SCRS/2018/133 Collect. Vol. Sci. Pap. ICCAT, 75(6): 1363-1372 (2019)

A MIXTURE MODEL INTERPRETATION OF STOCK OF ORIGIN DATA FOR ATLANTIC BLUEFIN TUNA

T. Carruthers[1] and D. S Butterworth[2]

SUMMARY

Stock of origin data provide uncertain inferences for the stock composition of catches even in natal spawning areas. When conditioning operating models that account for stock mixing, these data must be interpreted in as unbiased a manner as possible in order to prevent overestimation of the extent of East-West mixing which leads to an overestimation of Western stock size because of the imbalance between the sizes of the two stocks. Here we provide a data summary of the distribution of current genetics and otolith microchemistry data, and identify an approach using mixture distributions to derive less biased inferences (compared to approaches suggested earlier) from stock of origin data for conditioning operating models. The estimated mixture distributions provide a reasonable fit to assignment data and remove fewer raw data than approaches that discard data based on somewhat ad-hoc rules. It is recommended that operating models be modified to be based on a simpler 7-area model.

RÉSUMÉ

Les données sur le stock d'origine fournissent des conclusions incertaines sur la composition du stock des prises, même dans les zones de frai natales. Lors du conditionnement des modèles opérationnels qui tiennent compte du mélange des stocks, ces données doivent être interprétées de manière aussi neutre que possible afin d'éviter une surestimation de l'ampleur du mélange est-ouest qui entraîne une surestimation de la taille du stock occidental en raison du déséquilibre entre les tailles des deux stocks. Le présent document résume la distribution des données actuelles sur la génétique et la microchimie des otolithes et identifie une approche utilisant des distributions de mélanges pour aboutir à des conclusions moins biaisées (par rapport aux approches suggérées précédemment) à partir des données sur le stock d'origine pour conditionner les modèles opérationnels. Les distributions estimées du mélange fournissent un ajustement raisonnable aux données d'attribution et suppriment moins de données brutes que les approches qui éliminent les données en fonction de règles quelque peu ad hoc. Il est recommandé de modifier les modèles opérationnels afin de les faire reposer sur un modèle plus simple à 7 zones.

RESUMEN

Los datos del stock de origen proporcionan conclusiones inciertas sobre la composición por stock de las capturas incluso en las zonas de desove natales. Al condicionar los modelos operativos que tienen en cuenta la mezcla entre los dos stocks, estos datos deben interpretarse de la manera menos sesgada posible para impedir la sobrestimación del alcance de la mezcla este-oeste que conduce a una sobrestimación del tamaño del stock occidental a causa del desequilibrio entre los tamaños de ambos stocks. Se presenta un resumen de la distribución de los datos actuales de genética y de microquímica de otolitos, y se identifica un enfoque utilizando distribuciones mezcladas para sacar conclusiones menos sesgadas (en comparación con enfoques sugeridos con antelación) de los datos del stock de origen para condicionar los modelos operativos. Las distribuciones de la mezcla estimadas proporcionan un ajuste razonable a los datos de asignación y eliminan menos datos en bruto que enfoques que descartan datos basados en normas algo ad hoc. Se recomienda modificar los modelos operativos para que se basen en un modelo de 7 áreas más simple.

KEYWORDS
Otolith microchemistry, genetics, bluefin tuna, operating model, stock mixing

[1] IOF, 2202 Main Mall, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z4. t.carruthers@oceans.ubc.ca
[2] Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, South Africa. Email: doug.butterworth@uct.ac.za

1. Introduction

When fitting multi-stock models to data, the stock composition of catches (stock of origin) are among the most informative data regarding relative size of the stocks and the extent of their spatial mixing (SCRS/2015/180). The principal sources of these data for Atlantic Bluefin Tuna are otolith microchemistry (Fraile *et al.* 2014) and genetics analysis (Puncher *et al.* 2017). The chemistry and genetics data were collected by several research institutes that use varying approaches to process the data into a single percentage assignment score (to belonging to the Eastern stock) per fish (e.g. random forests, SCRS/2017/021; Bayesian clustering, Pritchard *et al.* 2000). The exact interpretation of these assignment scores is not always completely clear, though Puncher *et al.* (2017) use an arbitrary 70% cut-off below which individuals are considered 'poorly assigned' (i.e. not of Eastern origin). In the analysis presented here we compare our results to a version of Puncher *et al.*'s (2017) approach that discards all assignments in the interval [30%, 70%] ('30-70 cut-off'; assignments above 70% are considered Eastern origin and those below 30% are considered Western origin, those between 30-70 are not used).

We summarize the distribution and coverage of the available stock of origin data across time and ocean areas (**Figure 1**) and highlight two problems associated with the current methods to use assignment scores. Firstly, that the 30-70 cut-off removes a substantial number of data points. Secondly, there is a high degree of uncertainty in the assignment scores even in areas thought to be exclusive to Eastern and Western stocks.

Given this uncertainty in assignment scores, we develop a mixture modelling approach to interpret these data based on the distribution of assignment data in the stock-specific natal areas of the Gulf of Mexico and Mediterranean. We also describe an appropriate likelihood function for these processed data for the conditioning of a multi-stock operating model. Finally, we list some recommended modifications to the Atlantic bluefin tuna Operating model specifications.

2. Methods

Assignment scores are highly variable for both otolith chemistry and genetics data (**Figure 2**). For the more numerous otolith microchemistry data, the median assignment score is not as strong for the Mediterranean (thought to be 100% Eastern origin fish) as it is for South Atlantic and East Atlantic areas (84% median assignment score compared with 87% respectively) (**Table 1**).

The principal problem with the distributions presented in **Figure 2** is that if taken at face value, they imply a greater extent of mixing than is credible given other sources of data such as electronic tagging (for example many assignment scores <60% in the Mediterranean and >40% in the Gulf of Mexico.

Given this variability in assignment scores, we have developed a mixture distribution approach that, given numerous assignment scores, determines the fractional contribution of Eastern and Western fish based on the assignment distributions in the stock-specific natal spawning areas. This approach can be applied only to the Otolith microchemistry data (82% of the assignment data, **Table 2**) since these have distributions for both the Mediterranean and Gulf of Mexico areas to define the signals for "pure" Eastern and Western fish (assignments scores from genetics are not available for the Gulf of Mexico, **Table 1**).

First, we calculated the fraction of assignment data in 5% bins (i=1, 2, 3… n) across the range of 0-100% for both the Gulf of Mexico (GOM) and Mediterranean areas (**Figure 3**, top panel). The fraction of scores corresponding to the i^{th} bin of the Western (GOM) distribution and the Eastern (Mediterranean) distribution are referred to as W_i and E_i, respectively, and the distributions of these fractions is taken to characterise each of these stocks[3]. For any strata (year, quarter, area, age class combinations) s with more than 5 assignment data points, the fraction of assignment data O_i was calculated.

For any mixed strata s, the predicted fraction in each bin i, for the mixture distribution $P_{s,i}$ was calculated from a ratio R_s - the fraction of Eastern fish:

$$P_{s,i} = R_s E_i + (1 - R_s W_i) \qquad (1)$$

For each strata s, R_s was then estimated numerically by minimizing the objective function:

[3] Effectively we are treating these fractions in the same way that separate alleles are treated in genetic analyses. The approach developed is then identical to that used to estimate the proportion of one stock present in a mixture of two stocks from the distribution of alleles for this mixture when data for the distribution of alleles for each stock separately is available.

$$\theta_s = \sum_i ln(\sigma_s) + \frac{(\sqrt{O_{s,i}} - \sqrt{P_{s,i}})^2}{2\sigma_s^2} \quad (2)$$

where σ is the observation variance term. The square root transformation was used as this renders a binomial-like distribution approximately normal.

3. Results

By accounting for the uncertainty in assignments seen in stock-specific natal areas, the mixture modelling approach provides more certain estimates of stock-of-origin for mixed areas (**Table 3**, **Figure 5**). For example, for the East Atlantic (EATL) where the raw assignment scores and the 30-70% cut-off approach include 9/10 assignments in the ranges [55%, 97%] and [71%, 97%], the mixture model infers a much lower proportion of Western fish with 9/10 assignments in the range of [98%, 99%].

The new approach discards considerably fewer of the otolith chemistry data than the 30-70 cut-off approach, losing around an additional 1% of the raw data (once Genetics data are discarded) in contrast to 18% for the 30-70 cut off approach.

In general, the mixture model provides a reasonable fit to the assignment data of mixed strata (**Figure 4**). The approach does however struggle to fit unimodal distributions of assignment data where the mode was of intermediate magnitude (e.g. between 20% and 80%, **Figure 4**). This is a consequence of having two natal distributions *E* and *W* that are strongly skewed, for which it is difficult to develop a combination distribution with a single mode of intermediate magnitude.

4. Discussion

The current assignment data arising from genetics and otolith microchemistry analyses are highly variable even in areas that are thought to be exclusive to each stock (i.e. the Gulf of Mexico and the Mediterranean). Not only does this imply a greater degree of stock mixing than is credible given other data, but importantly the impact on a multi-stock operating model could be relatively large. This is particularly true in the case of Atlantic bluefin tuna where conventional assessments estimate the Western stock to be an order of magnitude smaller than the Eastern stock. In such a case, any 50% assignment score implies equal numbers of Western and Eastern fish in that stratum, leading to an overestimation of the size of the smaller stock that is much larger in proportion to stock size. In other words, the impact of overestimating mixing is likely to be a large positive bias in the estimation of Western stock size.

The mixture model approach applied here removes fewer data than the 30-70 cut-off approach and leads to more precise estimates of stock of origin in each of the 7 geographic areas that are more consistent with observations arising from electronic and conventional tagging data, for example suggesting that the East Atlantic contains a relatively low fraction of Western fish.

The model used to estimate the mixture model was parameterized to estimate *r*, the logit transformation of fraction east: $r = \ln(R/(1-R))$. The standard error σ_r in estimates of *r* were obtained by inverting the Hessian matrix returned from the numerical optimizer (using the R-software function *optim* under default settings).

This made it possible to derive values r_s and their approximate standard errors $\sigma_{r,s}$ for each stratum (with more than 5 assignment data points), which can then be used to condition operating models.

In this case the negative log likelihood contribution to the global objective function of the operating model is calculated by a Gaussian likelihood function (without constants) comparing model predicted $\hat{r}$ with *r* derived using the mixture model:

$$-lnL_{SOO} = \sum_s ln(\sigma_{r,s}) + \frac{(\hat{r}_s - r_s)^2}{2\sigma_{r,s}^2} \quad (3)$$

where the operating model estimated logit fraction $\hat{r}_s$ is calculated from the operating model predicted ratio of Eastern fish in the catch $\hat{R}_s$: $\hat{r}_s = ln(\hat{R}_s/(1-\hat{R}_s))$.

Data to inform stock composition of catches have been collected over the spatial range of Atlantic bluefin stocks (i.e. for all 7 ocean areas of **Figure 1**); however, there is a high degree of inconsistency in spatial sampling intensity (**Tables 1 and 4**) with the majority of data collected and processed in Western areas such as the Gulf of Mexico, West Atlantic and Gulf of St Lawrence (~75%).

The seasonal coverage is also patchy, having for example no observations in quarter 1 (Jan – Mar) in four of the 7 areas (GSL, NATL, EATL, MED) (**Table 4**). If more than 5 observations are required to inform on the stock composition of catches, then around a third (10 of the 28) of quarter-area strata are missing the required data. When fitted to a range of data types, operating models must inform stock composition (mixing) in these strata from the substantially less informative electronic tagging data. It remains to be seen whether the operating model must be simplified in order for all parameters to be informed by the data available.

5. Conclusions

The following recommendations arise from these analyses.

i) Given the efficient use of available otolith microchemistry data and the more precise and credible information regarding spatial stock composition, it is recommended that the new mixture modelling approach be used for the conditioning of operating models.

ii) Revisiting the available otolith chemistry data reveals that these are patchy in terms of seasonal-spatial coverage even for a 7-area model (**Figure 1**). Since these are the primary data informing mixing and stock size, it is recommended that operating models move to this simpler spatial model of 7 areas.

iii) We also recommend a reconfiguration of the movement estimation which has similar movement among the three age classes unless there is evidence to suggest otherwise (i.e. seasonal movements missing data default to the seasonal movements of other age classes that have data rather than a prior of full mixing).

iv) Rather than impute data for the North Atlantic area, the stock mixing should be informed by the available electronic tagging data. If this is problematic, imputation of stock assignment data or the simplification of spatial strata should be considered as a secondary measure.

6. Acknowledgements

Many thanks are due for the provision of data and feedback by Haritz Arrizabalaga, Alex Hanke and David Secor. This work was carried out under the provision of the ICCAT Atlantic Wide Research Programme for Bluefin Tuna (GBYP), funded by the European Union, several ICCAT CPCs, the ICCAT Secretariat and by other entities (see: http://www.iccat.int/GBYP/en/Budget.htm). The contents of this paper do not necessarily reflect the point of view of ICCAT or other funders and in no way anticipate ICCAT future policy in this area.

References

Fraile, I., Arrizabalaga, H., Rooker, J. 2014. Origin of Atlantic Bluefin tuna (*Thunnus thynnus*) in the Bay of Biscay. ICES. J. Mar. Sci. doi:10.1093/icesjms/fsu156

Pritchard, J. K., Stephens, M., Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155(2), 945–959.

Puncher, G.N. *et al.* 2018. Spatial dynamics and mixing of Bluefin tuna in the Atlantic Ocean and Mediterranean Sea revealed using next-generation sequencing. DOI: 10.1111/1755-0998.12764

Table 1. Summary of the assignment scores by area.

Type	Percentile	GOM	WATL	GSL	SATL	NATL	EATL	MED
Otolith microchemistry	**5th**	0%	1%	4%	14%	6%	48%	32%
	Median	**7%**	**27%**	**23%**	**87%**	**75%**	**87%**	**84%**
	95th	48%	97%	89%	99%	97%	98%	97%
Genetics	**5th**	0%	0%	0%	4%	9%	29%	40%
	Median	**0%**	**45%**	**56%**	**82%**	**96%**	**98%**	**99%**
	95th	94%	100%	100%	100%	100%	100%	100%

Table 2. The frequency of SOO data by type and ocean area. These are assignment scores for years before 2016 (the current operating model runs to 2015 after which spatial total catch data 'CATDIS' are not available) that have covariate area data and either age or length data.

	GOM	WATL	GSL	SATL	NATL	EATL	MED	Total	%
Otolith Chemistry	319	2537	864	382	409	335	245	5091	79.7%
Genetics	214	123	34	294	172	82	380	1299	20.3%

Table 3. The assignment scores by ocean area for the raw, the 30-70 cut-off and the new mixture distribution approach. The % raw data is lower than 100% for the raw approach since genetics assignment data are not included. The mixture distribution is lower than this again as it excludes any strata with fewer than 5 observations (too few to characterize a distribution). The mixture distribution approach assumes that GOM and MED fish are 0% and 100% Eastern respectively.

Method	Percentile	GOM	WATL	GSL	SATL	NATL	EATL	MED	% raw data
Raw	5th	0%	1%	3%	8%	7%	45%	37%	
	Median	**4%**	**28%**	**23%**	**85%**	**81%**	**90%**	**95%**	**100.0%**
	95th	67%	98%	91%	100%	100%	100%	100%	
30-70 cut-off	5th	0%	0%	3%	7%	5%	72%	71%	
	Median	**3%**	**16%**	**15%**	**91%**	**88%**	**92%**	**96%**	**79.9%**
	95th	72%	98%	92%	100%	100%	100%	100%	
Mixture distribution	5th	0%	15%	46%	10%	64%	89%	100%	
	Median	**0%**	**29%**	**46%**	**90%**	**90%**	**97%**	**100%**	**98.8%**
	95th	0%	76%	46%	98%	97%	98%	100%	

Table 4a. Seasonal-spatial coverage of the raw otolith chemistry data. Orange shaded cells represent quarter-area strata for which there are no stock of origin data available for the mixture model approach.

Quarter	GOM	WATL	GSL	SATL	NATL	EATL	MED	Total	%
1: Jan-Mar	52	347	0	37	0	0	0	436	8.6%
2: Apr-Jun	267	303	0	259	0	22	154	1005	19.7%
3: Jul-Sept	0	1565	604	54	8	291	91	2613	51.3%
4: Oct-Dec	0	322	260	32	401	22	0	1037	20.4%
Total	319	2537	864	382	409	335	245		
%	6.3%	49.8%	17.0%	7.5%	8.0%	6.6%	4.8%		

Table 4b. Seasonal-spatial coverage of the raw genetics data. Orange shaded cells represent quarter-area strata for which there are no stock of origin data available for the mixture model approach.

Quarter	GOM	WATL	GSL	SATL	NATL	EATL	MED	Total	%
1: Jan-Mar	214	0	0	58	0	0	0	272	20.9%
2: Apr-Jun	0	0	0	139	0	0	223	362	27.9%
3: Jul-Sept	0	84	26	34	26	82	41	293	22.6%
4: Oct-Dec	0	39	8	63	146	0	116	372	28.6%
Total	214	123	34	294	172	82	380		
%	16.5%	9.5%	2.6%	22.6%	13.2%	6.3%	29.3%		

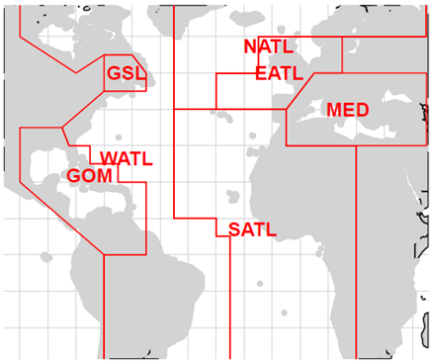

Figure 1. The spatial definitions for the seven areas used in this analysis.

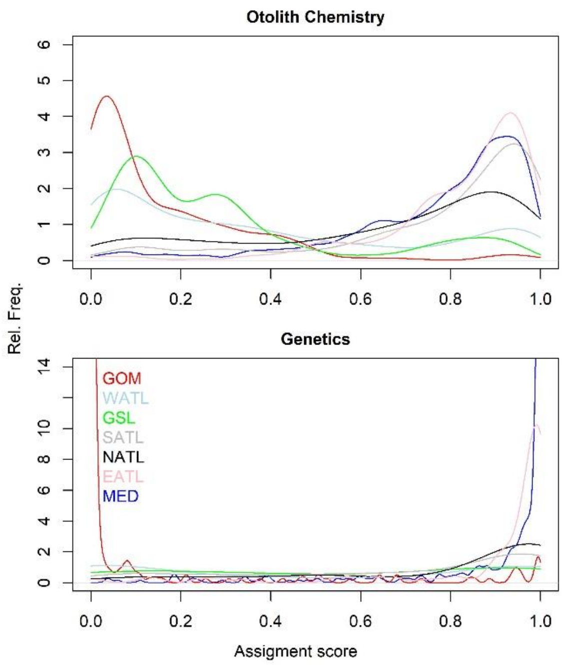

Figure 2. The distribution of assignment scores in each area by type (otolith chemistry, genetics).

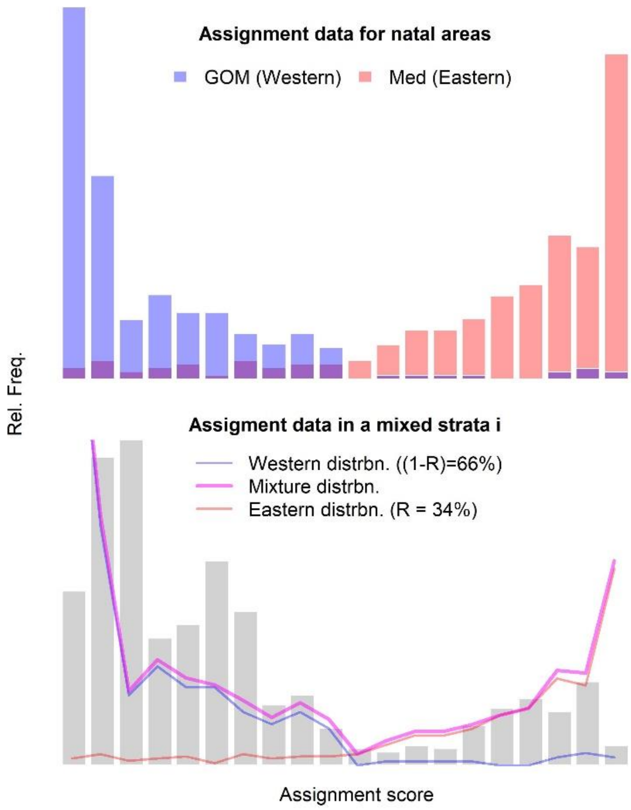

Figure 3. An illustrative example of the mixture distribution approach that combines fractions of assignment distributions from natal areas (W, blue; E red) (top panel) to describe a mixture distribution (P) (purple line, bottom panel) that fits the observed assignment data (O) (grey bars) in a mixed stratum.

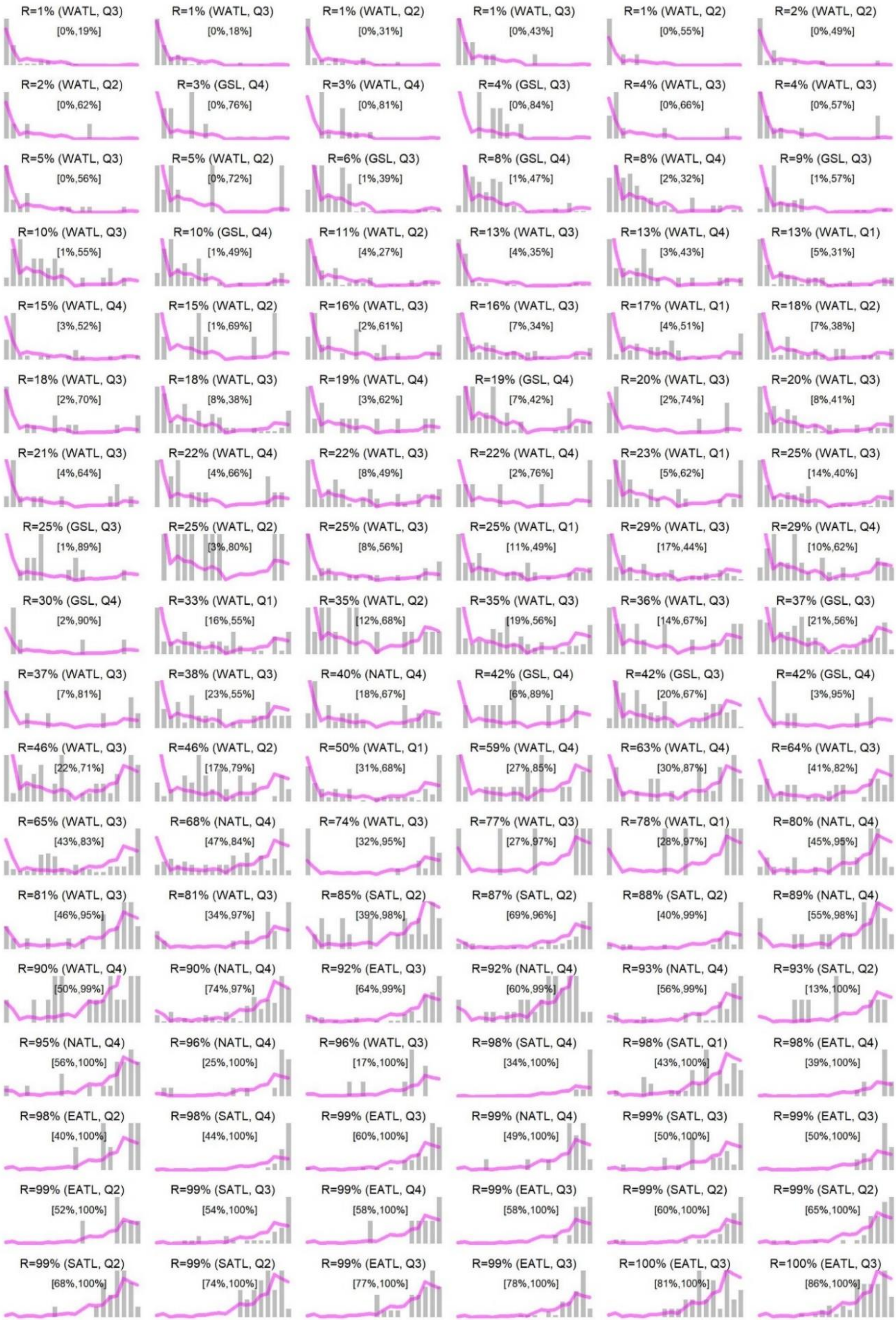

Figure 4. Fitted mixture distributions for all 108 strata (years, areas, quarters and age classes) for which more than 5 assignment data were available and derived by otolith microchemistry analysis. Grey bars are the observed frequencies of assignment data (*O*), purple lines are the fitted mixture distributions (*P*). In each panel the fraction of Eastern fish is reported (*R*). The square bracketed numbers represent the 95% confidence interval for *R*.

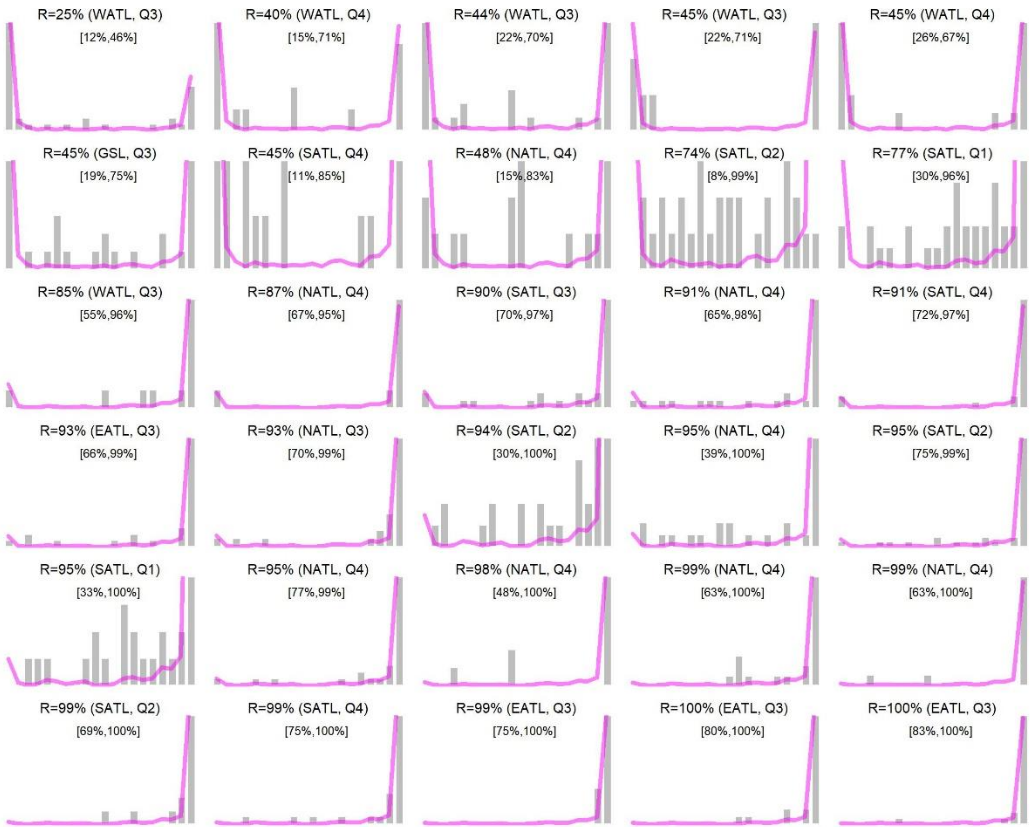

Figure 5. As Figure 4 but based on the genetics assignment data.

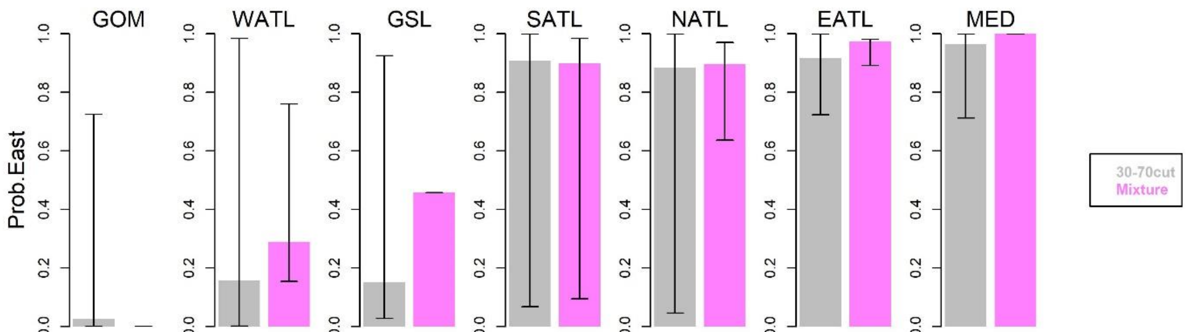

Figure 5. The assignment scores (of **Table 3**) by oceans areas for the 30-70 cut-off approach and the mixture model. Black bars represent the 5th and 95th percentiles. For the mixture model these are approximate, and based on the assumption that the square root transformation used in equation (2) provides distribution normality.