Exploration of Japanese size data and historical changes in data management.

Simon Hoyle[1], Kaisuke Satoh[2] and Takayuki Matsumoto[2].

Executive Summary

One of the primary datasets in tropical tuna assessments is the size frequency data obtained from the fisheries, and the Japanese longline fishery provides the longest and most valuable size dataset for the bigeye and yellowfin tuna assessments. In 2016 the IOTC Working Party on Methods recommended work to improve the understanding of the fishery, including the size data.

In this working paper we explore the Japanese size data, in order to describe and characterise the types and sources of size data, so that analysts can understand the patterns in the data; and to check the validity of rounding direction assumptions and look for changes through the time series.

In this working paper we provide figures showing the types of data available (spatial resolution, commercial vs research & training, measurement unit, and sampling type), for each species and by time period and location.

We also show that there was a previously unsuspected change in rounding practices for 2cm bins in 1970, from rounding up to rounding down. The current practice is to round up, so there must have been a further change after 1988. Further investigation is recommended to determine when the later change occurred. We recommend exploring the implications of these changes for other size datasets used by IOTC and other RFMOs. We further recommend exploring how size data biases noted by Satoh et al (2016) in the Eastern Pacific may affect Indian Ocean data.

Introduction

Integrated stock assessments are very dependent on understanding the fisheries that provide the data. Two of the primary datasets in tropical tuna assessments are the catch and effort data used to generate indices of abundance, and the associated size frequency data. The Indian Ocean Tuna Commission's 7th Working Party on Methods (IOTC-2016-WPM07-R) noted concern about a step change in the Japanese CPUE in the late 1970s, which affects the joint indices (Hoyle et al. 2016) and therefore the assessments. The IOTC-WPM recommended work to improve the understanding of the fishery, including the factors that created the discontinuity in the bigeye (and to a lesser extent yellowfin) CPUE 1976-80, and the associated size data.

Size sampling is designed to document the sizes of fish in the sea at the time and location of sampling, and the size selectivity of the fisheries. However there are processes that may interfere with this documentation, including fish sorting prior to measurement, biased fish selection for sampling (e.g. grab sampling), measurement errors, measurement rounding, and transcription errors.

For example, recent work by Japanese scientists and the Inter-American Tropical Tuna Commission (Satoh *et al.* 2016) identified differences in size distribution between fish measured on commercial vessels and those measured on research and training vessels, yet the sources were not identified

[1] IOTC Consultant, Nelson, New Zealand. Simon.hoyle@gmail.com.
[2] National Research Institute for Far Seas Fisheries, Shimizu, Japan.

when submitting data to the IATTC. They also found that some length data submitted to the IATTC had been converted from weight data with incorrect assumptions about the state of the fish.

Interpreting size data correctly therefore requires an understanding of the sampling processes and how they may have changed through time, but historical sampling methods in many fisheries have been poorly recorded and are often not well understood. The data managers in charge of historical sampling are no longer alive. Where sampling manuals exist, it is not always clear whether fishermen have followed the correct approaches.

Without reliable documentation, analysts must explore the data to identify times when sampling or recording processes may have changed. For example, Japanese size databases record lengths at standard resolutions of 1 cm, 2 cm or 5 cm, but samplers are believed to have mostly measured and recorded fish at 1 cm resolution, after which data entry operators rounded the measurements (Hiroaki Okamoto personal communication). These sizes have been assumed to be rounded up, so the mean sizes of fish reported at 2 cm resolution are assumed to be the size bin minus 1 cm, but if rounded down the mean size would be the size bin plus 1 cm, or 2 cm larger. With 5 cm bins, an error would similarly change the mean size by 5 cm.

In this working paper we explore the Japanese size data with the following objectives:

- To characterise the types and sources of size data so that analysts can understand the patterns in the data;
- To check the validity of rounding direction assumptions and look for changes through the time series.

Data description and analysis methods

The data (Table 1) were imported into R (R Core Team 2016), and prepared by allocating to location and time strata. The data were used as provided, with no additional range checking or cleaning.

Data were provided for multiple species but here we focus on bigeye, yellowfin, and albacore. For 1952-1964 only yellowfin data were provided, and were measured in very large numbers (Figure 1). Other species were measured during this period but are held in a different database which was not provided.

Samples were obtained from commercial vessels and from training and research vessels (Figure 2). Training and research vessels are combined in one category. Prior to 1965 all samples came from commercial vessels, but training and research vessel sampling began in 1965 and soon dominated the dataset. After 1980 the training and research vessel sampling began to decline, and by 2000 was a very small proportion of sampling. After 2004 the dataset included only commercial samples.

Measurements included both lengths and weights. Weight samples were all recorded in 1 kg bins, while length samples were recorded in 1, 2, and 5 cm bins (Figure 3). Albacore length measurements were all at 1 cm resolution, and weight samples were only a small proportion. Bigeye and yellowfin initially included mostly 2 cm and 1 kg bins, but 1 cm bins increased and comprised most of the dataset by 2016. Samples at 2 cm resolution were available for every year from 1952-1988, with smaller numbers from 1999-2001, and a few more in 2005 and 2006. All three species included a few samples at 5 cm resolution, mostly in the early 2000s.

Measured fish were allocated to location strata with resolutions (degrees latitude by longitude) 10 x 20, 5 x 10, and 1 x 1, with a few allocated to 5 x 5 (Figure 4). All samples were recorded at 5 x 10 resolution until 1965, when 10 x 20 resolution was introduced. This was followed by 1 x 1 resolution in 1967, which quickly became dominant. After 1995 all samples were reported at 1x1 spatial resolution.

Sampling locations were first reported in 1986 (Figure 5). Until 2002 most measurements were made by fishers on vessels, but from 2003 these numbers were surpassed by observers on vessels. Small numbers were reported by port samplers in Yaizu, Shimizu, and Tokyo.

Sex was first reported for measured fish in 1987 (Figure 6). By 1999 the 'not recorded' category was replaced by 'Unknown'. From 1997 the majority of measured bigeye and yellowfin were sexed, but few albacore were sexed at any time.

Sampling locations by period and spatial resolution are shown in figures 7 to 14. Before 1955, almost all sampling was in the north-eastern areas, but by the end of the decade had spread to cover the whole Indian Ocean (Figure 7). By the 1970's most bigeye and yellowfin samples were obtained from tropical areas, particularly east of 75 °E, with relatively few samples from further south (Figures 7 to 14).

We checked assumptions about rounding direction by plotting the numbers per size class by 5-year period and for each measurement unit. We noted the positions of the peaks that often occur at 5 and 10 cm, which in rounded data will indicate the direction of rounding.

After identifying changes in rounding practices through time, we generated adjusted bins that represented the midpoint of each bin, which depended on the binning interval and the rounding direction by year.

Results

Measurements per size class for fish measured at 1 cm length resolution showed a pattern often seen in size frequency data, with peaks appearing consistently in classes ending in 0 and 5 cm, due to rounding by samplers (figure not shown).

For bigeye and yellowfin measured at 2 cm length resolution, a change in data treatment occurred in 1970 (Figures 15 and 16). The 2 cm length measurements from 1950 – 1969 are reported at even numbered bins, but from 1970 – 1988 are reported at odd numbered bins. There were no 2 cm length measurement data in the period 1989 – 1998. Moderate numbers of fish were reported at 2 cm bins in 1999 – 2000, with almost all reported at even numbered bins. Some fish were reported at 2cm bins in 2001, 2004, 2005, and 2006, at a mixture of odd and even bins.

Data from 1952-1960, believed to have been collected by fisheries scientists (ref?), show no clear peaks in classes ending in 0 and 5 cm. In the 2 cm resolution data, from 1960 to 1969 there are peaks in classes ending in 0 and 6 cm, suggesting that measurements were rounded up, with fish recorded at 5 cm rounded to 6 cm. From 1970 until 1988, with the units shifted to odd numbers, the peaks occur at classes ending in 5 and 9. This indicates that measurements were rounded down, with fish recorded at 0 cm rounded to 9 cm. Similarly, the 5 cm resolution billfish data, which mostly occur from 1970-1988, appear to be rounded down (figures not shown). These data are recorded with classes ending in 1 and 6, and classes ending in 6 appear slightly higher on average than those ending in 1.

Discussion

It appears that there was a change in rounding practices for 2cm bins in 1970, from rounding up to rounding down. The current practice is to round up, so there must have been another change after

1988. There are very limited 2cm bin data after this time, with lower sample sizes and recorded at both odd and even numbers, so it is not possible to tell when the change occurred. Further investigation is required to identify when the final rounding direction change occurred, perhaps by comparing data collected by different samplers and methods.

Most samples after 1989 are recorded at 1 cm or 1 kg measurement units, which makes the rounding direction less important for large species like bigeye and yellowfin tuna. More important is for data managers and assessment scientists to be aware of the changes in both class boundaries and rounding direction in 1970. We recommend that the IOTC Secretariat note the issue and adjust their software where necessary. We recommend that other RFMO Secretariats are notified of the issue, which seems likely to apply to all Japanese size data.

Further work is needed to explore the extent of the changes in rounding; such as whether they apply to all data in the longline size database, and whether they apply to other size datasets such as Japanese purse seine and pole and line data. Changes in rounding direction for 1cm bins will be relatively more important for smaller species such as skipjack and albacore.

Further work is needed to carefully explore the size database, and to review aspects of size data collection through time, including fish selection protocols, data sheet formats, data entry protocols, screening, and post-processing methods across all size databases. There is also a need to explore how biases noted by Satoh et al (2016) for the Eastern Pacific may affect the Indian Ocean data. The differences in selectivity they observed between research and training vessels versus commercial vessels may be a problem for the Indian Ocean. Similarly we recommend comparison of weight data with length data, to identify the likely processing state of the weighed fish so that the weight data can be used appropriately.

References

Hoyle, S., Kim, D., Lee, S., Matsumoto, T., Satoh, K., Yeh, Y. (2016) Collaborative study of tropical tuna CPUE from multiple Indian Ocean longline fleets in 2016. DOI: 10.13140/RG.2.2.22918.16962

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Satoh, K., Minte-Vera, C.V., Vogel, N.W., Aires-da-Silva, A., Lennert-Cody, C., Maunder, M.N., Okamoto, H., Uosaki, K., Matsumoto, T., Semba, Y. (2016) An exploration into Japanese size data of tropical tuna species because of a prominent size-frequency residual pattern in the stock assessment model. *Inter-Amer. Trop. Tuna Comm., 7th Scient. Adv. Com. Meeting. SAC-07-03d*.

Tables

Table 1: Fields available in the size dataset, variable descriptions, and the meaning of each code.

Variable	Description	Codes	Meaning
Species		1	bluefin tuna
		2	southern bluefin tuna
		3	albacore
		4	bigeye
		5	yellowfin
		6	swordfish
		7	striped marlin
		8	blue marlin
		9	black marlin
		10	sailfish
		11	shortbill spearfish
		12	skipjack
level	Spatial resolution in degrees of latitude x longitude	1	10 x 20
		2	5 x 10
		3	5 x 5
		4	1 x 1
latitudec	Latitude type	1	N
		2	S
longitudec	Longitude type	1	E
		2	W
fleet	Set type	1	Longline
		2	Longline (night setting)
vesselc	Type of vessel	1	Commercial vessel
		>2	Training and research vessel
M_unit	Measurement unit	1	kg
		2	1 cm
		3	1 kg
		4	2 kg
		5	5 kg
		6	1 cm
		7	2 cm
		8	5 cm
place	Sampling location	1	On board by fishermen
		2	Port sampling Kagoshima
		3	Port sampling Katsuura
		4	Port sampling Yaizu
		5	Port sampling Shimizu
		6	Port sampling Tokyo
		7	Port sampling Shiogama
		8	Port sampling Kesennuma
		9	SBT monitor ship
		10	Port sampling Sakai-minato
		11	Port sampling Kamaishi
		12	Port sampling Misakai
		13	On board observer
sex		0	unknown
		1	female
		2	male

Figures

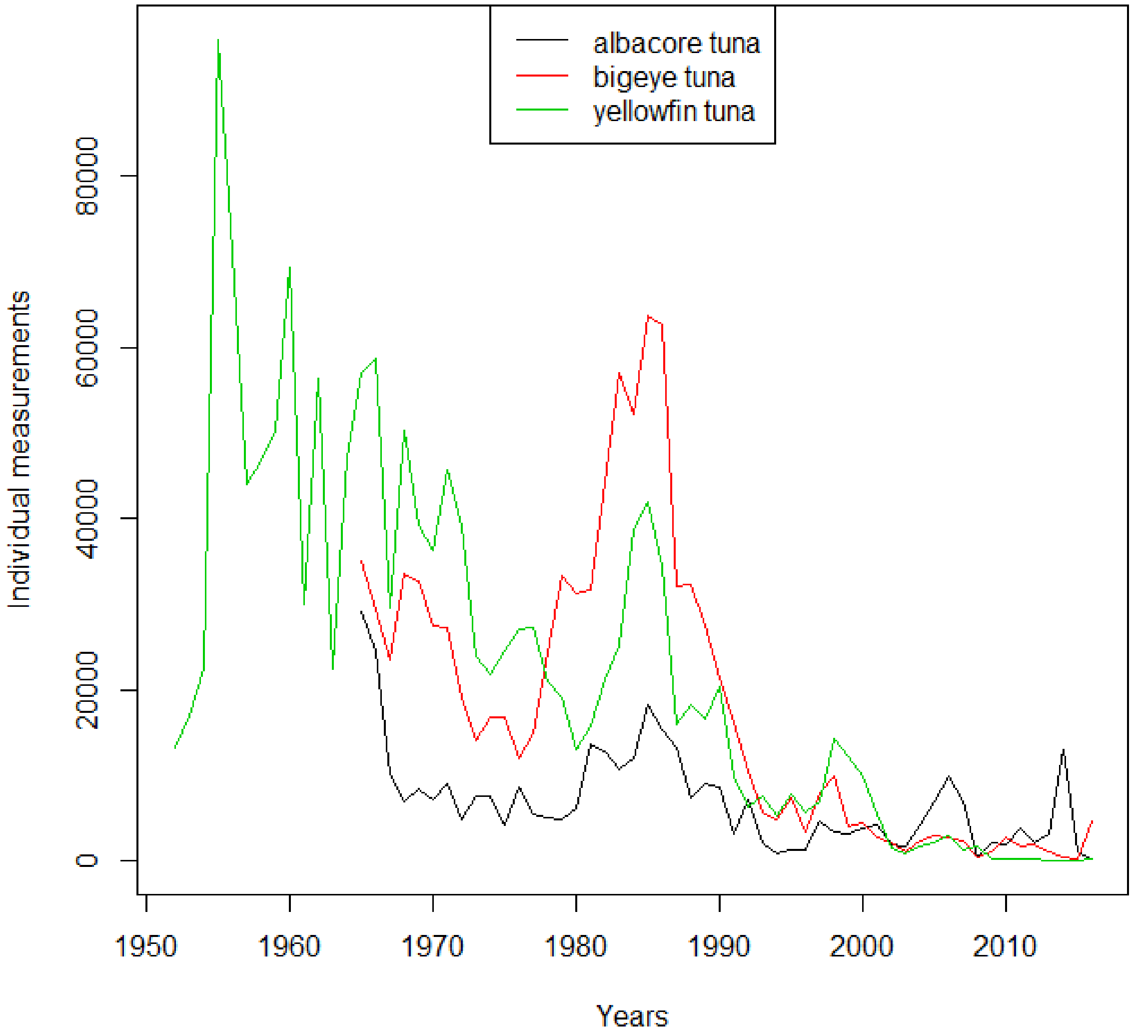

Figure 1: Number of fish measured and weighed by species and year.

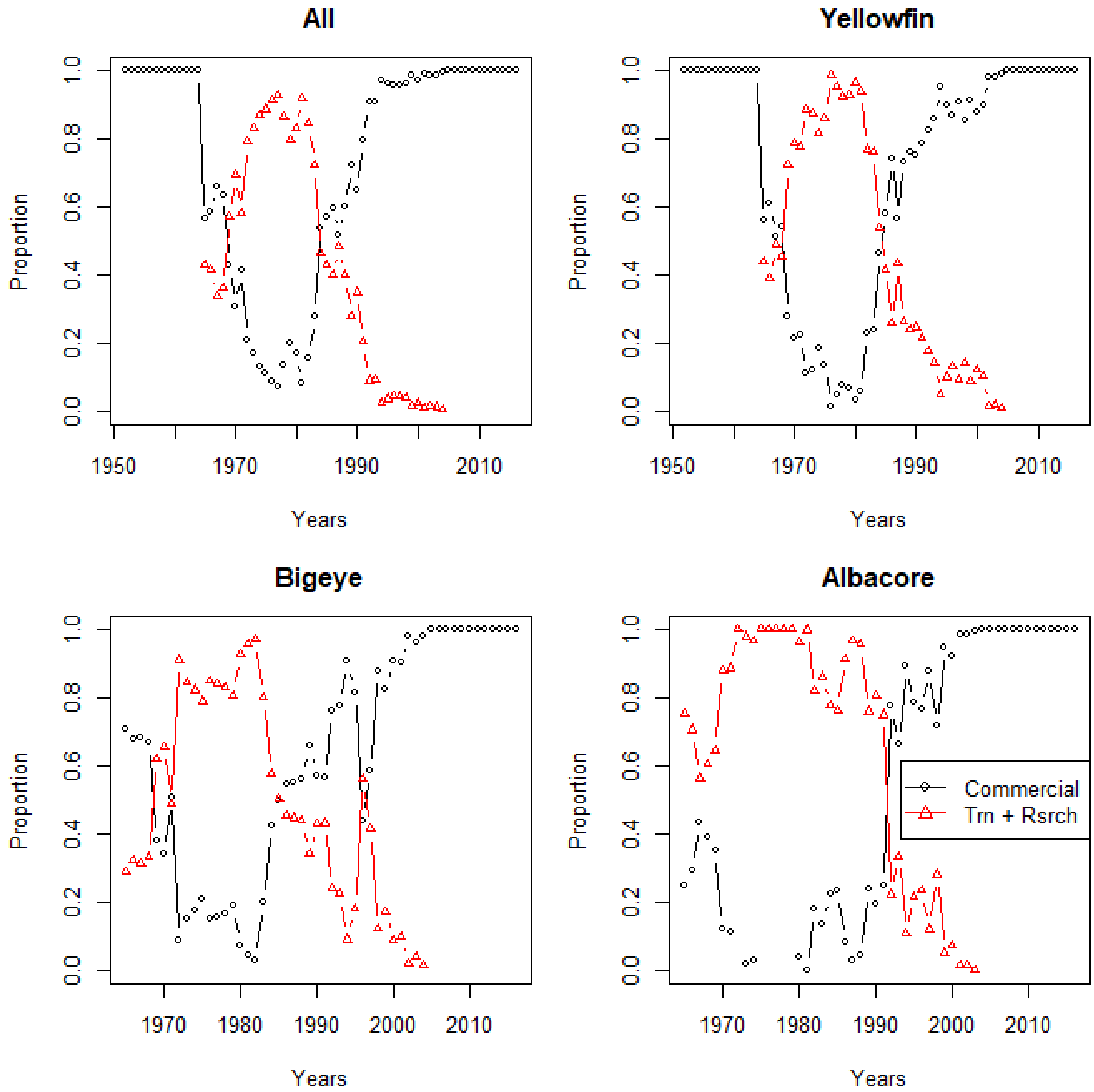

Figure 2: Proportions of measurements by type of vessel and year, for all species combined and separately for bigeye, yellowfin, and albacore.

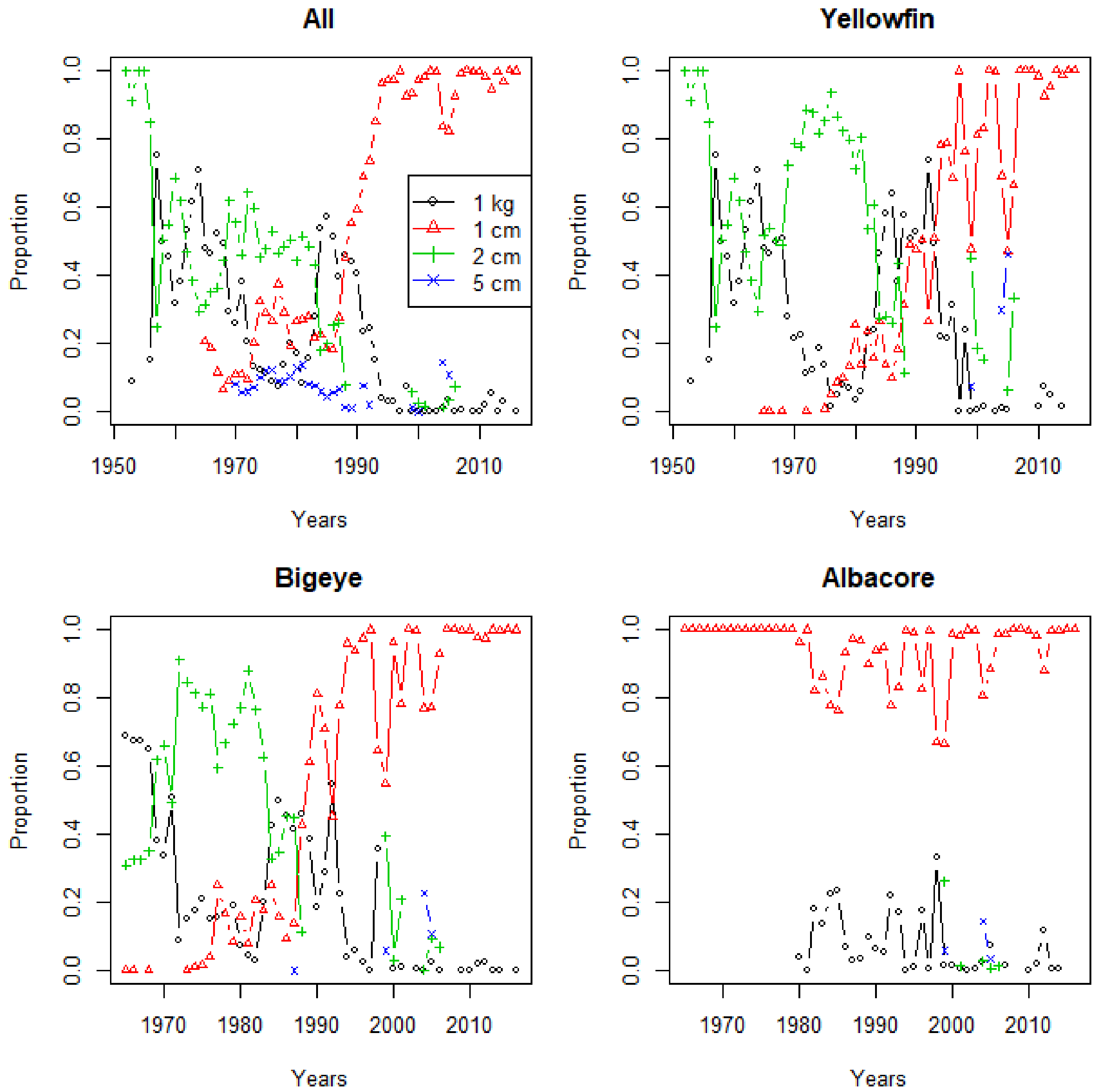

Figure 3: Proportions of measurements by measurement unit and year, for all species combined and separately for bigeye, yellowfin, and albacore.

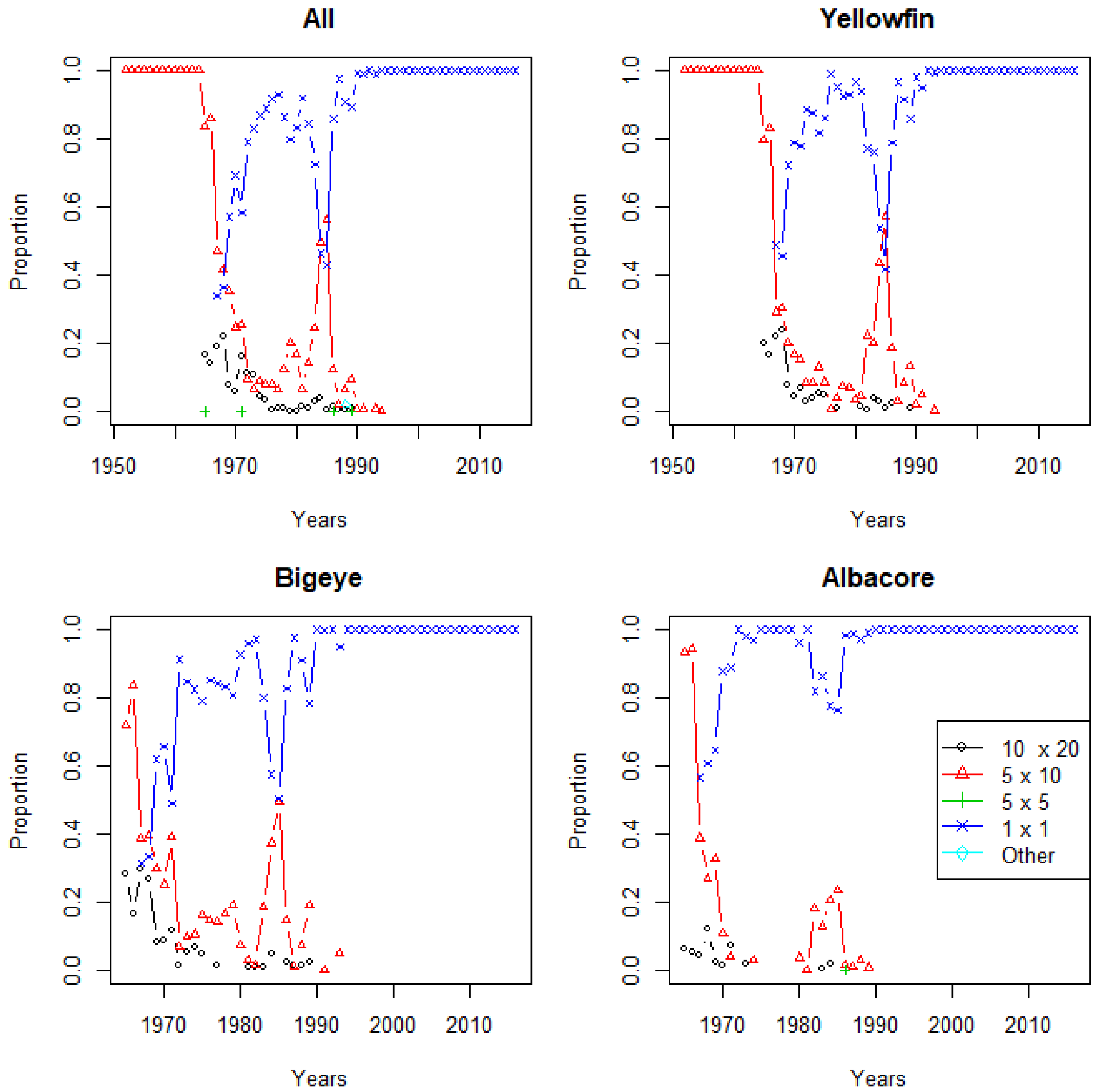

Figure 4: Proportions of measurements by spatial resolution and year, for all species combined and separately for bigeye, yellowfin, and albacore.

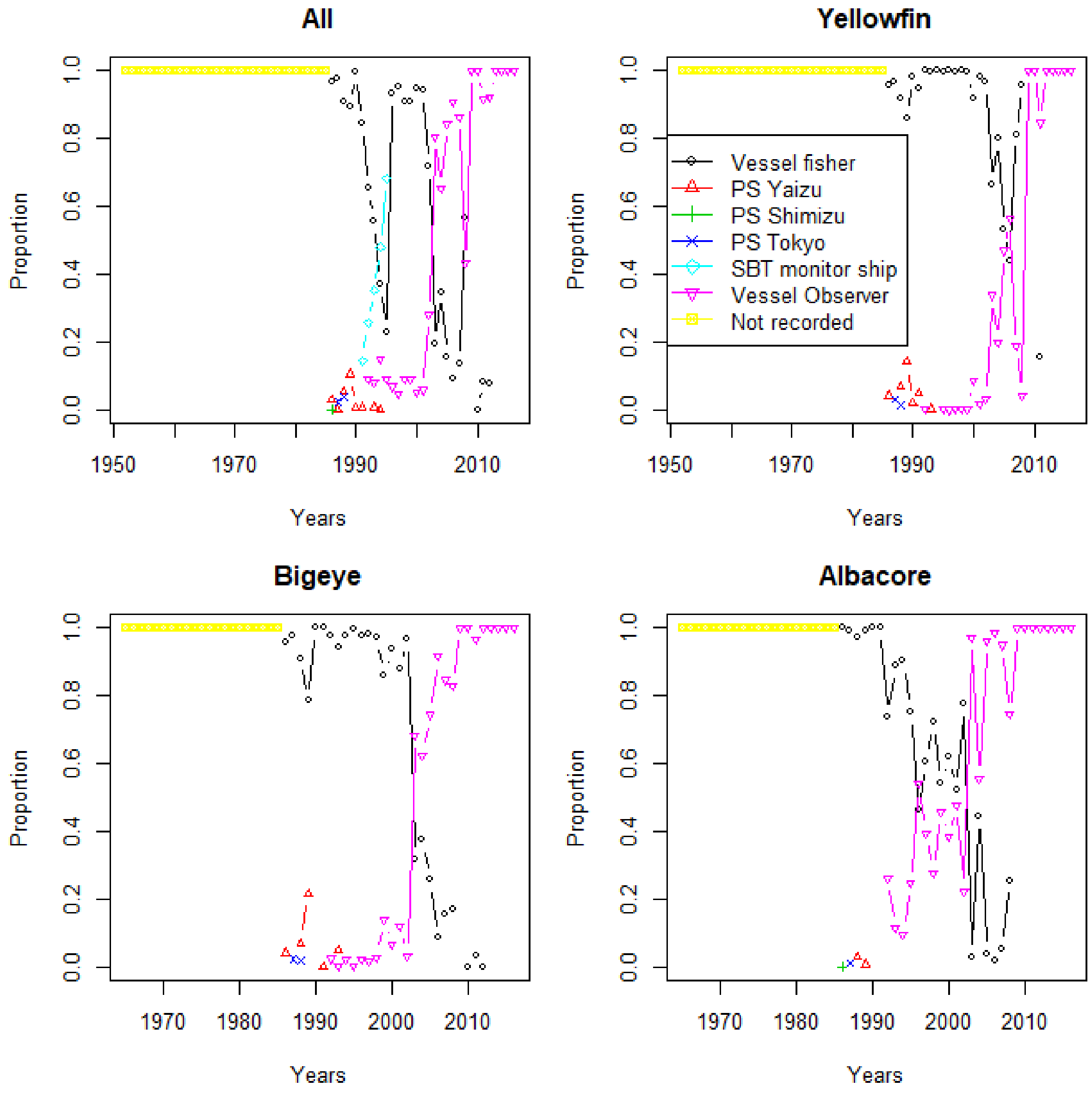

Figure 5: Proportions of measurements by sampling location and year, for all species combined and separately for bigeye, yellowfin, and albacore.

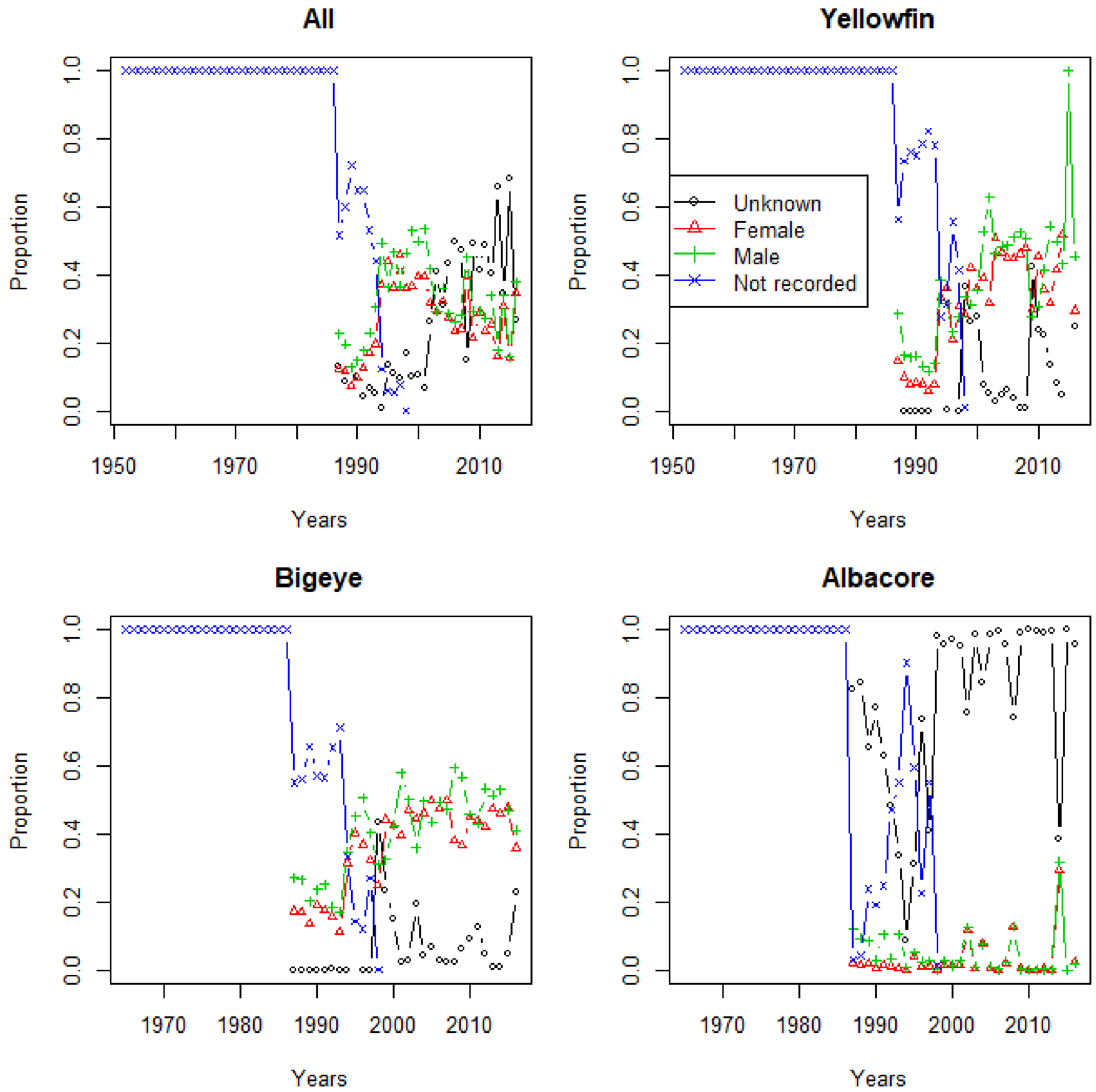

Figure 6: Proportions of measurements by sex and year, for all species combined and separately for bigeye, yellowfin, and albacore.

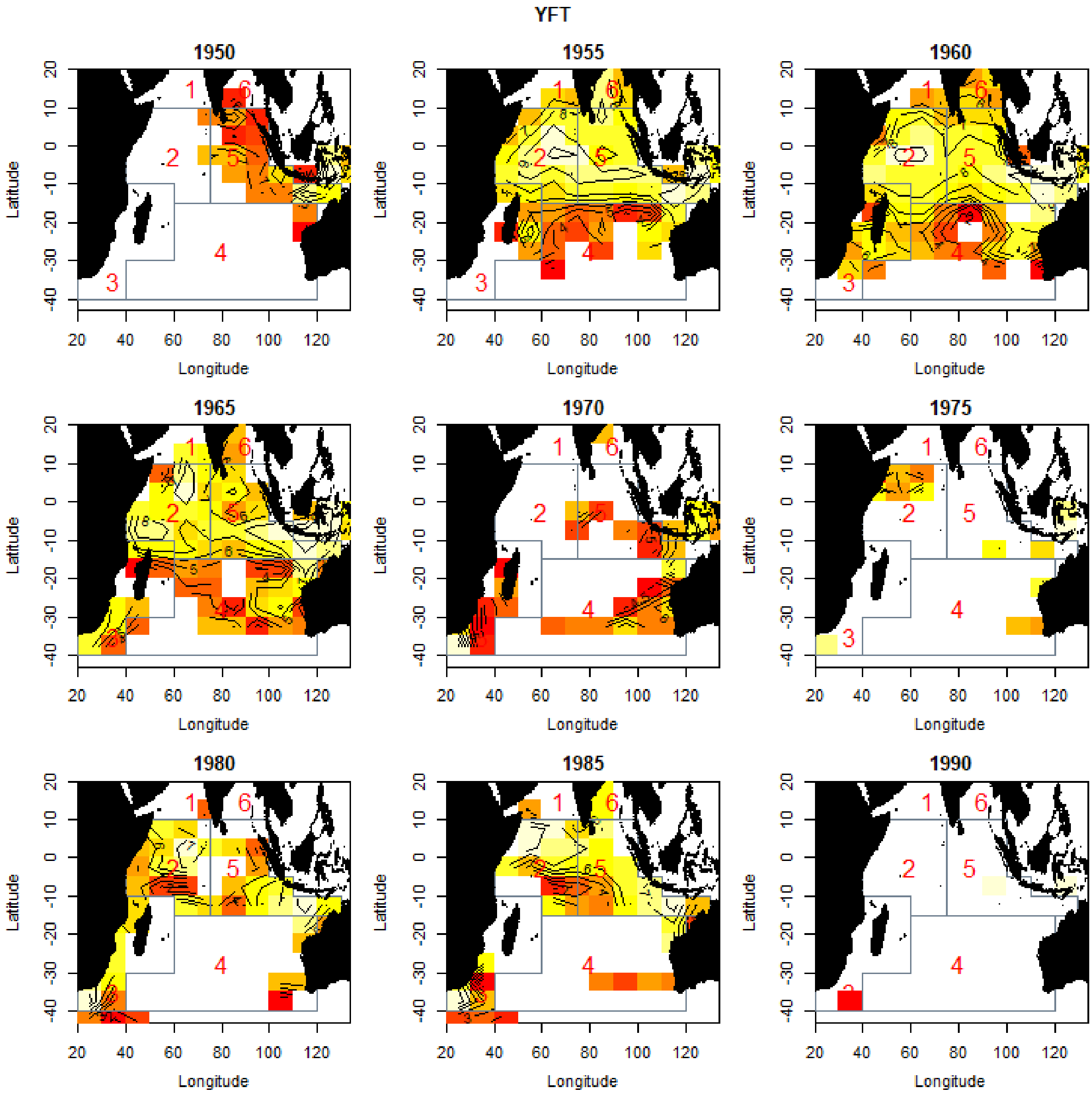

Figure 7: Yellowfin sampling locations and intensity for 5 x 10 data by 5-year period.

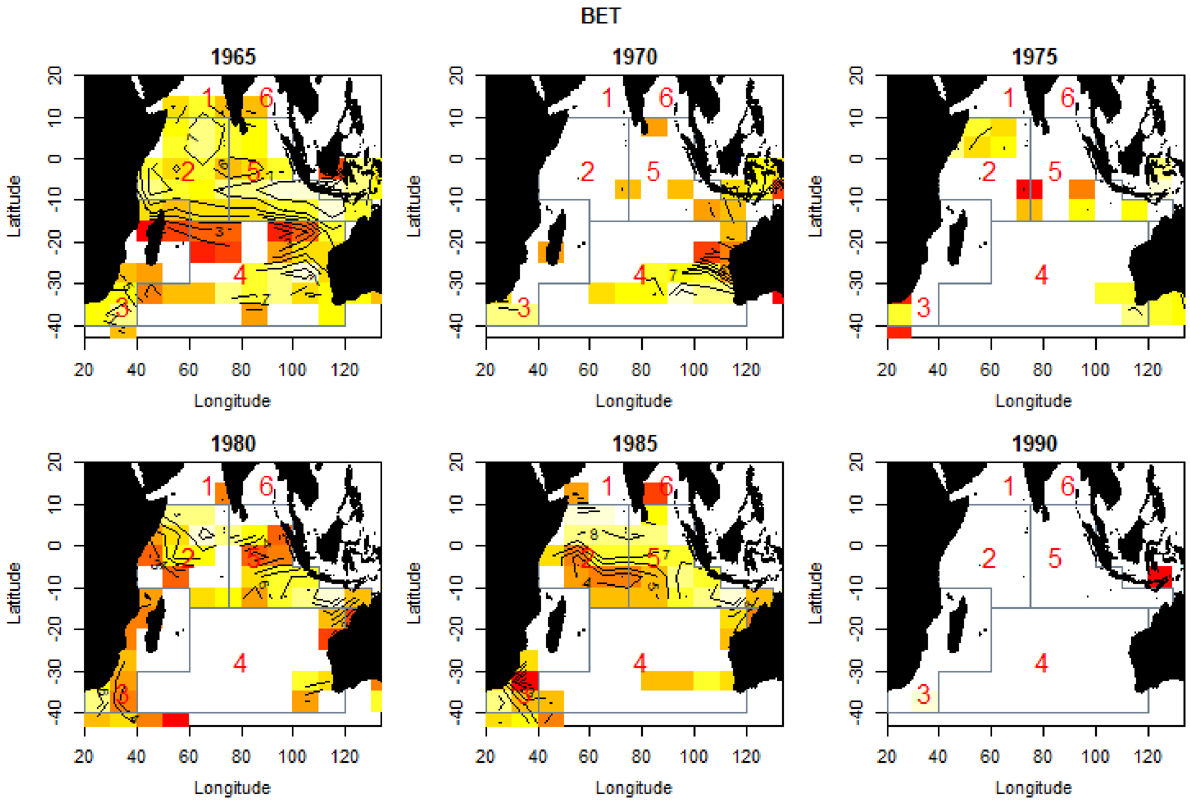

Figure 8: Bigeye sampling locations and intensity for 5 x 10 data by 5-year period.

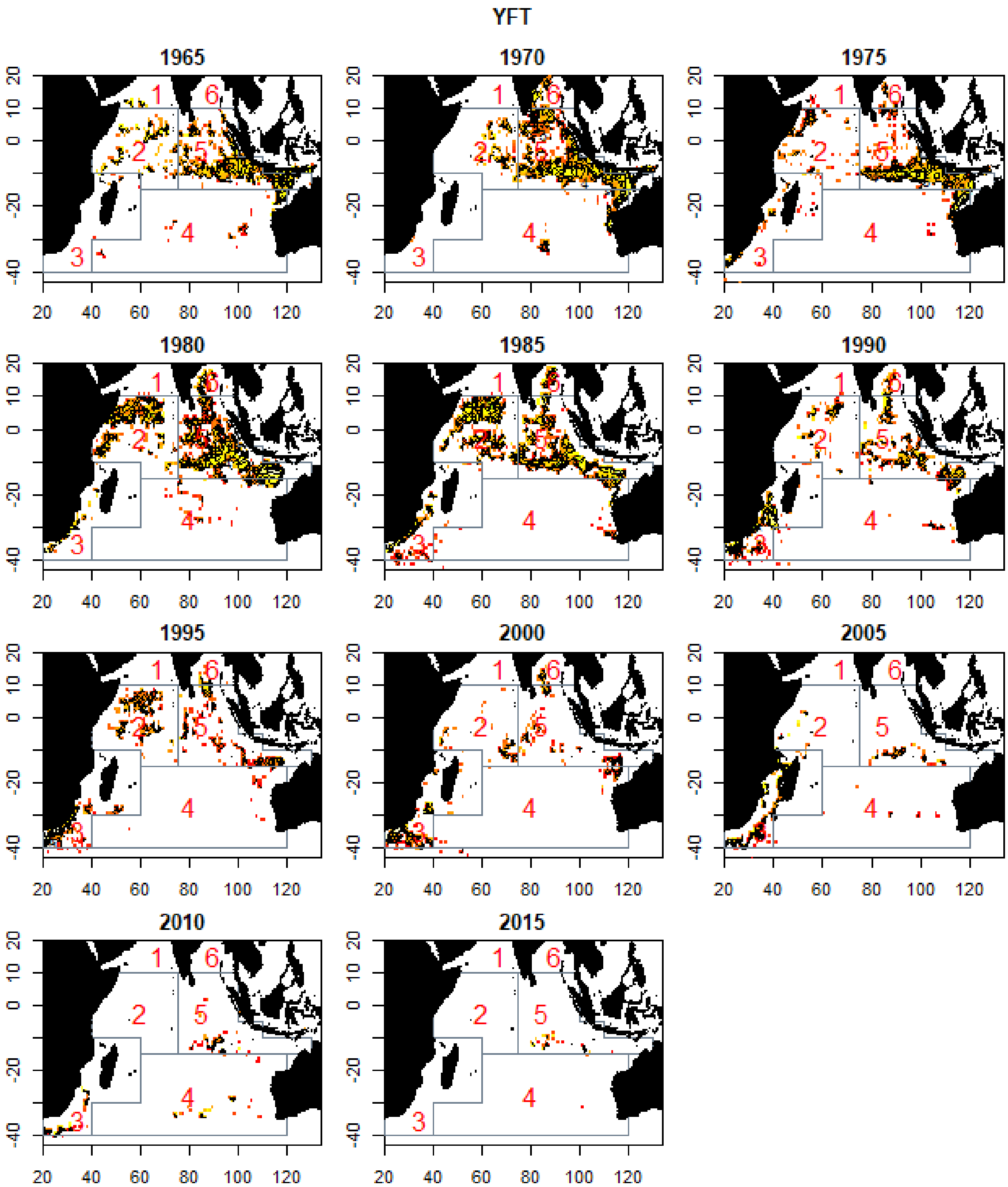

Figure 9: Yellowfin sampling locations and intensity for 1 x 1 data by 5-year period.

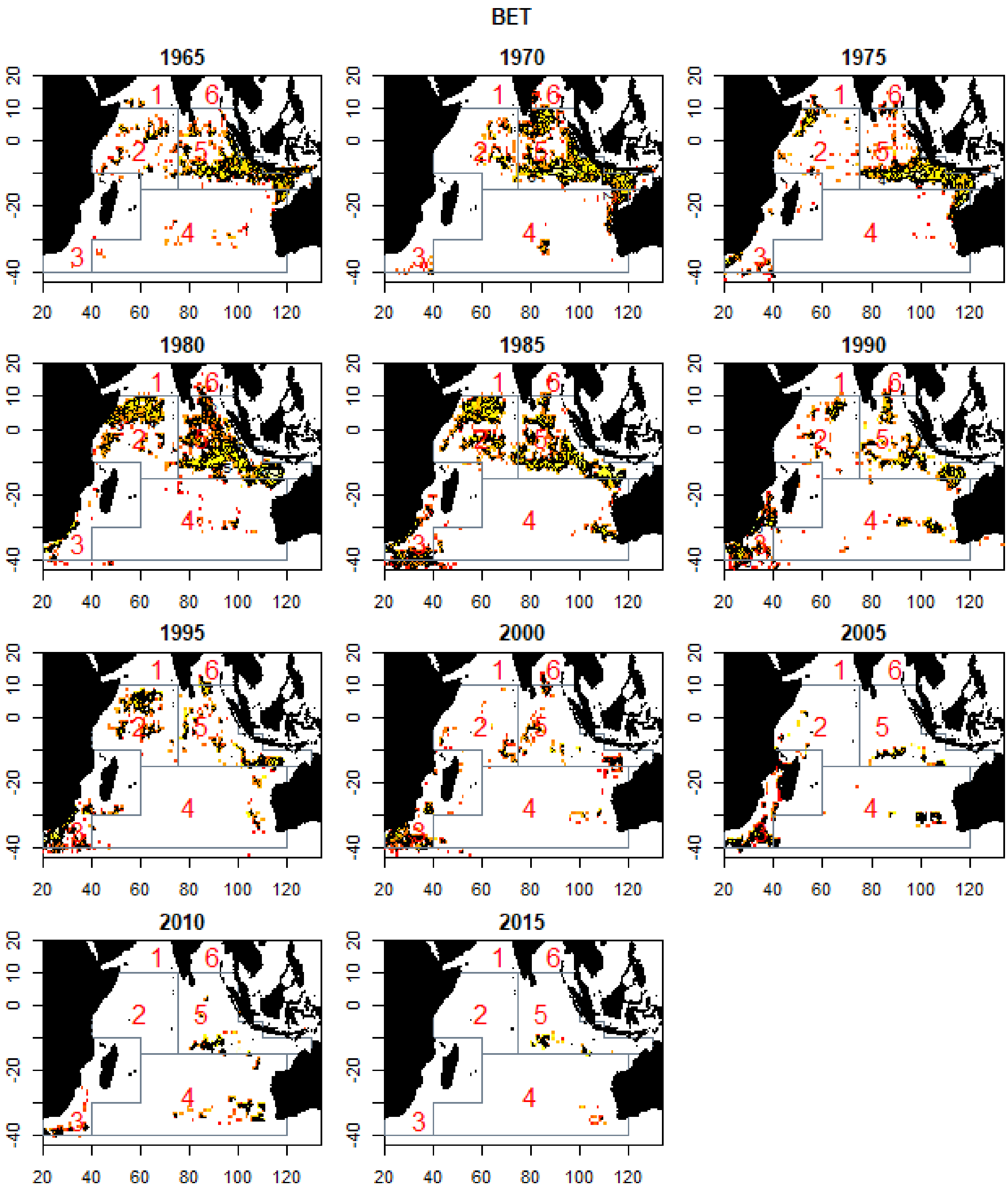

Figure 10: Bigeye sampling locations and intensity for 1 x 1 data by 5-year period.

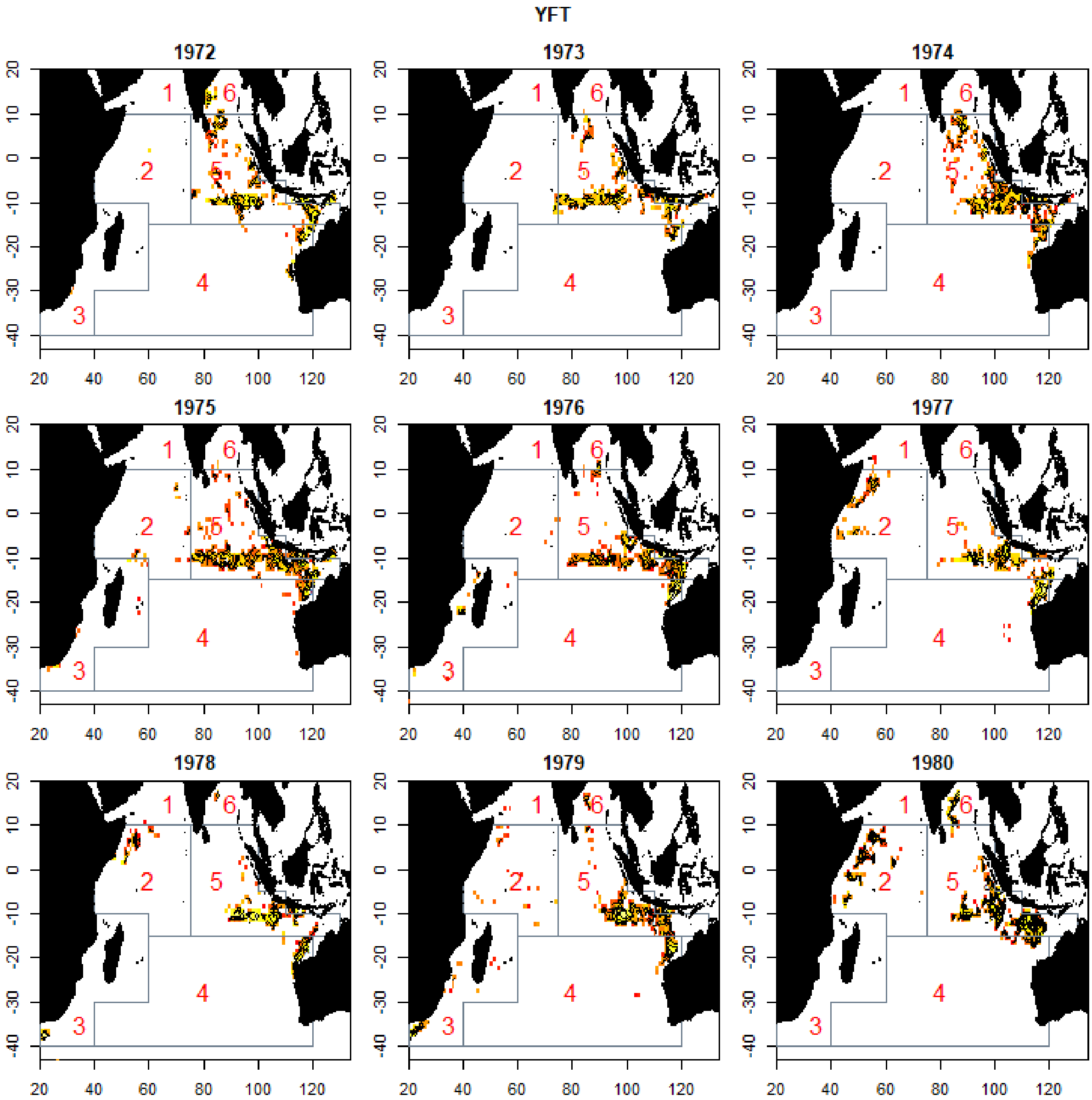

Figure 11: Yellowfin sampling locations and intensity for 1 x 1 data by year, from 1972-1980.

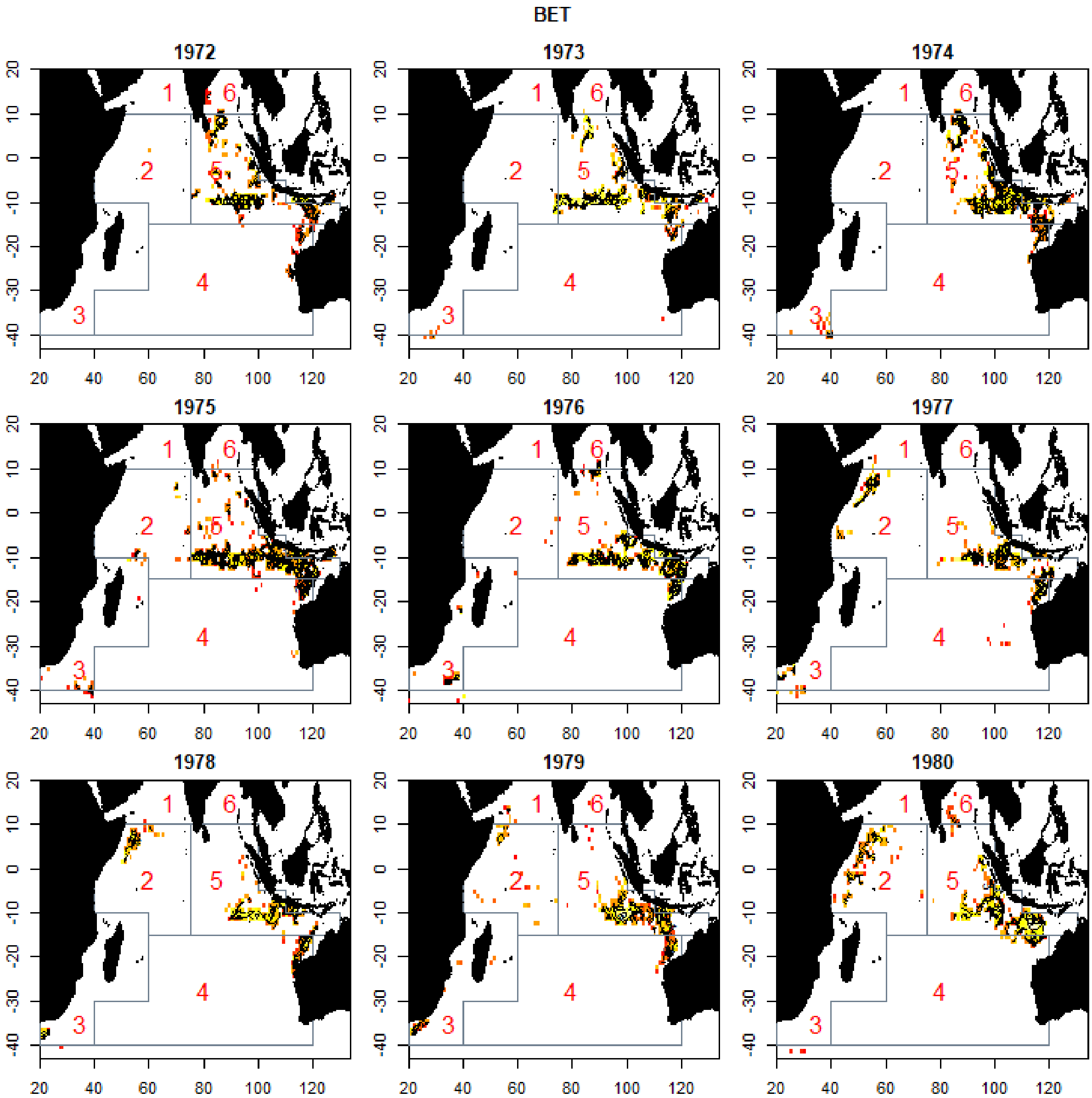

Figure 12: Bigeye sampling locations and intensity for 1 x 1 data by year, from 1972-1980.

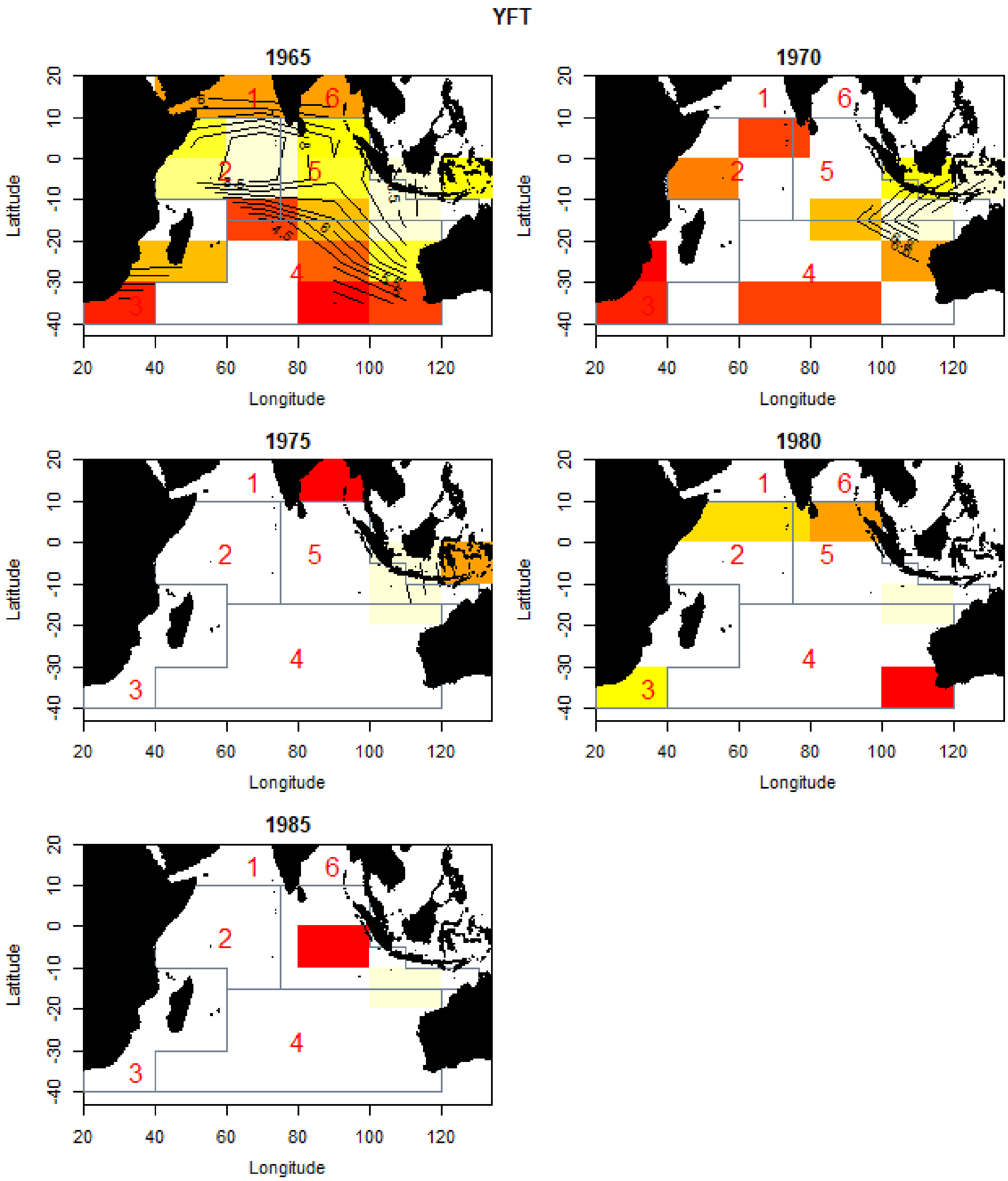
YFT
1965
1970
1975
1980
1985
Latitude
Longitude
1
2
3
4
5
6

Figure 13: Yellowfin sampling locations and intensity for 10 x 20 data by 5-year period.

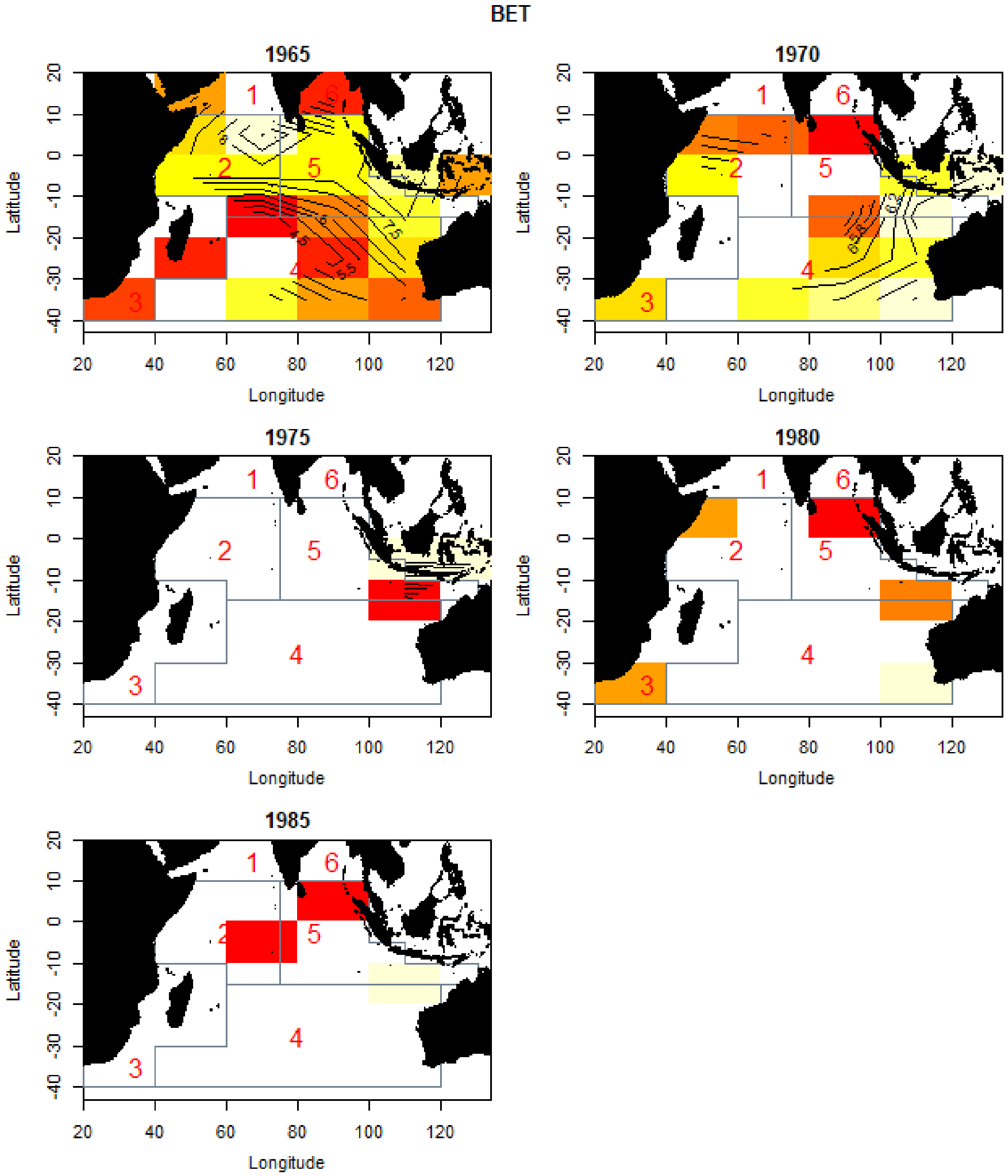

Figure 14: Bigeye sampling locations and intensity for 10 x 20 data by 5-year period.

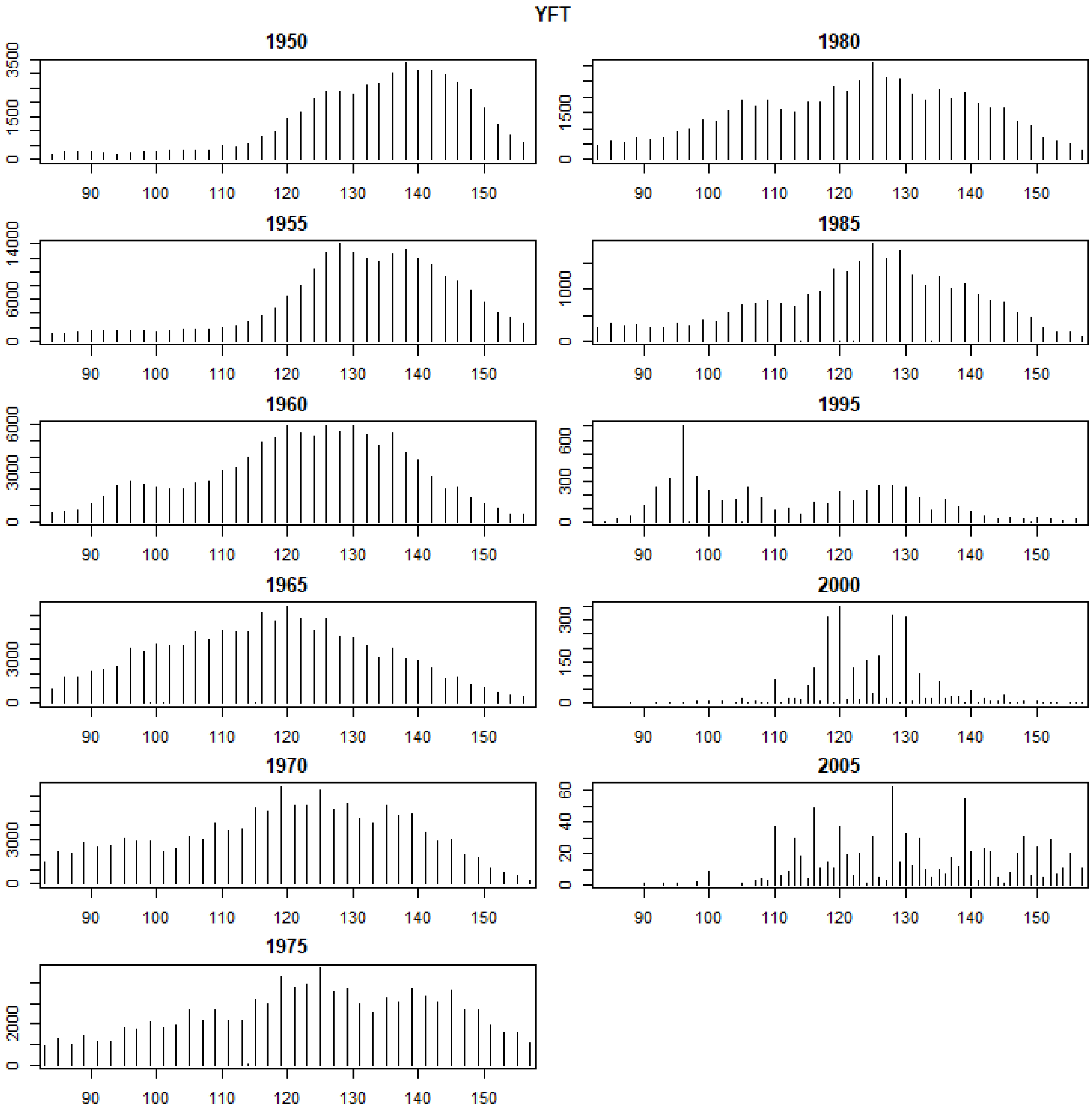

Figure 15: Yellowfin samples per length class with the measurement unit indicating 2cm length bins.

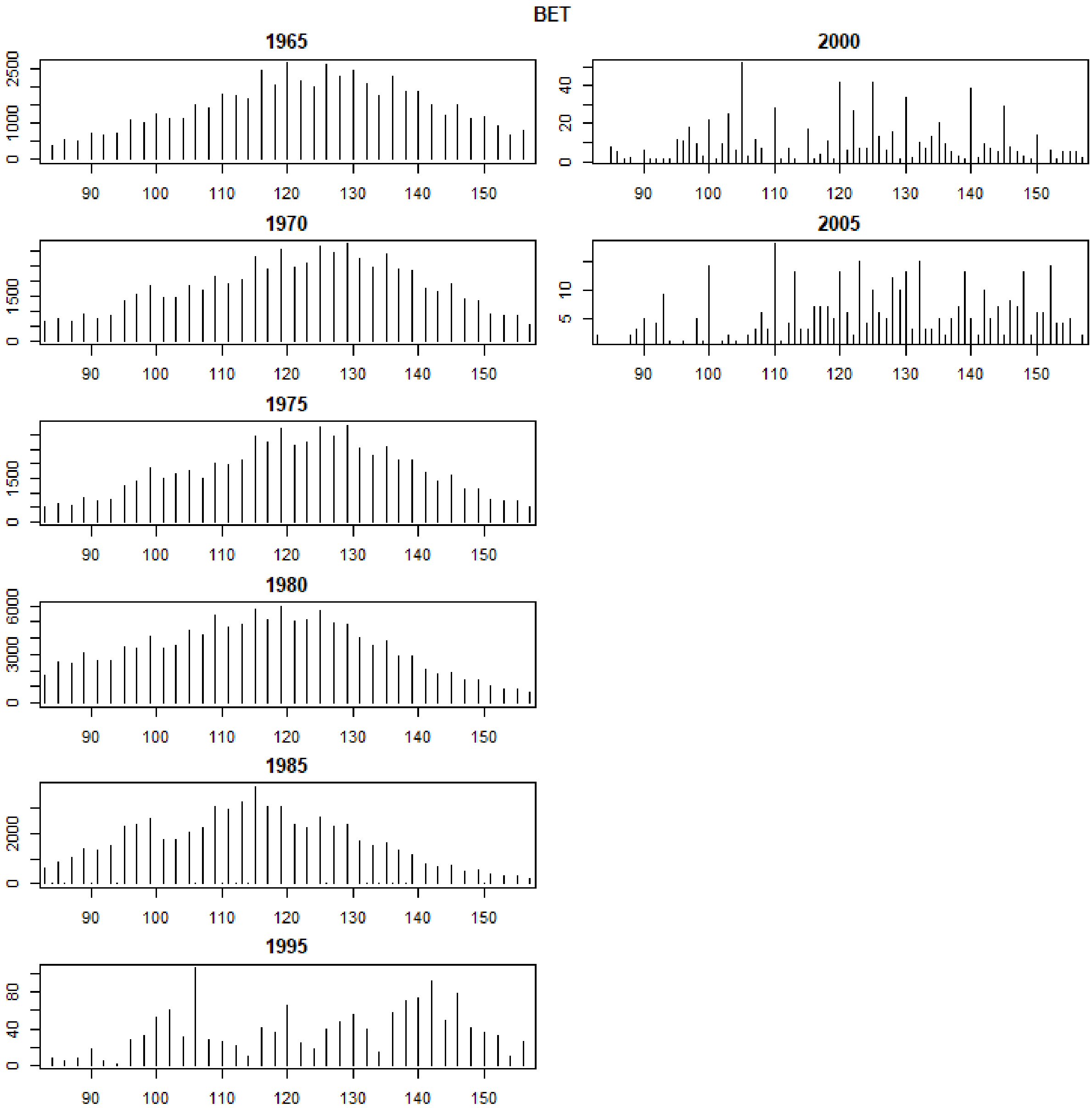

Figure 16: Bigeye samples per length class with the measurement unit indicating 2cm length bins.