

令和 6(2024)年度 資源評価におけるモデル診断の手順と診断結果の提供指針

2024年6月21日¹

資源評価高度化作業部会

(文責) 八木達紀・濱邊昂平・市野川桃子

モデル診断の概要

我が国の1系資源の評価には主に VPA が用いられているが、近年では多くの資源で複数の資源量指数を導入したチューニング VPA が資源量推定に利用されている。チューニング VPA では、最終年の漁獲係数 (F) の推定方法・資源量と資源量指数との間の非線形関係・複数資源量指数間の分散 (σ^2) の推定の有無・リッジ VPA (最終年の F の推定値をペナルティとして尤度に加えた VPA) 導入の有無など多岐にわたるオプションが利用できる。毎年の資源評価で提案されるモデル (ベースケースモデル) は、これらのオプションを組み合わせたものであるが、ベースケースモデルが統計的に見て妥当なものであるか、また、資源評価結果がモデルの仮定に対してどの程度頑健なものであるかをきちんと説明する (モデル診断) が必要である。

本指針では、特にチューニング VPA で実施可能なモデル診断手法を列挙し、これらを実施するための R の関数 (パッケージ frasyr, <https://github.com/ichimomo/frasyr>) を紹介する。それをふまえ、資源評価で用いられるベースケースモデルについて、資源評価会議時に報告書やドキュメントの形で提供することが求められるモデル診断の内容を示す。VPA の手法によって適用できるモデル診断手法は異なるが、それぞれの手法で実施可能なものについてはできるだけ実施し、情報提供することが望ましい。また、VPA 以外の資源評価手法を導入する場合でもここで示された考え方に基づき同様のモデル診断結果を示す必要がある。

モデル診断の内容と資源評価報告書への掲載の有無

モデル診断結果は定量化できない部分があり、「問題ない」と「問題あり」の閾値を明確に決定することができない。したがって、担当者が「問題ない」と考えても、他の人 (査読者や有識者) は「問題あり」と捉える可能性があり、その逆もあり得る。そこで、「問題あり」の結果のみを示すのではなく、基本的には、計算できる全てのモデル診断の結果はドキュメントとして資源評価前に共有することが望ましい。また、資源評価ピアレビュー委員会によるピアレビューを受ける際にもそのドキュメントを事前にレビューアーに共有しておくとよい。それにより、資源評価結果に対する質問 (たとえば非線形性パラメータ (b) を推定しているが、推定しない場合はどうなるのか? など) に、すぐに回答することができる。あるいは、ドキュメントを配布しておくことでそのような質疑応答を減らし、より本質的な議論をおこなう時間を確保することができる。その上で、資源評価報告書には「一般的に重要なモデル診断結果」と「問題があって今後の課題にしたいなど、評価報告書内で言及される結果」を「補足資料 2 計算方法」における補足図 2-1 に載せる。あらかじめ、さまざまなパターンを網羅したドキュメントを作っておくことにより、報告書に新たなモデル診断結果を急遽載せることになったとしても、その場で報告書の修

¹ 令和 5 年度版からの変更点：編集上の語句の修正

正について議論することができるため、修正した報告書を再度回覧して合意を改めてとるような労力を省くことができる。

表 1 に、ドキュメントとして示すべき内容の例を網羅的に列挙した。その中で、資源評価報告書の補足資料として掲載すべきものは、基本的には、下線をひいたもの、すなわち、残差分析とレトロスペクティブ解析の結果になる。また、上述しているように資源評価報告書内で言及される重要な結果が他にもあれば、それも適宜補足資料に入れ込む。これらのモデル診断の説明と R のコマンドは Appendix のコードで実行ができる。なお、最新版のコードについては問い合わせに応じて共有を行う。

モデル診断手法の詳細

(1) パラメータ推定が適切になされているかの確認：do_estcheck_vpa()

モデル診断においては、推定されたパラメータが適切なものになっているかをまず確認する必要がある。適切なパラメータ推定がなされていると判断するための基準として、ここでは、解が収束している、大域解が得られている、ヘッセ行列の逆行列が正定値であることと考える（図 1）。

frasyr の do_estcheck_vpa 関数はこれらの項目を確認するための関数である。do_estcheck_vpa は初めにヘッセ行列の逆行列が正定値となっているかどうかと、パラメータ推定が収束しているかどうか（局所最適解が得られている）を確認し、これらが満たされていない場合にはその旨を伝えるメッセージとともに関数の実行が止まる。この場合は、初期値を変えてみる、あるいは推定するパラメータ数を減らすなどの工夫をする必要がある。この条件がクリアされると、do_estcheck_vpa はジッター解析を行う。ジッター解析はパラメータ推定をするときの初期値を様々なものに変えて、それぞれの初期値からパラメータ推定を複数回実施する。それにより、推定値が大域解であるかを確認できる (Carvalho et al., 2021, 図 2)。ジッター解析の結果は図 (図 2) とともに出力され、関数に与えた VPA の推定結果の最大対数尤度とジッター解析で得られた最大対数尤度が等しいことを確認する必要があり、ジッター解析完了後に R のコンソール画面に両者の最大対数尤度が表示される。

(2) 残差分析：plot_residual_vpa()

次に、与えられたデータが資源評価モデルによってどの程度説明されているか、また、パラメータ推定の際に与えた仮定にどの程度整合しているかを確認するため、残差分析を行う。チューニング VPA においては、年齢別漁獲尾数に誤差は仮定されていないが、資源量指数には対数正規分布の観測誤差が仮定されている。そのため、チューニング VPA では、資源量指数の対数から資源量指数の予測値を引いた残差の平方和を最小にする最小二乗法によってパラメータ推定がなされている。これは、残差の分布が正規分布に従うと仮定した場合の最尤推定値と同等である。そのため、チューニング VPA における残差は、平均が 0、分散が σ^2 の正規分布に従い、 σ^2 は一定（等分散）、かつ、個々の残差が独立であることが仮定されている。

このような仮定が満たされているかどうかを検討するため、残差を様々な形でプロットしたものを残差プロットと呼び、モデル診断に用いる。例えば、残差を時系列でプロットした残差の時系列プロットにおいて、ある一定の年代で残差が大きく、別の年代で残差が小さくなるような傾向にある場合、等分散の仮定が満たされていないかもしれない。また、過去は残差が負の値を、最近は残差が正の値を取るような

傾向にある場合には、独立の仮定が満たされていないかもしれない。独立の仮定が満たされているかどうかは、例えば、残差の自己相関係数の大きさや有意性を見ることによって数値化できる。

関数 `plot_residual_vpa()` では、各資源量指数の観測誤差の大きさ (σ) と 1 次の自己相関係数 (ρ) が残差プロット中に併記され、自己相関が有意である場合には*がつく (図 3 左上)。ただし、残差の自己相関は、用いられている資源量指数だけにバイアスがあることを示すわけではなく、資源量指数が正しいとしても、モデル内の他の部分での仮定の破れ等によって残差に自己相関傾向が現れていることもある。そのため、残差に自己相関がみられたとしても当該資源量指数をすぐに取り除くのではなく、モデル内の他の仮定の妥当性など、モデル全体の設定を総合的に確認し、見直す必要がある。しかし、VPA というモデルの構造上の制約や知見の不足等により、すぐにこのような問題を解決できないことが多い。

現実的な対処法としては、外れ値の影響を考慮したり、資源量指数が十分に資源動向を反映しているかどうかを確認し、複数の資源量指数がある場合には重みづけの方法の変更や、資源量指数と資源量の非線形的な関係を考慮、漁獲能率 (q) が年代によって変化している可能性などを検討する。その上で、まだ残差に自己相関が残っているようなモデルを用いる場合には、残差の自己相関が有意であることを注釈に加えつつ、当該資源量指数を除いた時の感度分析やレトロスペクティブ解析などの結果とともに、その推定結果を用いることもある。その場合は、残差に自己相関が見られる理由 (例えば、資源量指数を参照する資源年齢が異なるから等) について考察の上、本文中に記載する必要がある。残差に自己相関が見られる場合には、レトロスペクティブパターンが出やすいため、レトロスペクティブ解析の結果とも併せてモデルの使用の判断を行うことが望ましい。

`plot_residual_vpa()` の残差の時系列プロットでは観測値の 1.28σ (80%) 区間と 1.96σ (95%) 区間も同時に出力される。 1.96σ 区間は観測値の 95% がこの区間内に入ると考えられる範囲であるため、この範囲から逸脱しているデータはパラメータ推定への影響力が大きすぎる可能性があり、そのデータの影響力の大きさや外れ値である可能性について慎重に検討する必要がある。

また、モデルが資源量指数をどの程度説明できているのかを判断するために、資源量指数とモデルの予測値の両方を時系列にプロットした図 (図 3 右上) も `plot_residual_vpa()` から出力される。資源量指数がモデルによって十分に説明されていない場合、海外の資源評価ではリジェクトされる事例もある (Punt et al., 2020)。

資源量と資源量指数には $CPUE=qB$ という比例関係が仮定されている。これは資源量が大きい場合、資源量指数も大きくなることを意味する。しかし、資源量指数によっては線形関係が認められない場合があり、非線形性のパラメータ (b , $CPUE=qB^b$) を推定することでモデルと資源量指数のフィットが向上する可能性がある (Hashimoto et al., 2018)。 b が 1 より小さく、資源量の減少に対して資源量指数の減少が鈍感で、ある枯渇レベルに達すると資源量指数が急減する現象を *hyperstability* と呼ぶ。一方、 b が 1 より大きく、資源量の減少に対して、資源量指数がそれ以上に大きく減少する現象を *hyperdepletion* と呼ぶ。一般に、本パラメータは推定が難しいことが知られているため、本パラメータを推定した場合は、資源量と資源量指数の関係のプロット (図 3 左下) を通して推定結果が妥当であるか確認する必要がある。

(3) レトロスペクティブ解析：do_retrospective_vpa()

レトロスペクティブ解析は、推定の最終年から1年ずつデータを取り除いて資源評価モデルによる推定を繰り返し行う解析手法である。1年ずつデータを取り除いたモデル（レトロモデル）の推定値と、得られている全ての時系列データを用いたモデル（フルモデル）の結果の推定値を比較することで、新しいデータが追加されたときの資源量推定値の修正度合いが確認できる（図4）。比較のために用いる資源量推定値としては資源量・漁獲係数・加入量・産卵親魚量などが目的に応じて用いられる。フルモデルからの推定値に対して、各レトロモデルからの推定値が大きく異なっている場合、毎年の資源評価の頑健性が低いことが示唆される。また、フルモデルからの推定値に対して、レトロモデルからの推定値が毎年同じ傾向を持ってずれているような現象を、レトロスペクティブパターンと呼ぶ。このレトロスペクティブパターンが大きい場合は、フルモデルからの推定結果がバイアスしている可能性が高い。

レトロスペクティブパターンの強さを表す指標として Mohn' s ρ という尺度がある (Mohn, 1999) :

$$\rho = \frac{1}{N} \sum_{n=1}^N \frac{\theta_{R,T-n} - \theta_{F,T-n}}{\theta_{F,T-n}}$$

ここで n はレトロスペクティブ解析で遡った年数（最大 N 年まで遡る）、 T はフルモデルの推定最終年、 $\theta_{R,T-n}$ はレトロモデルの推定最終年 ($T-n$ 年) の推定値、 $\theta_{F,T-n}$ はフルモデルの $T-n$ 年の推定値をそれぞれ表す。レトロスペクティブパターンの強さを判断するうえで Mohn' s ρ の大きさの絶対的な指標は存在しないが、経験則的な事例はまとめられている。Hurtado-Ferro et al. (2015) ではレトロスペクティブパターンが問題視される Mohn' s ρ の経験則について、寿命の長い種では -0.15 以下または 0.2 以上、短い種では -0.22 以下または 0.30 以上としている。ただし、この値は長寿命種は yellowtail flounder、短寿命種は Pacific sardine の生物学的パラメータを基に、シミュレーションによって算出されたものであり、魚種が異なる場合はこの限りではない点に注意が必要である。レトロスペクティブパターンが認められた場合、これらの値を参考に許容範囲内であるかどうかを議論する必要がある。

また、レトロスペクティブパターンが大きく表れる一つの要因は、モデルに与えられた複数のデータがそれぞれ異なる情報を持つ場合であることが知られている。チューニング VPA の場合には、年齢別漁獲尾数と資源量指数の2種類の質の異なるデータが与えられているが、それぞれが資源状態に対して異なる情報を持つ場合、資源量指数を用いないチューニングなしの VPA ではレトロスペクティブパターンが出ないのに、チューニング VPA でレトロスペクティブパターンが大きく表れることがある。特にチューニング VPA はモデルの構造上、近年の資源量推定値は資源量指数に大きく影響を受ける一方で、過去年にさかのぼるほど資源量推定値は年齢別漁獲尾数によって決定論的に決まるため、資源量指数と年齢別漁獲尾数がそれぞれ矛盾した情報を持つ場合にレトロスペクティブパターンが出やすくなると考えられる。このようなケースでは、資源量指数がバイアスしている可能性だけでなく、年齢別漁獲尾数が何らかの理由でバイアスしている可能性も排除できない。そのため、一概にレトロスペクティブパターンだけを理由に用いる資源量指数を選択したり、資源量指数を用いなかったりすることは、資源量指数が正しくて、年齢別漁獲尾数がバイアスしている、ひいては、資源量推定が潜在的にバイアスしている可能性を無視していることになる。したがって、Mohn' s ρ だけでなく、データの特性や管理に与える悪影響などを総合的に加味した慎重な判断が必要である。レトロスペクティブパターンの大きいチューニング VPA の対応の一例として、リッジ VPA (Okamura et al., 2017) を導入することで、レトロスペクティブ

パターンを減らすことが可能と考えられる。

(4) 感度分析 : `do_sensitive_vpa()`

感度分析は、資源評価モデルにおける様々な仮定の変化が、推定資源量などにどのような影響を与えるかを確認するための診断である。VPA においては、生物パラメータ（年齢別成熟率や年齢別体重、自然死亡係数など）や最高齢と最高齢-1 歳の漁獲係数の比 (α)、最終年の F の仮定、チューニング指標値間の重み付けなど、様々な仮定が置かれている。これらの仮定を変えて資源量推定を実施した場合、推定結果が変わることが予想されるが、それが意図した程度の変化量なのか、資源評価におけるその仮定の影響について確認する（図 5）必要がある。特に、資源特有の仮定や資源評価モデルのパラメータを大きく変更する場合には、変更前の設定を用いた感度分析の結果を示すことが望ましい。

`frasyr` では `do_sensitive_vpa` 関数で実行可能だが、感度分析をしたいパラメータとその値の設定については、パラメータに応じて引数の指定方法が異なる。詳細な引数の設定方法については、Appendix (1. 感度分析(チューニングあり・なし)) を確認されたい。

(5) ジャックナイフ法 : `do_jackknife_vpa()`

一般的なジャックナイフ法は、推定に用いたデータを 1 つずつ取り除いてパラメータ推定を繰り返すことである。それにより、各データがパラメータ推定に与える影響の強さを調べることができる。`frasyr` の `do_jackknife_vpa` では、複数の資源量指数を用いている場合、それぞれの資源量指数の影響を評価するために、資源量指数の系列を 1 つずつ取り除いて資源量推定を行うことを繰り返す（図 6）。複数の資源量指数がありそれぞれの傾向が異なる場合に、どのデータが推定にどのような影響を与えているか確認したい際に有用である。特に、ジャックナイフ法で推定値に対する影響力が大きいと判断された資源量指数の傾向が、その他の資源量指数の傾向と全く異なる場合には、特に影響力が強い資源量指数の信頼性を確認し、必要があれば、データ間の重みづけの方法を変更するなどの手法を通して、代表性の高い資源量指数の傾向を推定結果に反映させるような検討もおこなう。

また、単一の系列の資源量指数を用いていた場合にも、極端な資源量指数を示す年を除いて資源量推定を行うことで、そのような値に対して資源量推定の感度が高いかを確認することができる。

(6) ブートストラップ法 : `plot_resboot_vpa()`

ブートストラップ法はパラメータの信頼区間を推定するために用いる方法である。データ、ここでは資源量指数の疑似データを生成し資源量推定を繰り返すことで、パーセンタイル点としてブートストラップ信頼区間を計算することができる（図 7 左）。`frasyr` で提供されている `plot_resboot_vpa` では 2 通りの疑似データの生成方法が実装されている。ノンパラメトリックブートストラップ法では、資源量指数とモデルから計算された予測資源量指数の残差をリサンプリングする手法である。一方、パラメトリックブートストラップでは対数を取った予測資源量指数を平均に、推定された観測誤差を標準偏差に持つ正規分布から、疑似データを乱数生成する。R 関数 `plot_resboot_vpa()` ではパラメトリックブートストラップ法をデフォルトとしている。

この他に、ブートストラップサンプル分の観測誤差や年齢別漁獲係数、加入量や産卵親魚量などの推定

値が得られる。これらをプロットすることで、パラメータ間の相関の強さを把握することができる（図7右）。パラメータ間の相関が強い場合、パラメータ間で影響しあうことで推定が難しい場合がある。

引用文献

- Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirrip, M., Kitakado, T., Yamane, 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 assessment. *Fisheries Research* 240, 105959
- Hashimoto, M., Okamura, H., Ichinokawa, M., Hiramatsu, K., Yamakawa, T. (2018). Impacts of the nonlinear relationship between abundance and its index in a tuned virtual population analysis. *Fish Sci* 84 (2), 335-347
- 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 Journal of Marine Science* 72, 99–110.
- Mohn, R., (1999). The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. *ICES Journal of Marine Science* 56, 473–488.
- Okamura, H., Yamashita, Y., Ichinokawa, M. (2017). Ridge virtual population analysis to reduce the instability of fishing mortalities in the terminal year. *ICES Journal of Marine Science* 74 (9), 2427-2436
- Punt, A. E., Tuck, G. N., Day, J., Canales, C. M., Cope, J. M., de Moor, C. L., De Oliveira, J. A.A., Dickey-Collas, M., Elvarsson, B., Haltuch, M. A., Hamel, O. S., Hicks, A. C., Legault, C. M., Lynch, P. D., Wilberg M. J. (2020). When are model-based stock assessment rejected for use in management and what happens then? *Fisheries Research* 224, 105465.

表1. ドキュメントとして資源評価報告書とともに提出することが推奨されるモデル診断の候補の例。状況に応じて示すべきもので、すべてを示さないといけないわけではないが、下線を引いた項目は資源評価報告書の補足資料2の図2-1として基本的には示すようにする（マアジ太平洋系群の例：https://abchan.fra.go.jp/wpt/wp-content/uploads/2024/03/details_2023_03.pdf）。

1. 収束の確認とジッター解析 (jitter analysis)
 - ・失敗している場合にはパラメータ推定がうまくいっていないことを示すため、これについては必ず成功している結果を載せることになる（特に問題がなければ載せなくて良いが、確認して問題なかったことは記述しておく）
2. 残差プロット
3. レトロスペクティブ解析
 - ・基本的に本文に載せることとする
 - ・特にレトロスペクティブパターンが問題になる場合や、それによってリッジ VPA などを導入した場合
4. 生物パラメータやチューニング方法についての感度分析
 - ・系群特有のパラメータや仮定に対して感度分析したい場合
 - ・チューニング方法などで特に昨年度から手法を変えた場合、昨年度の手法を用いたときの結果と比較する
 - ・自然死亡係数 (M) や最高齢と最高齢-1 歳の漁獲係数の比 (α) など、明確な根拠はないものの毎年利用している値について他の候補の値が指摘された場合、管理基準値を含めた資源量推定結果に大きな影響を及ぼしそうな場合
 - ・リッジ VPA を用いている場合にはリッジ VPA を用いない場合、資源量指数 の非線形性パラメータ (b) を推定している場合にはbを推定しない場合など、より単純な仮定をおいたときの結果と比較（リッジ VPA やb推定の効果をすぐに確認することができるため）
5. ジャックナイフ
 - ・複数の資源量指数の中で傾向の異なる資源量指数がある場合
 - ・残差分析などから外れ値の多い資源量指数がある場合
6. ブートストラップ信頼区間
 - ・資源量推定の不確実性を評価したい（信頼区間を確認したい）場合

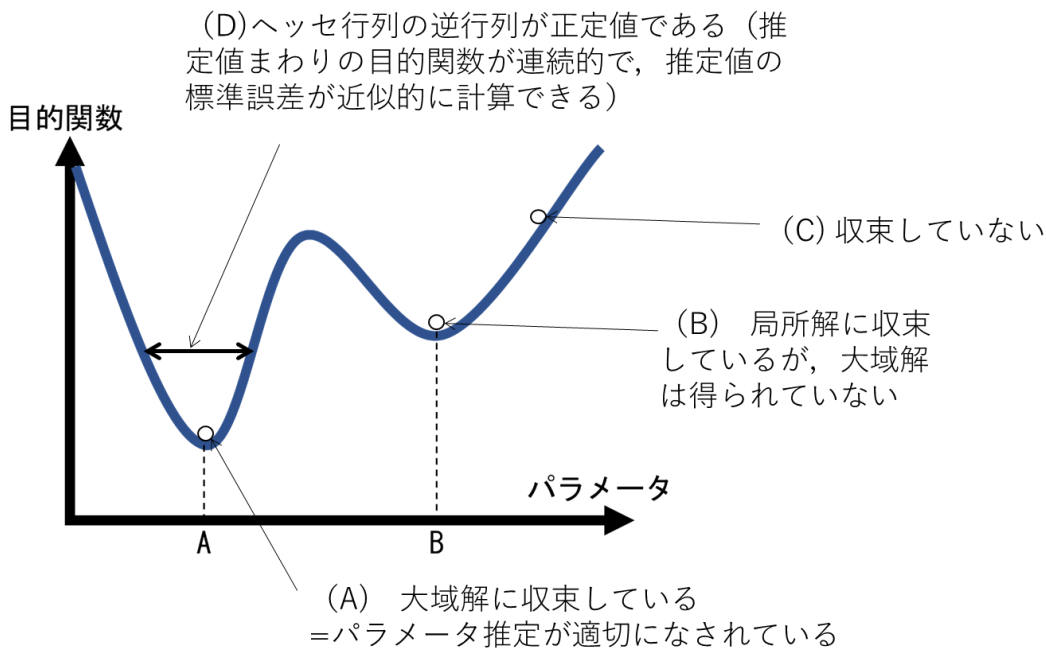


図1. 最適化を用いたパラメータ推定概念. ここで縦軸の目的関数は、最小二乗法の場合には残差平方和、最尤法の場合には負の尤度となり、目的関数が最も小さくなる時のパラメータの値（図のA）がいわゆる「推定値」（最尤推定値）となる. 適切な推定値が得られている場合というのは、解が収束していること（目的関数の微分がゼロになるような値に解が推定されているか）だけでなく、現実的なパラメータの範囲内で最も目的関数の値が小さくなるようなところの解が得られている（大域解が得られている）必要がある. 例えば、図のBは、解が収束はしているが局所解に陥ってしまっており、大域解が得られていない. また、図のCの状態は、解が収束もしていない. さらに、実用上は、大域解を得た上で、ヘッセ行列の逆行列が正定値であること（図のD）も適切なパラメータ推定の条件として求められる.

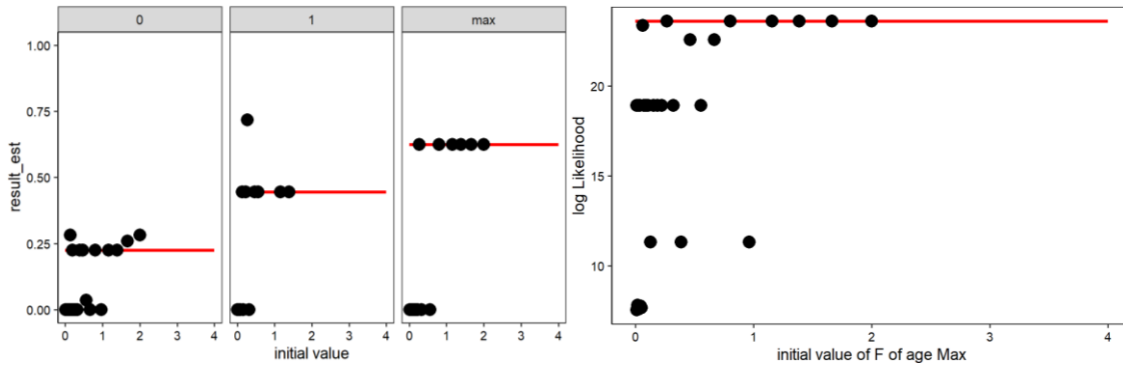


図 2. frasyr でジッター解析 (do_estcheck_vpa) を実行した時の作図. 左図は初期値 (横軸) を変えてパラメータ推定した場合のパラメータ推定値 (縦軸) を表す. 赤線はデフォルトの初期値を用いた場合のパラメータ推定値. ここでは 0 に近い初期値を用いると推定値は極端に小さな値をとる傾向がみられる. 右図は最終年最高齢の F の初期値 (横軸) と対数尤度 (縦軸) を示し, 左図右側の “max” の横軸と対応する. この結果から, 0.5~2 までの間の初期値を取れば, 多くの場合で最尤推定値を得られていることが確認でき, 考えうる初期値の範囲でデフォルトの初期値から推定された対数尤度 (右図赤線) よりも大きい結果がないことから, デフォルトの初期値の結果が大域解であることが確認できる.

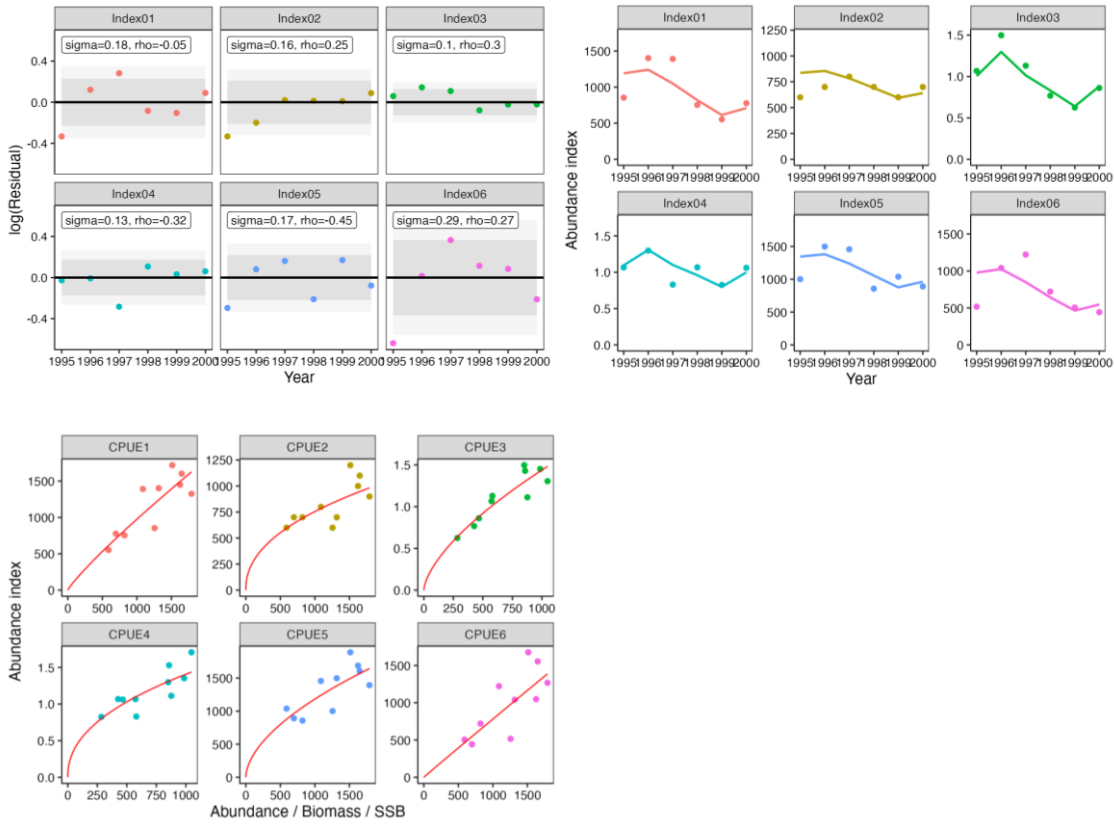


図 3. frasyr で残差解析 (plot_residual_vpa) を実行した時の作図結果. 残差の時系列プロット (左上図), 資源量指数の予測値と観測値の時系列プロット (右上図) と資源量指数に対する推定資源量指数のプロットの図 (左下図). 左上図において, sigma は観測誤差, rho は残差の自己相関係数, 薄い灰色は 1.96σ 区間, 濃い灰色は 1.28σ 区間を表す.

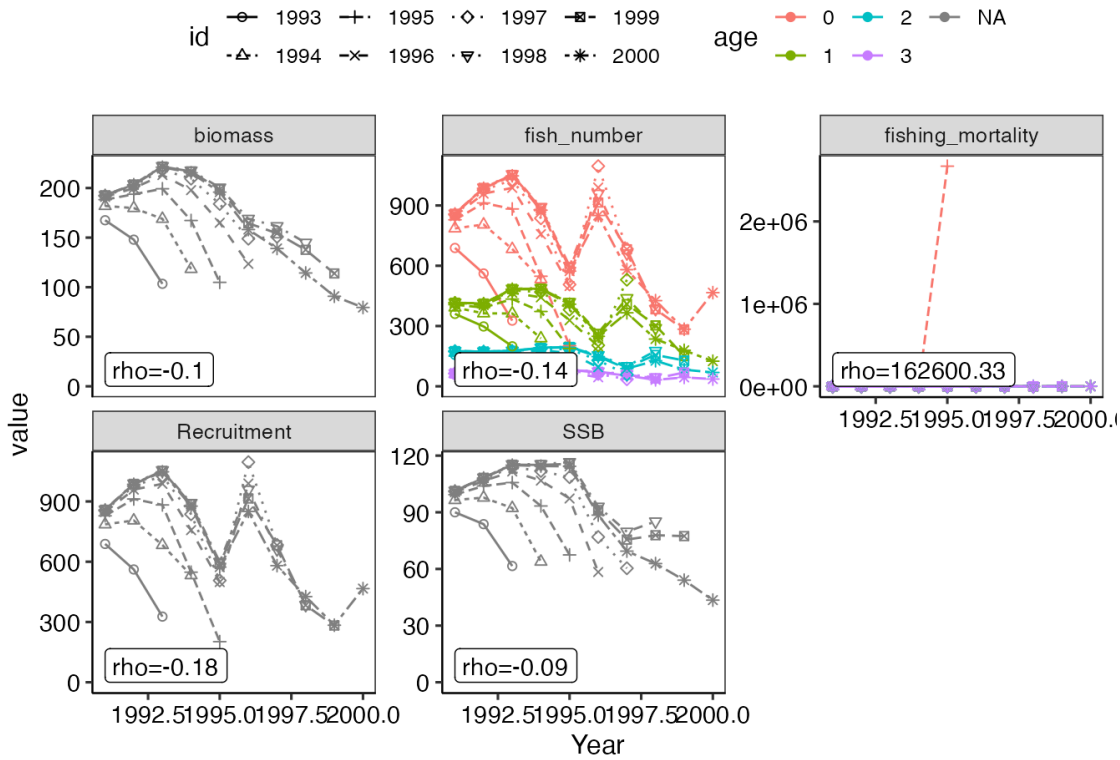


図 4. frasyr でレトロスペクティブ解析 (do_retrospective_vpa) を実行した時の作図. 各パラメータ (biomass:資源量, fish_number:年齢別資源尾数, fishing_mortality:年齢別漁獲係数, Recruitment:加入尾数, SSB:産卵親魚重量) のフルモデルとレトロモデルの動態, および Mohn' s ρ (rho) を表す.

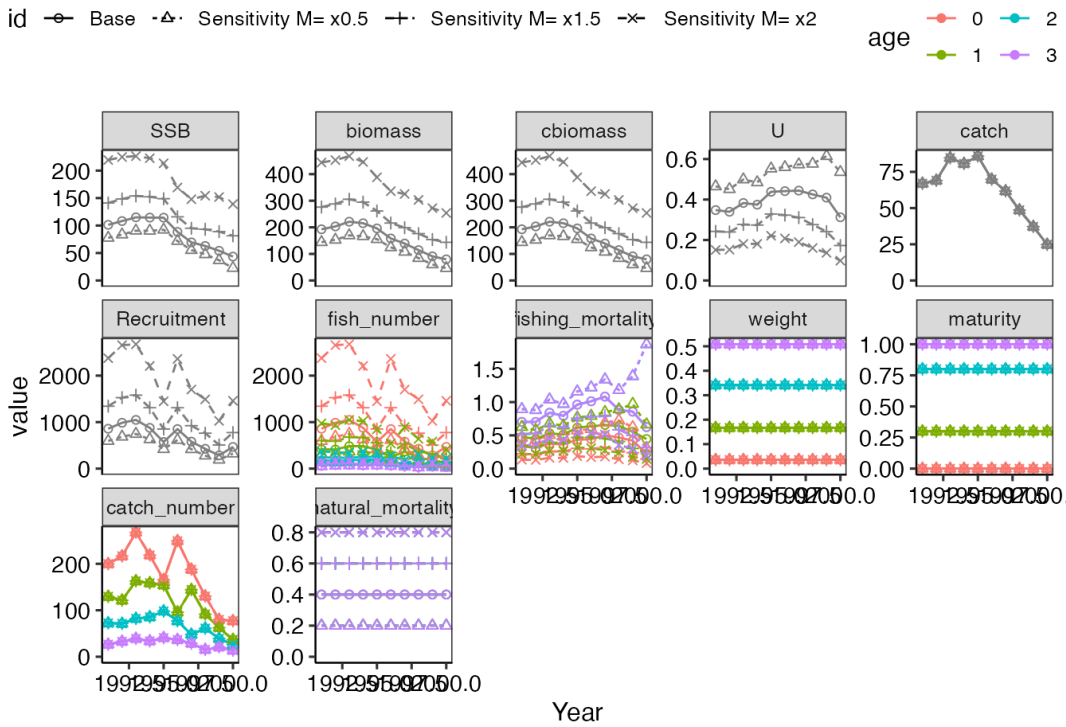


図 5. frasyr で自然死亡係数 (M) についての感度分析 (do_sensitive_vpa) を実行した時の作図. 各パラメータ (cbiomass: 漁獲資源量, U: 漁獲割合, catch: 漁獲量, weight: 年齢別体重, maturity: 年齢別成熟率, catch_number: 年齢別漁獲量, natural_mortality: 年齢別自然死亡係数) の動態結果を表す.

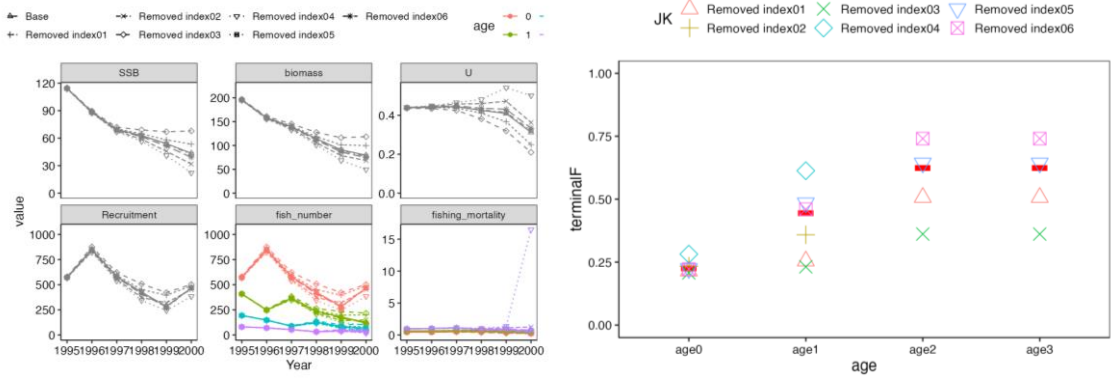


図 6. frasyr でジャックナイフ法 (do_jackknife_vpa) を実行した時の作図. 左図は資源量指数の種類を 1 つずつ取り除いた場合の各パラメータの動態結果を表す. 右図は各年齢の最終年の漁獲係数の推定値を表す. 左図の漁獲係数を見ると Index4 を取り除くと極端な推定値を取っていることから, Index4 の推定への貢献度が高いことが示唆される.

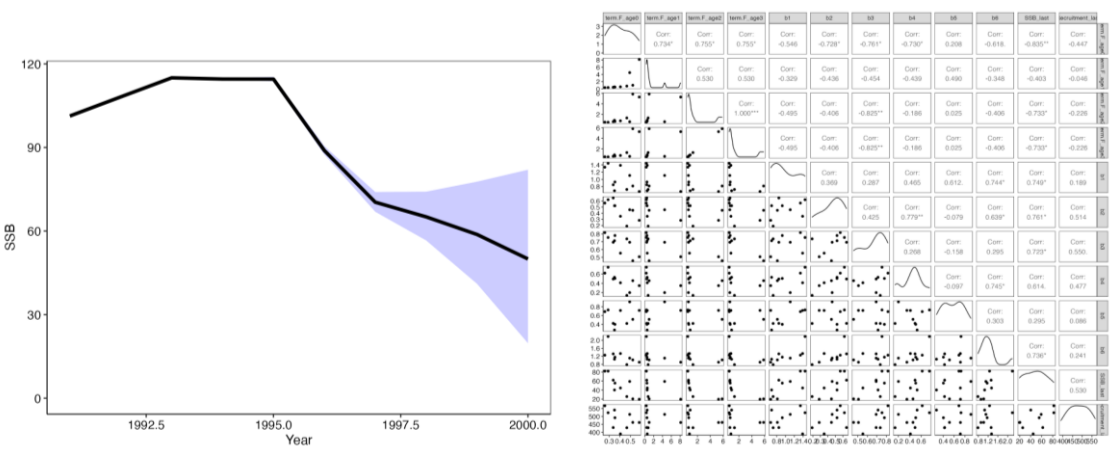


図 7. frasyr でブートストラップ法 (plot_resboot_vpa) を実行した時の作図. 左図は SSB の推定値 (黒実線) と 95%信頼区間 (青色) を表す. 右図は各推定パラメータの関係をプロットしたものである.