

統計モデルとデータマイニング手法の水産資源解析への応用

庄野 宏 (遠洋水産研究所)

本論文では、水産資源解析学における様々な問題、特に魚の資源密度に対応し、相対的な資源量を表す CPUE (catch per unit effort: 単位努力当たり漁獲量) の解析に関する様々な問題について、遠洋域に生息するまぐろ類・関連種の漁業データや計算機によるシミュレーション実験を利用し、統計モデルおよびデータマイニング的なアプローチにより問題解決するための手法を提案した。CPUE は漁獲量を投下した努力量で割ることによって定義される、漁獲効率を表し、資源密度に比例することから相対資源量に対応する重要な概念である。しかし、漁船などの加工されていない CPUE は、季節・海区・漁具など資源密度以外の様々な時空間的な要因や環境要因などを含んでおり、資源の年変動を知るためにはこれらの影響を取り除く必要がある。そこで、CPUE の自然対数を応答変数に設定し、正規誤差の下で考えられる要因効果を説明変数に組み込んだ共分散分析モデル (CPUE Log-Normal モデル) や、離散変数である Catch を応答変数と設定し、Poisson 分布や負の二項分布などを仮定した一般化線形モデル (Catch Poisson モデル、Catch Negative-Binomial モデルなど) を用いて年の要因効果を推定することが伝統的に行われてきた。この作業を CPUE 標準化と呼び、統計モデルに加えて近年では樹形モデルやニューラルネットワーク等のデータマイニング的なアプローチも用いられるようになってきている。本研究では、この水産資源解析における主要な問題である CPUE 標準化を論文のメインテーマとし、以下の3つの課題について取り上げて詳しく検討した。

- 1) CPUE 標準化を想定した分散分析型モデルにおける、様々な情報量規準や stepwise 検定を通じた要因効果の取捨選択、モデルの性能評価 (第3章)
- 2) ニューラルネットワークによるミナミマグロの操業がない時空間の CPUE 予測および簡便な要因分析法 (CPUE 年トレンド抽出法) の提案 (第4章)
- 3) ゼロ・キャッチを多く含む場合の、Tweedie モデルの性能評価、および従来手法 (ad hoc な共分散分析モデル・Catch モデル) との比較検討 (第5章)

本論文の第1章は序論であり、研究の背景と目的、論文の構成を記述した。第2章では、CPUE 標準化の

現状について、統計モデル・データマイニング手法・漁業資源特有の問題に分けて整理し、レビューを通じて明らかになった問題点について、特に本研究で取り上げる3つの主要課題に関して、概説した。

第3章では、CPUE 標準化に対応する一般化線形モデルを用いて、小標本の場合、大標本の場合など様々なケースを取り上げ、水産分野で広く知られている情報量規準 AIC の他に、BIC, CAIC, c-AIC, HQ, TIC などを使用し、実際の漁業データを用いて利用する情報量規準によってモデル選択結果が異なること、および複数の候補モデルの中から定めた真のモデルから乱数を発生させて正しいモデルを選ぶという選択パフォーマンスをシミュレーションにより計算し、情報量規準の良さを評価した。なお、ネスト構造モデルでは、カイ二乗検定や F 検定などの stepwise 検定も使用可能であり、計算機実験を通じて情報量規準と stepwise 検定の性能を比較した。この変数選択の問題は、CPUE に影響を与えている要因効果を統計的に取捨選択するという意味において重要であるが、使用する情報量規準や stepwise 検定によるモデル選択結果が、推定された CPUE 年トレンドという要因分析結果の違いを引き起こし、これらをチューニング指標として組み込んだモデルでの資源の絶対量推定結果の大きな差異となることもあり、極めて本質的な問題であると考えられている。なお、本章の具体的な研究成果は、次の通りである。

- 小標本の場合や未知パラメーター数の標本数に占める割合が高い場合に、AIC に有限修正を施した規準である c-AIC によるモデル選択結果が AIC などによるそれと異なることを例示し、さらに分散分析型のシミュレーションを通じて、c-AIC の選択パフォーマンスが AIC のそれに比べて高くなることを証明した。
- 大標本の場合に AIC が偏りを持つ可能性があることを示し、使用する規準により選択結果に差が生じること、および漸近的に望ましい性質である一貫性を持つ情報量規準 (BIC, HQ and CAIC) が AIC に比べて全体として優れていることを、それぞれ漁業データによる実例および回帰分析型の実験により示した。合わせて、HQ における定数項 c の検討を行い、推奨値と推奨式を提案した。
- ネスト構造を持つモデルにおいて、従来性能が良かったと言われてきた AIC の精密評価である TIC が正規誤差を持ちかつ連結関数が恒等写像であるような一般化線形モデルでは AIC と同等になることを理論的に証明し、合わせて TIC と AIC の選択パフォーマンスにはほとんど差がないことを、

計算機実験により示した。

- ネストモデルにおいて、計算機実験により情報量規準と stepwise 検定の比較を行い、一般に前者が多少優れていること、後者で有意水準を小さく設定した場合にパラメータ数が少ない単純なモデルが選ばれがちであることを示した。

第4章では、ミナミマグロ資源における CPUE 解釈の問題、すなわち操業がない時空間の CPUE 予測の問題を取り上げて、ニューラルネットワークを利用した解析を行った。CPUE を相対資源量の観点から捉えた場合、標準化された CPUE に相対的な面積指数を掛け合わせたものとして考えることが自然であり、これを資源量指数 (AI: abundance index) と呼んでいる。ミナミマグロ資源では過去から現在にかけて漁場が縮小しており、このような過去に漁獲があり現在操業がないセルの CPUE をどのように設定するか、極論すれば周囲と同じと考えるかそれとも 0 と仮定するか資源量指数の計算に影響してくる。ひいては、資源量指数から得られた CPUE 年トレンドの違いとなって表れる。そこで、本論文では、このような欠測セルの CPUE を教師付きニューラルネットワークの代表的なアルゴリズムである誤差逆伝播法を用いて予測を行い、合わせて得られた予測値から CPUE 年トレンド抽出を行うための簡便な要因分析法を提案した。ニューラルネットワークの精度評価のために、クロス・バリデーションにより同じ条件での MCMC 法に基づく EM algorithm との比較を行った。n-fold cross-validation により観測値と予測値の相関係数および MSE (平均二乗誤差) に基づき、モデルの性能評価および比較検討を行った。結果として、ニューラルネットワークによる CPUE 予測値に基づく、操業がないセルの CPUE と操業が行われたセルの CPUE 比は、0.8~1 前後を推移しており、1998年から2000年にかけて局所的に行われた日本の調査漁獲における CPUE 比 (年、季節、エリアは非常に局所的であるが0.7前後を記録) と比べ極端な矛盾は見られない。また、ニューラルネットワークによる CPUE の予測性能は、全く同じ条件での解析である MCMC 法に基づく EM algorithm によるそれよりも格段に高く、CPUE 予測値を元に算出された CPUE 年トレンドは一般化線形モデル (共分散分析) によるそれと比較的良く似ていた。このことから、ニューラルネットワークの予測性能の良さ、および提案した簡便な要因分析法の妥当性が言える。

第5章では、まぐろはえ縄漁業で混獲されるサメ類などを想定し、ゼロ・キャッチ問題と呼ばれる漁獲がゼロであるデータが含まれる場合に、CPUE の

自然対数を取ったものを応答変数とする共分散分析モデルが使用出来ない問題について、詳細に議論した。Tweedie 分布と呼ばれる、複合 Poisson 分布の拡張であるゼロ・データを統一的に取り扱えるモデル使用し、ゼロ・キャッチ率が10%程度と低い日本のはえ縄商業船によるインド洋キハダ資源の CPUE 解析、およびその割合が80%以上と高い日本のはえ縄公庁船による北太平洋クロトガリザメ資源の CPUE 標準化を行った。実際には、Tweedie モデルと全ての CPUE に定数項を加える ad hoc な共分散分析モデル、Catch Negative-Binomial モデルに基づく CPUE 年トレンドを比較した。その結果、ゼロ・キャッチ率が低いターゲット種のインド洋キハダ資源では Tweedie モデルと ad hoc な方法で年トレンドに極端な違いが見られなかったのに対し、ゼロ・キャッチ率が高い混獲種の北太平洋クロトガリザメ資源では、Tweedie モデルからの CPUE 年トレンドが、Catch モデルや ad hoc な方法からのトレンドと異なっていた。また、ニューラルネットワーク解析と同様に、n-fold validation を利用した観測値と予測値の相関係数や MSE に基づくモデルの性能評価を行ったところ、いずれの例においても、両方の指標に関して Tweedie モデルの精度が良かった。クロス・バリデーション結果から判断すると、ゼロ・キャッチ率が低い場合には Tweedie モデルの有意性が顕著に表れず、ad hoc な方法を使用しても実用上さほど問題が生じないと考えられる一方、ゼロ・キャッチの割合が高い場合には、Tweedie モデルの精度が他に比べて非常に高くなり使用が推奨される。なお、ゼロ・キャッチ率が高い場合には、相関係数が Tweedie モデル、Catch モデル、ad hoc な方法の順、MSE は Tweedie モデル、ad hoc な方法、Catch モデルの順に優れていたが、ad hoc な方法では観測値の大きさにかかわらず予測値が極端に小さくなることもあり、バイアスの大きさを考慮すると、サメ類などのゼロ・キャッチ率が高い場合には適用すべきでない、と結論付けられる。

最後の第6章は、本論文の結論部であり、今回取り上げた3つの課題に関する研究成果について、水産資源学の観点から、および応用統計学の視点からに分類して再度系統的に整理し、合わせて、今後の研究課題について記述した。

No.22, 1-85 (2008)

日本産マイワシの初期発育と産卵生態に関する研究

松岡正信 (瀬戸内海区水産研究所)

資源最大期から減少期における九州周辺海域のマイワシの初期発育過程と産卵生態について、生物学的側面から調査・研究した。

調査船上で人工受精を行い、ふ化後1.5日まで飼育し、観察した。天然卵を用いた実験では、全ふ化率、正常ふ化率、摂餌開始期生残率および摂餌率は10.8℃以下および23.5℃以上ではこれらが低下し、明らかに悪影響が認められた。塩分濃度は17.4区では、正常ふ化率が極めて低かった。ふ化直後の仔魚は全長(TL) 3.44 mmであった。ふ化後4日目(約5.7 mm TL)には摂餌開始期に達し、S型シオミズツボムシを捕食した。17日目に平均9.87 mm TL、29日目に15.37 mm TL(最大20.0 mm TL)となった。絶食条件下においたものでは、3～4日で瀕死状態となった。人工飼育魚と天然魚のシリーズ標本を用いて、全ての軟骨と硬骨の発達過程を記載した。それを基にして、骨格系の発達過程からみた摂餌・遊泳機能の発達を明らかにすると共に、4期8相から成る発育段階を区分した。重要な運動器官である体側筋について、赤色筋、桃色筋、白色筋および tonic-like fibers の発達過程を調べた。感覚器官の内、眼、嗅覚器、味蕾、側線器官および内耳の発達過程を調べた。体側筋系および感覚器官系の発達過程を基に、仔稚魚の生残に最も重要と考えられる摂餌・遊泳機能について骨格系を含めて総合的に検討した。また、発育段階についても、単なる区分ではなく発育の変化の集積として総合的に検討した。

産卵盛期は3月で、雌のGSIの最大値は20以上であった。吸水卵または産卵当日の排卵後濾胞を持つ個体の割合から算出される本種の産卵周期は4～6日と推定された。ネット曳網の結果、本種の産卵時刻は20時前後であった。また、同時に採集されたカタクチイワシ卵の分析の結果、両種の産卵時間帯は全く重複しなかった。産卵後間もない卵割前のAA卵は水深40～60 mで多く採集された。また、カタクチイワシの産卵中心は水深0～20 mであった。卵の発育過程を観察したところ、未受精卵でも囲卵腔が発達して胚盤を形成するものが認められた。これらは後に内部が崩壊して沈降した。薩南海域における産卵調査でもこのような卵が多数採集されたことから、天然海域でも多くの未受精卵が存在するものと推察された。9定点において、1時間～1週間程度の間採集卵数の短期変動を調べた結果、大きな変動が認められた。1979～1995年の九州周辺海域におけるマイワシの産卵量と卵

分布について調べた。産卵量の最低値は1995年の47兆粒、最大値は1987年の2,873兆粒であった。資源の水準により産卵期がずれる傾向があった。また、資源が低水準の時には九州西海域から山陰西部が産卵場となり、高水準の時には薩南海域が主産卵場となることが示された。産卵場の変化に伴い産卵水温も変化することが示された。

現在、本種の資源水準はごく低位であるが、資源変動機構を解明するために基礎的研究の継承が必要であろう。

No.22, 87-183 (2008)