

VALIDATION OF THE BIOMASS DYNAMIC STOCK ASSESSMENT MODEL FOR USE IN A MANAGEMENT PROCEDURE

Laurence Kell¹, Haritz Arrizabalaga², and Paul De Bruyn¹, Gorka Merino², Iago Mosqueira³, Rishi Sharma⁴ and Jose-Maria Ortiz de Urbina⁵

SUMMARY

mpb is an R Package for modelling Management Procedures, based on biomass dynamic stock assessment models and empirical harvest control rules. A Management Procedure is the combination of predefined data, together with an algorithm to estimate stock trends or status and to set management measures such as a total allowable catch. Management Procedures are simulation tested by conducting Management Strategy Evaluation. To validate the biomass stock assessment algorithms incorporated in the mpb package and used in the Management Procedure, the fits obtained with the biodyn assessment class are compared to the estimates made using another biomass dynamic model (ASPIC) by the SCRS for North Atlantic albacore.

RÉSUMÉ

mpb est un paquet R servant à modéliser les procédures de gestion, sur la base de modèles d'évaluation de la dynamique de la biomasse du stock et de règles de contrôle de l'exploitation empiriques. Une procédure de gestion est une combinaison de données prédéfinies et d'un algorithme servant à estimer les tendances du stock ou son état et d'établir des mesures de gestion telles que le total des prises admissibles. Des procédures de gestion sont testées par simulation au moyen d'une évaluation de la stratégie de gestion. Afin de valider les algorithmes de l'évaluation de la biomasse du stock intégrés dans le paquet mpb et utilisés dans la procédure de gestion, les ajustements obtenus avec le modèle d'évaluation Biodyn sont comparés aux estimations réalisées au moyen d'un autre modèle de dynamique de la biomasse (ASPIC) par le SCRS pour le germon de l'Atlantique Nord.

RESUMEN

El mpb es un paquete R para la modelación de procedimientos de ordenación, basado en modelos de evaluación de stock de dinámica de biomasa y normas empíricas de control de la captura. Un procedimiento de ordenación es la combinación de datos predefinidos, junto con un algoritmo para estimar las tendencias o el estado del stock y para establecer medidas de ordenación como un total admisible de captura. Los procedimientos de ordenación son probados mediante simulación realizando una evaluación de estrategias de ordenación. Para validar los algoritmos de evaluación de biomasa del stock incorporados al paquete mpb y utilizados en el procedimiento de ordenación, se comparan los ajustes obtenidos con la clase de evaluación biodyn con las estimaciones realizadas utilizando otro modelo de dinámica de biomasa (ASPIC) por el SCRS para el atún blanco del Atlántico norte.

KEYWORDS

Biomass Dynamic, Management Procedure, Management Strategy Evaluation, Stock Assessment, Validation

1 ICCAT Secretariat, C/Corazón de María, 8. 28002 Madrid, Spain; Laurie.Kell@iccat.int; Phone: +34 914 165 600; Fax: +34 914 152 612.

2 AZTI -Tecnalia, Herrera Kaia Portualdea, 20110, Pasaia, Spain

3 European Commission Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen (IPSC), Maritime Affairs Unit, Via E. Fermi 2749, 21027, Ispra, VA, Italy

4 U.S. National Marine Fisheries Service, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL, 33149-1099, USA

5 Instituto Español de Oceanografía (IEO- CO Málaga).

Introduction

mpb is an R Package for modelling Management Procedures, based on biomass dynamic stock assessment models and empirical harvest control rules. Where a Management Procedure is the combination of predefined data, together with an algorithm to estimate stock trends or current status and to set management measures such as a total allowable catch. Management Procedures are simulation tested by conducting Management Strategy Evaluation. To validate the biomass stock assessment algorithms used in the Management Procedure (i.e. in the mpb package) the fits obtained are compared to the estimates made by the SCRS for North Atlantic albacore.

The software used by the SCRS in 2013 to conduct the biomass based assessment was ASPIC (Prager, 1994). ASPIC is already in the ICCAT software catalogue and the mpb assessment routines are validated by replicating the SCRS assessment.

Material and Methods

Material

In the 2013 assessment seven scenarios were considered based on different combinations of the available CPUE series (**Table 1**). The time series used in 2013 are shown in **Figure 1**; points are the standardised indices, the red line is a common loess smoother fitted to all series and the blue a loess smoother fitted by series.

Figure 2 is a visual representation of a hierarchical cluster analysis of the series; the depth of blue indicates the strength of the positive correlation between pairs of series. Two clusters are seen, i.e. the Japanese longline and the split Chinese-Taipei indices and Combined Troll.

Methods

First the estimates obtained by the aspic stock assessment procedure in the mpb package were compared with those obtained by the albacore stock assessment working group. Profiles of the objective function were calculated by fixing carrying capacity (K) at a range of values and then estimating the other parameters (Kell et al., 2014a). This allowed to check that the assessments ran in 2013 had converged to a unique solution.

The population model was then validated by projecting the stock for the reported catch and estimated parameters. Following which the biodyn biomass dynamic assessment procedure (Kell *et al.*, 2014b) was fitted to the data and the times series and objective functions, obtained from aspic and biodyn, compared.

Finally a jack knife was conducted to evaluate the stability of the assessments, i.e. whether the addition or the change of a single point have a big effect or not.

In a biomass dynamic model the stock next year (B_{t+1}) is the sum of the current biomass (B_t) less the catch (C_t) plus the surplus production (P_t) i.e.

$$B_{t+1} = B_t - C_t + P_t$$

There are various forms of production functions (P_t), e.g. the symmetric logistic (Hassell, 1975) or the generalised Pella and Tomlinson (1969) forms. The logistic production function is probably not appropriate for tuna species, due to high steepness (Maunder, 2003) and a Pella-Tomlinson form with $B_{MSY} < 0.5B_0$ is perhaps more realistic, e.g.

$$\frac{r}{p} \cdot B \left(1 - \left(\frac{B}{K}\right)^p\right)$$

where (r) is the intrinsic rate of increase, (K) the carry capacity (p) the shape of the surplus production function. If $p < 1$ then the curve is skewed to the left.

The dynamics, i.e. productivity and reference points and the response of the stock to perturbations, are determined by r and the shape of the production function p. If $p = 1$ then MSY is found halfway between 0 and K, as p increases MSY shifts to the right.

To fit the model requires time series of catch and relative abundance, e.g. catch per unit effort (CPUE), and a fitting criteria. In the albacore assessment the least absolute value (LAV) was used as the objective function, where the sum of the absolute values was minimised.

The working group also used the logistic form of the production function and in the validation procedure the Pella Tomlinson production function was fitted with $p=1$. i.e. a symmetric shape was assumed. The logistic function can be fitted either assuming harvest rate (catch/stock per year) or instantaneous mortality. Both assumptions were evaluated.

Jack knife

A jack knife was conducted to evaluate the impact of individual data point on the estimates. In a jackknife a single point is removed in turn and the model refitted. A jack knife also provides an estimate of bias.

Hindcast

A major uncertainty in stock assessment is the difference between models and reality. The validation of model predictions is difficult, however, as fish stocks can rarely be observed and counted. We therefore conducted a hindcast to prediction skill. In a hindcast a model is fitted to the first part of a time series and then projected over the period omitted in the original fit. Prediction skill can then be evaluated by comparing the projection predictions with observations (Kell et al. accepted).

Results

The fits obtained in 2013 are shown in **Figure 3**; the first step in a stock assessment is to check that the fitting algorithm has converged, i.e. a solution has actually been found. Therefore profiles of the LAV are plotted in **Figure 4**; the blue dots are the working group estimates and the red dots the estimates obtained by the aspic method in the mpb package. The bottom row shows scenario 4 on different scales. In two cases (scenario 3 and 4) a global minimum was not found.

In all cases, other than scenario 4, the aspic algorithm in mpb found the same solution as the standalone executable. In scenario 4 a "better fit" was found, due to better starting conditions. This is because in scenario 4 the objective function is bimodal (the blue and red lines) and the fitting algorithm flips between them (grey lines). Therefore there are many local minima.

To evaluate the population dynamic model the stock was projected by scenario for the reported catch and the 2013 ASPIC stock assessment parameter estimates (**Figure 5**); red lines are the projections obtained using the mpb package and black are the ASPIC fits. Two options are available for the population model in biodyn; either all the catch is taken after recruitment to the stock at the beginning of the year (i.e. a harvest rate) or fishing is modelled as the instantaneous value throughout the year. In the former case catches cause the stock to crash in some cases, in the latter both aspic and ASPIC give the same results.

The biodyn procedure was then fitted to the data, for the two harvesting assumptions. The estimates of the stock time series are shown in **Figure 6** and the objective function values (LAV) in **Table 2**. For harvest rate in some cases the stock collapsed. Assuming that fishing mortality was instantaneous results in similar fits to ASPIC, although these were not identical. Reference to **Table 2** shows that this is because in all cases, other than scenario 4, biodyn obtains better fits, although the differences are small.

To test the stability of the assessments a jack knife was performed for two objective functions i) sum of squares and ii) LAV. **Figures 7 and 8** show the results for r and K for fitting using sum of squares. The plots of r are the inverse of K and so for LAV (**Figure 14**) only the results for K are shown. These figures show the estimates of r and K for assessment scenario (row) and CPUE series (column). In the case of the assessment based on the Troll series, a single point can have a large effect, i.e. up to 10%. This translates into a very large bias in the estimates.

In the case of LAV similar magnitudes of bias were seen, however, there is also flipping between positive and negative bias. This indicates that the fits are likely to be highly unstable and unreliable, i.e. can not be replicated if a single point changes.

The reliability of advice is examined in **Figures 9,10,11,12 and 13** for fitting using sum of squares and in **Figures 15,16,17,18 and 19** using LAV. Ideally the lines should coincide; otherwise advice is likely to be unreliable.

When using sum of squares to fit, absolute values of biomass are more reliable than relative values; as can be seen by comparing **Figures 9 and 10**. Using sum of squares also appears to be more reliable than using LAV. F/F_{MSY} also appears to be more stable than biomass.

Discussion and Conclusions

The intention of the paper was validate the stock assessment algorithms *biodyn* and *aspic* in the *mpb* package, however, problems were found in the fits. This was due to lack of convergence and instability. In that not all the stock assessments had actually found the best solution and the results were sensitive to small changes in the dataset.

Ideally an assessment should be stable, i.e. the addition or removal of a single point should not have much effect on the fits. Least squares regression is unstable if the datasets show collinearity, while least absolute deviation (LAD or LAV) can exhibit instability at datasets that are far from collinear (e.g. Ellis, 1998). Using conflicting CPUE series, or series with little signal can therefore produced unstable and unreliable results when LAV, as in this case, was used as the objective function.

Instability is undesirable for several reasons, e.g. you do not want to find that updating an assessment with additional data point change the results.

References

- S. P. Ellis. Instability of least squares, least absolute deviation and least median of squares linear regression. *Statistical Science*, pages 337–344, 1998.
- M. Hassell. Density-dependence in single-species populations. *The Journal of animal ecology*, pages 283–295, 1975.
- L. Kell, P. De Bruyn, M. Maunder, K. Piner, and I. Taylor. 2014a. Likelihood component profiling as a data exploratory tool for North Atlantic albacore. *ICCAT Collect. Vol. Sci. Pap.*, 70(3):1288–1293.
- L. Kell, D. B. P. M. I. Ortiz de Urbina, J., and A. Magnusson. 2014b. An evaluation of different approaches for modelling uncertainty in ASPIC and biomass dynamic models. *ICCAT Collect. Vol. of Sci. Pap.*, 70(5):2111-2119.
- M. P. Laurence Kell. ASPIC, 2013. URL <http://CRAN.R-project.org/package=ALKr>. R package version 0.5.3.
- M. N. Maunder. Is it time to discard the Schaefer model from the stock assessment scientists toolbox? *Fish. Res.*, 61(1):145–149, 2003.
- J. Pella and P. Tomlinson. A generalized stock production model. *Inter-American Tropical Tuna Commission*, 1969.
- M. Prager. A suite of extensions to a nonequilibrium surplus-production model. *Fishery Bulletin*, 92(2): 374–389, 1994.

Table 1. CPUE series by assessment scenario

Scenario	cpue
1	Troll Composite
2	CT Old CT New
3	CT All
4	JLL Old JLL New
5	Troll Composite JLL Old JLL New CT Old CT New
6	Troll Composite JLL Old JLL New
7	Troll Composite CT Old CT New

Table 2. Objective function values

Scenario	aspic	biodyn
1	19.67	19.63
2	5.08	5.08
3	5.18	5.18
4	12.54	12.94
5	40.28	39.85
6	34.55	34.39
7	25.44	25.27

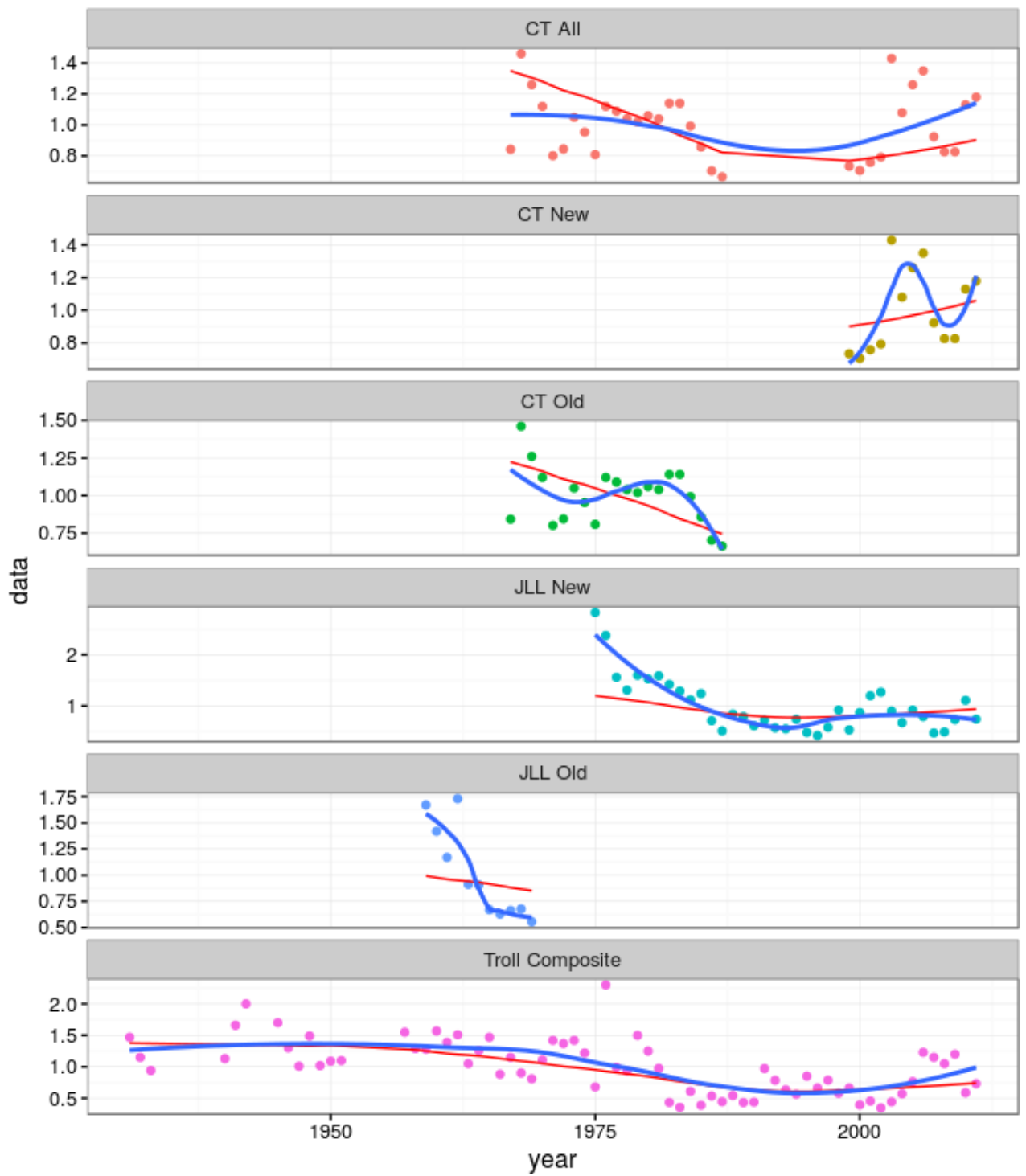
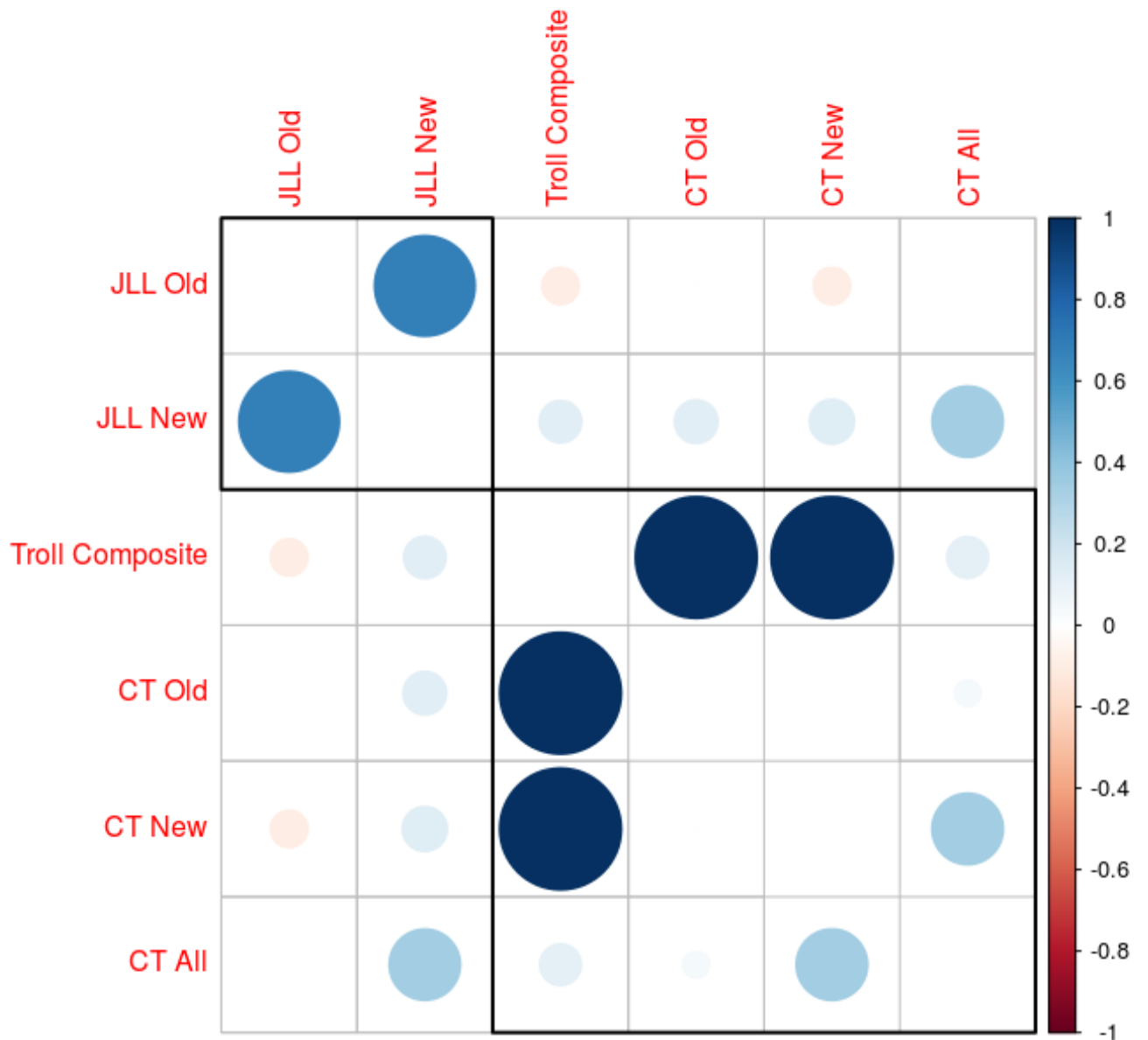


Figure 1. CPUE Series used in the 2013 albacore ASPIC assessment. Points are the standardised indices, the red line is a common loess smoother fitted to all series and the blue a loess smoother fitted by series.



NULL

Figure 2. Hierarchical cluster analysis of CPUE series.

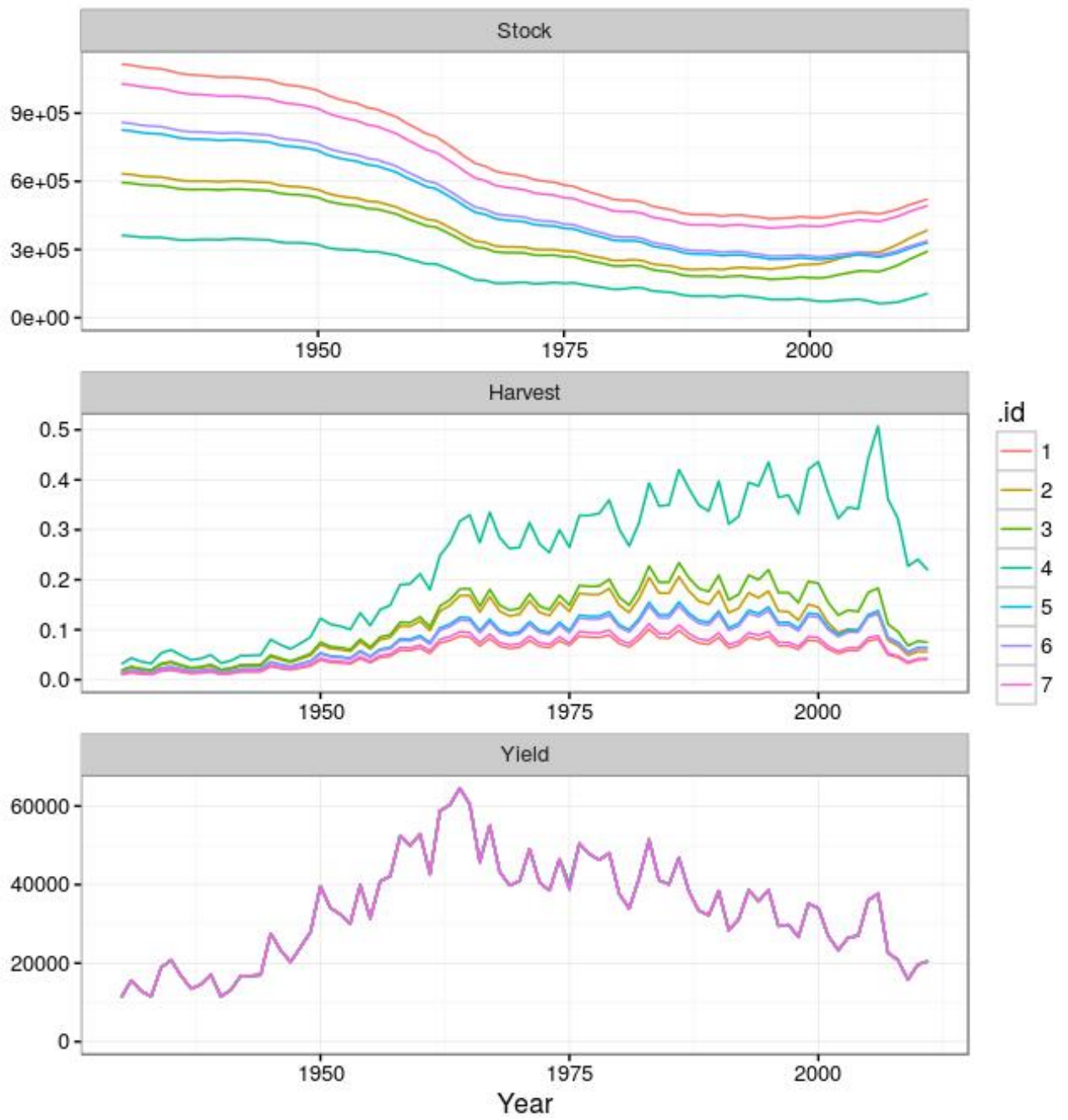


Figure 3. Fits to North Atlantic albacore, by scenario.

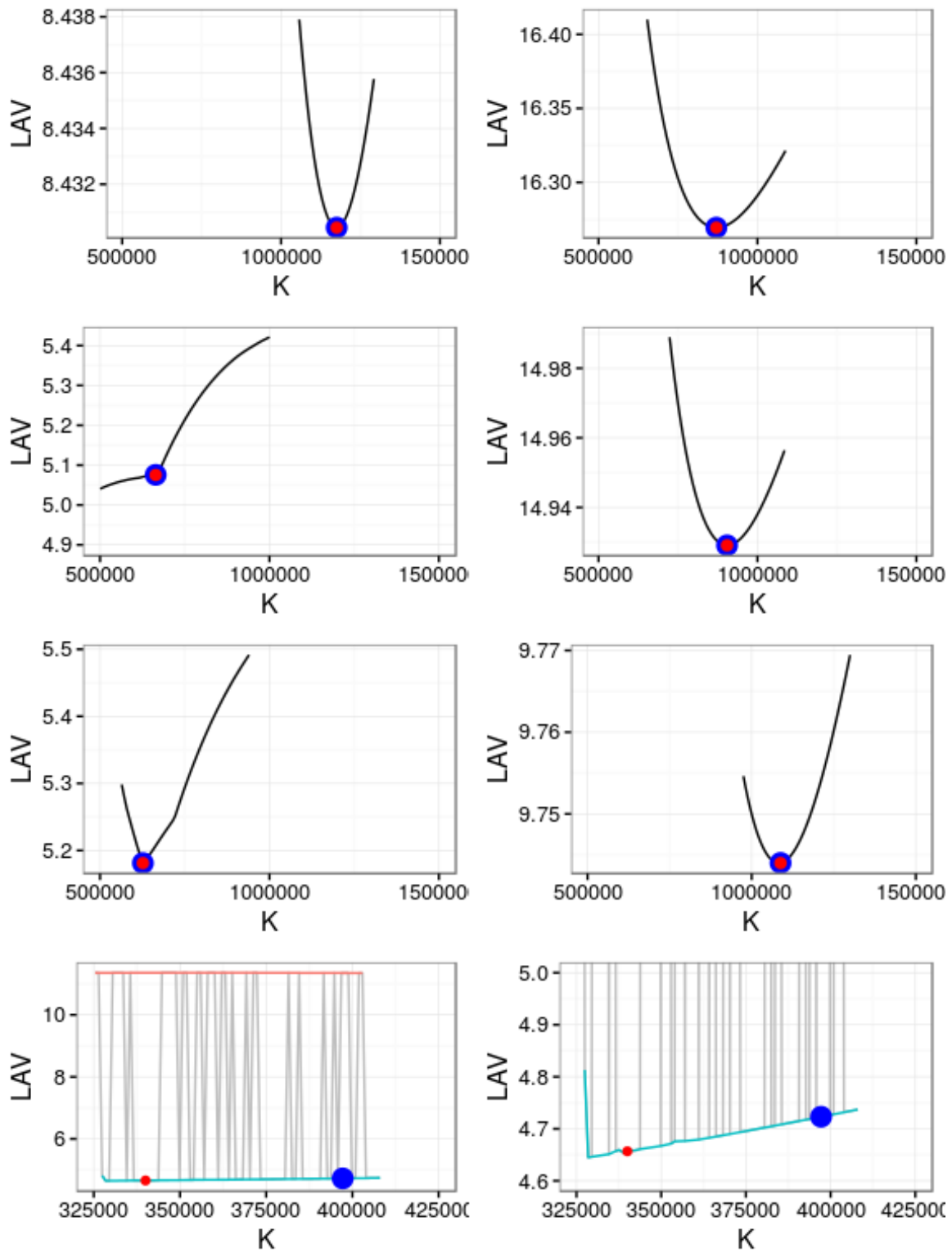


Figure 4. Likelihood profiles of K for North Atlantic albacore. The bottom row shows Scenario 4 on two different scales.

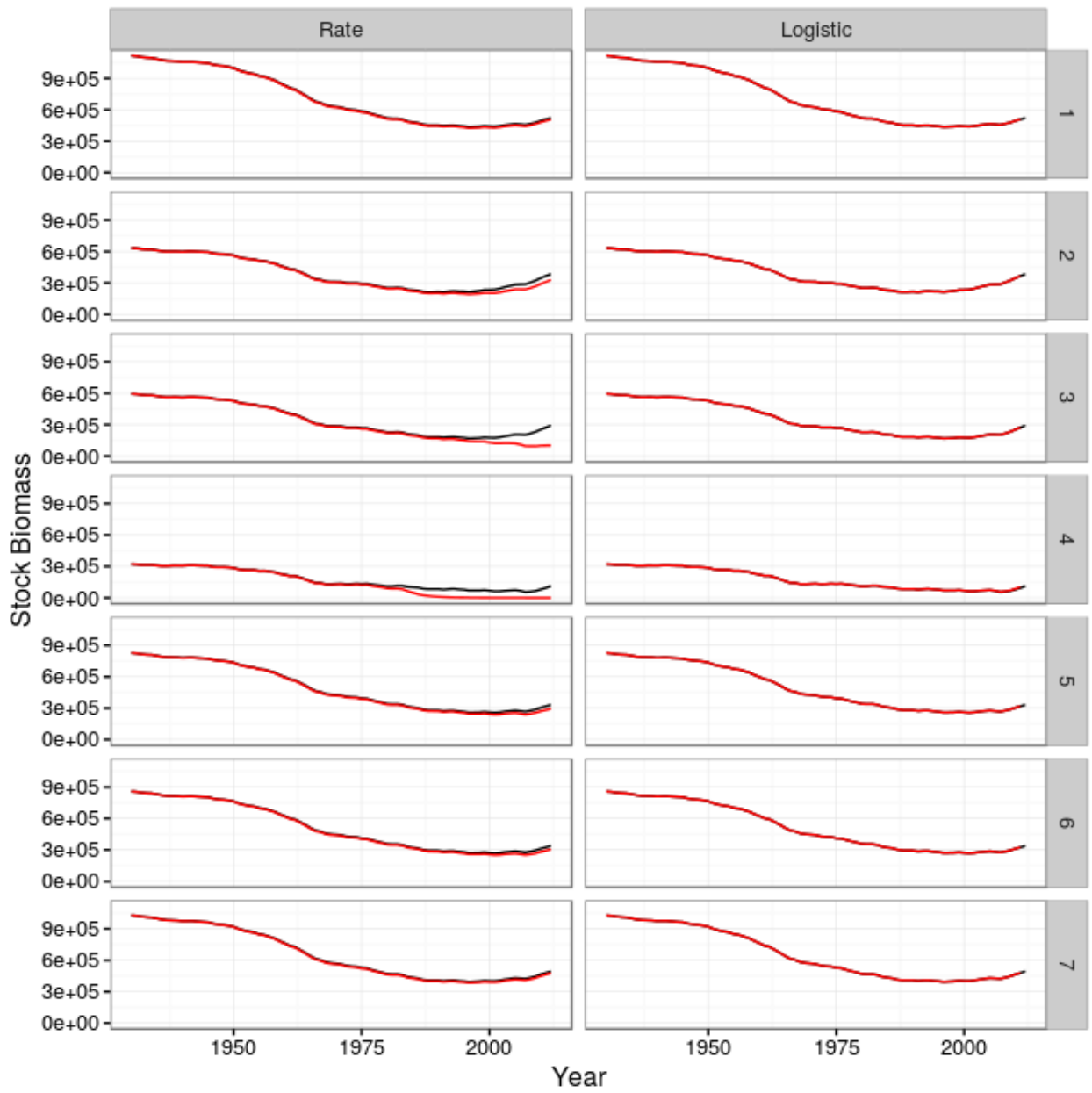


Figure 5. Projected time series of biomass, using biodyn and the aspic estimates.

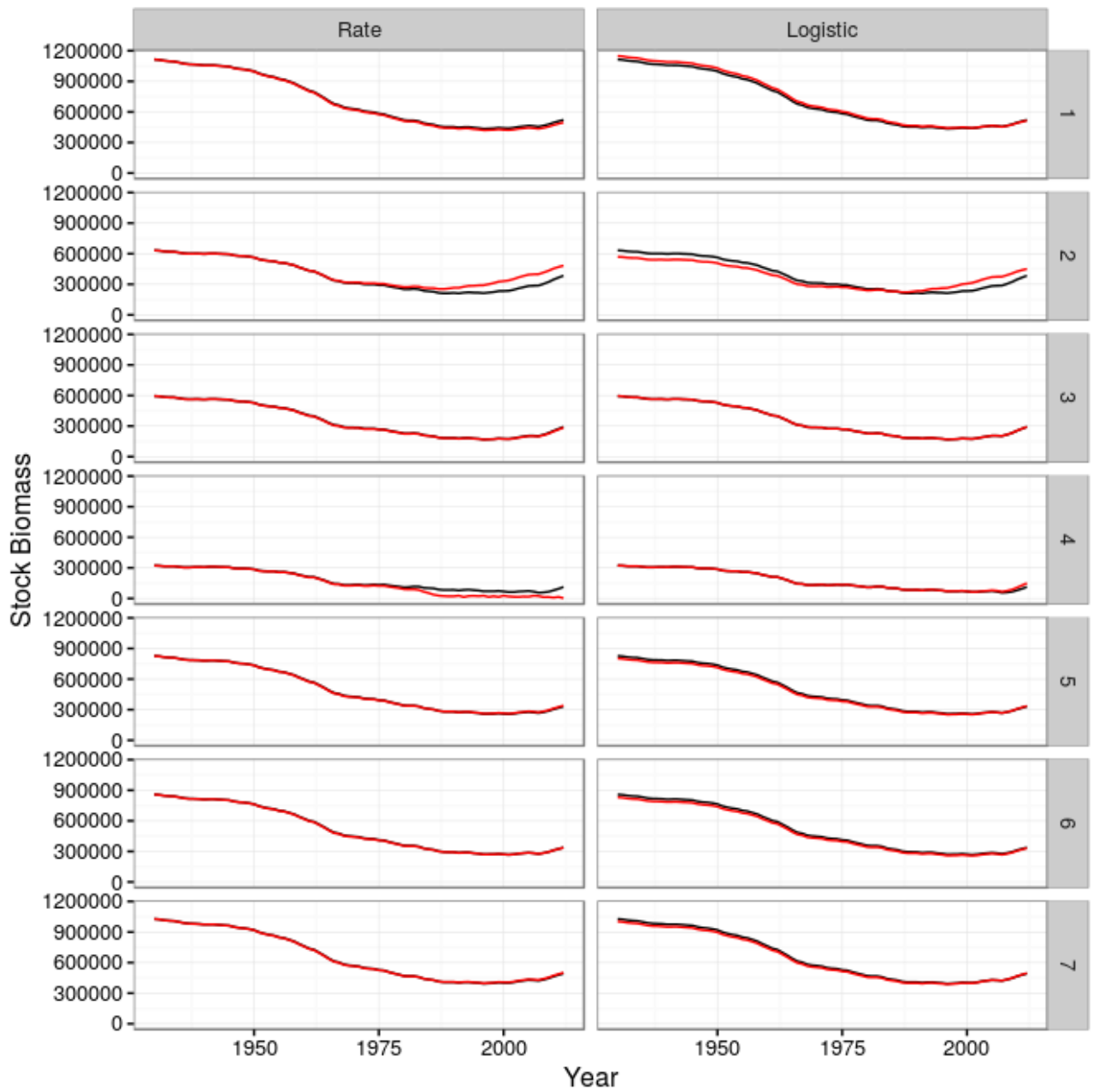


Figure 6. Fitted time series of biomass.

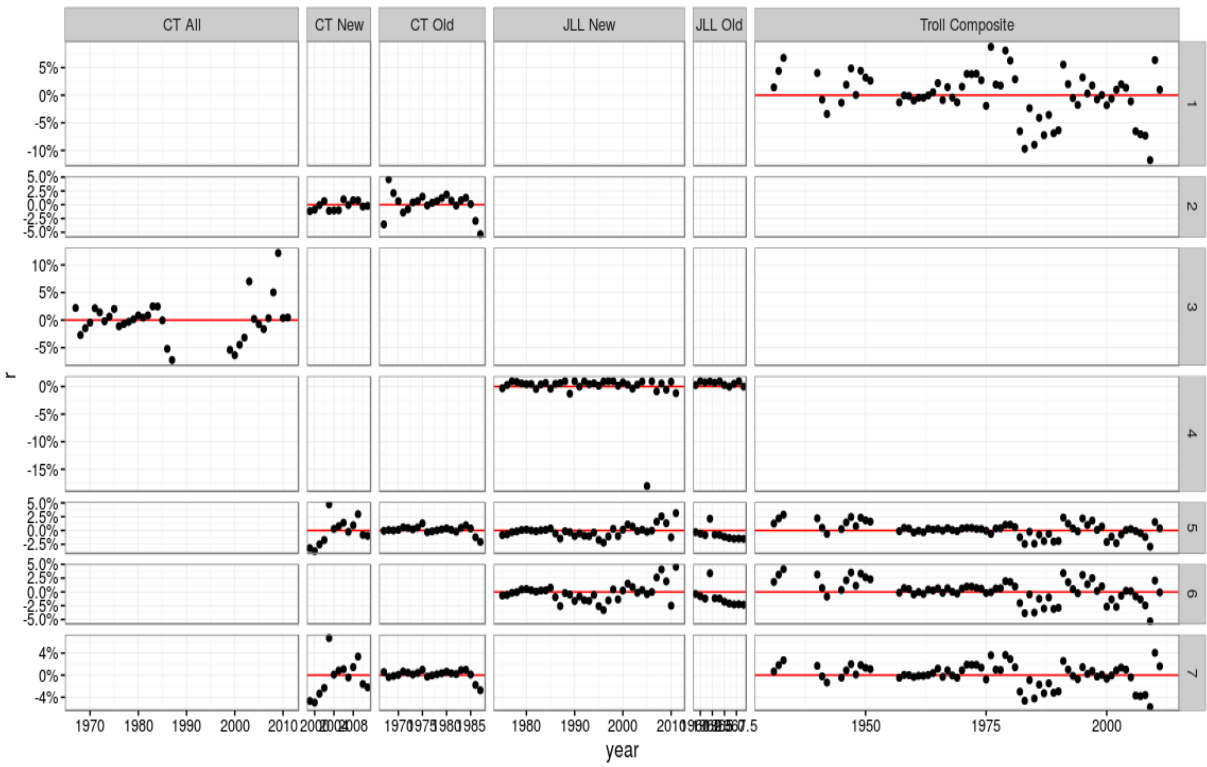


Figure 7. Jackknife estimates of r .

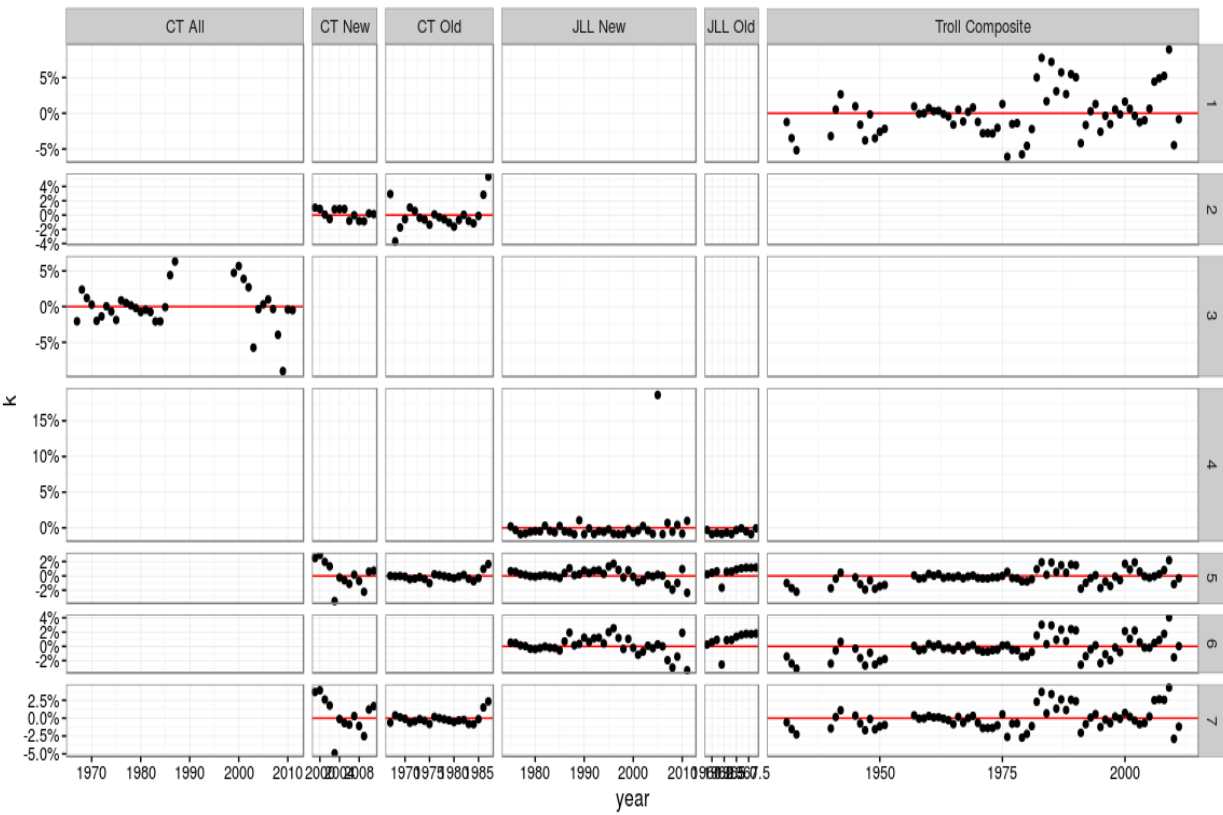


Figure 8. Jackknife K .

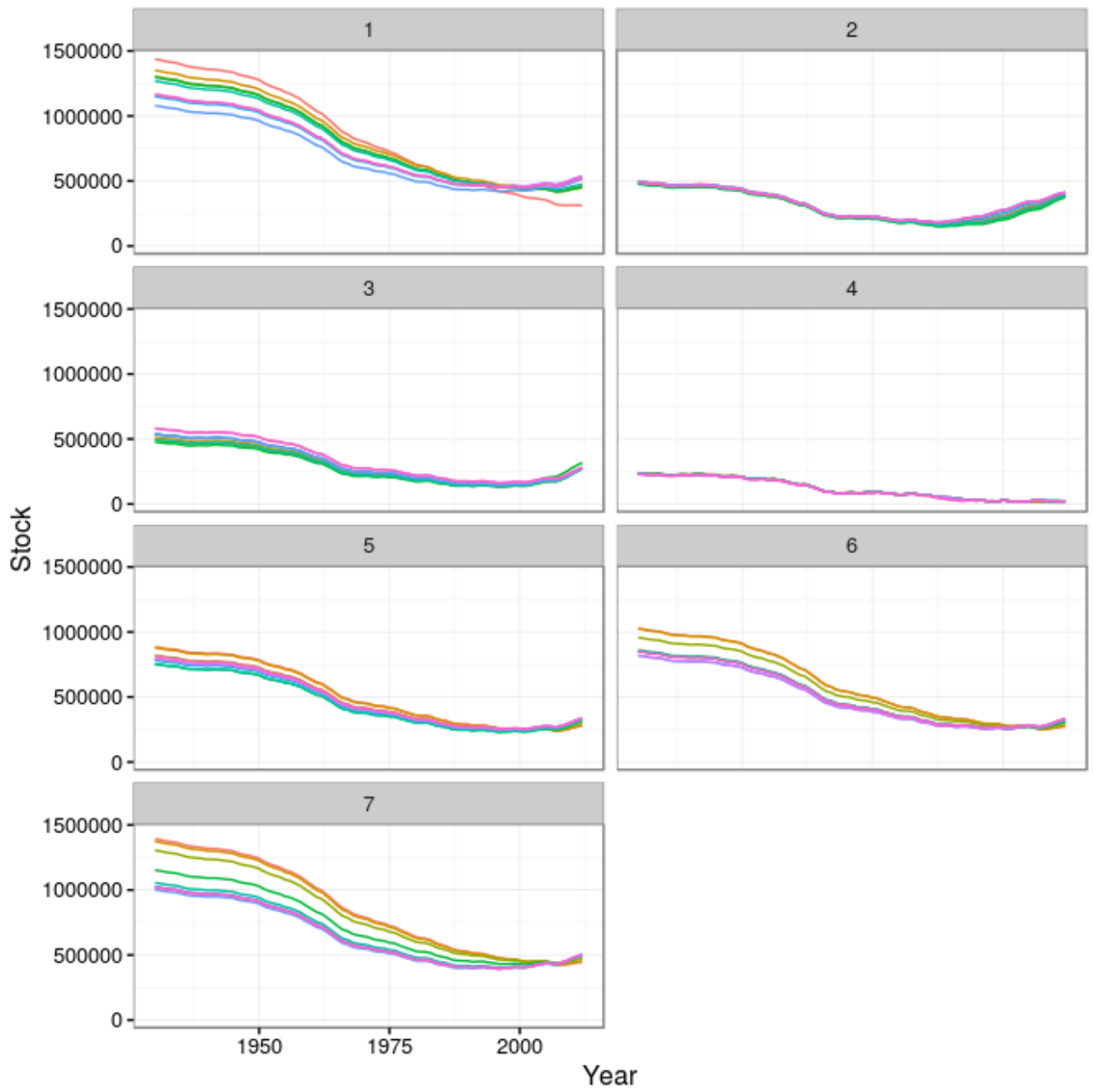


Figure 9. Hindcast

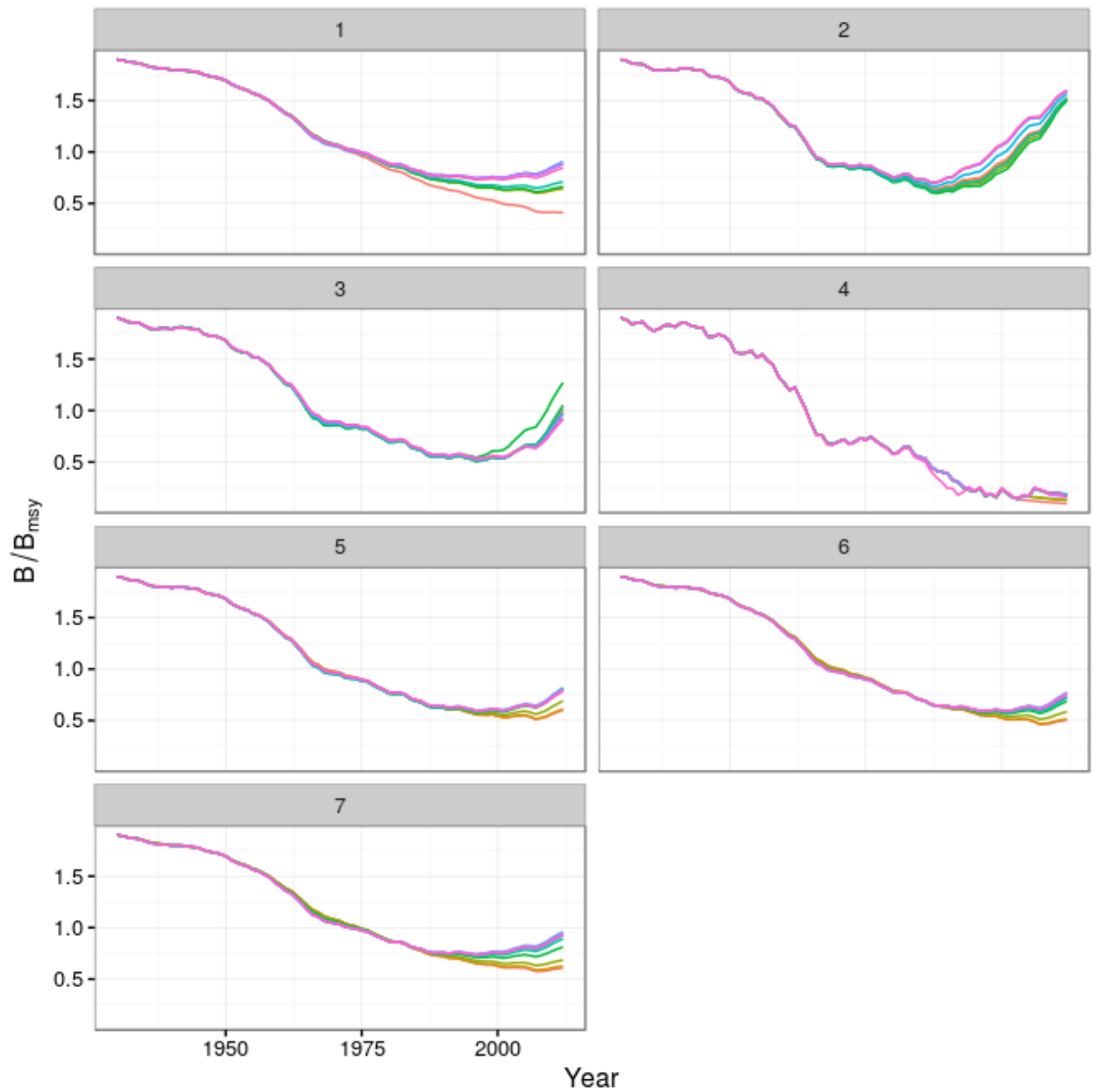


Figure 10. Hindcast

Warning: Removed 9 rows containing missing values (geom_path).

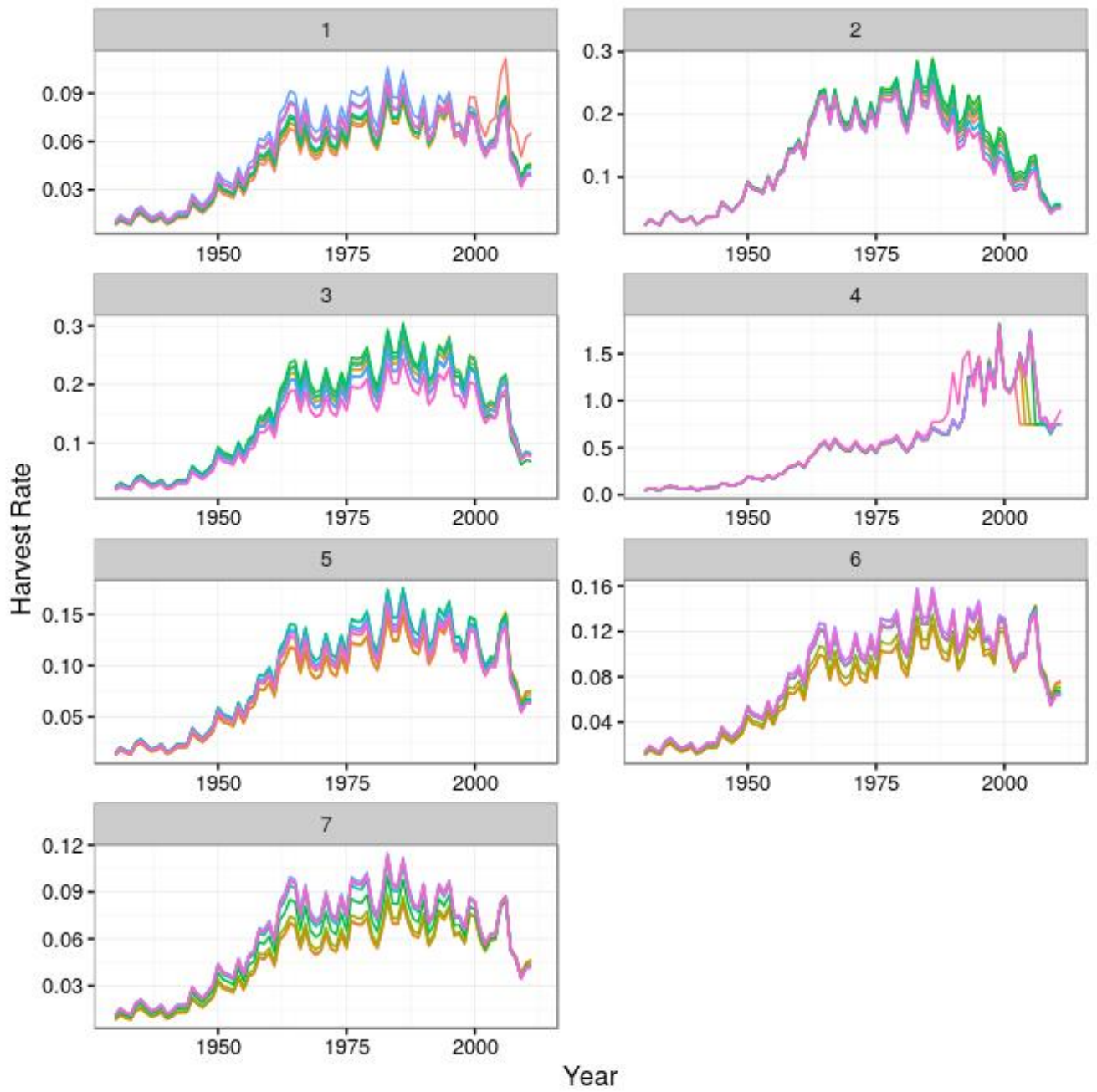


Figure 11. Hindcast
 Warning: Removed 9 rows containing missing values (geom_path).

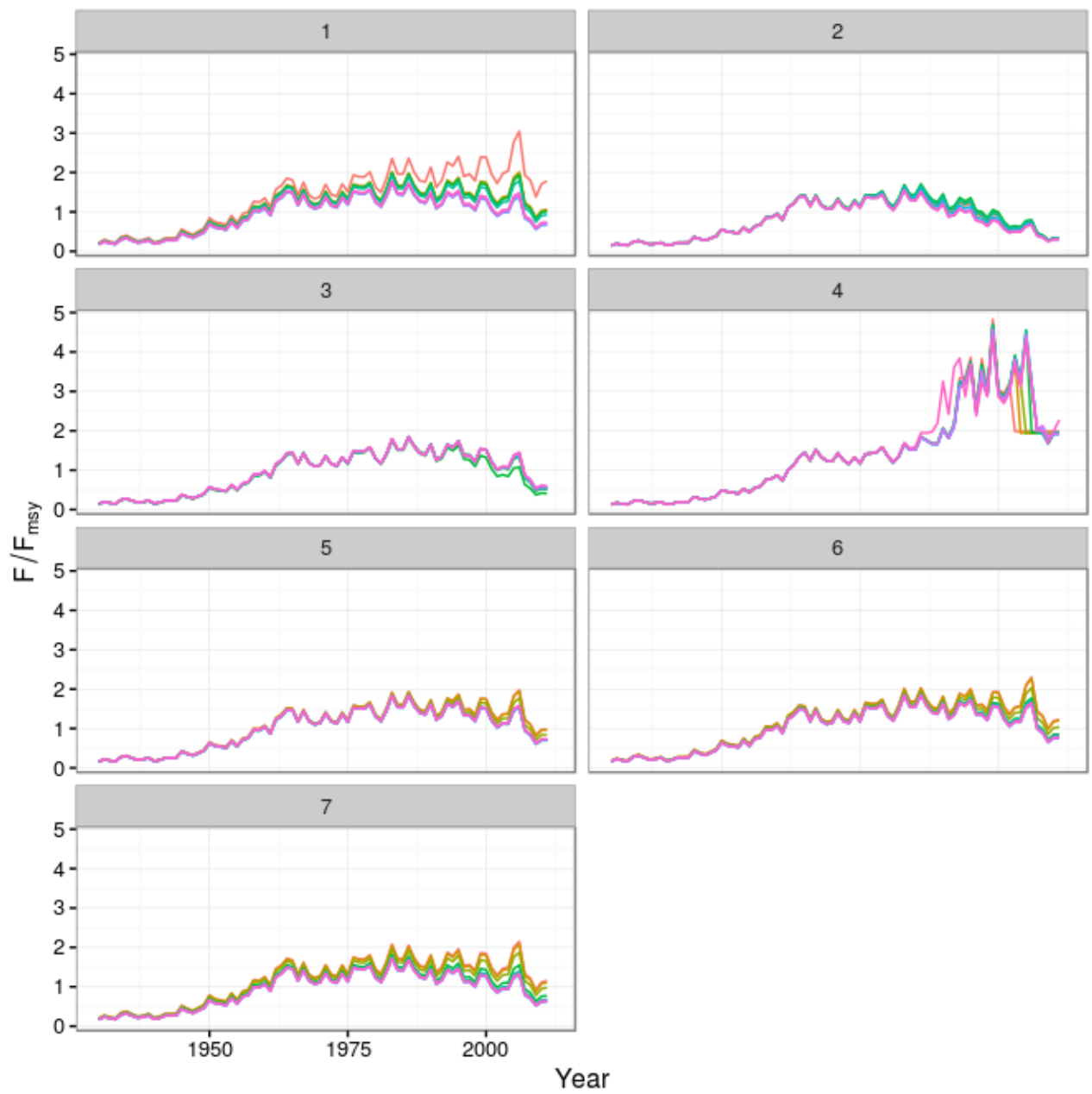


Figure 12. Hindcast

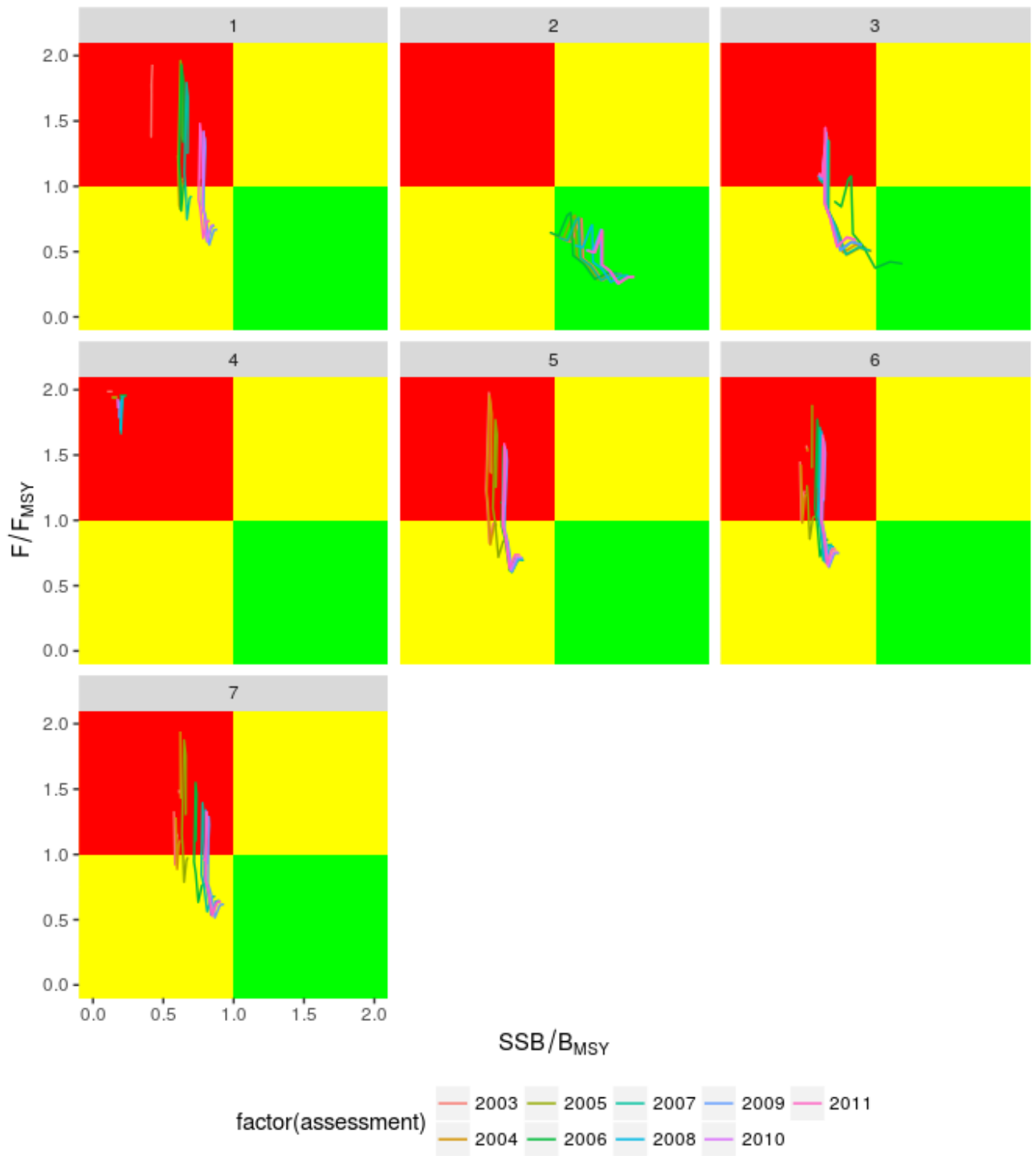


Figure 13. Hindcast

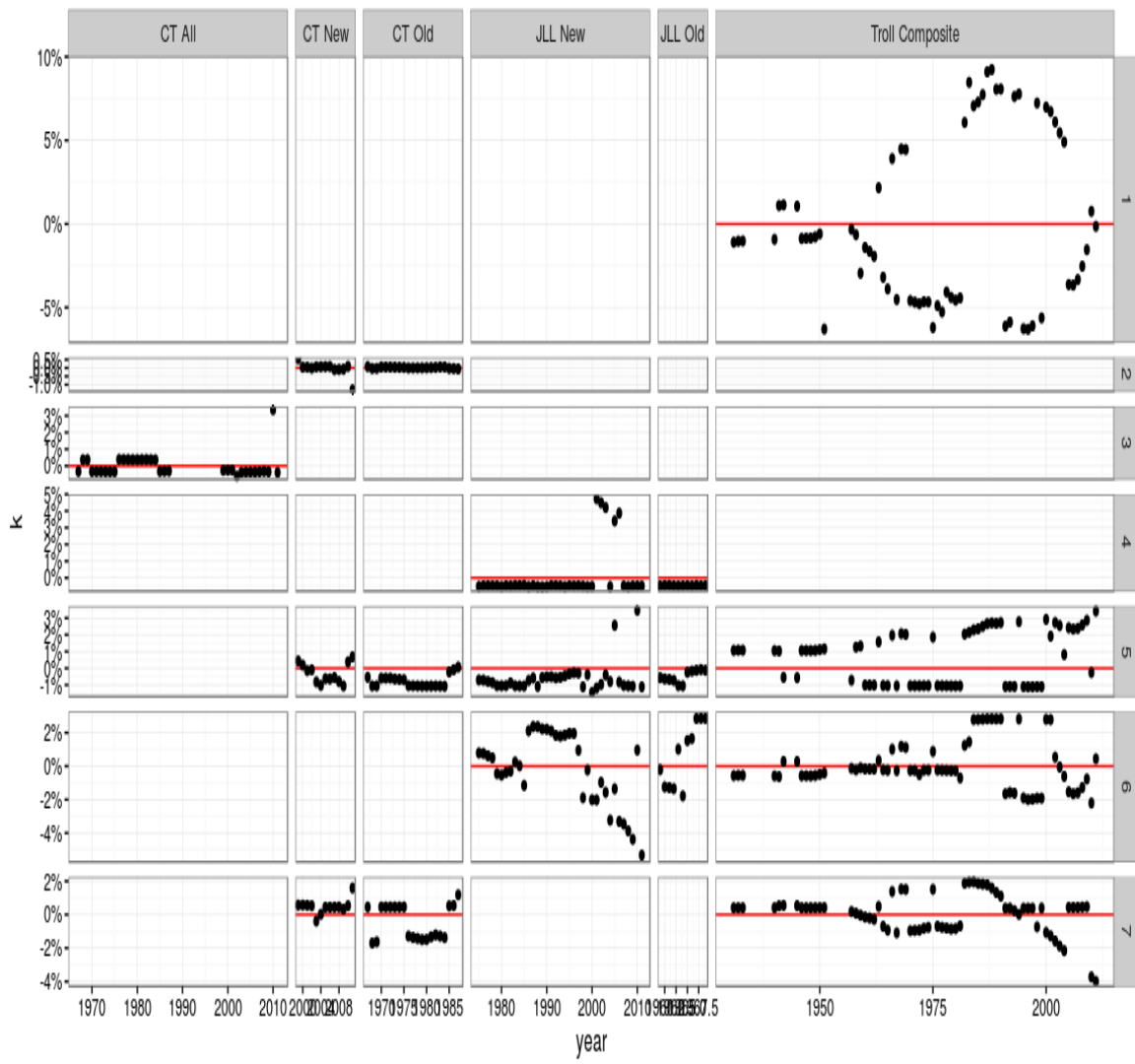


Figure 14. Jackknife K.

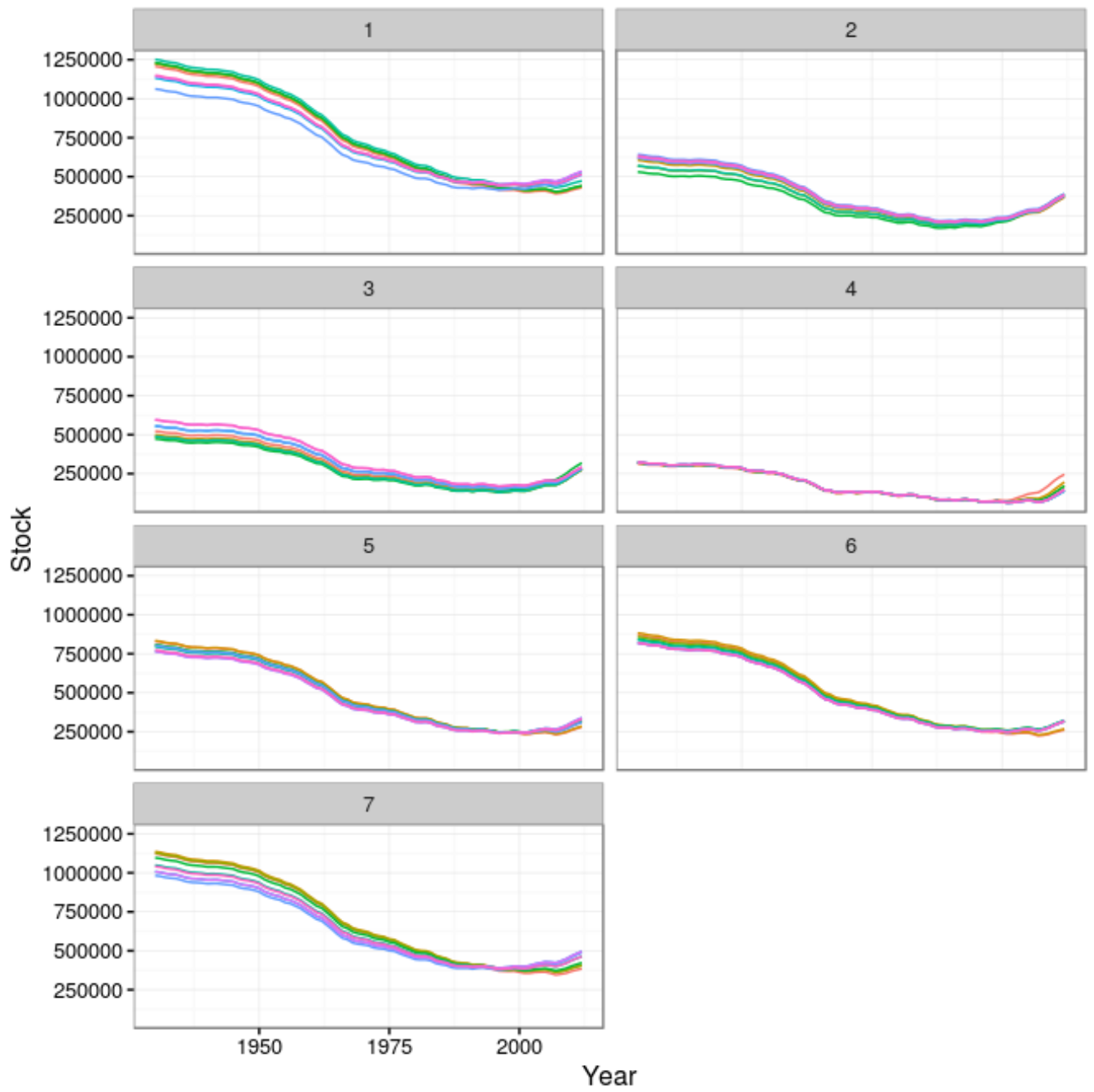


Figure 15. Hindcast

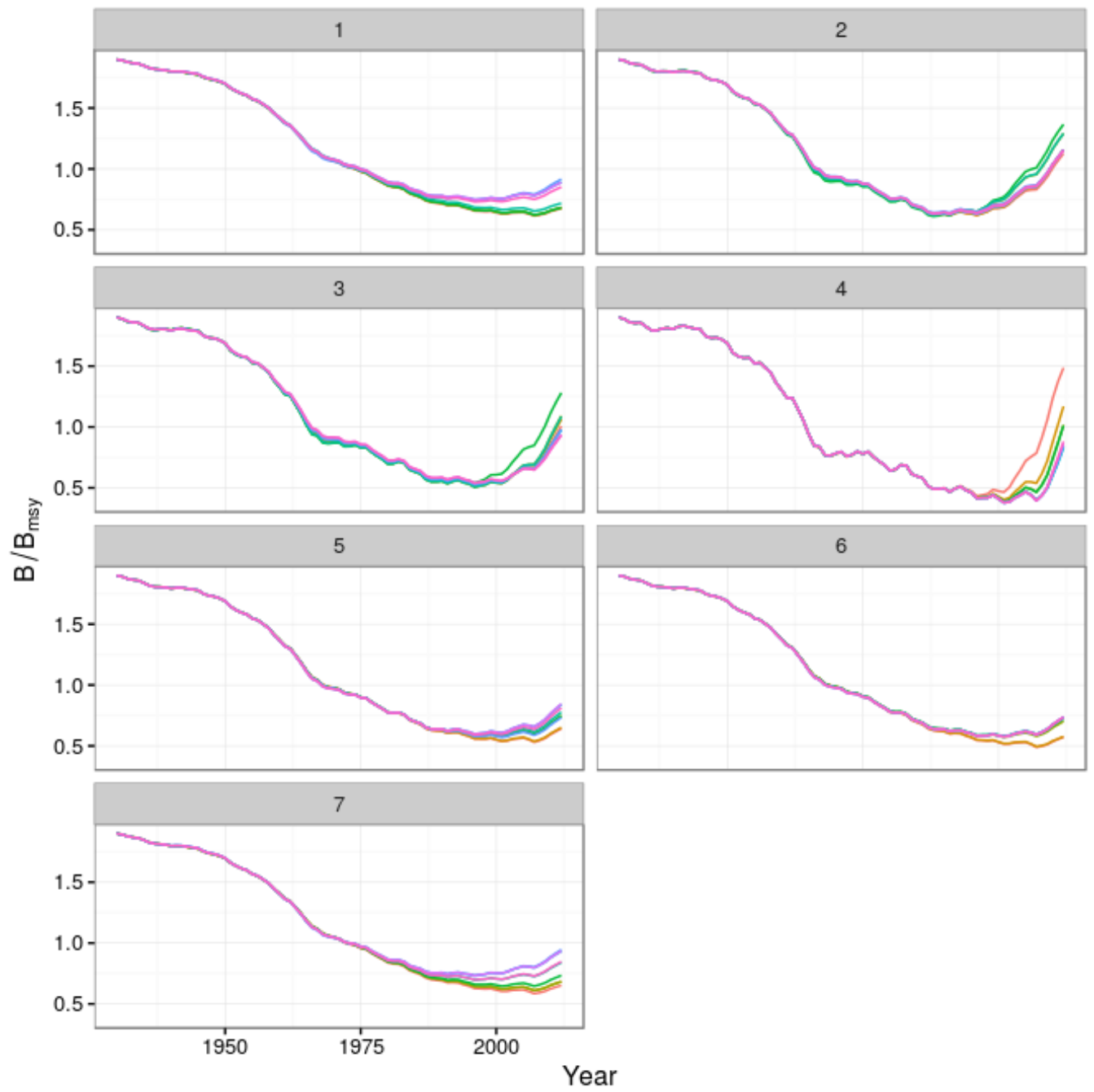


Figure 16. Hindcast
 Warning: Removed 9 rows containing missing values (geom_path).

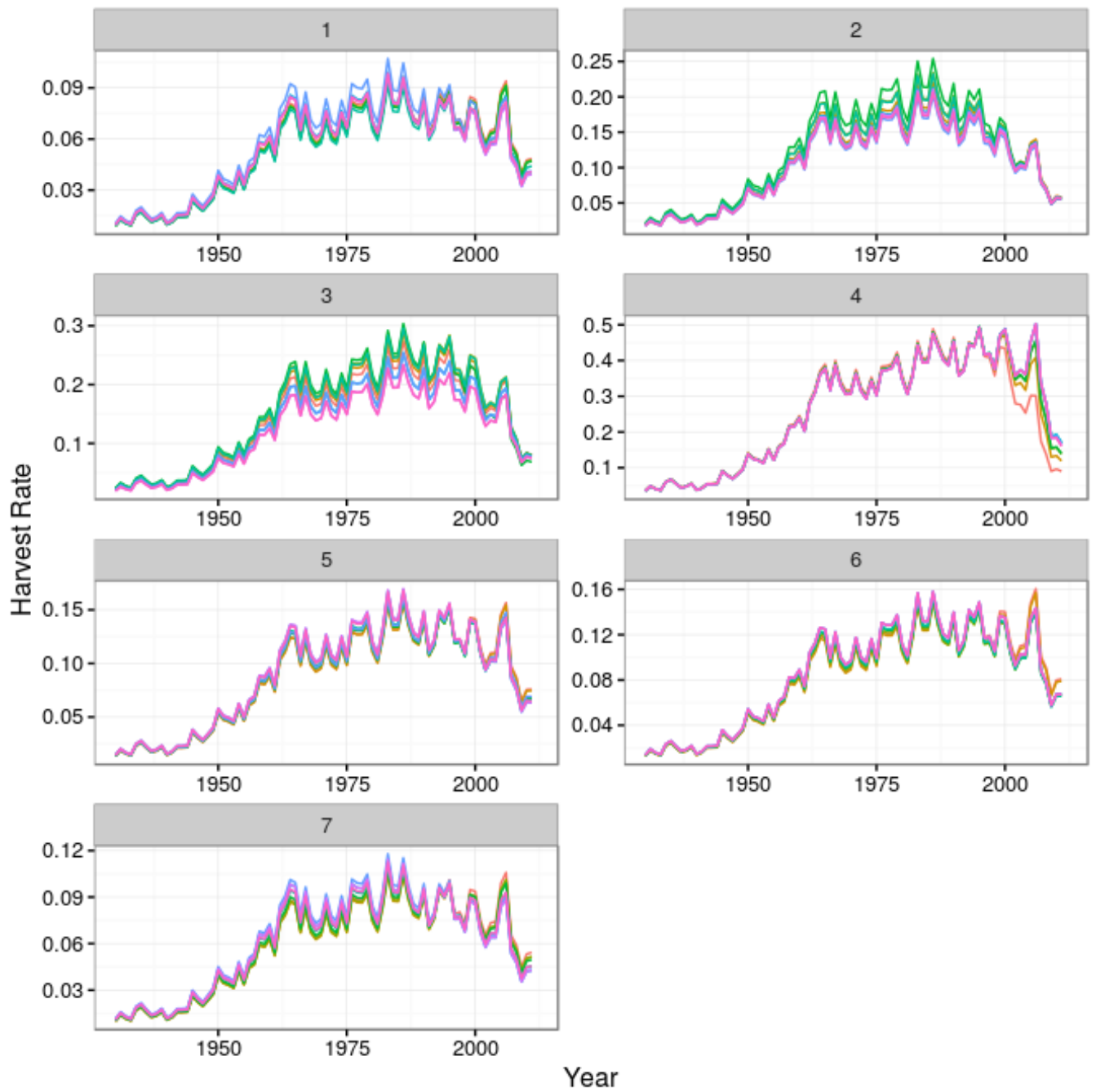


Figure 17. Hindcast
 Warning: Removed 9 rows containing missing values (geom_path).

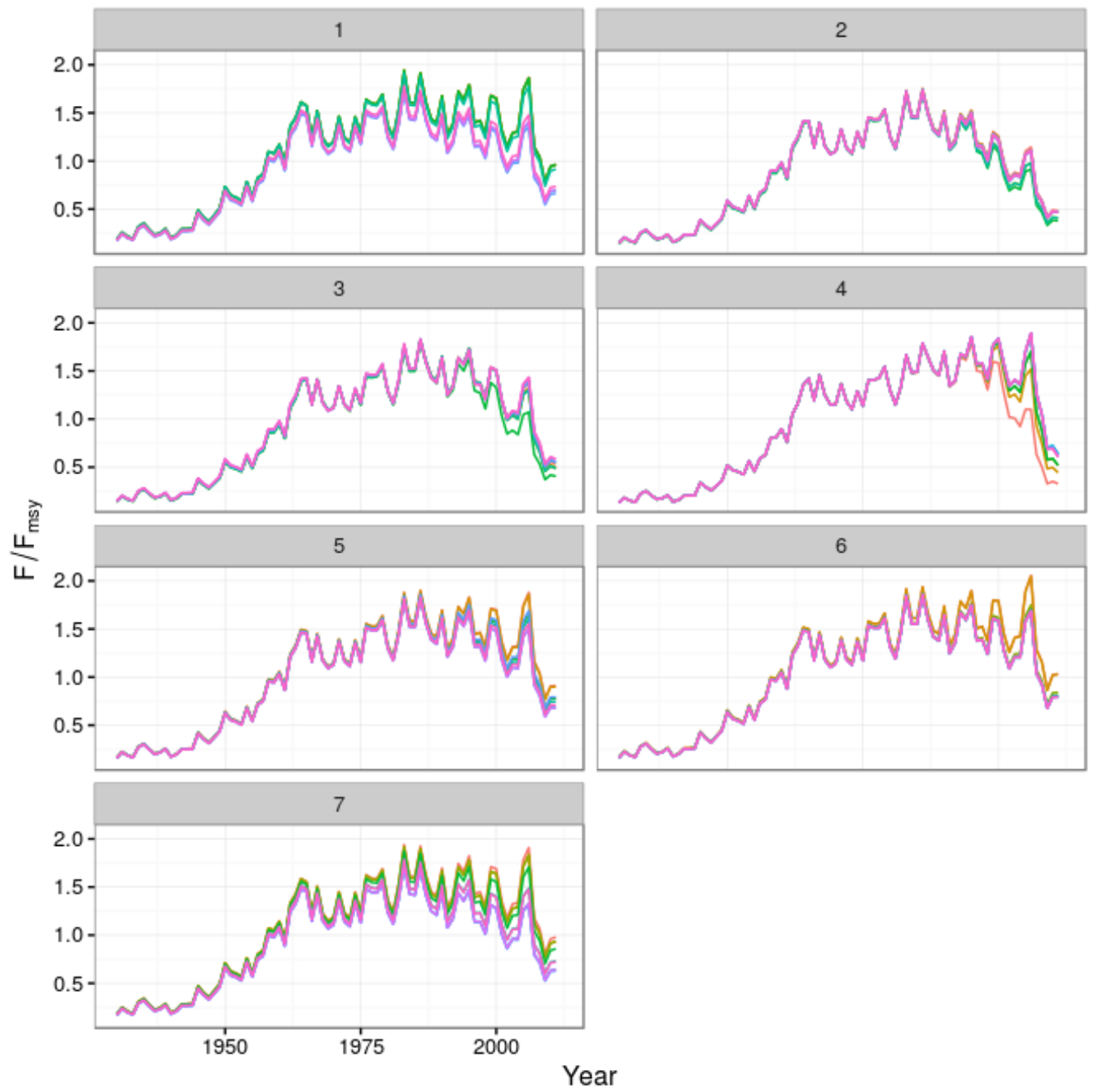


Figure 18. Hindcast

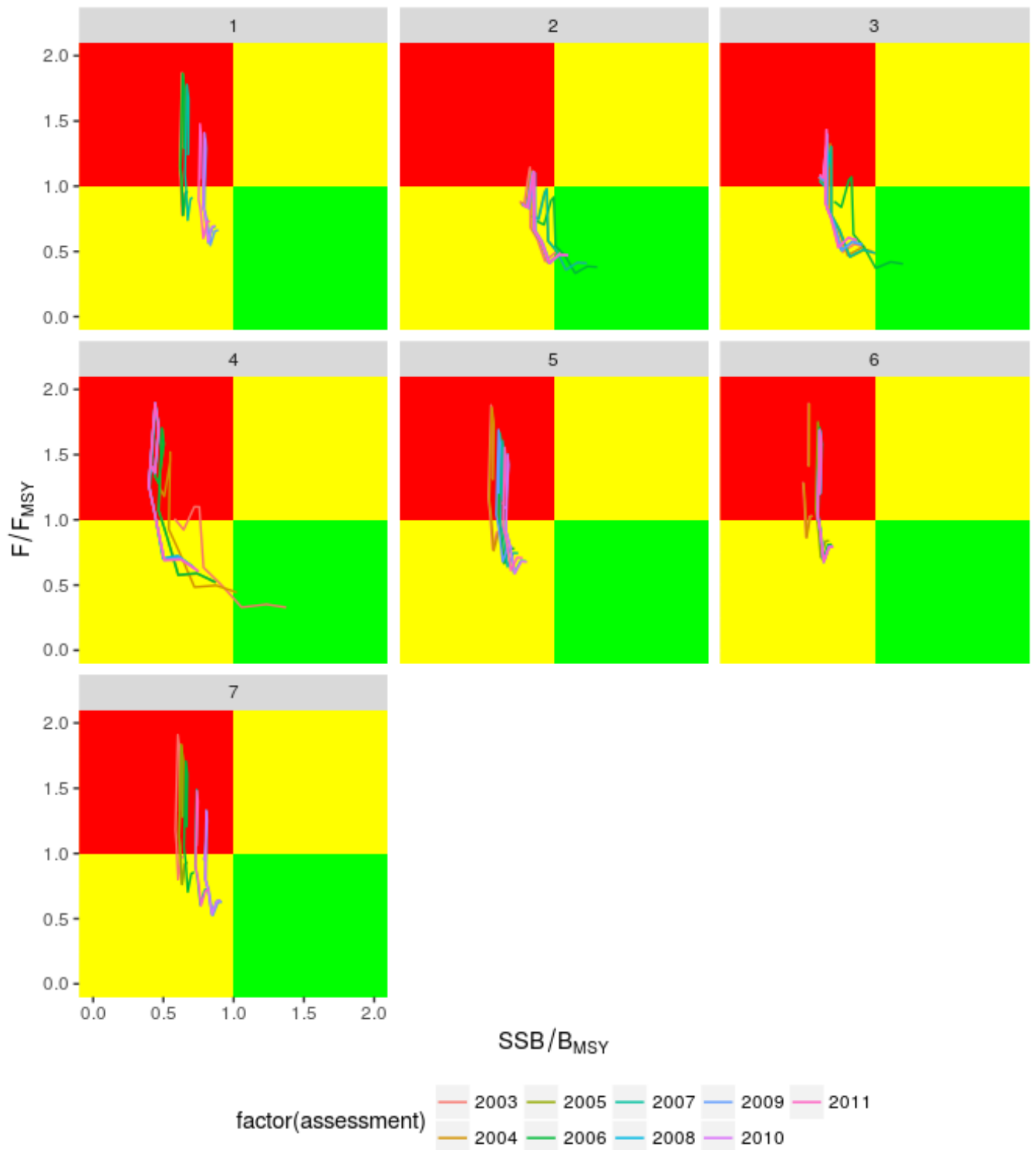


Figure 19. Hindcast