Evaluating the Uncertainty Grid Using the 2024 stock assessment: Applying Diagnostic Tools

Huihua Lee[a], Desiree Tommasi[a,b]

a: NOAA Fisheries, Southwest Fisheries Science Center,
La Jolla, CA, USA
b: Institute of Marine Sciences, University of California Santa Cruz,
Santa Cruz, CA, USA

April 2024

Working document submitted to the ISC Pacific bluefin tuna Working Group, International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean (ISC), from 9-10 April 2024, Webinar.

Summary

Fishery management can rely on robust management strategy evaluations (MSE) to inform decision-making in the face of uncertainties. MSE assesses feedback-control management strategies by simulating future scenarios, considering uncertainties in the system. For parameter uncertainty, productivity parameters such as length at age 3, natural mortality for age 2 and older, and the steepness of the stock-recruitment relationship greatly impacted the historical trajectory of Pacific bluefin tuna spawning stock biomass. In this working paper, a comprehensive evaluation of multiple diagnostic criteria provided valuable insights. Jitter analyses guided the exclusion of grids with 0% successful runs in subsequent diagnosis and selection processes. The assessment of goodness-of-fit provided inconsistent grid profiles among data sources, leading to exclusion from the selection process. Consistency in R_0 profiles and retrospective analyses further emphasized the need to exclude grids with data conflicts and unfavorable Mohn's ρ values. ASPM-R models reinforced the significance of avoiding grids with statistically significant degradation in NLLs. Ensemble diagnostic results consolidated these findings, recommending only grids passing three or more diagnostics for selection. The conflicting information observed underscores the necessity of a comprehensive approach to ensure the robustness and reliability of selected grids for subsequent modeling applications.

Introduction

Fishery managers and decision-makers rely on the outcomes of management strategy evaluation (MSE) to determine which management strategies will be implemented in the future. One notable benefit of MSE is its ability to assess management strategies under a range of uncertainties in the system. Uncertainties are five fold in MSE: (1) process uncertainty, (2) parameter uncertainty, (3) model uncertainty, (4) errors in data and observation systems when conducting assessments, and (5) implementation uncertainty, as outlined by Punt et al. in 2016.

The ISCPBF working group identified productivity parameters as the most influential and uncertain factors among the examined uncertainties, which include model uncertainty and errors in data and observation systems (ISC 2022). These productivity parameters include length at age 3 (L_2), natural mortality for age 2 and older (M_2^+), and the steepness of the stock-recruitment relationship (h).

The uncertainty grid associated with the identified productivity parameters and their plausible values was previously examined by Lee and Tommasi in 2023 using the

2022 assessment model. In this working paper, we incorporated the fishery data into each grid model and applied the diagnostic tools using the 2024 benchmark stock assessment, including jitter analyses, goodness-of-fit, likelihood profile on R_0, retrospective analyses, and ASPM-R, to eliminate underperforming grids as described in Lee and Tommasi (2023).

Methods

Diagnostics on the grid model

1. *Convergence and stability*

To evaluate convergence towards a global minimum, we conducted 25 jitter analyses for each grid. This process involved randomly perturbing the initial values of all parameters by 10% and subsequently re-running the model. The primary objective of these jittering analyses was to ensure that none of the randomly generated starting values of parameters led to a solution with a lower total negative log-likelihood (NLLs) compared to the reference model. The final reference model had the lowest total NLL and a positive-definite Hessian matrix. These analyses served as a quality control procedure to confirm that the model was not converging towards a local minimum.

2. *Goodness-of-fit*

We used total NLLs to guide our assessment of the goodness-of-fit for both data components (abundance indices and size composition). We utilized the NLL values from the 2024 stock assessment as the basis to determine whether the grid models fit each data component better or worse. A statistically significant worse fit in the alternative grid from the base grid was defined when the increase in NLLs exceeded 1.92 units.

3. *Model consistency*

3.1. *R_0 likelihood profile*

The R_0 likelihood profile served as a tool for assessing which data sources provided information on a global scale and for pinpointing regions where conflicts arose among these sources (Lee et al. 2014). The profile involved running a series of models, where the ln(R_0) parameter was fixed (not estimated) at a range of values both above and below the estimated derived within the model. This process quantifies the extent of loss of fit for each data component resulting from changing the population scale. Data components rich in information on population scale will exhibit substantial degradation in fit when the population scale deviates from the best estimate.

Following the completion of all profile runs, the degradation in fit was computed by subtracting the overall and component's minimum NLL (or best fit) across all profile runs from the overall and component's NLL from each specific profile run, respectively. We calculated the 95% confidence interval for the changes in NLL around R_0^{MLE} (R_0 at the minimal total likelihood estimates), corresponding to half of the chi-squared values for p=0.95 with 1 degree of freedom. Ultimately, if R_0^c for the data component at the minimal likelihood estimates falls outside the 95% confidence interval for R_0^{MLE}, it indicates a conflict with the overall model. Conversely, if R_0^c for the data component at the minimal likelihood estimates falls inside the 95% confidence interval for R_0^{MLE}, the data component aligns with the overall model on a global scale. This entire process is iterated for each model grid.

3.2. Retrospective analyses

A retrospective analysis was used to examine consistency of model output once recent data were systematically removed from each of the potential grid models. The underlying assumption is that estimates of historical abundance using all data are more accurate than estimates from retrospective models that ignore recent data. Therefore, this analysis reveals potential biases within model estimates. A 7-year retrospective analysis was conducted across all model grids by sequentially removing one year of data at a time. Subsequently, the Mohn's rho statistic (Hurtado-Ferro et al. 2014) was calculated to quantify the severity of retrospective patterns. A greater absolute Mohn's rho indicates a consistently obvious pattern of change in the retrospective models.

3.3. Age-structured production model with recruitment (ASPM-R)

The age-structured production model diagnostic (ASPM; Maunder and Piner 2015) served as a diagnostic tool to evaluate the current state of the production function and to identify potential misspecifications in the system dynamics (Carvalho et al. 2017). To account for cohort growth, we modified the ASPM, introducing the ASPM-R model, which allows for recruitment deviations to be specified at previously estimated value in addition to selectivities.

Initially, each grid model was fitted to catch, size compositions, and abundance indices (adult and recruitment indices) as in the assessment model, but with alternative productivity assumptions. Subsequently, the ASPM-R model was conducted, incorporating recruitment deviations and selectivities specified at the estimates from the full dynamics model. The ASPM-R model estimated scaling parameters (ln(R0) and R1) and the initial fishing mortality rates, fitting to catch and adult abundance indices.

Comparison between the ASPM-R model with the alternative grids and the base grid was then conducted. Statistical degradation was defined when the total likelihood in the ASPM-R model with the alternative grid was more than 1.92 likelihood units different from the total likelihood from the ASPM-R model with the base grid.

Results

1. *Convergence and stability*

When M_2^+ is 0.25, the percentage of jitter runs resulting in a positive-definite Hessian matrix generally increased with higher steepness values, regardless of L_2 (Table 1 and Figure 1). However, when M_2^+ is 0.193, the percentage of jitter runs resulting in a positive-definite Hessian matrix was low when steepness values are between 0.95 and 0.97. **Any grid with 0% of runs resulting in a positive-definite Hessian matrix will not be considered in the subsequent diagnostics and the selection process.**

2. *Goodness of fit*

The NLL values for the index data components suggest that most grids performed similar to or better than the base grid (yellow highlighted in Table 2). The NLLs for the size compositions indicate that more grids with a fit similar to or better than the base grid were achieved as L_2 decreased. For all data compositions, the NLLs conclude that more grids with a fit similar to or better than the base grid were achieved as L_2 decreased. **The index and size composition components provided inconsistent grid profiles and therefore, goodness-of-fit will not be considered in the selection process.**

3. *Model consistency*

3.1. R_0 profile

The R_0 profile plots for each grid are displayed in Figure 2. Only size components provided consistent estimates of the global scale (ln(R_0)) for the base grid, with $R_0^{c=size\ comp}$ (R_0 at the minimal likelihood estimates for the size data component) falling within the 95% confidence interval for R_0^{MLE}. This consistency, as in the base grid, was also observed in most of the grids (Table 3). **Any grid lacking the same consistency as in the base grid will not be considered in the selection process.**

3.2. Retrospective analyses

7-year retrospective analyses of spawning stock biomass for each grid are displayed in Figure 3. The Mohn's ρ value for spawning stock biomass from the base grid was 0.01 (Table 4). Other grids exhibited similar or smaller Mohn's ρ values compared to the base grid. When M_2^+ is 0.25, the retrospective pattern increased as *h* decreased,

accompanied by a larger absolute Mohn's ρ. However, M_2^+ is 0.193, the retrospective pattern decreased as *h* decreased with a smaller absolute Mohn's ρ. **Any grid with an absolute Mohn's ρ value larger than 0.1 will not be considered in the selection process.**

3.3. Age-structured production model with recruitment (ASPM-R)

Table 5 displays the total negative log-likelihood (NLL) values from the ASPM-R models for each grid. The NLLs generally deteriorated when *h* was smaller than the base value, regardless of M_2^+ or L_2 values. The selected range of *h* expanded when either M_2^+ or L_2 was larger. In the case of M_{2+}=0.25, the selected *h* values ranged from 0.99 to 0.999 when L_2 was 118.57, while the selected *h* expanded from 0.97 to 0.999 when L_2 was 119. **Any grid displaying a statistically significant degradation in NLLs, thus hindering the production relationship, will be excluded from the selection process.**

4. Ensemble diagnostic results

Table 6 represents a summary of selections based on the convergence, R_0 profile, retrospective, and ASPM-R analyses for each grid. The scores range from 0 to 4, with the highest score indicating successful passage of all four diagnostics. The scores reveal conflicting information across retrospective analyses, R_0 profile, and ASPM-R. Specifically, ASPM-R favored higher values for M_2^+ and *h*, while R_0 profile leaned towards lower values for *h*. **In summary, only grids that passed three or more diagnostics were recommended**.

The uncertainty range of the spawning biomass and spawning stock biomass ratio for the selected grids are shown in Figure 4.

Conclusion

This working paper relied on the methods established in Lee and Tommasi (2023). Lee and Tommasi (2023) provided the justification on selecting a range of productivity parameters based on data and life history information. They comprehensively evaluated multiple diagnostic criteria. Based on data structure and model structure from the 2024 stock assessment, we selected suitable grids for further consideration.

The analysis of convergence and stability highlighted the influence of M_2^+ and steepness values on the positive-definite Hessian matrix, guiding the exclusion of grids with 0% successful runs from subsequent diagnostics tests and selection processes. The assessment of goodness of fit, particularly in relation to NLL values, revealed inconsistent likelihood profiles between index and size compositions, necessitating the exclusion of goodness-of-fit considerations from the selection process. Model consistency, as

evaluated through R_0 profile plots and retrospective analyses, emphasized the importance of consistent estimates and patterns across various components, thereby excluding grids displaying clear data conflicts or worse Mohn's ρ values. The ASPM-R models further reinforced the significance of avoiding grids with statistically significant degradation in NLLs, as this implies a more poorly estimated production relationship. The ensemble diagnostic results provided a consolidated overview. We recommend that only grids passing three or more diagnostics be selected. The conflicting information observed across retrospective analyses, R_0 profiles, and ASPM-R underscores the importance of a comprehensive approach in ensuring the robustness and reliability of selected grids for subsequent modeling applications. This work serves as the basis for the ISCPBF working group to select the uncertainty range in productivity parameters to be considered for the MSE operating model(s) (i.e., 'conditioning' the operating model(s) to data).

References

Carvalho, F., Punt, A. E., Chang, Y.-J., Maunder, M. N., Piner, K. P., 2017. Can diagnostic tests help identify model misspecification in integrated stock assessments? Fisheries Research. 192: 28-40. https://doi.org/10.1016/j.fishres.2016.09.018

Hurtado-Ferro F., Szuwalski C.S., Valero J.L., Anderson S.C., Cunningham C.J., Johnson K.F., Licandeo R., McGilliard C.R., Monnahan C.C., Muradian M.L., Ono K., Vert-Pre K.A., Whitten A.R., Punt A.E., 2014. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. Ices J. Mar. Sci., 72(2015), pp. 99-110.

ISC 2022. Stock Assessment of Pacific Bluefin Tuna in the Pacific Ocean in 2022. Annex 13 22nd Meeting of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean. Available at https://isc.fra.go.jp/pdf/ISC22/ISC22_ANNEX13_Stock_Assessment_for_Pacific_Bluefin_Tuna.pdf

Lee, H.H., Piner, K.R., Methot, R.D., Maunder, M.N., 2014. Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: an example using blue marlin in the Pacific Ocean. Fish. Res. 158, 138–146.

Lee, H., Tommasi, D., 2023. Evaluating the uncertainty grid: Applying diagnostic tools ISC/23/PBFWG-2/12.

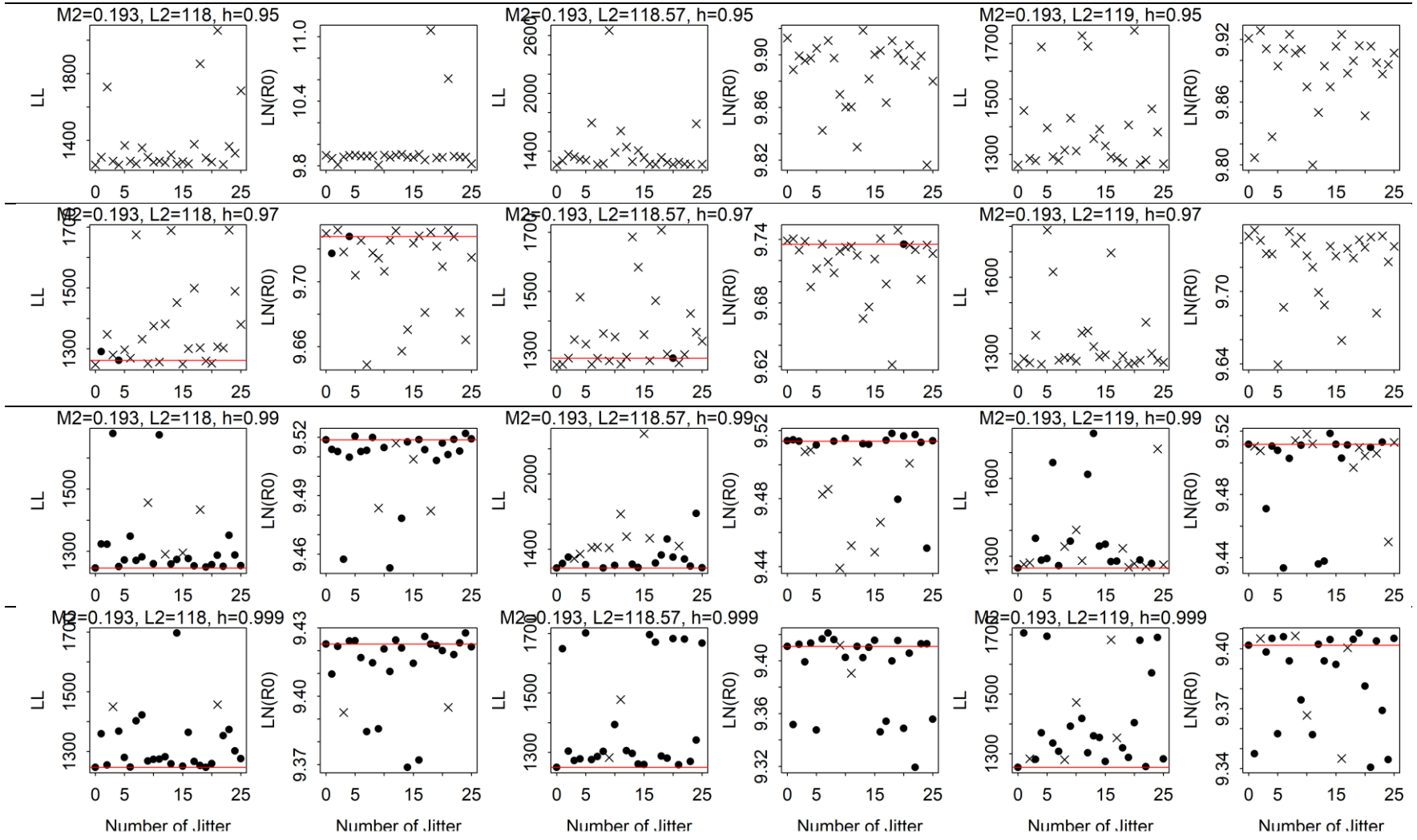
Maunder, M. N., Hamel, O. S., Lee, H., Piner, K. R., Cope, J. M., Punt, A. E., Ianelli, J. N., Castillo-Jordan, C., Kapur, M. S., Methot, R. D., 2023. A review of estimation methods

for natural mortality and their performance in the context of fishery stock assessment. Fisheries Research. 257: 106489. https://doi.org/10.1016/j.fishres.2022.106489

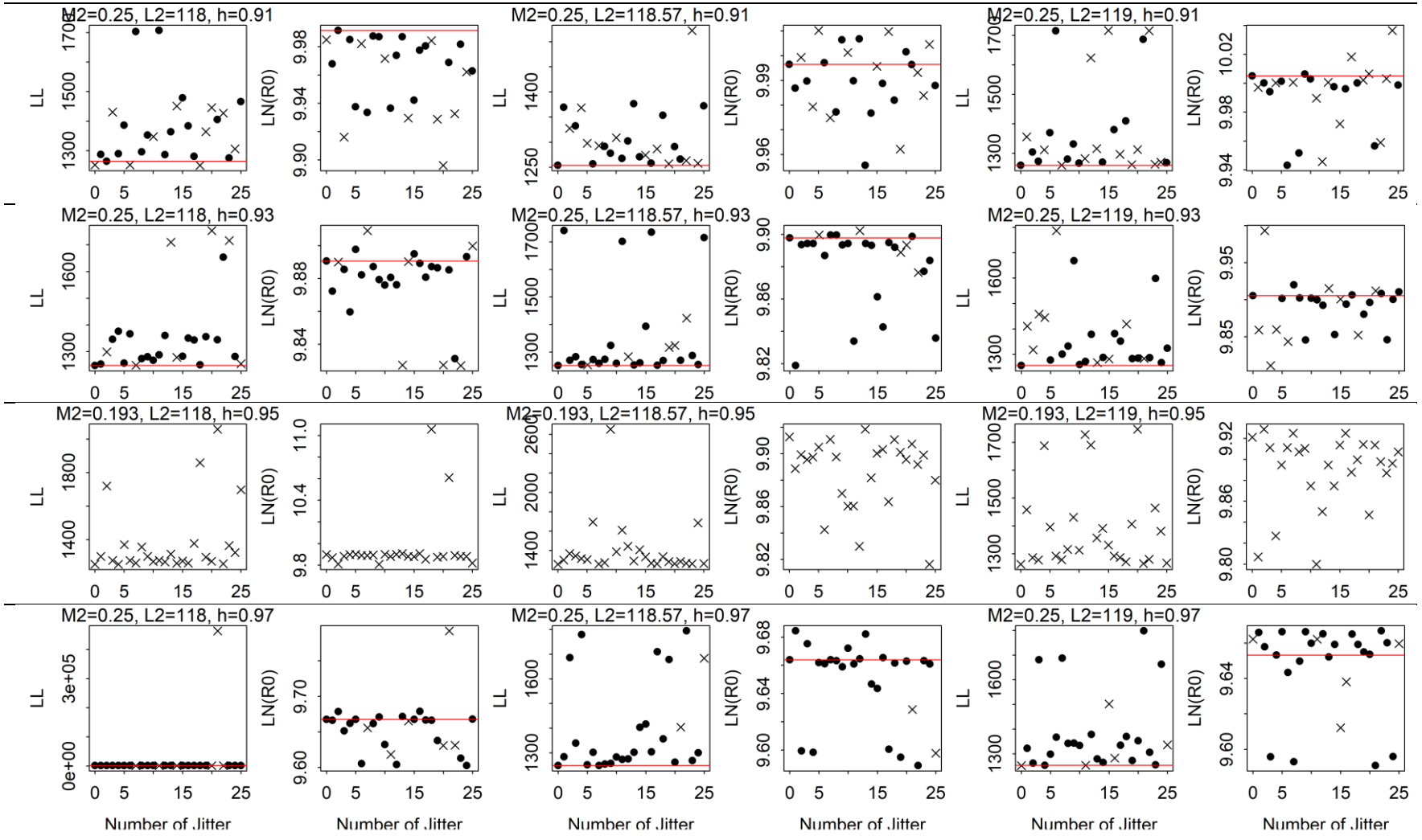
Table 1. The percentage of runs resulting in a positive-definite Hessian matrix in jitter analyses from models that varied by changing the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for ages 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Bold values represent the results of the base model (M_{2+}=0.25, L_2=118.57, and h=0.999). Models with 0% of runs having a positive-definite Hessian matrix are highlighted in red and were not considered in further diagnostics tests. N/A indicates the analysis was not conducted.

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	62%	58%	50%
	0.93	N/A	N/A	N/A	73%	81%	65%
	0.95	0%	0%	0%	88%	77%	65%
	0.97	8%	4%	0%	77%	92%	81%
	0.99	85%	62%	58%	92%	88%	92%
	0.999	92%	92%	81%	92%	**96%**	92%

a. M_{2+}=0.193


M2=0.193, L2=118, h=0.95
M2=0.193, L2=118.57, h=0.95
M2=0.193, L2=119, h=0.95
M2=0.193, L2=118, h=0.97
M2=0.193, L2=118.57, h=0.97
M2=0.193, L2=119, h=0.97
M2=0.193, L2=118, h=0.99
M2=0.193, L2=118.57, h=0.99
M2=0.193, L2=119, h=0.99
M2=0.193, L2=118, h=0.999
M2=0.193, L2=118.57, h=0.999
M2=0.193, L2=119, h=0.999
LL
LN(R0)
Number of Jitter

b. M_{2+}=0.25


M2=0.25, L2=118, h=0.91
M2=0.25, L2=118.57, h=0.91
M2=0.25, L2=119, h=0.91
M2=0.25, L2=118, h=0.93
M2=0.25, L2=118.57, h=0.93
M2=0.25, L2=119, h=0.93
M2=0.193, L2=118, h=0.95
M2=0.193, L2=118.57, h=0.95
M2=0.193, L2=119, h=0.95
M2=0.25, L2=118, h=0.97
M2=0.25, L2=118.57, h=0.97
M2=0.25, L2=119, h=0.97
LL
LN(R0)
Number of Jitter

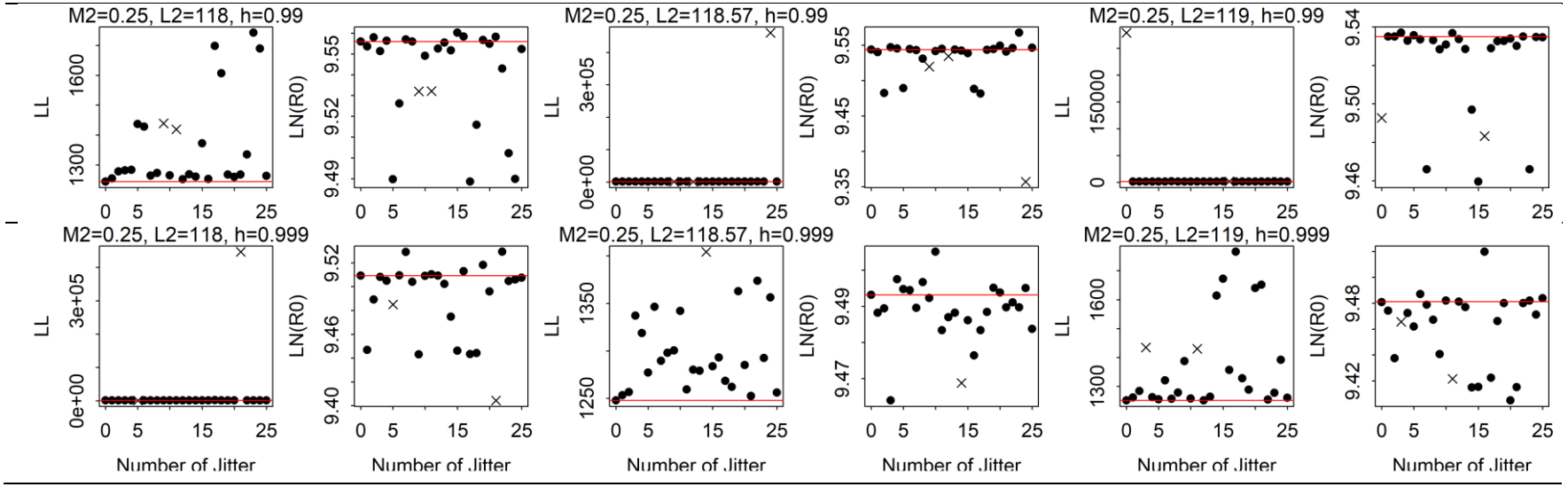


Figure 1. The 25 jitter runs were conducted using models that varied by changing the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for ages 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. In each grid, the left panel shows the total negative log-likelihood (NLL) and the right panel shows ln(R0) values on the y-axis. Dots represent positive-definite Hessian matrices, while crosses represent non positive-definite Hessian matrices. Red horizontal lines indicate runs with the lowest total NLL and positive-definite Hessian matrices.

Table 2. The negative log-likelihood values (NLLs) were derived from all components and the major data components: b) abundance indices (S1: Japan longline index, S4: Japan troll index, S5: Taiwan longline index) and c) all size compositions. These values are obtained from models that varied by changing the values of length at age 3 (L_2) and steepness (h), while keeping the natural mortality for ages 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. The bold values represent the results from the base model (M_{2+}=0.25, L_2=118.57, and h=0.999). Yellow highlights indicate changes in NLLs smaller than 1.92 likelihood units than the base model NLL. Missing values (.) indicate non-convergent models obtained through the jitter analyses (refer to Figure 1). N/A indicates the analysis was not conducted.

a) Total

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	1264	1254	1259
	0.93	N/A	N/A	N/A	1247	1251	1256
	0.95	.	.	.	1246	1249	1254
	0.97	1262	1274	.	1245	1248	1254
	0.99	1246	1250	1254	1244	1248	1252
	0.999	1246	1250	1254	1244	**1248**	1252

b) Indices

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	-83	-83	-83
	0.93	N/A	N/A	N/A	-82	-83	-84
	0.95	.	.	.	-83	-84	-84
	0.97	-83	-83	.	-83	-84	-85
	0.99	-82	-84	-83	-83	-84	-85
	0.999	-83	-83	-84	-83	**-85**	-85

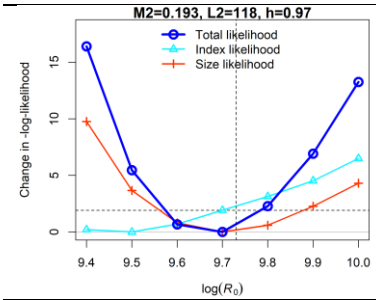
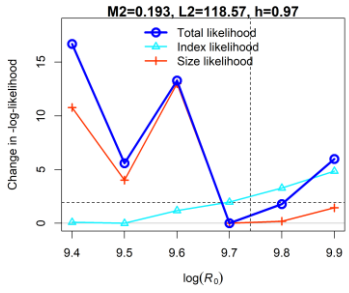
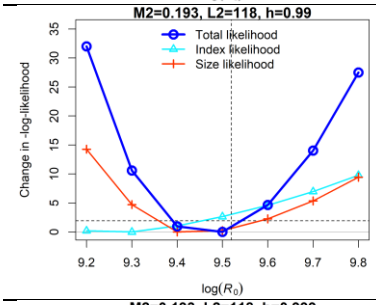
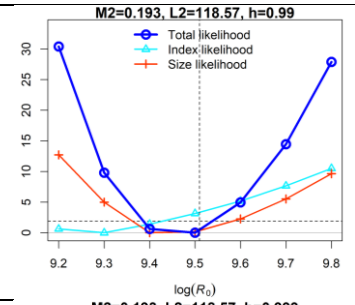
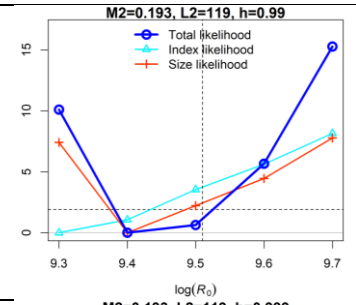
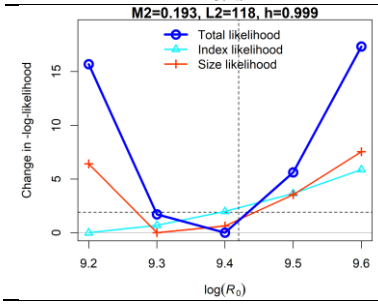
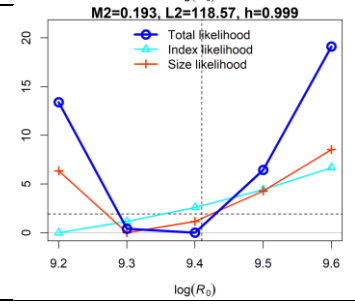
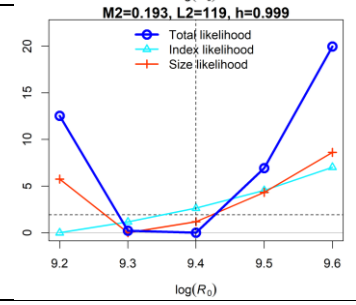
c) Size compositions

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	1321	1310	1316
	0.93	N/A	N/A	N/A	1305	1310	1314
	0.95	.	.	.	1305	1309	1315
	0.97	1319	1331	.	1305	1309	1316
	0.99	1304	1309	1313	1305	1309	1314
	0.999	1304	1308	1313	1305	**1309**	1314

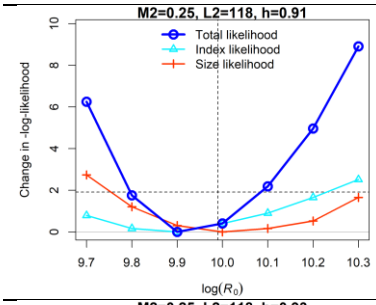
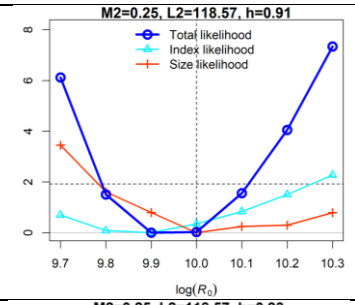
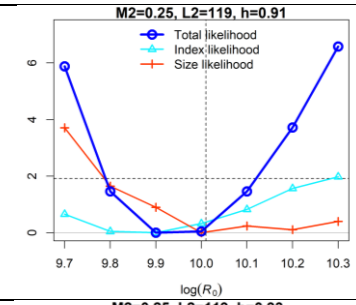
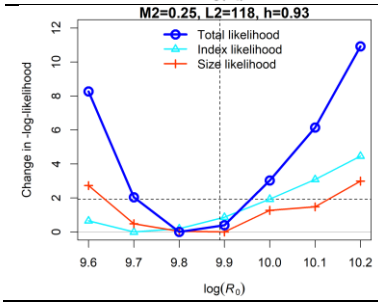
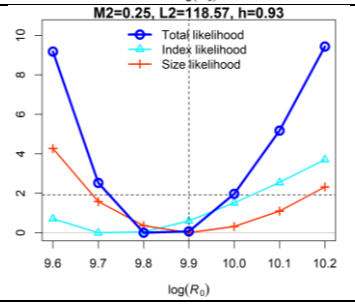
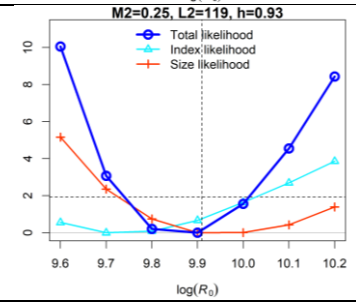
Table 3. The consistency of each likelihood component (indices or size compositions) with the total likelihood, as determined by the R_0 profile analyses conducted on models that change the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for age 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Bold text indicates that both the indices and size components are consistent with the total likelihood in terms of the global scale ($\ln(R_0)$) in the base model (M_{2+}=0.25, L_2=118.57, and h=0.999). Yellow highlights indicate consistency between size and the total likelihood, as in the base model. Missing values (.) indicate non-convergent models obtained through jitter analyses. N/A indicates the analysis was not conducted.

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	Indices & Size	Indices & Size	Indices & Size
	0.93	N/A	N/A	N/A	Size	Size	Size
	0.95	.	.	.	Size	Size	Size
	0.97	Size	Size	.	Size	None	Size
	0.99	Size	Size	Size	Size	None	Size
	0.999	Size	Size	Size	Size	**Size**	Size

a. M_{2+}=0.193

M2=0.193, L2=118, h=0.95 Hessian is not positive definite	M2=0.193, L2=118.57, h=0.95 Hessian is not positive definite	M2=0.193, L2=119, h=0.95 Hessian is not positive definite
		M2=0.193, L2=119, h=0.97 Hessian is not positive definite
		
		

b. M_{2+}=0.25

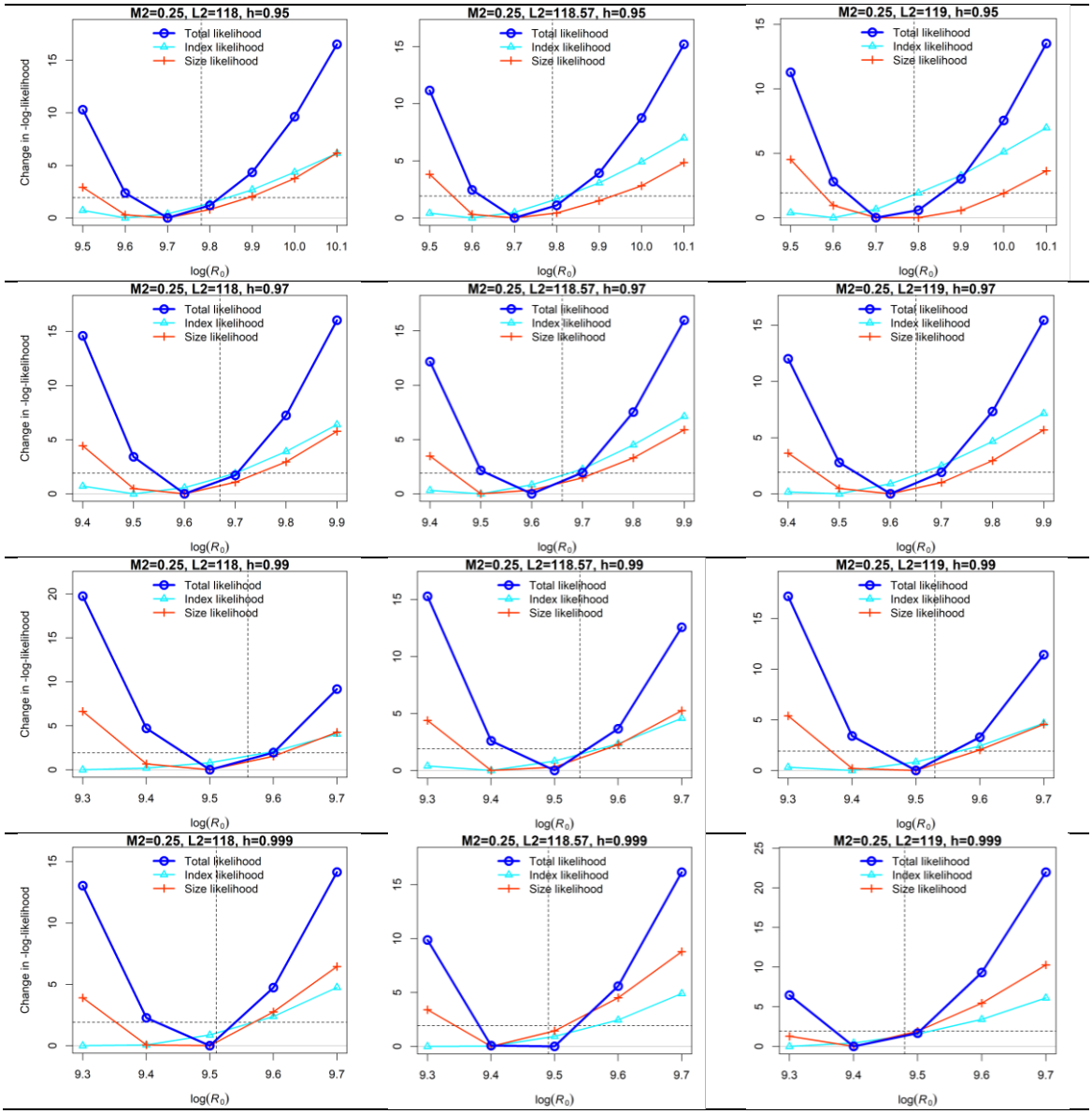
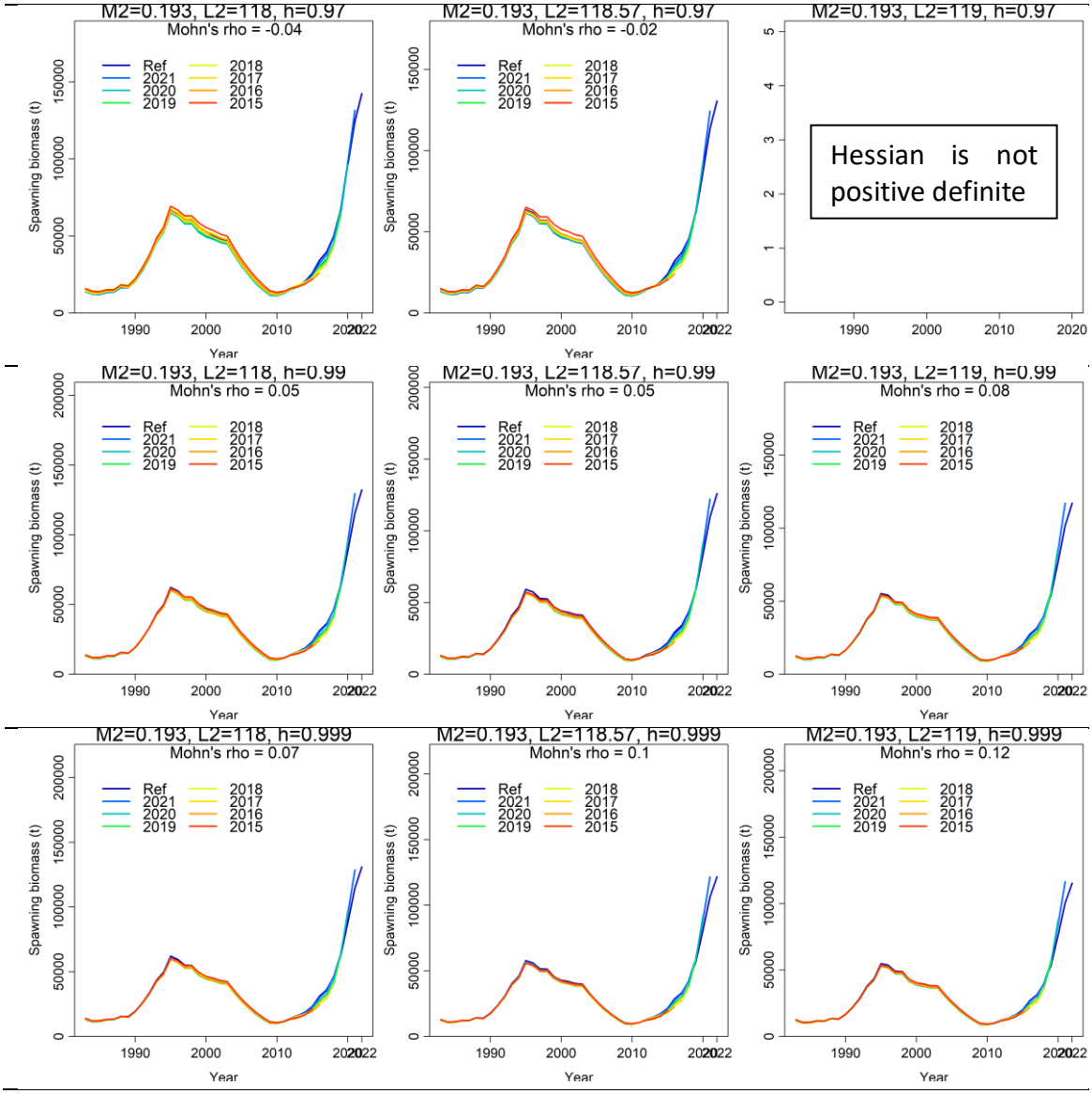


Figure 2. Changes in negative log-likelihood (NLL) for likelihood component across a range of R_0 in various models, achieved by altering the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for age 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Vertical dashed lines indicate R_0 at the minimal total likelihood estimates (R_0^{MLE}), and horizontal dashed lines indicate the 95% confidence interval for the changes in NLL around R_0^{MLE}, which corresponds to a half of the chi-squared values for p=0.95 with 1 degree of freedom. If R_0^c for the data component at the minimal likelihood estimates falls outside the 95% confidence interval for R_0^{MLE}, that data component conflicts with the overall model.

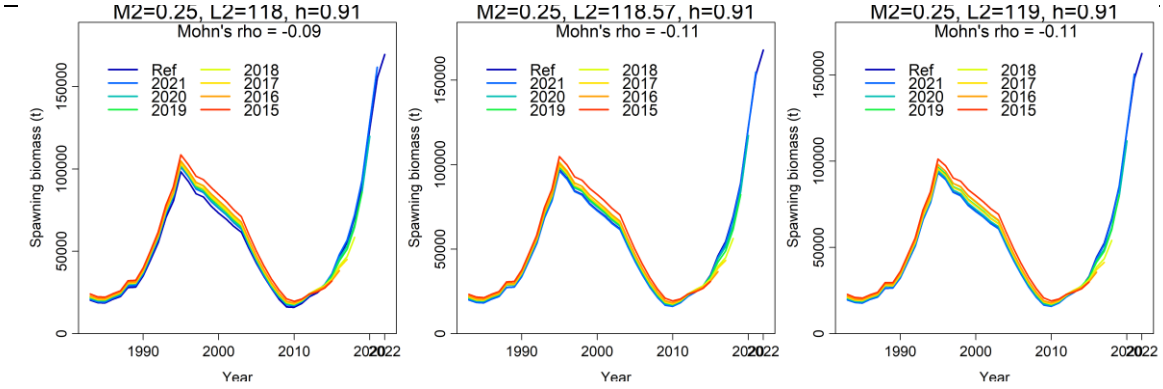
Table 4. Mohn's ρ values for spawning stock biomass from the 7-year retrospective analyses using various models that involve altering the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for age 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Bold value represents the Mohn's ρ from the base model (M_{2+}=0.25, L_2=118.57, and h=0.999). Yellow highlights indicate Mohn's ρ values smaller than 0.1. Missing values (.) indicate non-convergent models obtained based on the jitter analyses (refer to Figure 1). N/A indicates the analysis was not conducted.

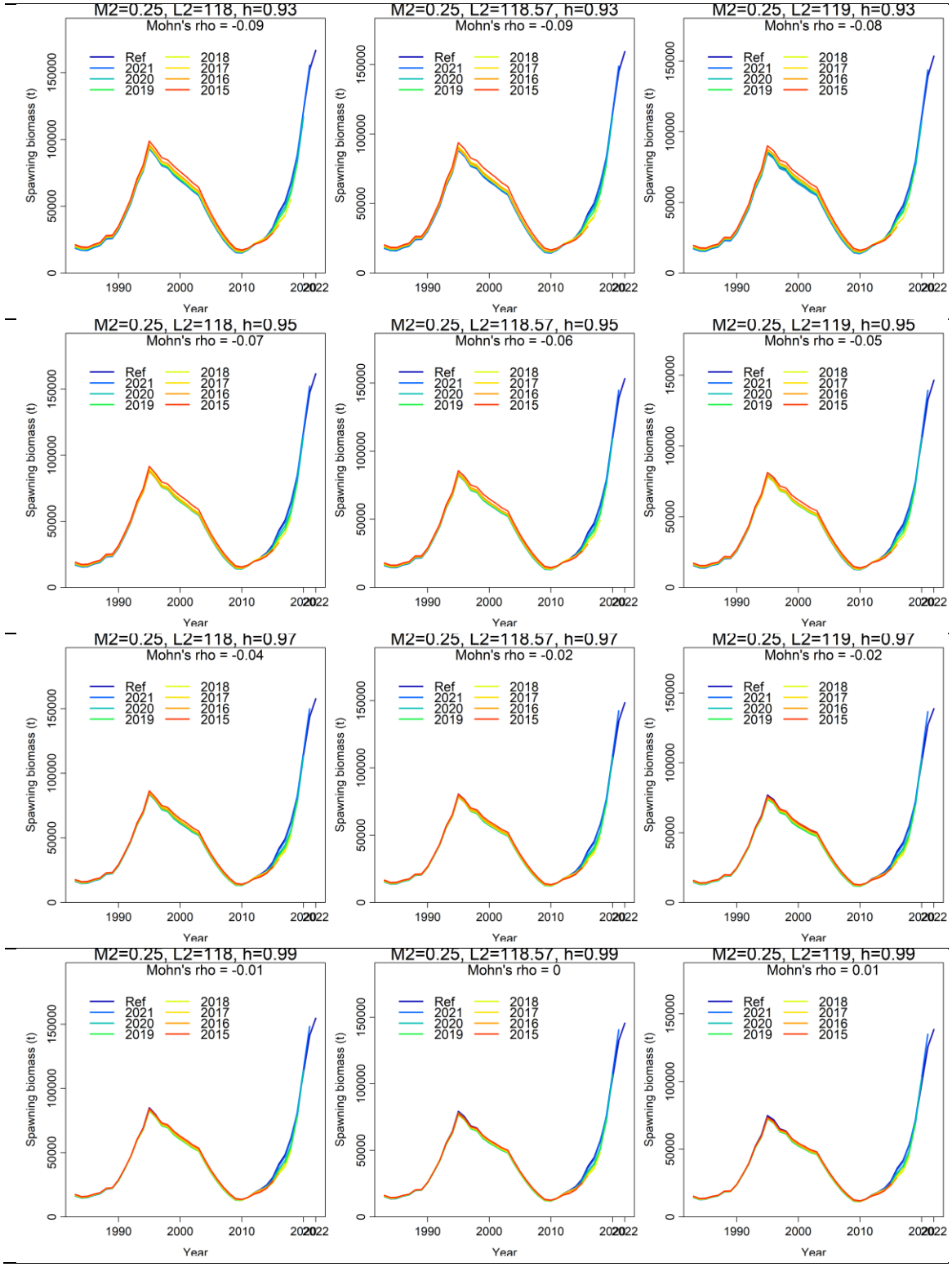
		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	-0.09	-0.11	-0.11
	0.93	N/A	N/A	N/A	-0.09	-0.09	-0.08
	0.95	.	.	.	-0.07	-0.06	-0.05
	0.97	-0.04	-0.02	.	-0.04	-0.02	-0.02
	0.99	0.05	0.05	0.08	-0.01	0	0.01
	0.999	0.07	0.1	0.12	0	**0.01**	0.03

a. M_{2+}=0.193



b. M_{2+}=0.25




M2=0.25, L2=118, h=0.93
Mohn's rho = -0.09
M2=0.25, L2=118.57, h=0.93
Mohn's rho = -0.09
M2=0.25, L2=119, h=0.93
Mohn's rho = -0.08
M2=0.25, L2=118, h=0.95
Mohn's rho = -0.07
M2=0.25, L2=118.57, h=0.95
Mohn's rho = -0.06
M2=0.25, L2=119, h=0.95
Mohn's rho = -0.05
M2=0.25, L2=118, h=0.97
Mohn's rho = -0.04
M2=0.25, L2=118.57, h=0.97
Mohn's rho = -0.02
M2=0.25, L2=119, h=0.97
Mohn's rho = -0.02
M2=0.25, L2=118, h=0.99
Mohn's rho = -0.01
M2=0.25, L2=118.57, h=0.99
Mohn's rho = 0
M2=0.25, L2=119, h=0.99
Mohn's rho = 0.01
Ref
2021
2020
2019
2018
2017
2016
2015
Spawning biomass (t)
Year

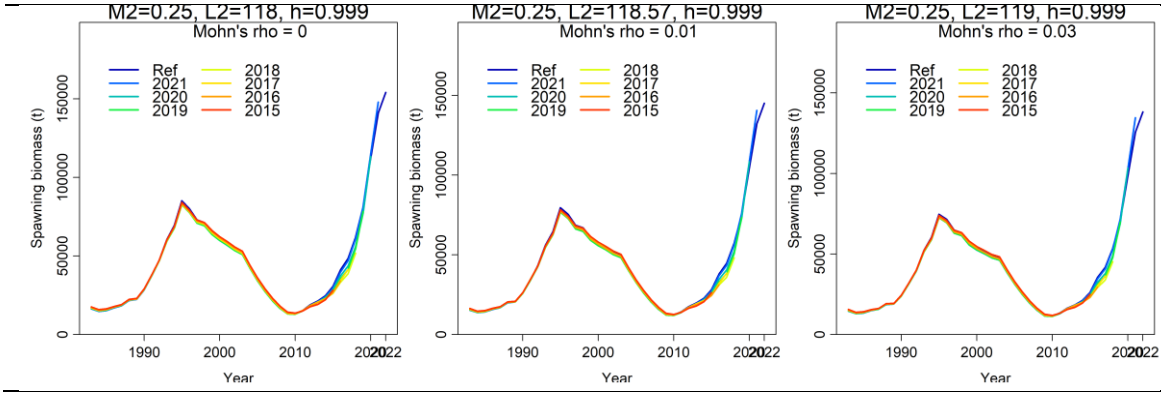


Figure 3. 7-year retrospective analyses of spawning stock biomass using various models that involve altering the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for age 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Mohn's ρ values are shown in each panel. Blank panels indicate non-convergent models obtained based on the jitter analyses (refer to Figure 1).

Table 5. The total negative log-likelihood (NLL) values from ASPM-R models varied by changing the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for age 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Bold value represents the total NLL value from the base ASPM-R model (M_{2+}=0.25, L_2=118.57, and h=0.999). Yellow highlights indicate the total NLL values that are either not statistically different (with no more than a 2-unit NLL degradation) or improved compared to the base ASPM-R model (with a smaller NLL value). Missing values (.) indicate non-convergent models obtained through the jitter analyses (refer to Figure 1). N/A indicates the analysis was not conducted.

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	-39.9	-39.1	-40.2
	0.93	N/A	N/A	N/A	-39.8	-40.6	-41.2
	0.95	.	.	.	-40.3	-41.3	-42.0
	0.97	-39.3	-39.6	.	-41.4	-42.2	-43.2
	0.99	-38.8	-39.8	-39.1	-43.5	-44.7	-45.3
	0.999	-40.5	-40.8	-41.1	-43.8	**-44.7**	-45.4

Table 6. Ensemble diagnostics scores from jitter (Table 1), R_0 profile (Table 3), retrospective (Table 4), and ASPM-R analyses (Table 5). The scores range from 0 (red) to 4 (green), with the highest score indicating successful passage of all four diagnostics.

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)	L_2=118 (L_{inf}=248.6)	L_2=118.57 (L_{inf}=249.9)	L_2=119 (L_{inf}=250.9)
Steepness	0.91	N/A	N/A	N/A	3	2	2
	0.93	N/A	N/A	N/A	3	3	3
	0.95	0	0	0	3	3	3
	0.97	3	3	0	3	2	4
	0.99	3	3	3	4	3	4
	0.999	3	2	2	4	4	4

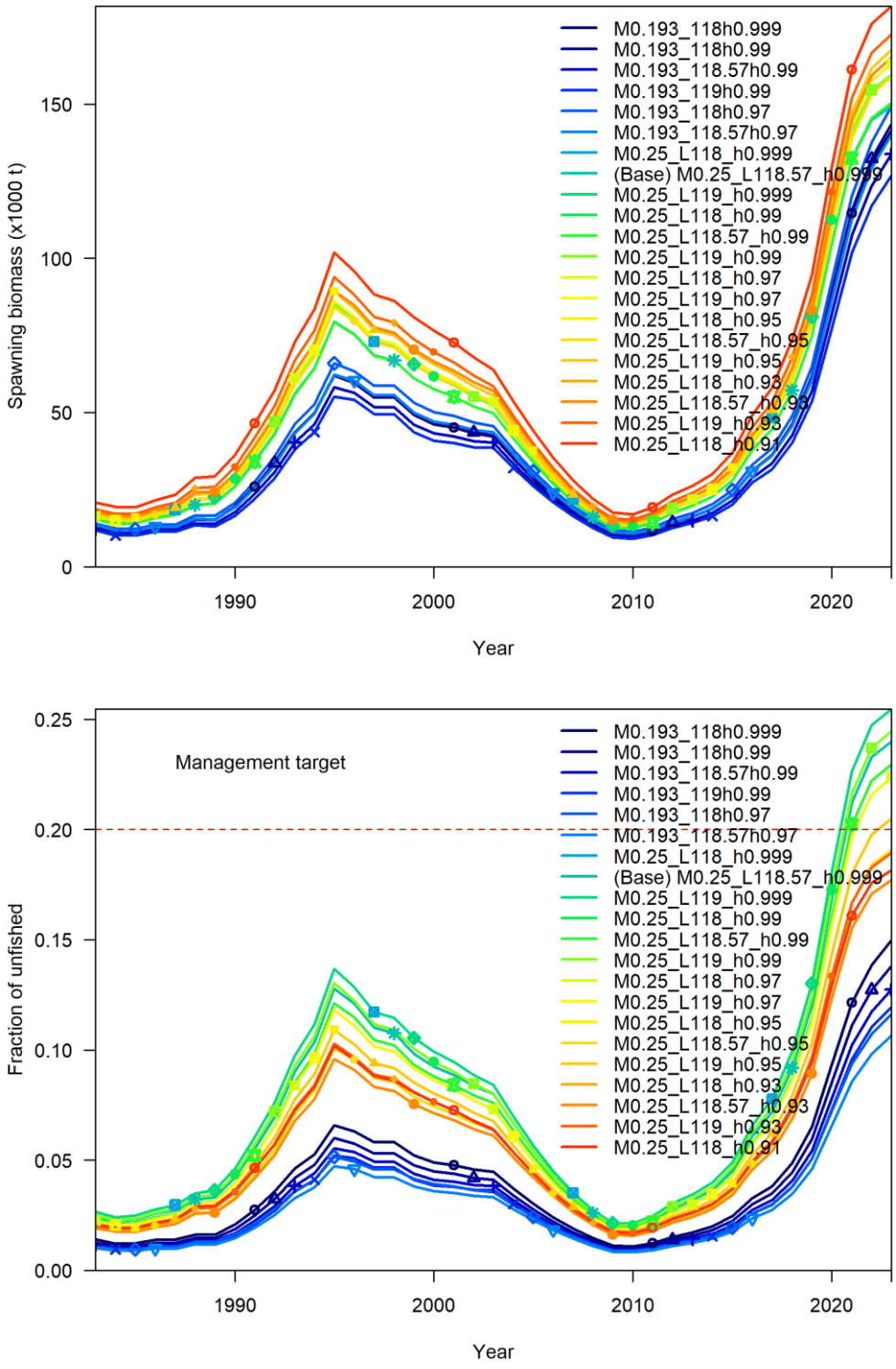


Figure 4. The trajectory of the spawning biomass (upper panel) and spawning stock biomass ratio (lower panel) estimated from all selected grid model with the score at 3 and 4 (referred to Table 6).