Evaluating the Uncertainty Grid: 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

November 2023

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 27 November to 1 December 2023, 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. These uncertainties include process uncertainty, parameter uncertainty, model uncertainty, errors in data and observation systems, and implementation uncertainty. 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 the 2022 assessment. Considering all possible combinations of these parameters is impractical. Therefore, a plausible uncertainty grid for productivity parameters was selected based on the following steps. We judiciously determined the range of productivity parameters using available data and life-history information. The 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 conflicting 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 today rely on the outcomes of management strategy evaluation (MSE) to determine which management strategies will be implemented in the future. MSE employs a forward simulation approach to assess the robustness of feedback-control management strategies in the face of uncertainties (Smith 1994). MSE takes into account the collection and use of future data and uncertainties within the managed system.

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. Process uncertainty refers to the random variation in parameters such as future recruitment and time-varying selectivity. Parameter uncertainty is associated with the uncertainty in the parameter values fixed in the operating models (e.g., steepness, natural mortality). Model uncertainty pertains to the uncertainty in the form of the biological relationships (e.g., whether the stock-recruitment relationship is Beverton-Holt or Ricker, whether fishery selectivity is asymptotic or dome-shaped). Errors in data and observation systems relate to collecting data, such as catches, size compositions, or surveys. Implementation uncertainty may arise from imperfectly implemented management actions.

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). Length at age 3 (L_2) was externally estimated from otolith data and set as a fixed value in the assessment model. Parameters M_2^+ and h were externally estimated using direct data from tagging and indirect information (e.g., through empirical relationships for M and reproductive ecology for h) and were specified within the assessment model as fixed values.

The uncertainty grid associated with the identified productivity parameters and their plausible values was previously examined by Lee et al. in 2023 through a sensitivity analysis in the assessment model. This analysis involved running the grid model with a wide range of parameter values for each parameter. Subsequently, the grid model was rerun using the age-structured production model with recruitment (ASPM-R) to assess the plausibility of each parameter combination for the stock, considering its fishing history and life-history traits, as described by Lee et al. in 2023. The uncertainty grid encompassed combinations of values for length at age 3 (L_2), steepness (h), and natural mortality rate for ages 2 and older (M_2^+).

In this study, we further examined the selection criteria for the grid model, taking into account the stock's fishing history, life-history traits, and grid model performance based on various model diagnostics (Carvalho et al. 2017; 2021). This goes beyond the use of the ASPM-R model, as in Lee et al. (2023). We first revisited the direct data and indirect information to determine the range of these parameters. Then, we incorporated the fishery data into each grid model. We finally applied diagnostic tools used in the assessment, including jitter analyses, goodness-of-fit, likelihood profile on R_0, retrospective analyses, and ASPM-R, to eliminate underperforming grids.

Methods

Determining the range of productivity parameters

A suite of empirical estimators for M_2^+ was previously used to investigate the range of natural mortality for mature fish (Lee et al. 2023). These estimators were based on the maximum age, the von Bertalanffy growth function, and the age at maturity etc. Subsequently, a meta-analysis was conducted to synthesize all these methods, assigning equal weight to each method for each estimator. In a recent review by Maunder et al. (2023), the methods for estimating *M* were examined, and they recommended focusing on the maximum observed age (*tmax*) as it provides a more direct relationship with *M*. Among the estimators based on *tmax*, the formula *M*=5.4/*tmax* was suggested by Then et al. in 2015, Hamel and Cope in 2022, and Maunder et al. in 2023 (equation T3.2.1). This equation is derived from models that assess the probability of a fish surviving to a particular age under a specific level of total mortality. Based on historical age data (Fukuda et al. 2015), the maximum observed age is 28, corresponding to an *M* value of 0.193 year^{-1}. The 2022 assessment assumed an M_2^+ value of 0.25 year^{-1} (ISC 2022), where *tmax* is about 22 years old according to the formula. We considered these values as bounding the potential uncertainty in M_2^+ and explored here alternative model structures with an M_2^+ specified at 0.193 or 0.25 year^{-1}.

The length-at-age data from otoliths, collected by Japanese and Taiwanese scientists between 1992 and 2014 (Fukuda et al. 2015), were used to explore the range of length-at-age 3 in the first quarter (L_2 = 3.0 years old). A total of 1,782 pairs of length-at-age were summarized, spanning from 1 to 27 years old and ranging from 70.5 cm to 271 cm in fork length. Ishihara et al. (2023) further bootstrapped these length-at-age data using different sampling methods and data points, revealing that the median of estimated length-at-age 3 ranged from 118.57 to 118.82. The 95% confidence interval for the estimated L_2 was within ±2 cm from the median. In the 2022 stock assessment model, L_2 was specified at 118.57 cm fork length, with the CV of L_2 at 4.4%. After synthesizing the model-based estimates and bootstrapped analyses, we consequently selected a range of potential L_2 as spanning from 118 to 119 cm. In the following analyses we tested alternative models with L_2 values of 118, 118.57, or 119 cm..

There is less information available to guide the choice of a range for parameter *h* than for M_2^+ or L_2 due to the lack of early life history data. Independent estimates of steepness that incorporate biological and ecological characteristics of the stock (Iwata

2012; Iwata et al. 2012b) reported that the mean of *h* was around 0.999. We explored a broad range of *h* values, ranging from 0.8 to 1.

Diagnostics on the grid model

1. *Convergence and stability*

In assessing the convergence of a model, several diagnostic checks were considered (Carvalho et al. 2021). These checks include examining parameters estimated at a bound, inspecting the correlation matrix for highly correlated parameter pairs, verifying the relative smallness of the final gradient, assessing parameters with high variance, and ensuring a positive-definite Hessian matrix. It is important to note that relying solely on these checks may not provide conclusive evidence of convergence to a global solution.

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*

Several methods and statistics were used to determine the goodness-of-fit of a model. The choice of statistics varied depending on the data type and statistical assumptions. Common statistics for the index of abundance included root mean square errors and runs tests, etc., while the composition data used effective sample sizes and runs tests, etc. To simplify, 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 2022 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 10-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 0.91 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 as M_2^+ and L_2 increased, more grids with a fit similar to or better than the base grid were achieved (yellow highlighted in Table 2). However, 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 conflicting 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. Both the indices and size components provided consistent estimates of the global scale (ln(R_0)) for the base grid, with R_0^c (R_0 at the minimal likelihood estimates for the data component, *c*) falling within the 95% confidence interval for R_0^{MLE}. This consistency, as in the base grid, was also observed in other grids as h and M_2^+ increased (Table 3). **Any grid lacking consistency in indices components will display clear data conflicts and will not be considered in the selection process.**

3.2. *Retrospective analyses*

10-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 -0.26 (Table 4). Other grids exhibited similar or smaller Mohn's ρ values compared to the base grid. The retrospective pattern decreased as h decreased, accompanied by a smaller absolute Mohn's ρ. **Any grid with a worse Mohn's ρ value 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 0.95 to 0.999 when L_2 was 118.57, while the selected h expanded from 0.93 to 0.999 when L_2 was 119. For each L_2 between 118 and 119 cm, the best fit was observed at a higher steepness value (0.99 - 0.999). **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, R_0 profile and ASPM-R favored higher values for M_2^+ and h, while retrospective analyses 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

MSE should consider influential uncertainties, if not all the uncertainties. However, not all productivity parameters' values are plausible given the fishing history and life-history of the stock. We first selected a range of productivity parameters based on data and life history information. We then showed that the comprehensive evaluation of multiple diagnostic criteria has provided valuable insights into the selection of suitable grids for further consideration for the reference set of operating models in our MSE framework.

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 contrasting trends between index and size compositions, necessitating the exclusion of goodness-of-fit considerations from the selection process due to conflicting grid profiles. 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

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M., Kitakado, T., Yemane, D., Piner, K. R., Maunder, M. N., Taylor, I., Wetzel, C. R., Doering, K., Johnson, K. F., Methot, R. D., 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research. 240: 105959. https://doi.org/10.1016/j.fishres.2021.105959

Fukuda, H., Yamasaki, I., Takeuchi, Y., Kitakado, T., Shimose, T., Ishihara, T., Ota, T., Watai, M., Lu HB. and Shiao, JC. 2015. Estimates of growth function from length-at-age data based on otolith annual rings and daily rings for Pacific bluefin tuna. Working paper submitted to the ISC Pacific bluefin tuna Working Group, 18-25 November 2015, Kaohsiung, Taiwan. ISC/15/PBFWG-2/11

Hamel, O.S., Cope, J.M., 2022. Development and considerations for application of a longevity-based prior for the natural mortality rate. Fisheries Research. 256: 106477. https://doi.org/10.1016/j.fishres.2022.106477

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

Ishihara, T., Tanaka, H., Fukuda, H., Tawa, A., Ashida, H., Tanaka, Y., 2023. Estimation of confidence intervals for the von Bertalanffy growth function parameters using the bootstrap method with a data set of direct age estimates from otoliths. ISC/23/PBFWG-1/10. https://isc.fra.go.jp/pdf/PBF/ISC23_PBF_1/ISC23_PBF_1_10.pdf

Iwata S. 2012. Estimate the frequency distribution of steepness for PBF. Working paper submitted to the ISC PBF Working Group Meeting, 10-17 November 2012, Honolulu, Hawaii, USA. ISC/12/PBFWG-3/01: 13p. Available at: http://isc.fra.go.jp/pdf/PBF/ISC12_PBF_3/ISC12_PBFWG-3_01_Iwata.pdf

Iwata S., Fukuda H., Abe O., and Takeuchi Y. 2012a. Estimation of steepness of PBFT -By using biological feature. Working paper submitted to the ISC PBF Working Group Meeting, 31 January-7 February 2012, La Jolla, California, USA. ISC/12/PBFWG-1/15: 9p. Available at: http://isc.fra.go.jp/pdf/PBF/ISC12_PBF_1/ISC12-1PBFWG15_Iwata.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., Fukuda, H., Piner, K., 2023. Evaluating productivity parameter uncertainty using the age-structured production model diagnostic with recruitment. ISC/23/PBFWG-1/09.

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

Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci., 56, pp. 473-488

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M. 2016. Management strategy evaluation: best practices. Fish and Fisheries 17, 303–334. https://doi.org/10.1111/faf.12104

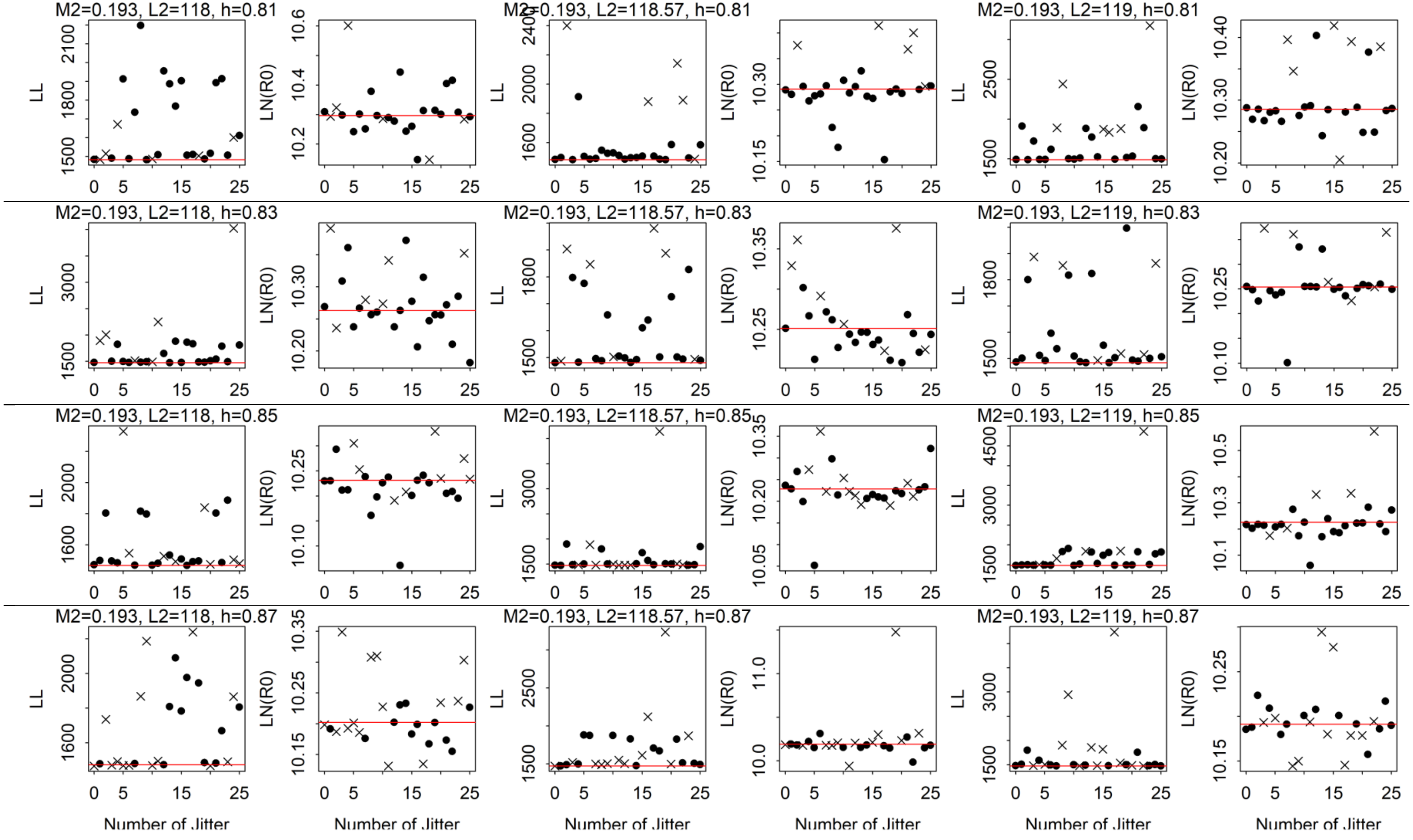
Smith, A.D.M. 1994. Management strategy evaluation: The light on the hill. In: D.A. Hancock (ed.), Population dynamics for fisheries management. Australian Society for Fish Biology, Perth, Western Australia, pp. 249-253.

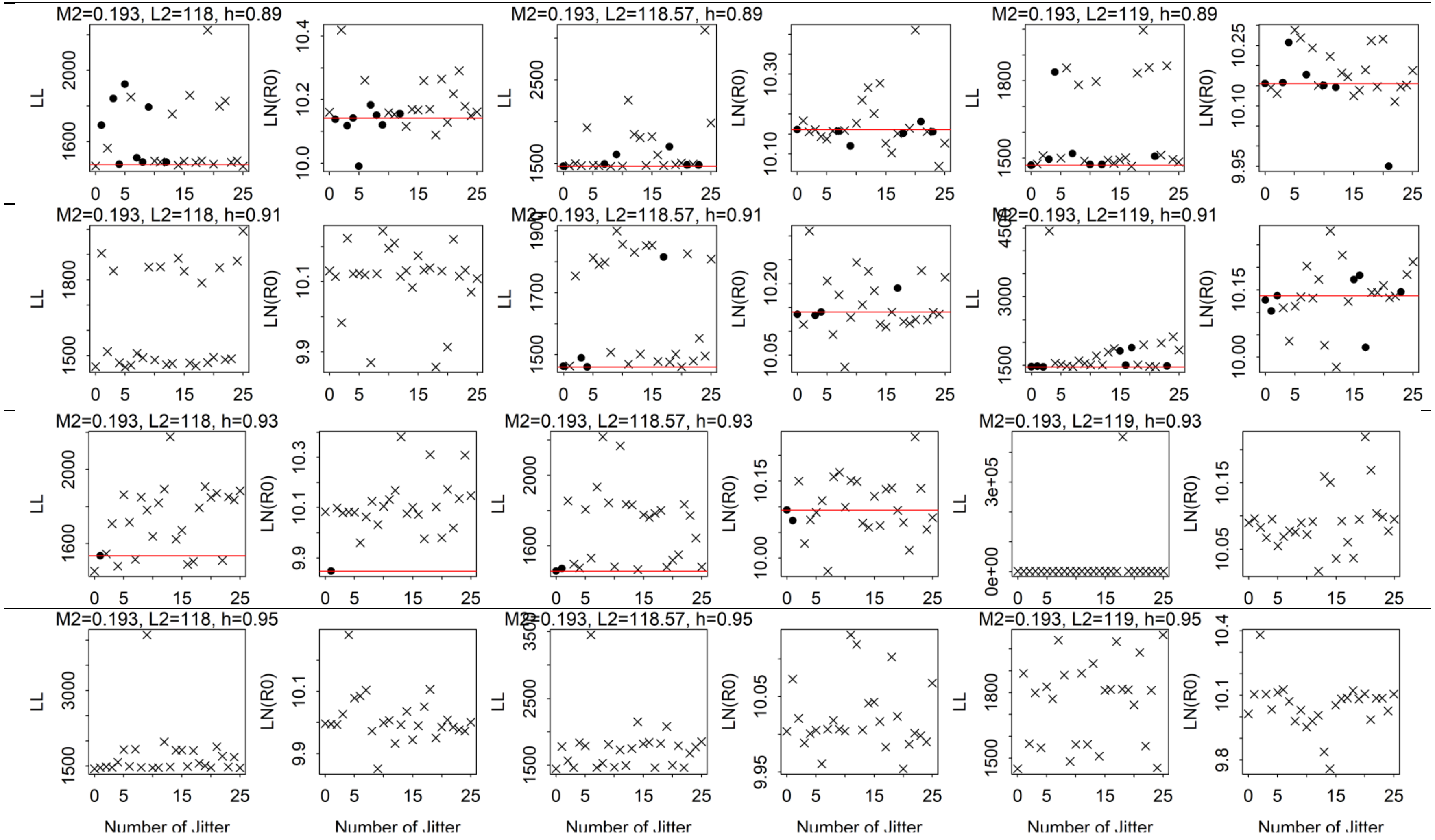
Then, A. Y., Hoenig, J. M., Hall, N. G., Hewitt, D. A., 2015. Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. ICES Journal of Marine Science 72(1): 82-92.

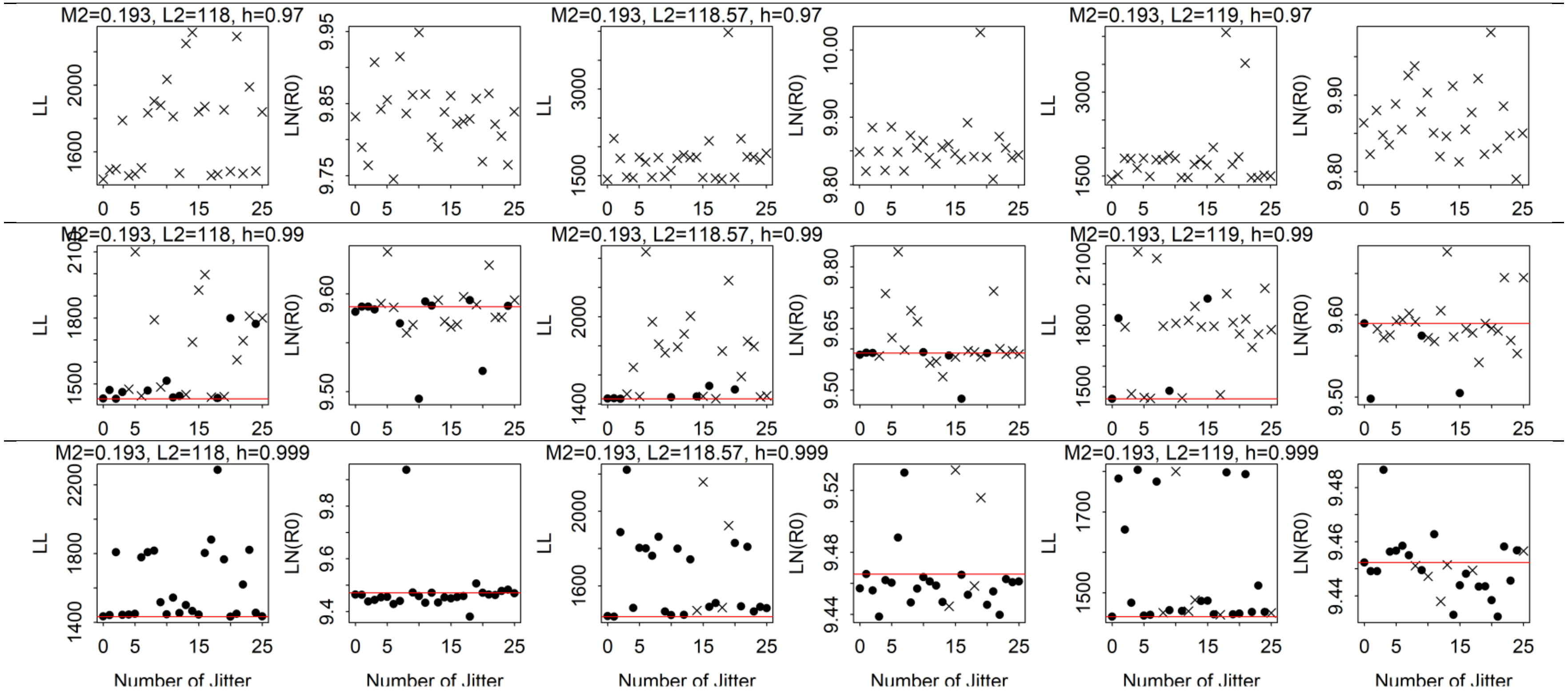
Table 1. The percentage of runs resulting in a positive-definite Hessian matrix in jitter analyses from models that varied by changing the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for ages 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Bold 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.

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118	L_2=118.57	L_2=119	L_2=118	L_2=118.57	L_2=119
		(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)	(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)
Steepness	0.81	77%	81%	77%	81%	65%	65%
	0.83	77%	73%	77%	42%	50%	62%
	0.85	69%	62%	81%	38%	46%	50%
	0.87	46%	54%	54%	23%	35%	42%
	0.89	31%	23%	27%	38%	23%	23%
	0.91	0%	15%	27%	19%	0%	19%
	0.93	4%	8%	0%	19%	23%	12%
	0.95	0%	0%	0%	50%	19%	19%
	0.97	0%	0%	0%	81%	62%	77%
	0.99	42%	27%	15%	100%	100%	100%
	0.999	100%	85%	77%	100%	**100%**	100%

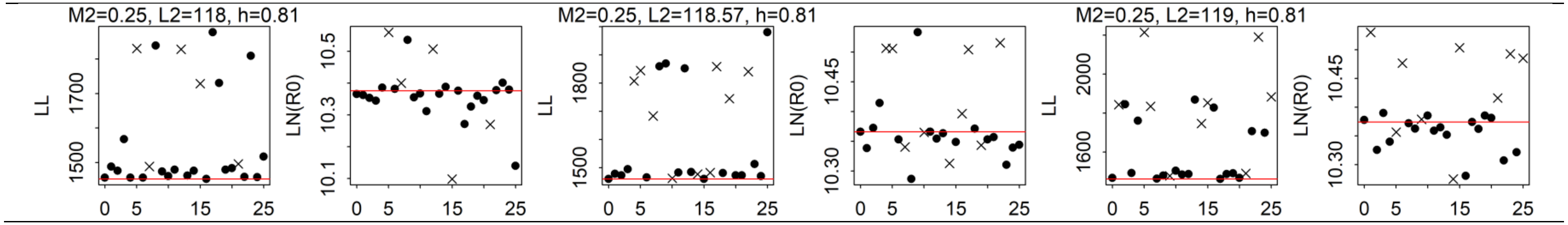
a. M_{2+}=0.193

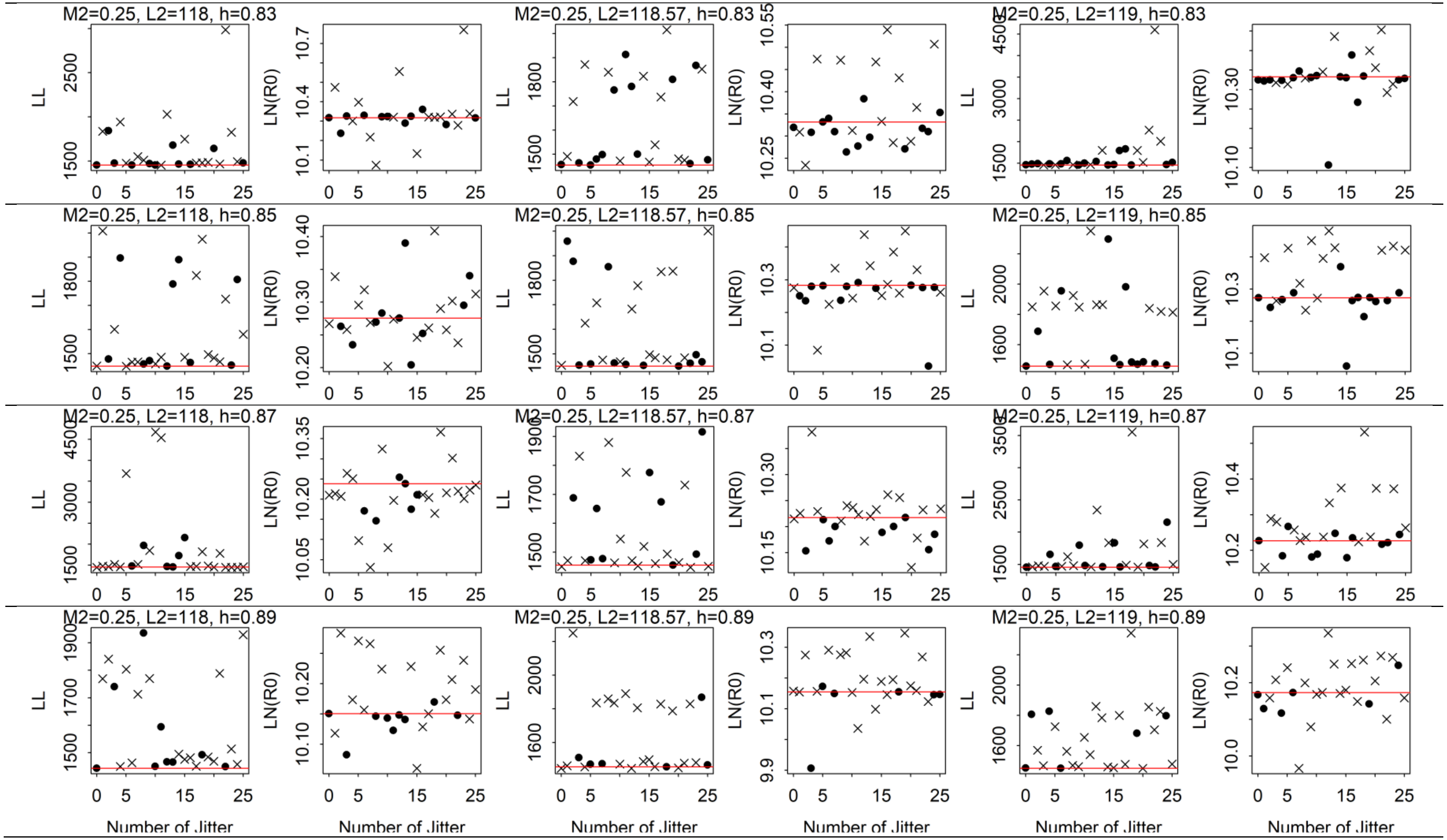

M2=0.193, L2=118, h=0.81
M2=0.193, L2=118.57, h=0.81
M2=0.193, L2=119, h=0.81
M2=0.193, L2=118, h=0.83
M2=0.193, L2=118.57, h=0.83
M2=0.193, L2=119, h=0.83
M2=0.193, L2=118, h=0.85
M2=0.193, L2=118.57, h=0.85
M2=0.193, L2=119, h=0.85
M2=0.193, L2=118, h=0.87
M2=0.193, L2=118.57, h=0.87
M2=0.193, L2=119, h=0.87
LL
LN(R0)
Number of Jitter
0
5
15
25

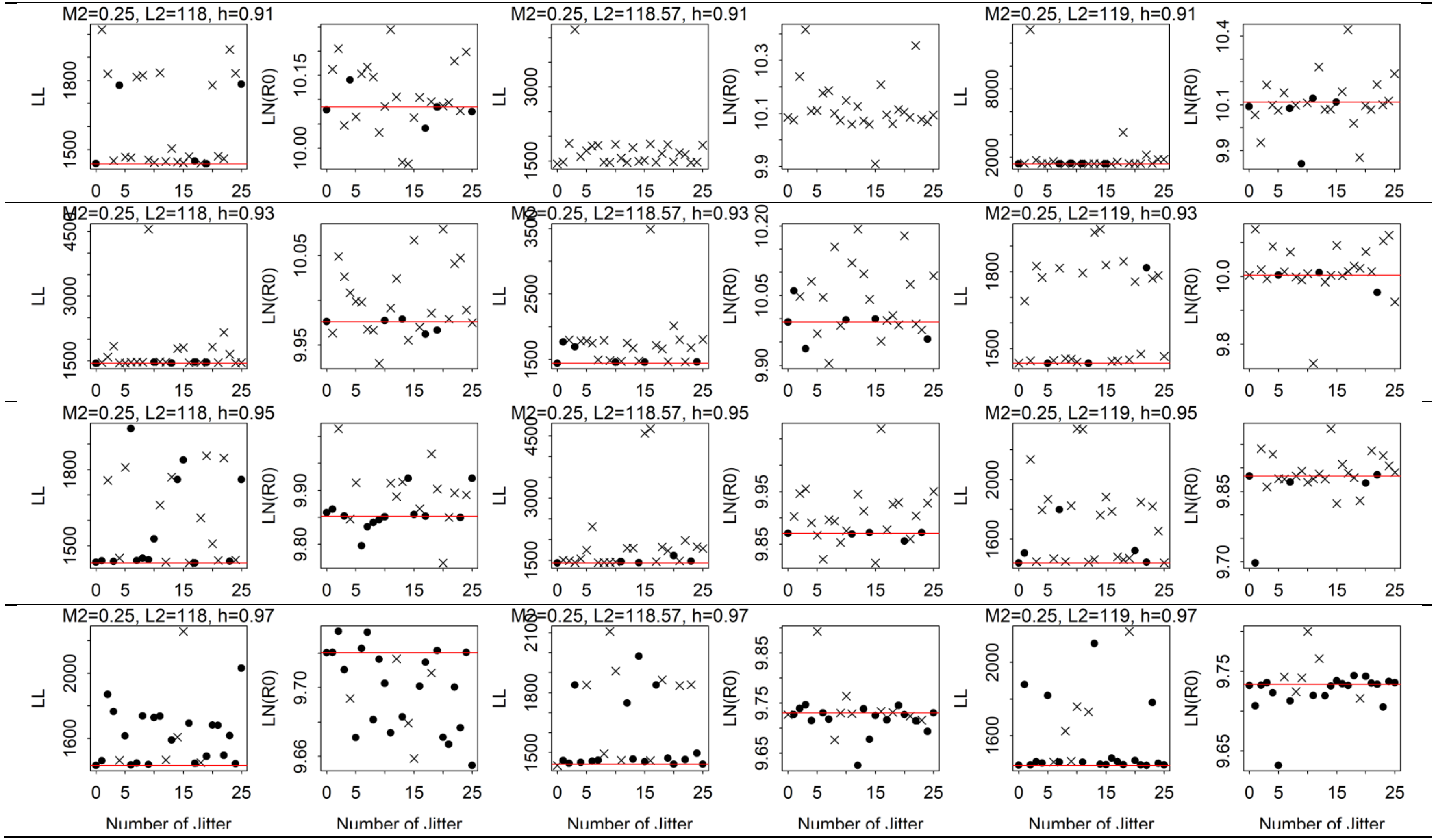

M2=0.193, L2=118, h=0.89
M2=0.193, L2=118.57, h=0.89
M2=0.193, L2=119, h=0.89
M2=0.193, L2=118, h=0.91
M2=0.193, L2=118.57, h=0.91
M2=0.193, L2=119, h=0.91
M2=0.193, L2=118, h=0.93
M2=0.193, L2=118.57, h=0.93
M2=0.193, L2=119, h=0.93
M2=0.193, L2=118, h=0.95
M2=0.193, L2=118.57, h=0.95
M2=0.193, L2=119, h=0.95
LL
LN(R0)
1600
2000
10.0
10.2
10.4
1500
2500
10.10
10.30
1800
9.95
10.10
10.25
9.9
10.1
1700
1900
10.05
10.20
3000
4500
10.00
10.15
10.3
3e+05
0e+00
10.05
3500
9.8
0
5
15
25
Number of Jitter


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)
0
5
15
25
Number of Jitter

b. M_{2+}=0.25


M2=0.25, L2=118, h=0.81
M2=0.25, L2=118.57, h=0.81
M2=0.25, L2=119, h=0.81
LL
LN(R0)
0
5
15
25


M2=0.25, L2=118, h=0.83
M2=0.25, L2=118.57, h=0.83
M2=0.25, L2=119, h=0.83
M2=0.25, L2=118, h=0.85
M2=0.25, L2=118.57, h=0.85
M2=0.25, L2=119, h=0.85
M2=0.25, L2=118, h=0.87
M2=0.25, L2=118.57, h=0.87
M2=0.25, L2=119, h=0.87
M2=0.25, L2=118, h=0.89
M2=0.25, L2=118.57, h=0.89
M2=0.25, L2=119, h=0.89
LL
LN(R0)
0
5
15
25
Number of Jitter
1500
2500
10.1
10.4
10.7
1800
10.25
10.40
10.55
3000
4500
10.10
10.30
10.20
10.30
10.40
10.3
1600
2000
10.05
10.20
10.35
1700
1900
10.15
10.30
3500
10.2
10.4
10.10
10.20
9.9
10.0


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=0.25, L2=118, h=0.95
M2=0.25, L2=118.57, h=0.95
M2=0.25, 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)
1500
1800
10.00
10.15
3000
9.9
10.1
10.3
2000
8000
10.4
4500
9.95
10.05
2500
3500
9.90
10.20
9.8
10.0
9.80
9.85
9.70
1600
9.66
9.65
9.75
2100
0
5
15
25
Number of Jitter

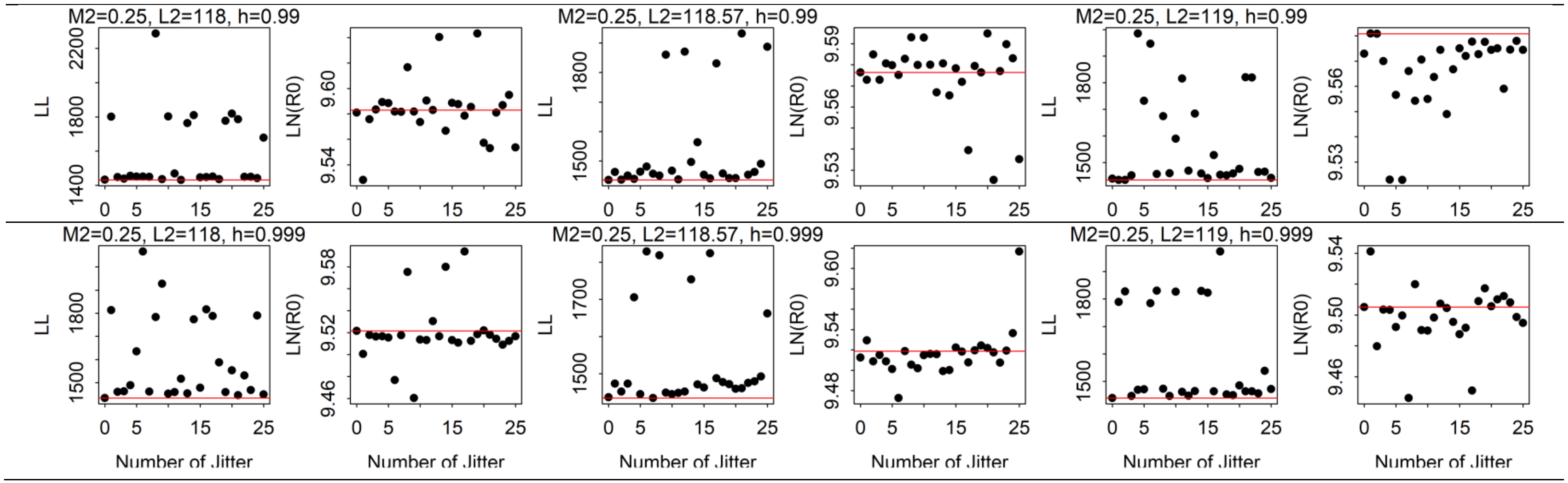


Figure 1. The 25 jitter runs were conducted using models that varied by changing the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for ages 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. In each grid, the left panel shows the total negative log-likelihood (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).

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	0.81	1482	1483	1488	1452	1460	1461
	0.83	1473	1481	1484	1453	1455	1457
	0.85	1466	1473	1479	1448	1450	1458
	0.87	1474	1471	1473	1454	1455	1454
	0.89	1472	1467	1472	1443	1454	1450
	0.91	.	1460	1463	1439	.	1444
	0.93	1533	1456	.	1437	1440	1443
	0.95	.	.	.	1433	1438	1441
	0.97	.	.	.	1434	1442	1438
	0.99	1433	1436	1441	1432	1436	1434
	0.999	1432	1434	1441	1433	**1434**	1439

b) Indices

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118	L_2=118.57	L_2=119	L_2=118	L_2=118.57	L_2=119
		(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)	(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)
Steepness	0.81	-58	-58	-58	-69	-69	-70
	0.83	-60	-60	-60	-70	-71	-71
	0.85	-62	-62	-62	-72	-73	-73
	0.87	-64	-64	-65	-72	-73	-75
	0.89	-66	-67	-67	-75	-74	-76
	0.91	.	-70	-70	-76	.	-77
	0.93	-65	-73	.	-77	-78	-78
	0.95	.	.	.	-78	-78	-79
	0.97	.	.	.	-78	-78	-79
	0.99	-77	-78	-78	-78	-79	-80
	0.999	-78	-78	-79	-79	**-80**	-80

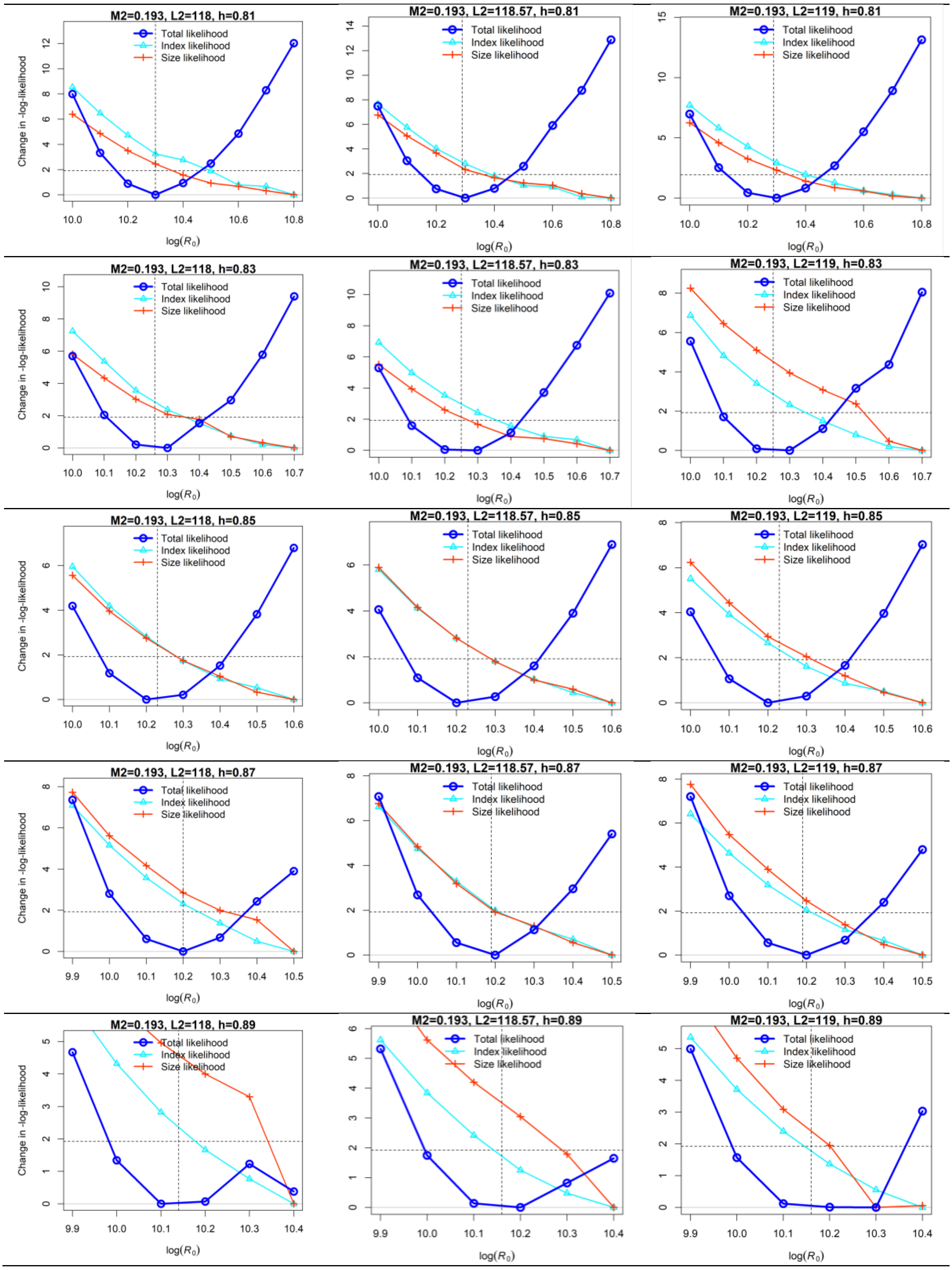
c) Size compositions

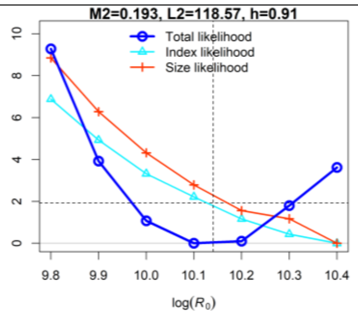
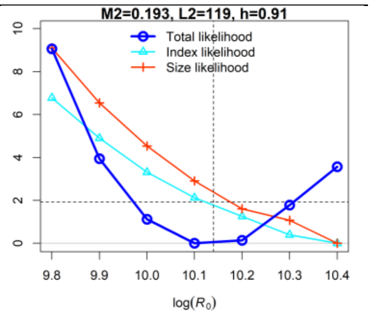
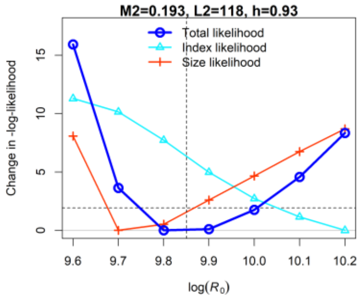
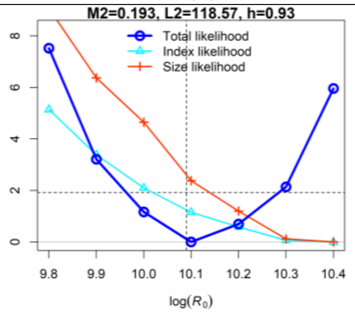
		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118	L_2=118.57	L_2=119	L_2=118	L_2=118.57	L_2=119
		(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)	(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)
Steepness	0.81	1513	1514	1520	1494	1503	1504
	0.83	1506	1514	1517	1497	1499	1502
	0.85	1501	1508	1514	1495	1497	1505
	0.87	1511	1508	1511	1502	1503	1503
	0.89	1511	1507	1512	1493	1504	1501
	0.91	.	1503	1505	1491	.	1496
	0.93	1576	1501	.	1491	1495	1498
	0.95	.	.	.	1489	1494	1498
	0.97	.	.	.	1491	1500	1496
	0.99	1488	1491	1496	1490	1495	1493
	0.999	1487	1490	1497	1491	**1493**	1498

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 indices and the total likelihood, as in the base model. Missing values (.) indicate non-convergent models obtained through jitter analyses.

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118	L_2=118.57	L_2=119	L_2=118	L_2=118.57	L_2=119
		(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)	(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)
Steepness	0.81	None	None	None	None	None	None
	0.83	None	None	None	None	None	None
	0.85	None	None	None	None	None	None
	0.87	None	None	None	None	None	None
	0.89	None	None	None	None	None	None
	0.91	.	None	None	Indices & Size	.	None
	0.93	None	None	.	Indices	Indices	Indices
	0.95	.	.	.	Indices & Size	Indices & Size	Indices
	0.97	.	.	.	Indices & Size	Indices & Size	Indices & Size
	0.99	Size	Size	Size	Size	Size	Size
	0.999	Size	Size	Indices & Size	Indices & Size	**Indices & Size**	Indices & Size

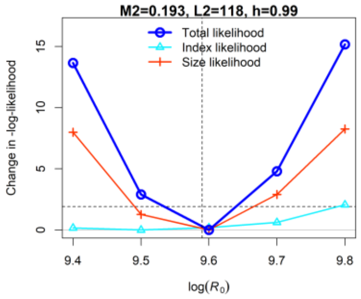
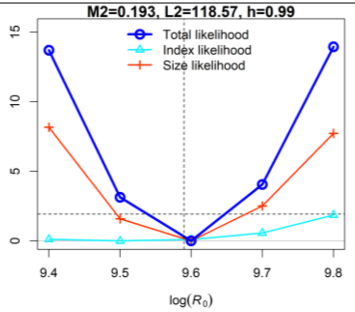
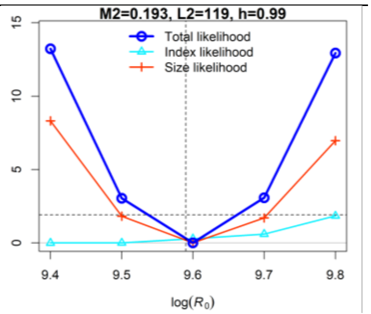
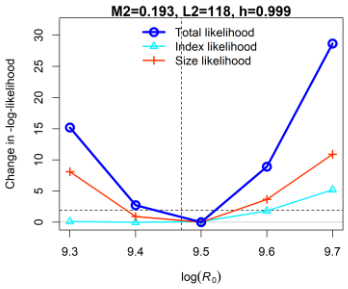
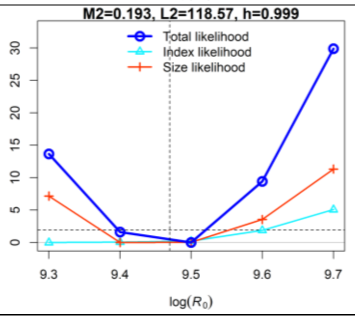
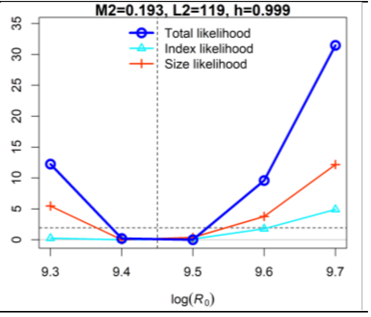
a. M_{2+}=0.193


M2=0.193, L2=118, h=0.81
M2=0.193, L2=118.57, h=0.81
M2=0.193, L2=119, h=0.81
M2=0.193, L2=118, h=0.83
M2=0.193, L2=118.57, h=0.83
M2=0.193, L2=119, h=0.83
M2=0.193, L2=118, h=0.85
M2=0.193, L2=118.57, h=0.85
M2=0.193, L2=119, h=0.85
M2=0.193, L2=118, h=0.87
M2=0.193, L2=118.57, h=0.87
M2=0.193, L2=119, h=0.87
M2=0.193, L2=118, h=0.89
M2=0.193, L2=118.57, h=0.89
M2=0.193, L2=119, h=0.89
Total likelihood
Index likelihood
Size likelihood
Change in -log-likelihood
log(R_0)

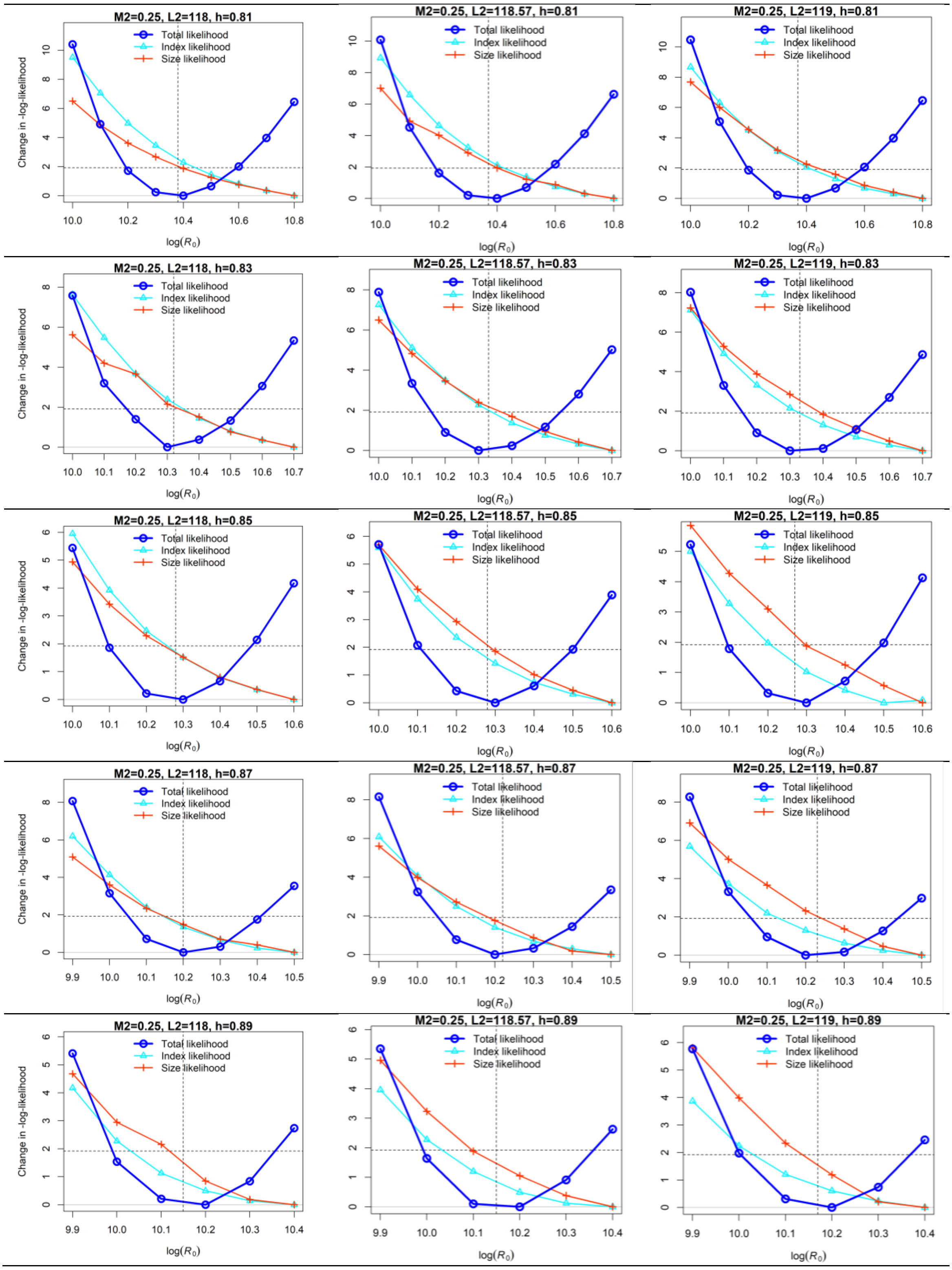
M2=0.193, L2=118, h=0.91 Hessian is not positive definite		
		M2=0.193, L2=119, h=0.93 Hessian is not positive definite

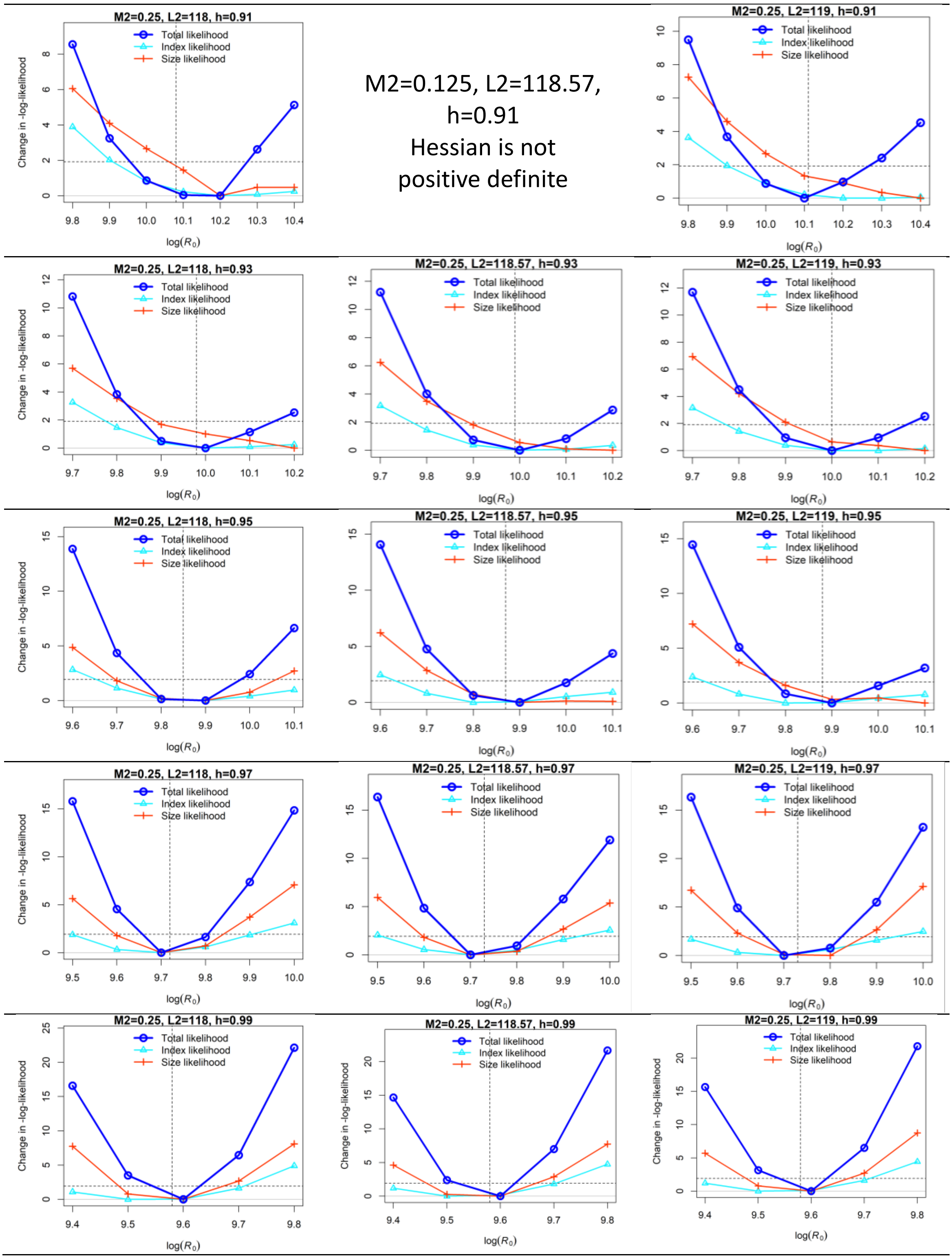
M2=0.193, L2=118, h=0.95
Hessian is not
positive definite
	M2=0.193, L2=118.57,
	h=0.95
	Hessian is not
	positive definite
		M2=0.193, L2=119, h=0.95
		Hessian is not
		positive definite

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

b. M_{2+}=0.25


M2=0.25, L2=118, h=0.81
M2=0.25, L2=118.57, h=0.81
M2=0.25, L2=119, h=0.81
M2=0.25, L2=118, h=0.83
M2=0.25, L2=118.57, h=0.83
M2=0.25, L2=119, h=0.83
M2=0.25, L2=118, h=0.85
M2=0.25, L2=118.57, h=0.85
M2=0.25, L2=119, h=0.85
M2=0.25, L2=118, h=0.87
M2=0.25, L2=118.57, h=0.87
M2=0.25, L2=119, h=0.87
M2=0.25, L2=118, h=0.89
M2=0.25, L2=118.57, h=0.89
M2=0.25, L2=119, h=0.89
Total likelihood
Index likelihood
Size likelihood
Change in -log-likelihood
log(R_0)


M2=0.25, L2=118, h=0.91
M2=0.125, L2=118.57, h=0.91
Hessian is not positive definite
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=0.25, L2=118, h=0.95
M2=0.25, L2=118.57, h=0.95
M2=0.25, 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
M2=0.25, L2=118, h=0.99
M2=0.25, L2=118.57, h=0.99
M2=0.25, L2=119, h=0.99
Total likelihood
Index likelihood
Size likelihood
Change in -log-likelihood
log(R_0)

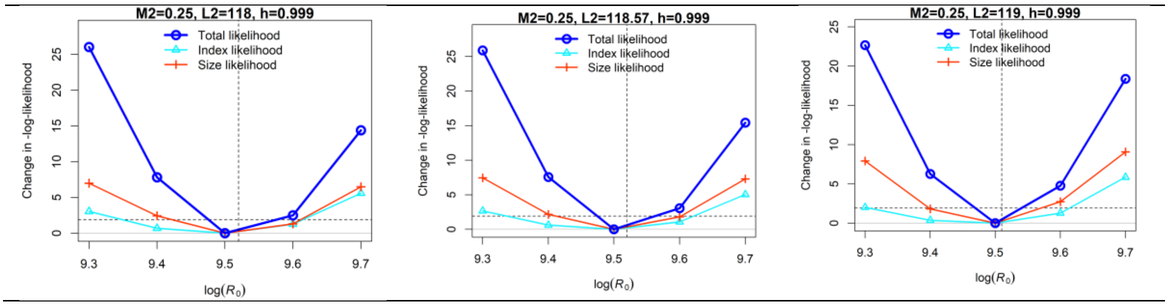
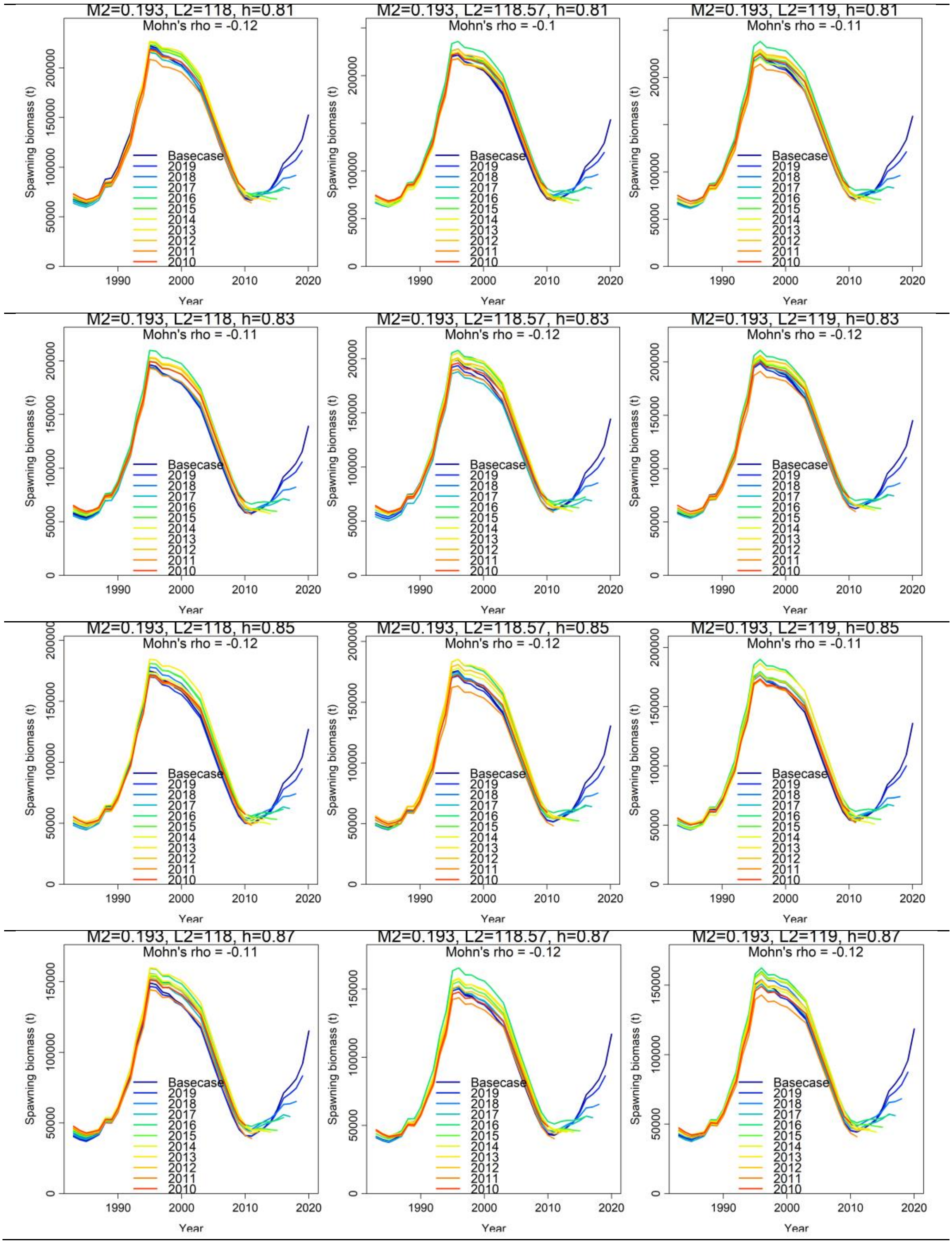


Figure 2. Changes in negative log-likelihood (NLL) for likelihood component across a range of R_0 in various models, achieved by altering the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for age 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. 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 10-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). Missing values (.) indicate non-convergent models obtained based on the jitter analyses (refer to Figure 1).

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118	L_2=118.57	L_2=119	L_2=118	L_2=118.57	L_2=119
		(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)	(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)
Steepness	0.81	-0.12	-0.10	-0.11	-0.17	-0.17	-0.17
	0.83	-0.11	-0.12	-0.12	-0.18	-0.17	-0.16
	0.85	-0.12	-0.12	-0.11	-0.18	-0.17	-0.18
	0.87	-0.11	-0.12	-0.12	-0.22	-0.19	-0.18
	0.89	-0.17	-0.12	-0.12	-0.19	-0.20	-0.19
	0.91	.	-0.13	-0.13	-0.21	.	-0.20
	0.93	-0.49	-0.13	.	-0.23	-0.23	-0.23
	0.95	.	.	.	-0.27	-0.25	-0.25
	0.97	.	.	.	-0.27	-0.28	-0.26
	0.99	-0.27	-0.26	-0.26	-0.28	-0.27	-0.27
	0.999	-0.27	-0.26	-0.26	-0.28	**-0.26**	-0.27

a. M_{2+}=0.193



M2=0.193, L2=118, h=0.81
Mohn's rho = -0.12

M2=0.193, L2=118.57, h=0.81
Mohn's rho = -0.1

M2=0.193, L2=119, h=0.81
Mohn's rho = -0.11

M2=0.193, L2=118, h=0.83
Mohn's rho = -0.11

M2=0.193, L2=118.57, h=0.83
Mohn's rho = -0.12

M2=0.193, L2=119, h=0.83
Mohn's rho = -0.12

M2=0.193, L2=118, h=0.85
Mohn's rho = -0.12

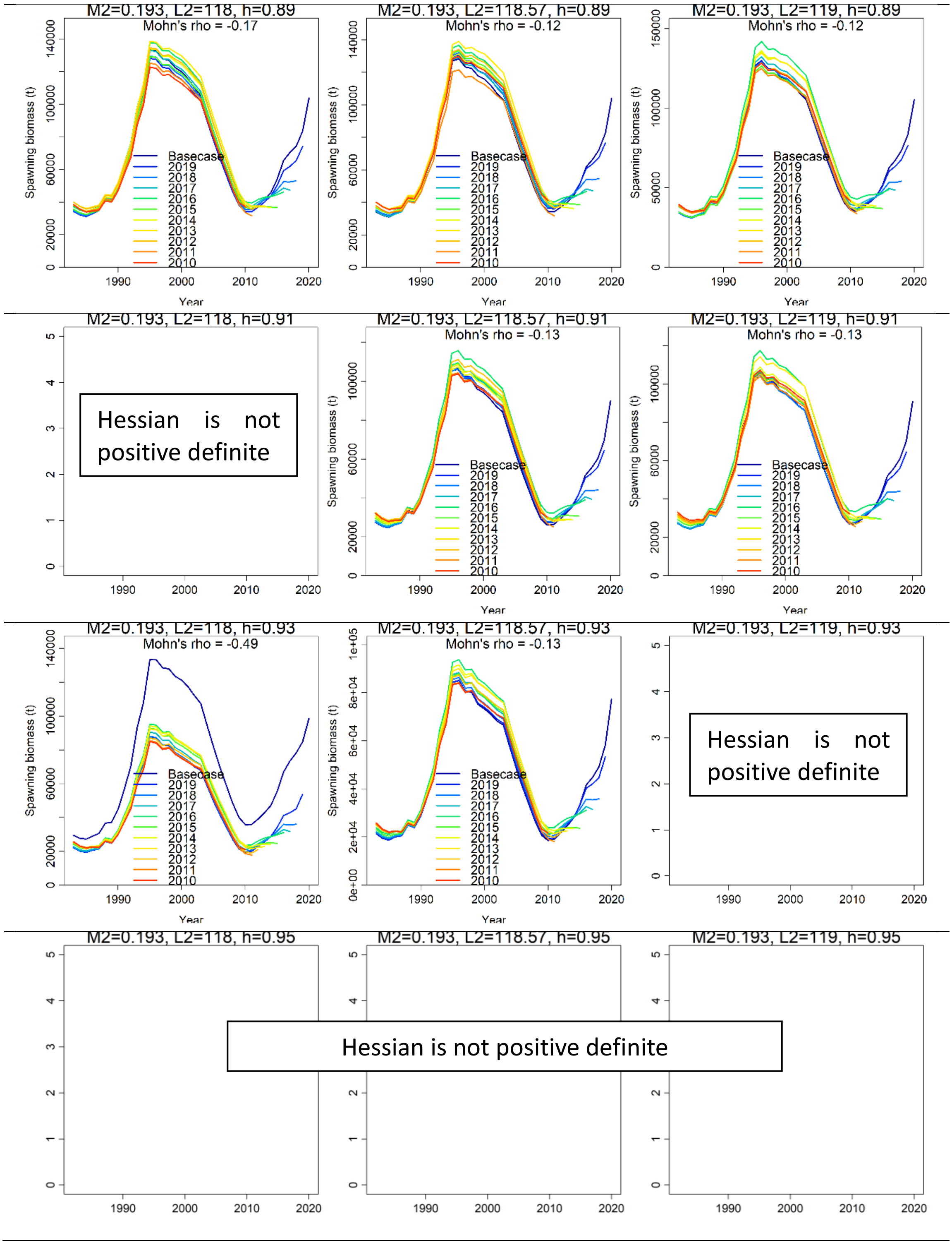
M2=0.193, L2=118.57, h=0.85
Mohn's rho = -0.12

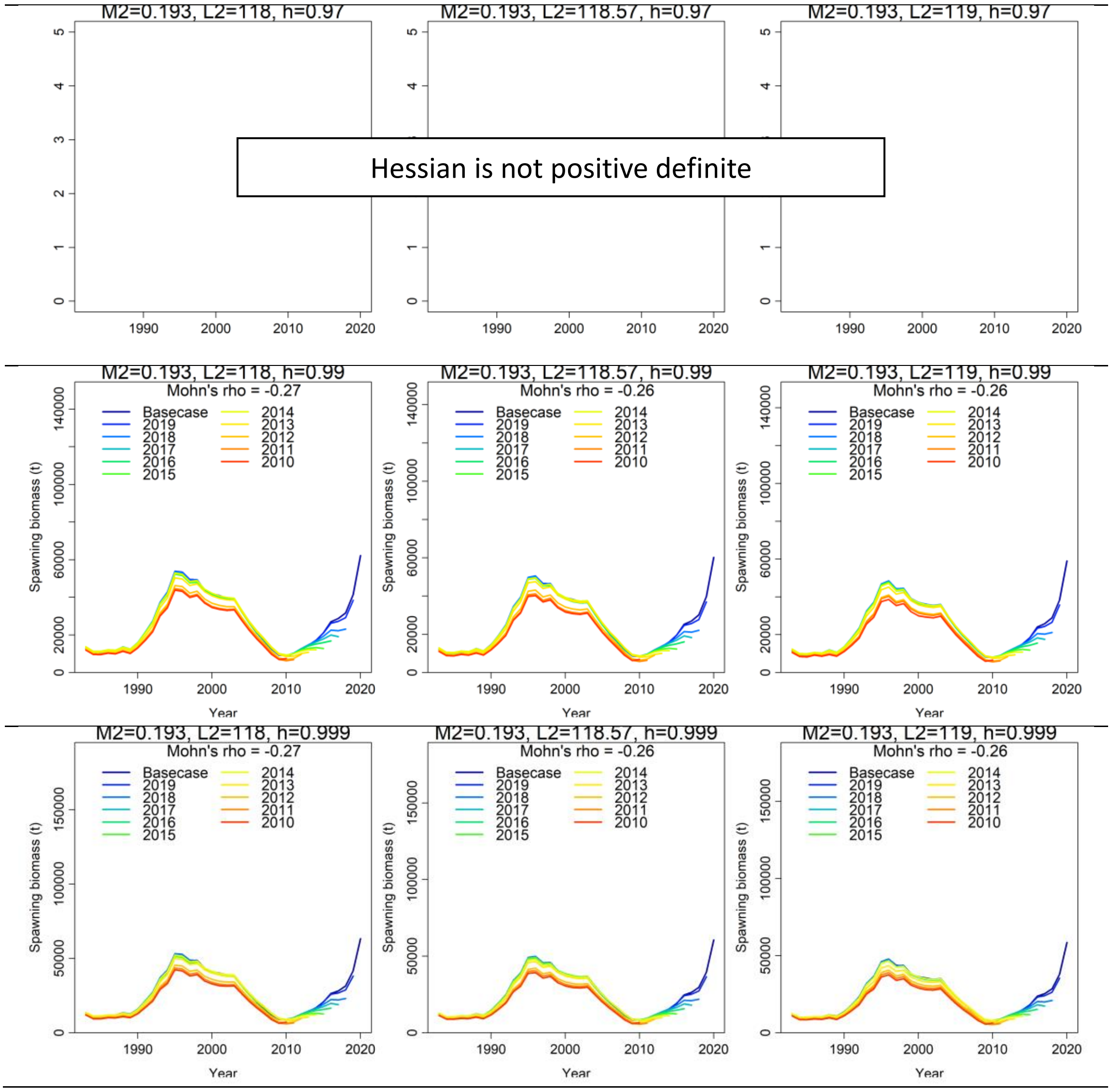
M2=0.193, L2=119, h=0.85
Mohn's rho = -0.11

M2=0.193, L2=118, h=0.87
Mohn's rho = -0.11

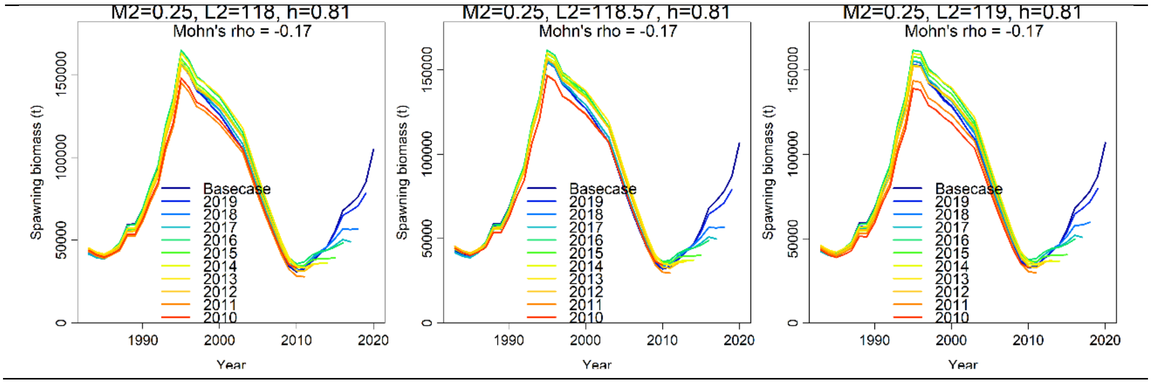
M2=0.193, L2=118.57, h=0.87
Mohn's rho = -0.12

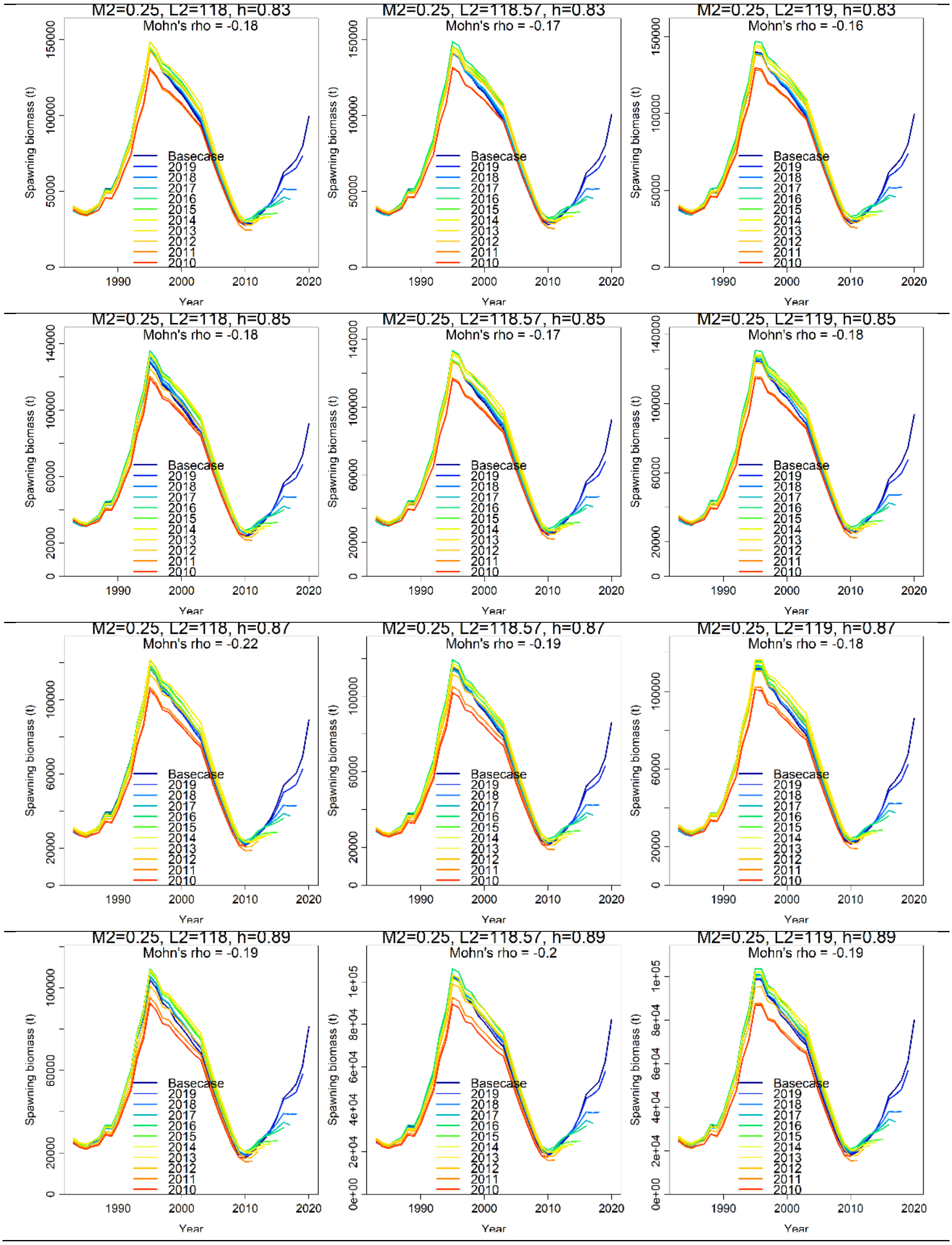
M2=0.193, L2=119, h=0.87
Mohn's rho = -0.12

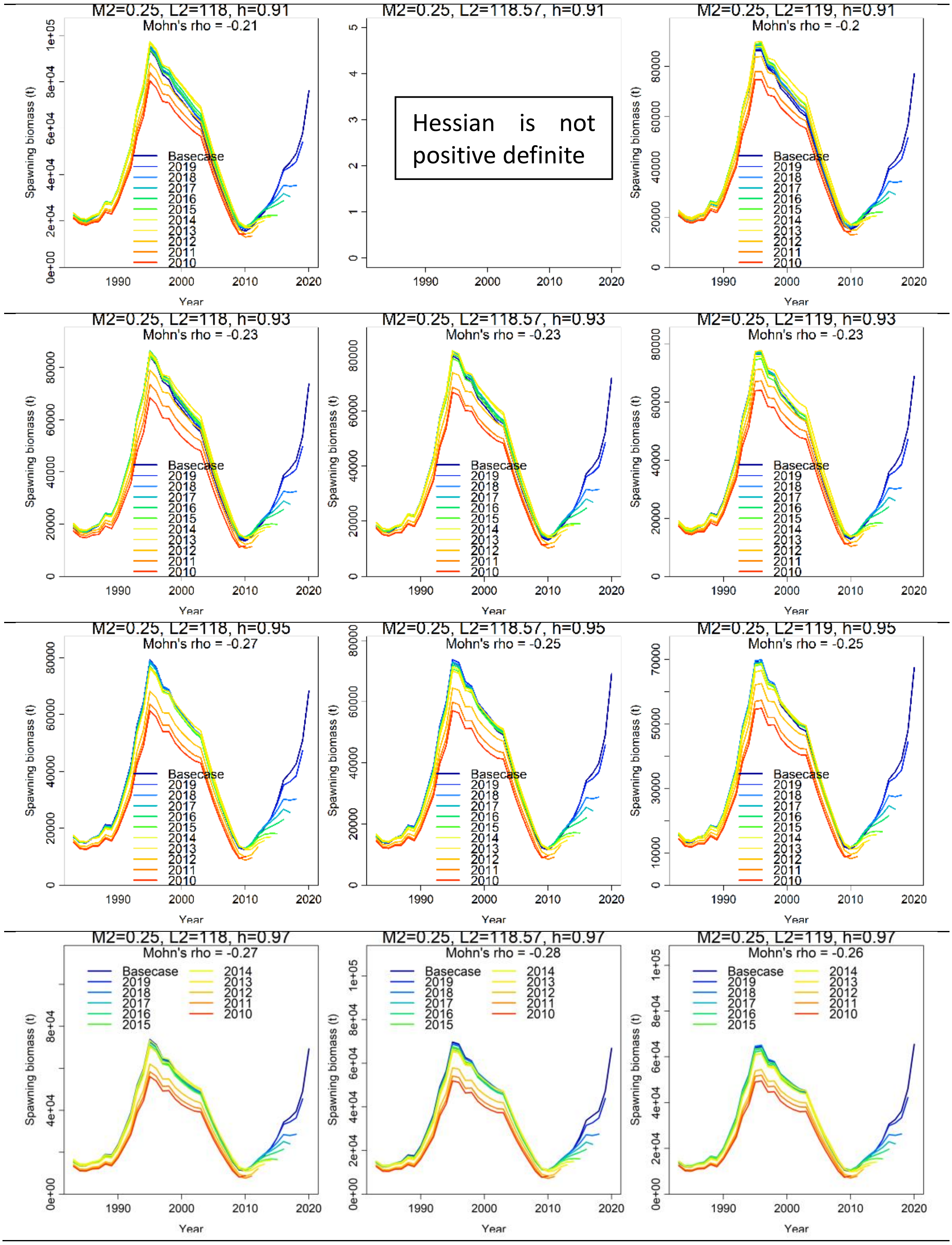

M2=0.193, L2=118, h=0.89
Mohn's rho = -0.17
M2=0.193, L2=118.57, h=0.89
Mohn's rho = -0.12
M2=0.193, L2=119, h=0.89
Mohn's rho = -0.12
M2=0.193, L2=118, h=0.91
Hessian is not positive definite
M2=0.193, L2=118.57, h=0.91
Mohn's rho = -0.13
M2=0.193, L2=119, h=0.91
Mohn's rho = -0.13
M2=0.193, L2=118, h=0.93
Mohn's rho = -0.49
M2=0.193, L2=118.57, h=0.93
Mohn's rho = -0.13
M2=0.193, L2=119, h=0.93
Hessian is not positive definite
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
Hessian is not positive definite
Basecase
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
Spawning biomass (t)
Year
1990
2000
2010
2020


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
Hessian is not positive definite
0
1
2
3
4
5
1990
2000
2010
2020
M2=0.193, L2=118, h=0.99
Mohn's rho = -0.27
M2=0.193, L2=118.57, h=0.99
Mohn's rho = -0.26
M2=0.193, L2=119, h=0.99
Mohn's rho = -0.26
Basecase
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
Spawning biomass (t)
0
20000
60000
100000
140000
Year
M2=0.193, L2=118, h=0.999
Mohn's rho = -0.27
M2=0.193, L2=118.57, h=0.999
Mohn's rho = -0.26
M2=0.193, L2=119, h=0.999
Mohn's rho = -0.26
0
50000
100000
150000

b. M_{2+}=0.25


M2=0.25, L2=118, h=0.81
Mohn's rho = -0.17
M2=0.25, L2=118.57, h=0.81
Mohn's rho = -0.17
M2=0.25, L2=119, h=0.81
Mohn's rho = -0.17
Basecase
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
Spawning biomass (t)
0
50000
100000
150000
1990
2000
2010
2020
Year


M2=0.25, L2=118, h=0.83
Mohn's rho = -0.18
M2=0.25, L2=118.57, h=0.83
Mohn's rho = -0.17
M2=0.25, L2=119, h=0.83
Mohn's rho = -0.16
M2=0.25, L2=118, h=0.85
Mohn's rho = -0.18
M2=0.25, L2=118.57, h=0.85
Mohn's rho = -0.17
M2=0.25, L2=119, h=0.85
Mohn's rho = -0.18
M2=0.25, L2=118, h=0.87
Mohn's rho = -0.22
M2=0.25, L2=118.57, h=0.87
Mohn's rho = -0.19
M2=0.25, L2=119, h=0.87
Mohn's rho = -0.18
M2=0.25, L2=118, h=0.89
Mohn's rho = -0.19
M2=0.25, L2=118.57, h=0.89
Mohn's rho = -0.2
M2=0.25, L2=119, h=0.89
Mohn's rho = -0.19
Basecase
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
Spawning biomass (t)
Year
1990
2000
2010
2020
0
50000
100000
150000
20000
60000
100000
140000
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05


M2=0.25, L2=118, h=0.91
Mohn's rho = -0.21
M2=0.25, L2=118.57, h=0.91
Hessian is not positive definite
M2=0.25, L2=119, h=0.91
Mohn's rho = -0.2
M2=0.25, L2=118, h=0.93
Mohn's rho = -0.23
M2=0.25, L2=118.57, h=0.93
Mohn's rho = -0.23
M2=0.25, L2=119, h=0.93
Mohn's rho = -0.23
M2=0.25, L2=118, h=0.95
Mohn's rho = -0.27
M2=0.25, L2=118.57, h=0.95
Mohn's rho = -0.25
M2=0.25, L2=119, h=0.95
Mohn's rho = -0.25
M2=0.25, L2=118, h=0.97
Mohn's rho = -0.27
M2=0.25, L2=118.57, h=0.97
Mohn's rho = -0.28
M2=0.25, L2=119, h=0.97
Mohn's rho = -0.26
Basecase
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
Spawning biomass (t)
Year
1990
2000
2010
2020

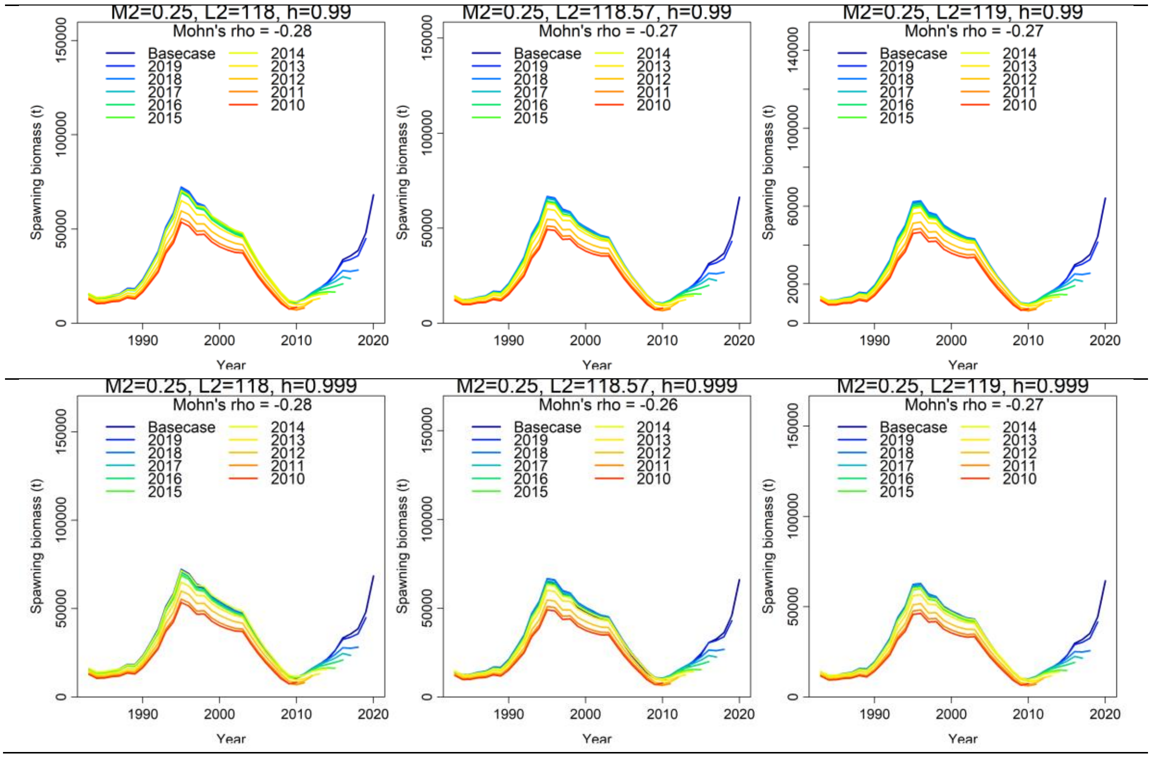


Figure 3. 10-year retrospective analyses of spawning stock biomass using various models that involve altering the values of length at age 3 (L_2) and steepness (h), while maintaining a constant natural mortality rate for age 2 and older (M_{2+}) at (a) 0.193 and (b) 0.25. Mohn's ρ values are shown in 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).

		M_{2+}=0.193			M_{2+}=0.25		
		L_2=118	L_2=118.57	L_2=119	L_2=118	L_2=118.57	L_2=119
		(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)	(L_{inf}=248.6)	(L_{inf}=249.9)	(L_{inf}=250.9)
Steepness	0.81	16.1	15.8	14.8	5.8	5.3	4.5
	0.83	14.4	14.6	13.7	4.8	3.6	3.0
	0.85	13.4	12.9	12.8	2.6	2.2	1.7
	0.87	11.1	11.0	10.9	2.4	1.4	0.6
	0.89	10.0	9.1	9.0	0.1	0.0	-0.8
	0.91	.	6.4	6.1	-1.8	.	-2.4
	0.93	7.8	3.3	.	-3.0	-3.9	-4.7
	0.95	.	.	.	-4.1	-5.0	-5.6
	0.97	.	.	.	-5.2	-5.3	-6.5
	0.99	-2.6	-3.4	-4.4	-5.5	-6.6	-7.4
	0.999	-3.8	-4.2	-4.7	-5.3	**-6.6**	-7.1

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	0.81	2	2	2	2	2	2
	0.83	2	2	2	2	2	2
	0.85	2	2	2	2	2	2
	0.87	2	2	2	2	2	2
	0.89	2	2	2	2	2	2
	0.91	0	2	2	3	0	2
	0.93	1	2	0	3	3	4
	0.95	0	0	0	2	4	4
	0.97	0	0	0	3	3	4
	0.99	1	2	2	2	2	2
	0.999	1	2	4	3	4	3

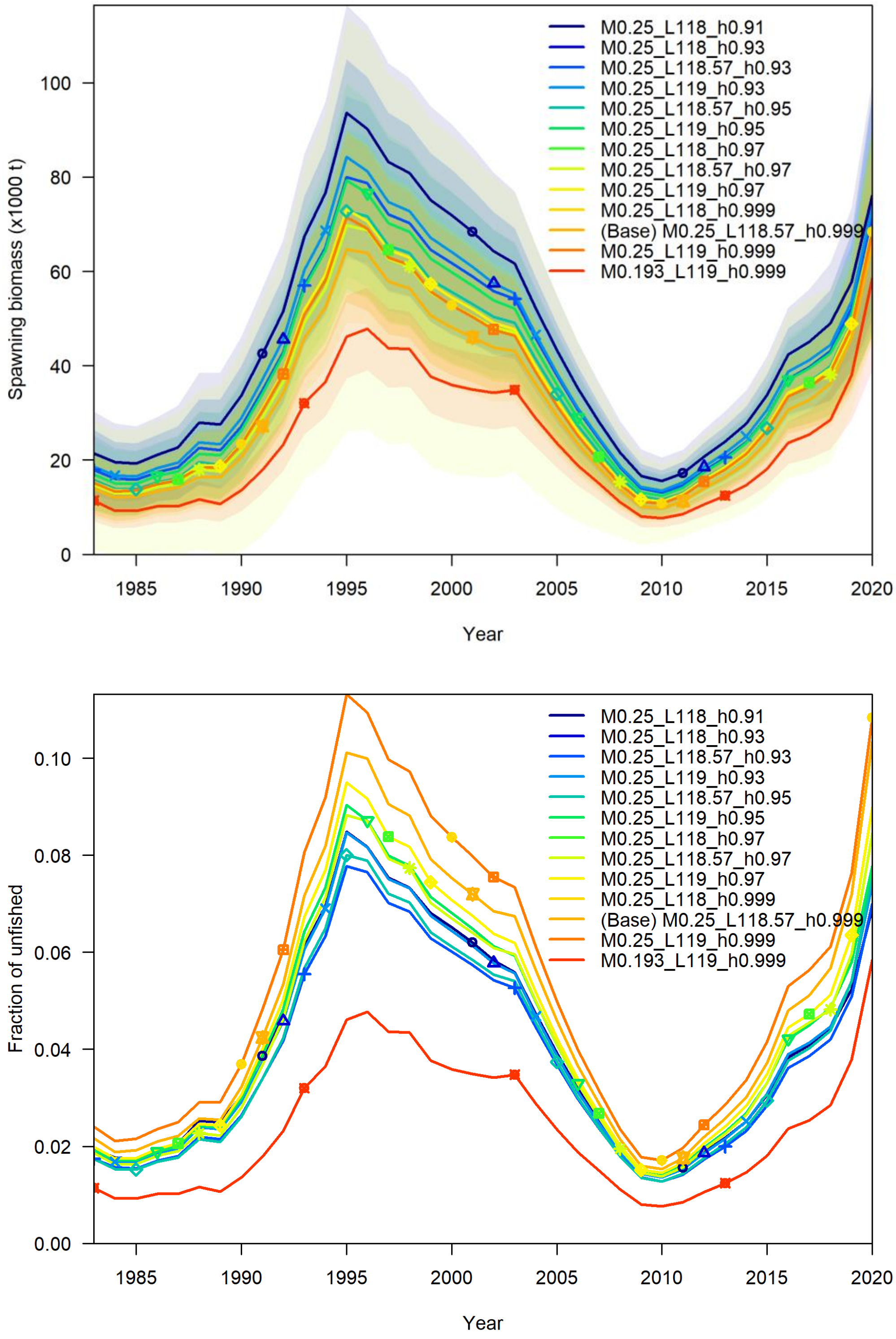


Figure 4. The trajectory of the spawning biomass (upper panel) and spawning stock biomass ratio (lower panel) estimated from all selected grid model with the score at 3 and 4 (referred to Table 6).