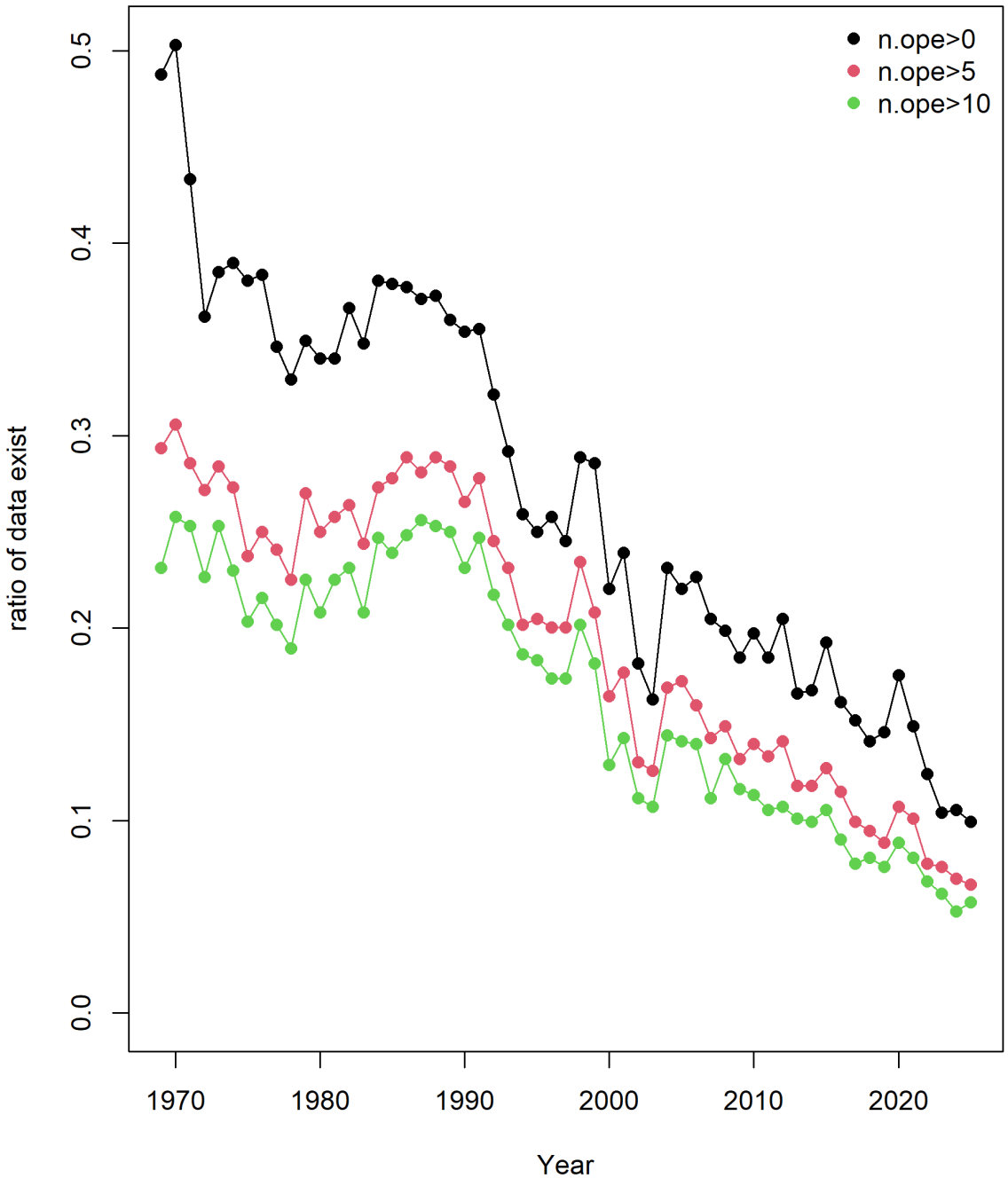


Fig. 1. Annual proportion of data with actual data in the dataset (dummy dataset) used to calculate the abundance index of southern bluefin tuna in CCSBT.

Color-coded by the actual number of operations.

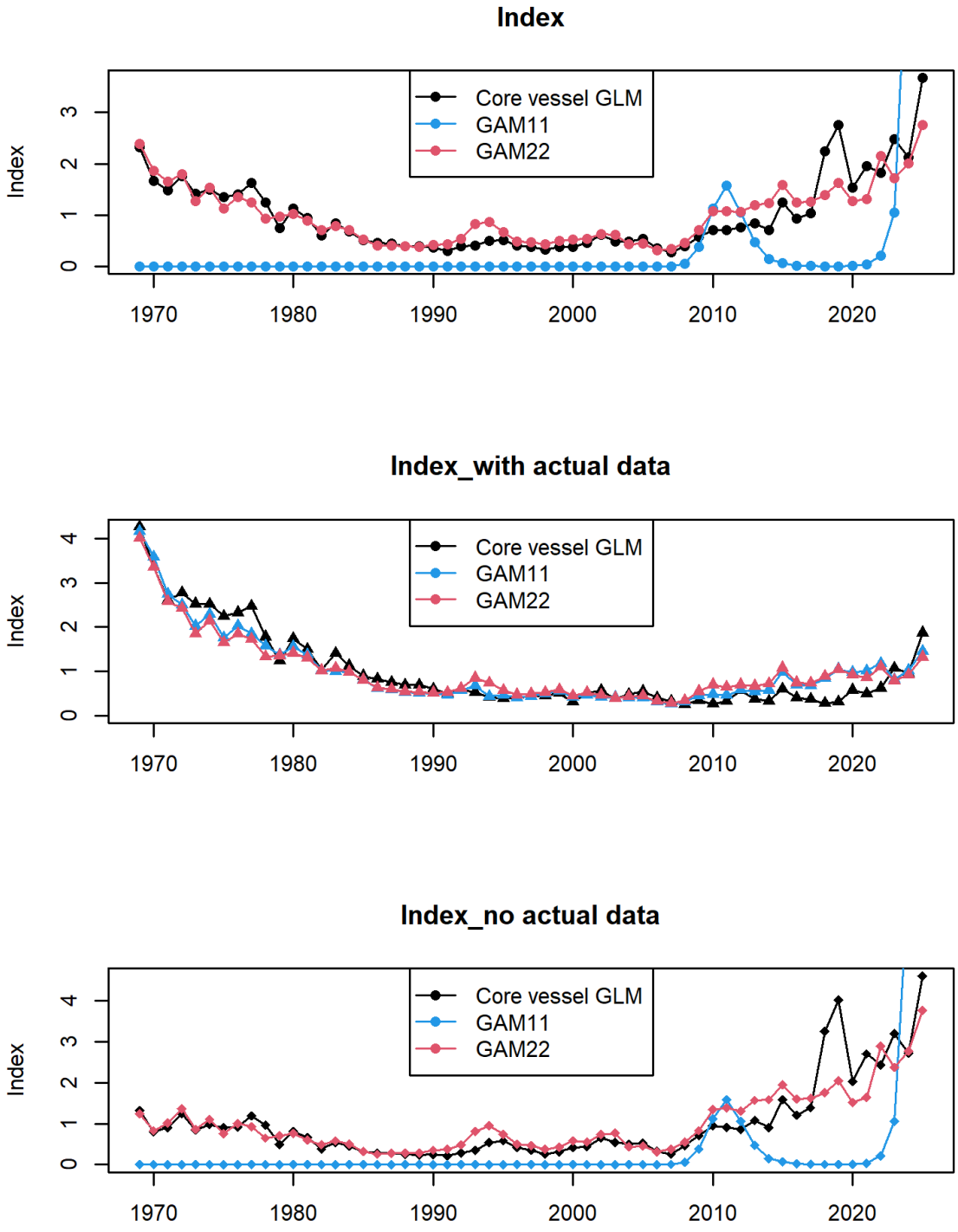


Fig. 2. Comparison of abundance indices.

(Top figure) Comparison of core vessel GLM, GAM11, and GAM22. (Middle figure) Comparison of the three abundance indices in a sub-dataset limited to the spatiotemporal strata in which actual operational data existed. (Bottom figure) Comparison of the three abundance indices in a sub-dataset limited to the spatiotemporal strata in which actual operational data did not exist.

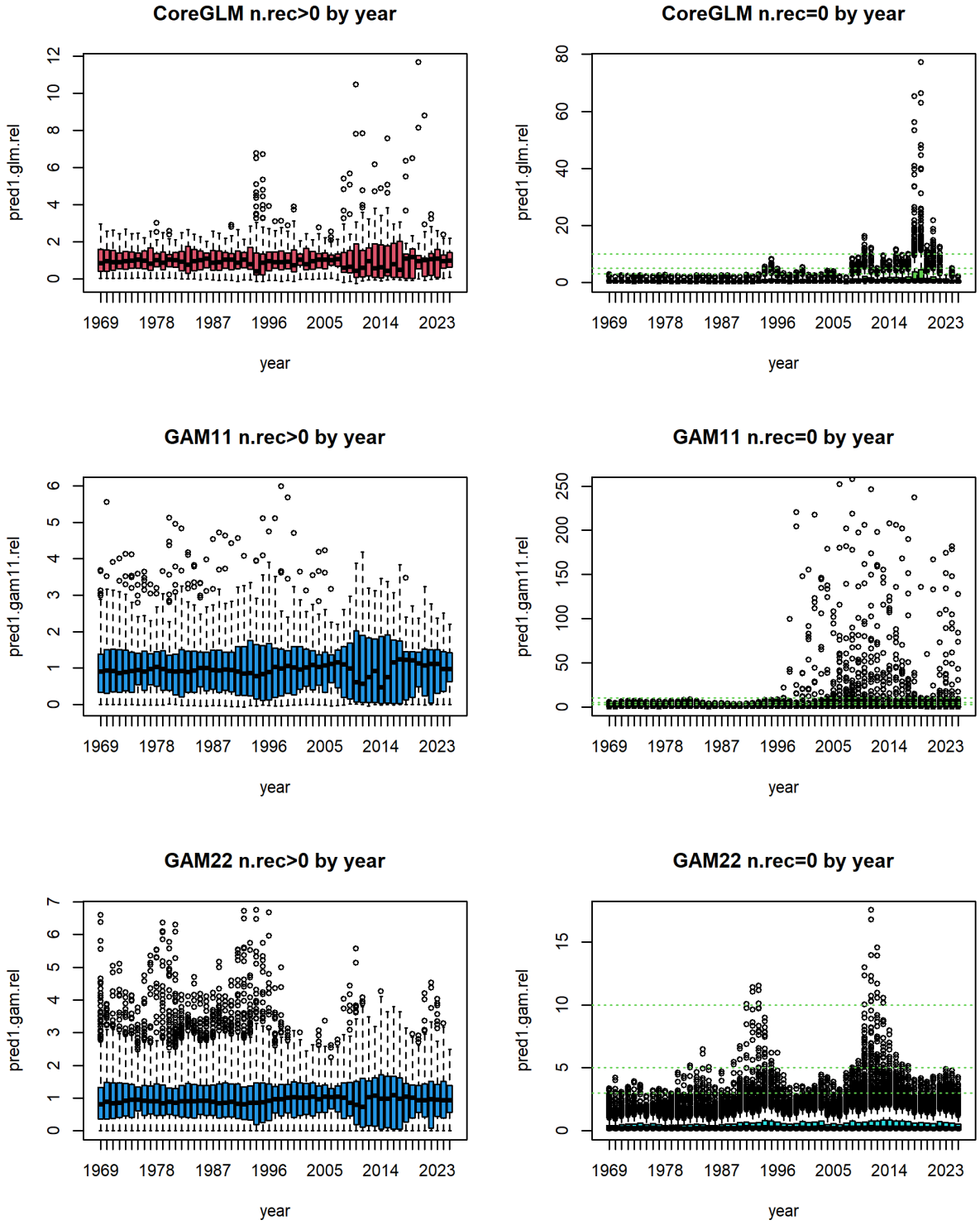


Fig. 3. Box plot of predicted values using dummy datasets.

The top, middle, and bottom figures show the core vessel GLM, GAM11, and GAM22, respectively. The left figure shows spatiotemporal data with actual data, and the right figure shows data without. The index values were normalized by the annual average values of the sub-datasets with actual operations. In the right figure, x_cut=3, 5, and 10 are shown by green dashed lines. In GAM11, there were points with multipliers of more than 250, but for readability, the Y-axis is limited to 250.

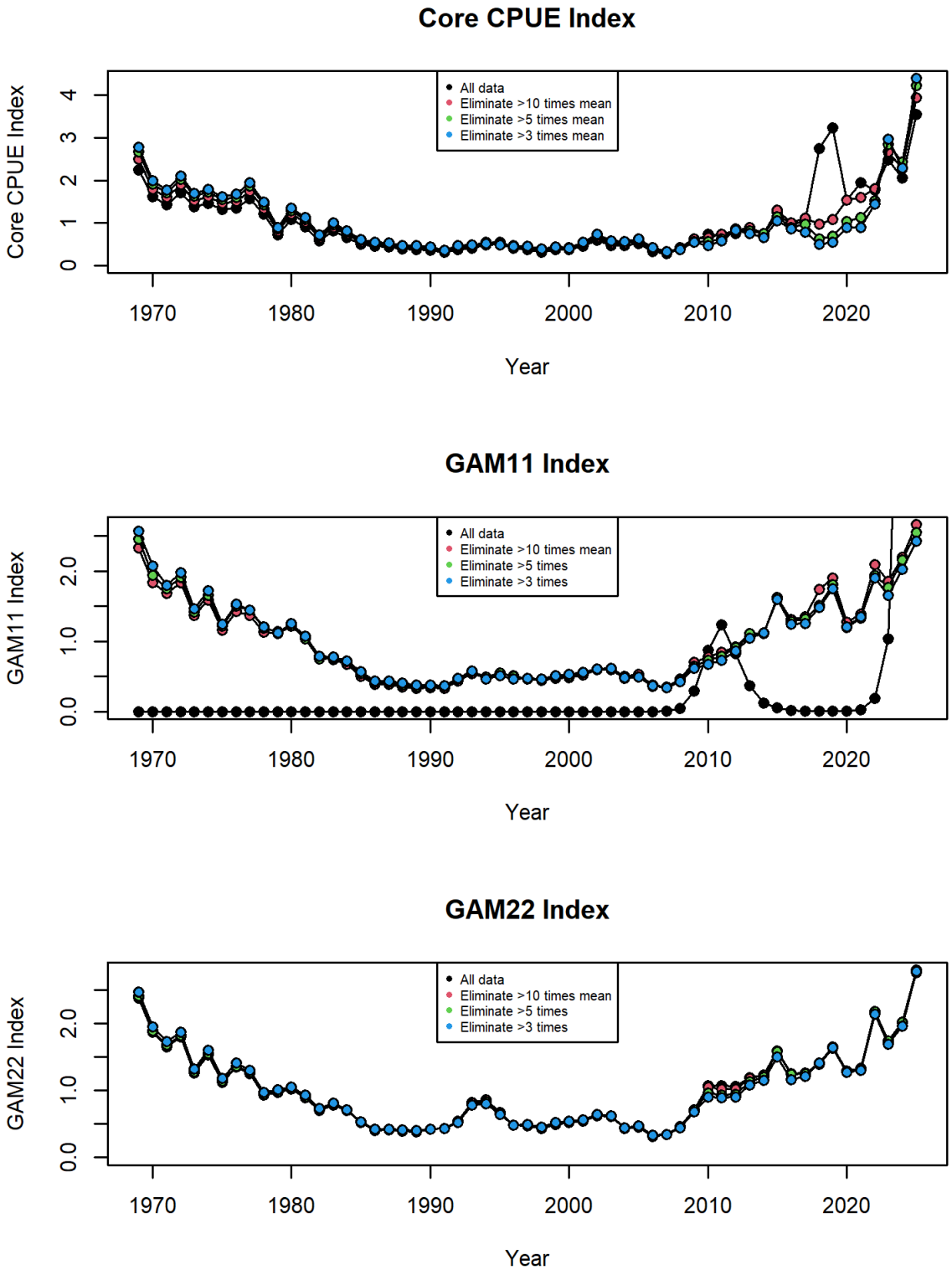


Fig. 4. Change in abundance index associated with x_cut.

x_cut is defined as a threshold multiple of the annual average value in the sub-dataset where actual operation data existed. For sub-datasets where no operations existed, records exceeding the x_cut value were removed before calculating the abundance index.

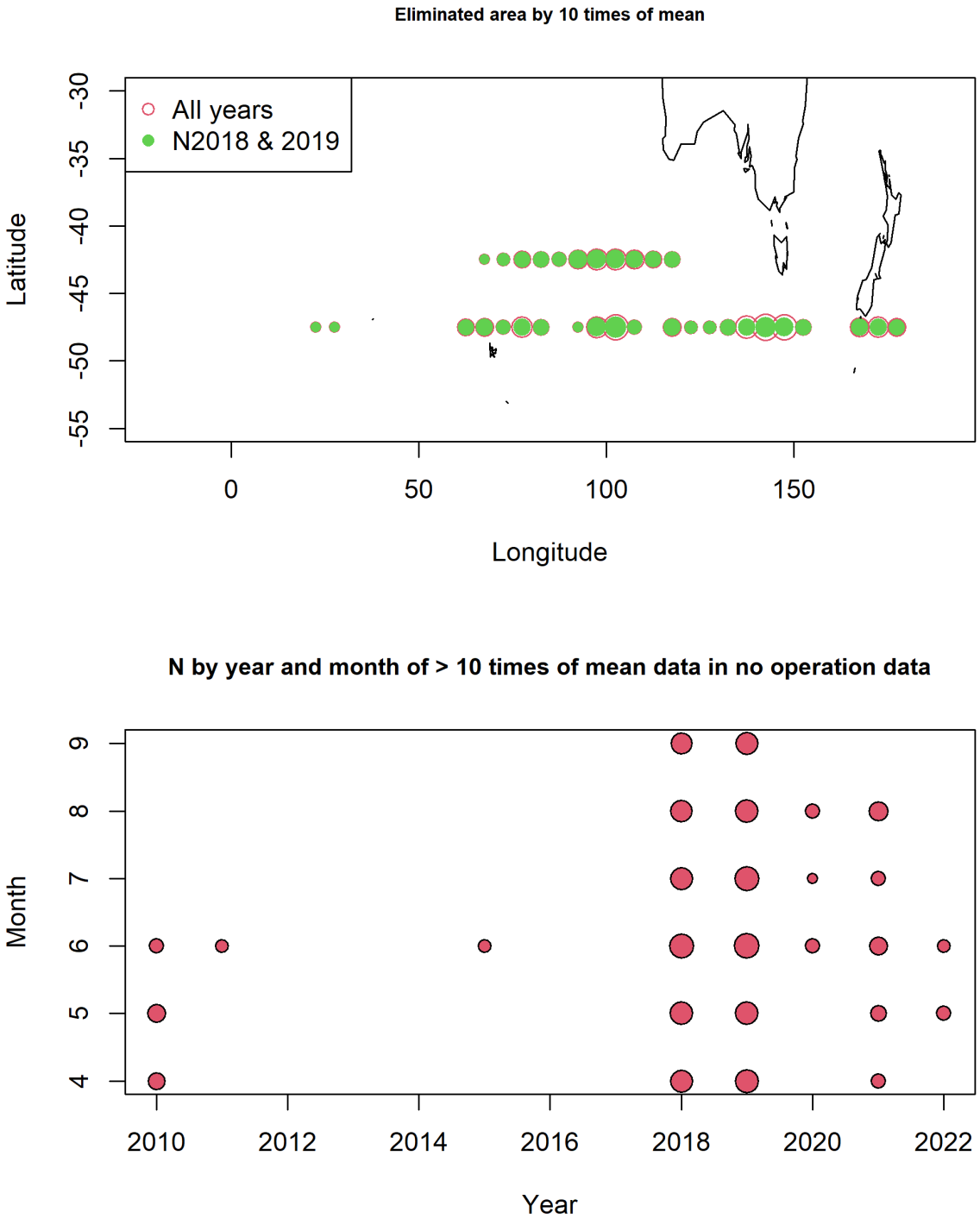


Fig. 5. Map of the number of records removed by x_cut=10 (top figure) and yearly/monthly distribution (bottom figure) in core vessel GLM.

In the top figure, red circles indicate the total number of records for all years, and green circles indicate those for 2018 and 2019.

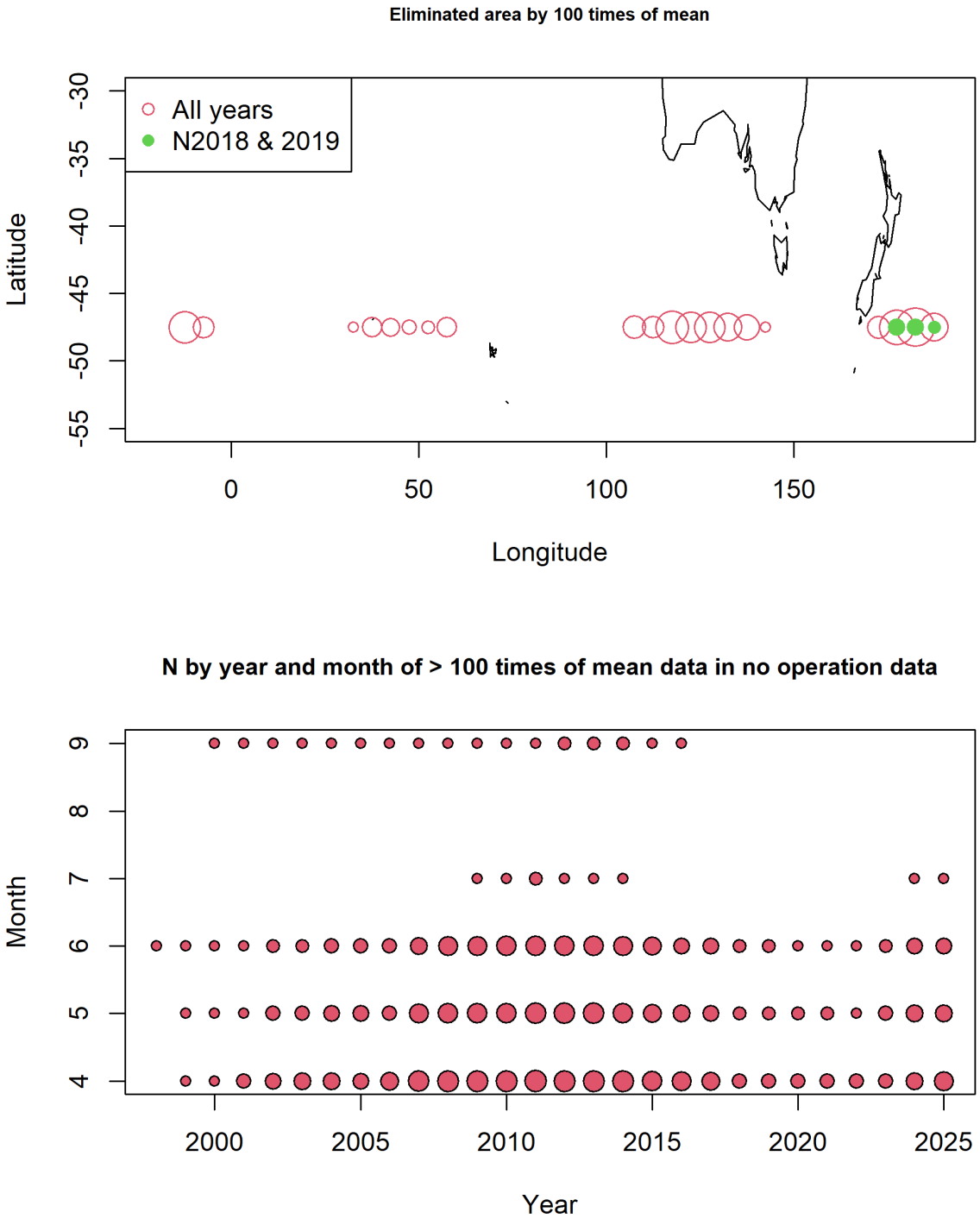


Fig. 6. Map of the number of records removed by x_cut=100 (top figure) and yearly/monthly distribution (bottom figure) in GAM11.

In the top figure, red circles indicate the total number of records for all years, and green circles indicate those for 2018 and 2019.