

寄稿

CPUE標準化に用いられる統計学的アプローチに関する総説

庄野 宏

A Review of Some Statistical Approaches Used for CPUE Standardization

Hiroshi SHONO

This paper outlines some approaches to CPUE standardization using statistical modeling, data mining technique and describes several proper problems of CPUE analyses in the fish population dynamics. In the statistical modeling, we mainly concentrate on the generalized linear model which is widely utilized for CPUE standardization. We describe the two typical fixed models (CPUE model and catch model), mixed model using random effect. The problems of variable selection and model comparison are also discussed. In the approach for data mining, this paper describes the applications of tree regression models, neural networks and generalized additive models to CPUE standardization as an alternative to statistical way. We mainly focus on the procedure to factorial experiment including the extraction of CPUE year trend and predict CPUE corresponding unsupervised data by these methods for data mining. As the proper problems of CPUE analyses, we express the problems of so called zero-catch, the definition of abundance index and other some issues such as habitat model and the handling of fishing effort in the purse seine fisheries.

Key words: CPUE standardization, generalized linear model, random effect, model selection, data mining, tree regression model, neural networks, generalized additive model, zero-catch problem, abundance index, habitat model

1. はじめに

水産資源解析における1つの重要な概念として、CPUE (catch per unit effort: 単位努力当たり漁獲量) (Russell, 1931) が挙げられる。この指標は、一般の漁業では

$$CPUE = \frac{\text{Catch}}{\text{Effort}} \quad (1.1)$$

と定義される。Catchは単一魚種の漁獲重量や漁獲尾数で考えることが多く、Effortとしては縄船では針数を、まき網船では操業日数や操業回数などを用いる。一般にCPUEは資源密度に比例していると考えられており、特別な計算を行わなくともCPUEの年トレンドを見るだけで相対資源量の増減傾向を把握することが可能である。また、CPUEは微分方程式をベースに個体群動態を表したプロダクションモデル (Pella and Tomlinson, 1969) や年別年齢別漁獲量を使用したVPA (virtual population analysis) (Gavaris, 1988) などの資源評価モデルにおけるチューニングインデックスとして用いられることも多く、水産資源評価にお

いて非常に重要な意味を持っている。しかし、商業船などの操業データによる加工していないCPUEは、資源密度以外の様々な要因 (季節、海区、漁具など) も含んでおり、資源の変動を正確に知るためにはこれらの影響を取り除く必要がある。Gavaris (1980) などは、このような資源の年変動に対応する部分を取り出す作業をCPUE標準化と呼んだ。その主な目的を挙げると以下ようになる。

- ・資源の年トレンド (相対資源量) の効果の抽出
- ・漁船に装備されている操業機器や環境要因などがCPUEに与える影響の測定
- ・VPAなどの資源評価モデルに対するチューニングインデックスとしての有用性の向上

CPUE標準化に関する一般的な事項については平松 (1995) の概説論文やHilborn and Walters (1992), Quinn and Deriso (1999) などの水産資源学に関する一般的なテキストに記載されているが、現状ではGLM (generalized linear models: 一般化線形モデル) (Dobson, 1990; McCullagh and Nelder, 1989) などの統計モデルを使用して、資源密度以外の要因効果を除去した何らかの基準化を行うことが多い。

本報告では、CPUE標準化の現状と課題に関して、まぐろ類の資源に関する題材を例として主に方法論に焦点を当ててレビューを行い、合わせて問題点について整理するこ

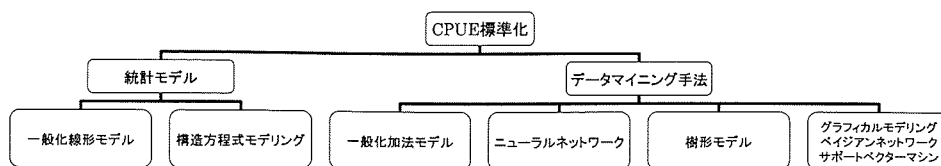


図1. CPUE標準化における各々の手法の関係を表した鳥瞰図

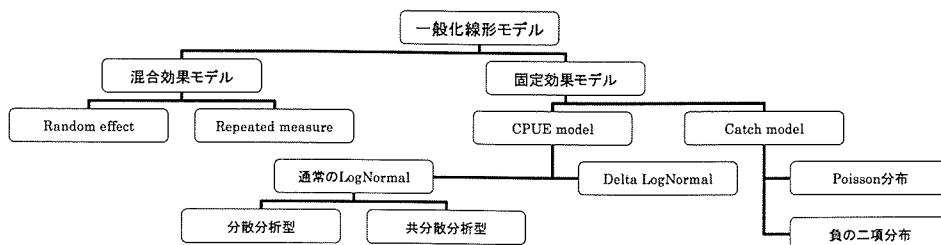


図2. 一般化線形モデルに関する鳥瞰図

とを目的とする。第2章では、CPUE標準化に広く使用されている統計モデル、中でも一般化線形モデルに焦点を当て、モデルの仮定を含めた標準化の現状およびモデル選択や要因分析の問題点について述べる。第3章では、最近適用例が増えてきているデータマイニング手法、中でも樹形モデルやニューラルネットワーク、一般化加法モデルを主に取り上げ、解析の現状とCPUE年トレンドの抽出などの問題点について記述する。第4章では、その他のCPUE標準化特有の問題点を紹介し、ゼロ・キャッチ問題と呼ばれる事項に関する対処法と相対面積指数による重み付けを行った資源量指数を中心に取り上げる。また、努力量の定義や不均一性、CPUE標準化に使用されるデータの性質やハビタットモデルなどの話題について簡単に触れる。

なお、CPUE標準化全体と一般化線形モデルの鳥瞰図はそれぞれ上図(図1, 図2)のようになる。

2. 統計モデル

CPUE標準化では、CPUEに影響を与える要因(年・季節・海区・漁船に装備されている操業機器・環境要因など)を説明変数に、CPUEあるいはCatchを応答変数としたGLM(回帰分析や分散分析・共分散分析モデルを含む)が伝統的に使用されている。代表的なモデル(Gavaris, 1980; Large, 1992; Reed, 1996)として以下のCPUEモデルとCatchモデルが挙げられるが、いずれもCPUEが各々の要因について効果の積の形で表されており、乗法モデル(multiplicative model)と呼ばれている。本章ではこれらの2つのモデルを中心に提起して、要因分析やモデル選択の現状について述べる。また、最近多く用いられつつある混合効果モデルについても簡単に触れる。

2-1. CPUEモデル

Robson (1966) 以来、CPUEに関して対数正規誤差を仮定した、いわゆるCPUE-LogNormalモデルが使用されることが多く、このモデルは(2.1)式のように表現される。データ解析において良く使用されるi.i.d. (independent and identical distributed: 独立で同一な分布に従うという意味)の条件をCPUE標準化においても仮定することが一般的であり、本論文についても特に断わらない限りこのi.i.d.の仮定を置くこととする。

$$E[\text{Log}(\text{CPUE})] = (\text{Intercept}) + (\text{Year}) + (\text{Area}) + \dots + (\text{EMT}) + (\text{Interactions}) \quad (2.1)$$

(但し $\text{Log}(\text{CPUE}) \sim N(\mu, \sigma^2)$ とし、 $E[\]$ は期待値を表す。 μ, σ^2 は未知パラメーターであり、それぞれ $\log(\text{CPUE})$ が従う正規分布の平均、分散を表す。)

(2.1)式での (Year) と (Area) は年や海区の効果を表し、(EMT) は操業機器や環境要因の効果を総称して表現している。CPUE標準化においては分布や回遊、系群などの情報をもとに海区分けを行うことが一般的であるが、区分された各々の海区内では資源が均一に分布していることを仮定している。これらの要因はカテゴリカル変数として扱われることが多いが、(EMT) については連続変数として用いられることもある。説明変数がカテゴリカル変数のみの場合と連続変数のみの場合は、それぞれ分散分析モデル・回帰分析モデルに対応しており、両者が混在している場合には共分散分析モデルとなる。

このモデルは取り扱いが容易なことと、観測誤差がCPUEの絶対値に比例するという仮定が合理的であると考

¹ 正確には、 (μ_1, \dots, μ_n) , $(\mu_1 = \dots = \mu_n = \mu)$, $\sigma^2 I_n$ (但し I_n : n 次の単位行列, n : データ数) と表記される。

えられていることもあり、まぐろ類のCPUE解析において様々な国際漁業委員会等で広く使用されている。CPUE-LogNormalモデルを用いたCPUE標準化の例は、古くはRobson (1966) に始まり数え切れないほど多く存在する。著者もこのCPUE-LogNormalモデルを使用して幾つかCPUE標準化を行っている (Shono and Ogura, 1999; Shono *et al.*, 2000, 2002)。

なお、ゼロキャッチ (CPUE=0となるデータ) が存在する場合にはCPUEの自然対数を取ることが出来ず、(2.1) 式のままでは計算が不可能である。そこで、解析を行うための幾つかの方法が提案されており、この問題については(4-1)節で議論する。

例1. CPUEモデルの例 (Shono *et al.*, 2002)

CPUE-LogNormalモデルの一例として、日本のはえ縄船によるインド洋キハダのデータを用いたCPUE標準化モデルを取り上げる。この例では、緯度と経度を5度毎に区切ったセルを1つの単位とし、月毎に集計したデータを使用している。

$$\begin{aligned} \text{Log}(\text{CPUE}_{ijkl}+0.1) = & (\text{Intercept}) + (\text{Year})_i + (\text{Month})_j + (\text{Area})_k \\ & + (\text{Gear})_l + (\text{SST}) + (\text{SOI}) + (\text{Year} * \text{Area})_{ik} \\ & + (\text{Month} * \text{Area})_{jk} + (\text{Month} * \text{Gear})_{jl} \\ & + (\text{Area} * \text{Gear})_{kl} + (\text{Area} * \text{SST})_k \\ & + (\text{Area} * \text{SOI})_k + \text{Error}_{ijkl} \end{aligned} \quad (2.2)$$

(但し、 $\text{Error}_{ijkl} \sim N(0, \sigma^2)$ とする)

(2.2) 式での各々の変数の定義は以下ようになり、SSTとSOIは連続変数として、その他の説明要因はカテゴリカル変数としてモデルに組み込んでいる。また、全てのCPUEに対して微小量 (0.1) を加えているのは、ゼロ・キャッチデータに対して自然対数を取ることが出来ない欠点を回避するためのものである (4-1) 節)。

CPUE : Catch (キハダの漁獲尾数)/Effort (はえ縄の針数 : 1000本を1と換算)

Intercept : 切片項

Year : 年の効果

Month : 月の効果

Area : エリア (海区) の効果

Gear : 枝縄数 (number of hooks between floats: NHF) の効果

SST : 表面水温 (sea surface temperature) の効果

SOI : 南方振動指数 (southern oscillation index) の効果

Year*Area : 年とエリアの交互作用 (以下同様にして、記号*は交互作用を表す。)

また、各々の添字の定義は下ようになり、月毎の41年分のデータに対して操業海域全体を6つに分けたサブエリアを使用しており、枝縄数についても4つのクラスに分類している。

i (Year): 1960–2000,

j (Month): 1–12,

k (Area): 1–6,

l (Gear): 1–4 (class 1: 5–6, class 2: 7–10, class 3: 11–14, class 4: 15–24)。

なお、(2.2) 式のモデルは解析の最初に仮定したものである。実際には統計的な方法に基づいて各々の要因効果の取捨選択を行う必要がある (2-4) 節)、説明変数を一つずつ減らしていくバックワードなステップワイズ検定 (2-4) 節) により最終的に選択されたモデルは、(2.2) 式からSOI指標の主効果のみが除かれ、(2.3) 式のようになった。

$$\begin{aligned} \text{Log}(\text{CPUE}_{ijkl}+0.1) = & (\text{Intercept}) + (\text{Year})_i + (\text{Month})_j + (\text{Area})_k \\ & + (\text{Gear})_l + (\text{SST}) + (\text{Year} * \text{Area})_{ik} \\ & + (\text{Month} * \text{Area})_{jk} + (\text{Month} * \text{Gear})_{jl} \\ & + (\text{Area} * \text{Gear})_{kl} + (\text{Area} * \text{SST})_k \\ & + (\text{Area} * \text{SOI})_k + \text{Error}_{ijkl} \end{aligned} \quad (2.3)$$

2-2. Catchモデル

Reed (1996) をはじめとして、Catchに対してPoisson分布もしくは負の二項分布 (negative binomial: NB) を仮定したモデルが近年用いられており、(2.4) 式のように定式化される。

$$\begin{aligned} E[\text{Catch}] = & (\text{Effort}) * \exp\{(\text{Intercept}) + (\text{Year}) + (\text{Area}) + \dots + (\text{EMT}) \\ & + (\text{Interactions})\} \end{aligned} \quad (2.4)$$

(但し Catch ~ Poisson (λ) あるいは Catch ~ NB (α, β) とする。ここで、漁獲尾数は努力量に比例することを仮定しており、この (Effort) の項は offset 変数² と呼ばれる。)

このモデルでの応答変数は整数値を取る必要があることから、Catchとしては重量ではなく尾数を用いられる。1990年代後半まではCatch-Poissonモデルが主流であったが、最近では負の二項分布モデルが使用されることも多い。この理由としては、統計パッケージSAS³の最新バージョン (Ver.8.2) で負の二項分布が標準装備されたことが挙げられる。

Poisson分布モデルでは、平均と分散が同じであるという仮定が現実の状況にそぐわない場合もあり、多くはover-dispersion parameter ϕ を使用している (Okamoto *et al.*, 2003)。すなわち、確率変数 X がパラメータ λ を持つPoisson分布に従う場合には、その平均と分散をそれぞれ λ , $\phi\lambda$ とする必要がある。しかし、over-dispersion parameterを導入すると想定する確率分布を何にしたかということ

² offset 変数は線形予測のための1つの項として使用されるが、係数パラメータが導入されない点が通常の変数と異なる。(2.4) 式では連結関数 (link function) を自然対数に設定しているため、(Effort) $^\alpha$ の形を仮定して、すなわち $\alpha * \log(\text{Effort})$ の形において $\alpha=1$ に固定したものと考えれば良い。なお、場合によってはこの指数部分のパラメータ α を推定することも可能である。

³ 現在はSASやS-Plus、SPSSなどの統計パッケージによる計算が主流になっているが、中でもSASは大規模データの定型的な処理に向いていることもあり、細かいカスタマイズが得意なS-PlusやRと並び、水産資源解析分野では多く使用されている。

表現出来ないため、通常の最尤推定は使用出来ない。そこで、疑似尤度 (Quasi-Likelihood) と呼ばれるフレームワークによりパラメーター推定が行われているが (Bayler, 1993), 理論的にはかなり複雑になる (Wedderburn, 1974; 椿, 1988)。このパラメーター推定の問題を回避するための1つのオプションとして、over-dispersion Poissonモデルの代わりに (平均と分散が異なる) 負の二項分布モデルを使用する方法が挙げられる。

2-3. CPUEの年トレンド抽出と要因分析

CPUE標準化における主な目的は、資源量 (資源密度) の年変動に対応する部分を取り出すことにある。交互作用を含まない場合には、CPUEモデルとCatchモデルのいずれにおいても年 (Year) 効果の推定値をそのまま取り出せば良い。しかし、年効果の交互作用を含む場合には、その解釈が難しい。そこで、交互作用を含む場合における要因分析のひとつの考え方として、LSMEAN (least squared mean) が挙げられる。まぐろ類のCPUE標準化では、通常は年効果のLSMEANを計算することによってCPUE年トレンドを抽出しており (Shono *et al.*, 2002; Matsunaga *et al.*, 2003), その定義は (2.5) 式で与えられる。

$$CPUE_i = \exp\{(\text{Intercept})_i + (\text{Year})_i + (\text{Year} * \text{Area})_i + (\text{Year} * \text{Season})_i + \dots\} \quad (2.5)$$

$$\begin{aligned} (\text{但し } (\text{Year} * \text{Area})_i &= \frac{1}{N_j} \sum_{j=1}^{N_j} (\text{Year} * \text{Area})_{ij}, (\text{Year} * \text{Season})_i \\ &= \frac{1}{N_k} \sum_{k=1}^{N_k} (\text{Year} * \text{Season})_{ik} \text{ などとする。これらの平均化され} \end{aligned}$$

た項は、単純平均ではなく各々のセルに属するデータ数に応じた重み付け平均とすることもある。)

なお、(2.5) 式では変数 (Year) の主効果とそれを含む交互作用についてのみ考慮すれば良い。また (2.6) 式で表される (Year*Area) のLSMEAN (高橋ほか, 1989) を考えることにより、海区別のCPUE年トレンドを抽出している (Shono *et al.*, 2002)。

$$CPUE_{ij} = \exp\{(\text{Intercept})_i + (\text{Year})_i + (\text{Area})_j + (\text{Year} * \text{Area})_{ij} + (\text{Area} * \text{Season})_j + \dots\} \quad (2.6)$$

$$(\text{但し } (\text{Area} * \text{Season})_j = \frac{1}{N_k} \sum_{k=1}^{N_k} (\text{Area} * \text{Season})_{jk}, \text{ などとする})$$

LSMEANの考え方を利用すると、交互作用を含む場合にも (2.7) 式を用いて漁船に装備されている操業機器や環境要因などの効果が検出可能であり、要因分析が行える。 (Shono *et al.*, 2000)。

$$\frac{CPUE_2}{CPUE_1} = \frac{\exp\{\overline{EMT_2 + (\text{Area} * \text{EMT})_2 + (\text{Season} * \text{EMT})_2 + \dots}\}}{\exp\{\overline{EMT_1 + (\text{Area} * \text{EMT})_1 + (\text{Season} * \text{EMT})_1 + \dots}\}} \quad (2.7)$$

$$\begin{aligned} (\text{但し } \overline{(\text{Area} * \text{EMT})_1} &= \frac{1}{N_j} \sum_{j=1}^{N_j} (\text{Area} * \text{EMT})_{j1}, \overline{(\text{Season} * \text{EMT})_1} \\ &= \frac{1}{N_k} \sum_{k=1}^{N_k} (\text{Season} * \text{EMT})_{k1} \text{ などとする。上式においても変} \end{aligned}$$

数 (EMT) の主効果とそれを含む交互作用についてのみ考えれば良い)

なお、(2.7) 式では説明変数 (EMT) が離散の場合を考えているが、連続変数の場合は推定されたパラメーターが1単位当たりのCPUEに対する変化量ととらえることが可能であり、回帰分析などの場合と同様にして要因分析を行うことが可能である (Shono *et al.*, 2002)。

2-4. 変数選択とモデル比較

本報告では、同一モデルにおける変数選択 (要因の取捨選択) と複数モデル間におけるモデル比較について取り上げる。

前者 (変数選択) は、モデルをひとつ固定したときに各々の説明要因がCPUEに影響を与えているか否かを統計的な手法に基づいて判断する作業であり、ステップワイズ検定や情報量規準を用いて行われることが多い。例えば、例1ではSOI指標の主効果がCPUEに影響を与えていないことがステップワイズ検定により確かめられたため、この項を残しておくことは統計的には意味を持たないことになる。このようにして、応答変数であるCPUEに影響を与えている要因効果のみを最終的に説明変数として含めることが一般的であり、本研究では同一モデルにおける変数選択と呼ぶことにする。

後者 (複数モデル間におけるモデル比較) は、候補となる複数のモデルを比較して一番良いモデルを選択する作業であり、CPUEモデルとCatchモデルの比較に代表される統計モデル間の比較のみならず、一般化線形モデルとニューラルネットワークの比較など統計的なアプローチとデータマイニング的なアプローチの比較を考える場合も含まれる。本報告では主に統計モデル間の比較 (CPUEモデル vs. Catchモデル) を想定しているが、現状ではこれといった有効な方法がなく、残差プロットなどを見て主観的に判断することが多い。

最初に変数選択に際して、次の2つの方法が広く用いられている。1つは、階層構造を持つモデルに対して使用可能な、尤度比に基づいたDevianceやPearsonカイ二乗統計量 (Dobson, 1990) に基づくステップワイズ検定であり、CPUEモデルではF検定が、Catchモデルではカイ二乗検定が用いられる (Shono and Ogura, 1999; Okamoto *et al.*, 2003)。ステップワイズ検定は総当たり法でないため、変数の数が

多い場合にも計算の手間が比較的少なく済む一方、検定のパスによって最終的な結果が異なる可能性がある(例2参照)(庄野, 2000)。

もう1つは情報量規準による方法であり、水産資源解析においてはAIC (Akaike's information criterion: Akaike, 1973)のみが長い間使用されてきた。しかし近年では一致性を持つBIC (Bayesian information criterion: Schwarz, 1978) やAICに有限修正を施したc-AIC (finite correction of AIC: Sugiura, 1978) などの情報量規準も状況に応じて用いられるようになってきている。特にc-AICは小標本の場合のパフォーマンスが良いこともあり(Shono, 2000)、国際委員会のICCATで多く使用されている。また、他の一致性を持つHQ (Hannan and Quinn, 1979)、BICと漸近的に同等であるMDL規準 (minimum description length criterion: Rissanen, 1983)、真のモデルが候補となるモデルを含まない場合の選択パフォーマンスを改善するTIC (Takeuchi's information criterion: 竹内, 1976)、AICにおけるペナルティ項の係数を変更したMAIC (Bozdogan, 1987) などの情報量規準も今後使用される可能性があると考えられる。なお、庄野 (2001) は正規誤差を持つ一般化線形モデル(回帰分析モデルや分散分析モデルを含む)において、TICがAICと漸近的に同等になることを証明した。

情報量規準による変数選択は、総当たり法で考えると一番良いモデルを一意に決めることが可能である。その一方、説明変数の数(主効果や交互作用など全て含む)が多い場合には、計算量が膨大となる欠点を持つ。また、over-dispersion Poisson モデルなど疑似尤度の枠組みで考える場合には、AICなど通常の情報量規準は使用することが出来ない。疑似尤度に関する情報量規準として、Burnham and Anderson (1998) はQ-AIC (Quasi-AIC) を提案したが、その信頼性や妥当性には疑問の余地が残っていることもあり、広くは知られていない。そのため、(over-dispersion parameterを仮定した場合には)フルモデルと候補となるモデルの対数尤度比統計量をベースにしたDeviance (逸脱度) やPearson統計量に基づくカイ二乗ステップワイズ検定を用いるのが一般的である。

なお、多くの主効果や交互作用を含む場合、ステップワイズ検定や情報量規準の値に基づいて主効果を取り除いた際に、関連する交互作用も一緒に取り除くべきか含めるべきかについては議論の分かれるところであり、ケース・バイ・ケースで判断することが多い。一般には、医薬品統計分野などでは薬の副作用の観点から残しておくことが多く、工学分野などでは推定精度の観点から取り除くことが多いと思われる。また、CPUEモデル(分散分析型モデル)におけるステップワイズF検定と情報量規準AICとの間には漸近的な関係が成り立っている(庄野, 2000)。

次に複数モデル間におけるモデル選択の問題として、CPUE標準化のための代表的なモデルであるCPUE-Log-

Normalモデル(式(2.1))とCatch-Poissonモデル(式(2.2))を比較する問題が挙げられる。現実問題としては、CPUE-LogNormalモデルとCatch-Poissonモデルの両方を使用したCPUE標準化の計算を行い、標準化残差の傾向をチェックすることが多い。しかし、その比較検討はあくまで主観的なものに過ぎない。そこで、Shono (2001) はCPUE-LogNormalモデルの応答変数をCatchに変更することにより応答変数を揃え、両者の情報量規準による比較を可能にした。

$$E[\text{Log}(\text{Catch})] = \text{Log}(\text{Effort}) + (\text{Intercept}) + (\text{Year}) + \dots + (\text{EMT}) + (\text{Interactions}) \quad (2.8)$$

(但し $\text{Log}(\text{Catch}) \sim N(\mu, \sigma^2)$ とする)

ただし、over-dispersion parameterを持つCatch-PoissonモデルではAICなどの一般的な情報量規準が使用出来ない。またQ-AICとAICとの比較は意味を持たないため、このままでは応答変数を変更した上のCPUE-LogNormalモデルとの比較が難しい。その場合には、Poisson分布モデルの代わりに負の二項分布を用いることが1つの解決策になる。なお、ゼロ・キャッチが存在する場合には、応答変数に対して一律に定数項を足し込む方法では(2.8)式に対応するモデルが複数考えられるため、4-2節で詳しく述べる。

例2. CPUEモデルにおける変数選択の例(庄野(2000)から抜粋して引用)

本報告では、Hilborn and Walters (1992)によるCPUE標準化の仮想例(表1、データは多少加工している)を取り上げて、2元配置分散分析型のシンプルなCPUEモデルにおける変数選択の実際について述べる。

まず、候補となる(2.9)式の4つのモデル(Model-1)~(Model-4)を考える。本来であれば、その他に(2.10)式で表現される候補モデル(Model-5)を考えることが可能である。しかし、(Model-5)ではデータ数よりもパラメーター数の方が多くなってしまっただけで推定が不可能なため、ここでは候補となるモデルから外した。

$$\begin{aligned} \text{Model-1: } & \text{Log}(\text{CPUE}) = \text{Intercept} + \text{Error} \\ \text{Model-2: } & \text{Log}(\text{CPUE}) = \text{Intercept} + \text{Year} + \text{Error} \\ \text{Model-3: } & \text{Log}(\text{CPUE}) = \text{Intercept} + \text{Class} + \text{Error} \end{aligned} \quad (2.9)$$

表1. CPUE標準化に関する仮想データ(4年間かつ異なった3つの漁船のクラスにおけるCPUEを表している)

Year	Class-1	Class-2	Class-3
1	0.63	1.03	1.22
2	0.48	0.56	1.26
3	0.33	0.67	0.89
4	0.54	0.48	1.01

Model-4: $\text{Log}(\text{CPUE}) = \text{Intercept} + \text{Year} + \text{Class} + \text{Error}$
 (但し $\text{Error} \sim N(0, \sigma^2)$ とする)

Model-5: $\text{Log}(\text{CPUE}) = \text{Intercept} + \text{Year} + \text{Class} + (\text{Year} * \text{Class}) + \text{Error}$ (2.10)

(2.9) 式の4つのモデル間の包含関係は次式 (2.11) のようになり、Backwardに変数を減らしていく形のステップワイズ検定におけるパスは2通り考えられる ((2.12) 式)。

$$\begin{matrix} (\text{Model-1}) \subset (\text{Model-2}) \\ \cap \qquad \cap \\ (\text{Model-3}) \subset (\text{Model-4}) \end{matrix} \quad (2.11)$$

Path-1: (Model-4) → (Model-2) → (Model-1)

Path-2: (Model-4) → (Model-3) → (Model-1) (2.12)

ステップワイズ検定の具体的な手順として、(Path-1) では最初に (Model-2) が真という帰無仮説に対して (Model-4) が真という対立仮説を考え、帰無仮説が棄却されたら (Model-4) を選択し、棄却されなければ次のステップに進む。次に (Model-1) が真という帰無仮説に対して (Model-2) が真という対立仮説を考え、帰無仮説が棄却されたら (Model-2) を選択し、採択されたら (Model-1) を選択する。このようにして、有意水準を1%としたステップワイズF検定により変数選択を行うと、Path-1では (Model-1) が選択された (庄野, 2000)。 (Model-1) は年によるトレンドの違いを持たない非常に単純なモデルである。 Path-2についても同様にして考えれば良く、結局 (Model-3) が選択された (庄野, 2000)。例2では、検定のパスによって最終的な変数選択の結果が異なってしまう。このようなステップワイズ検定におけるパスの一意性が問題になるケースでは、下のAICに代表される情報量規準を用いることが1つの解決策になる。

情報量規準による変数選択では、候補となる全てのモデルに対してAICなど情報量規準の値を計算し、その値が一番小さくなるモデルを選択すれば良い。表2に示したAICの値から判断すると、(Model-4) が最終的に選択された (庄野, 2000)。今回の例では、ステップワイズF検定による結果と情報量規準AICを用いた結果が一致しないが、一般にはこのような手順にてステップワイズ検定や情報量規

表2. 表1のCPUE標準化に関する仮想データを用いたAICによる変数選択結果。

Model	AIC
1	16.445
2	20.222
3	6.273
4	2.666

準によるモデル選択 (変数選択) を行うことが可能である。

2-5. 混合効果モデル

CPUE-LogNormalモデルやCatch-Poissonモデルにおける各々の要因は、通常は固定効果として扱われる。しかし、場合によっては (Year), (Area) などの主効果や交互作用を変量効果 (random effect) と考える場合もある (Verbeke and Molenberghs 1997; Little *et al.*, 1996; Searle *et al.*, 1992)。変量効果とは、該当する説明要因がある母集団からランダムに選択された標本であると考えられる効果であり、要因効果自身を確率変数と見なしている。これに対して、通常の分散分析型モデルでは、解析するデータ (要因実験でとられた因子水準) のみに推論の興味があり、この場合を含める主効果や交互作用を固定効果 (fixed effect) と呼んでいる。固定効果と変量効果は厳密に区別出来るものではなく、解析者が推論するときの関心によって仮定されることがある。なお、要因のいくつかが固定効果であり、それ以外のものが変量効果であるときのモデルを、一般には混合効果 (mixed effect) モデルと呼んでいる。要因のすべてを固定効果と設定した分散分析型モデルではカテゴリカル変数の水準数が大きくなることもあり、パラメーター数が多いゆえに推定精度が悪くなりがちであるが、変量効果を取り入れることによりパラメーター数が減少し、推定結果が安定することも多い。

まぐろ類のCPUE標準化に変量効果を導入する理由としては、要因効果自身を確率変数と見なすという本来の考え方ではなく、実用的な部分が大きいと考えられる。すなわち、漁業の分布などに基づいて分けられた海区を表すカテゴリカル変数と年の効果との交互作用 ((Year)*(Area)) の存在が、まぐろ類の広い範囲に及ぶ時空間的な移動が認められることから一般に強く示唆されるにもかかわらず、データの欠損ゆえに計算出来ない場合 (分けられたすべてのセルに対してデータが存在しない場合) に、これらの要因に関わる主効果や交互作用を混合効果に設定することが見受けられる (Yokawa and Shono, 2000)。

最近ではCatch-PoissonモデルやCatch-NBモデルなどの説明変数に対して変量効果を仮定するモデル、すなわち連結関数を恒等写像以外に設定して正規分布以外の確率分布を仮定するという、いわゆる一般化線形モデルのフレームワークの中での混合効果モデル (Fahrmeir *et al.*, 2001) も使用されるようになってきている。

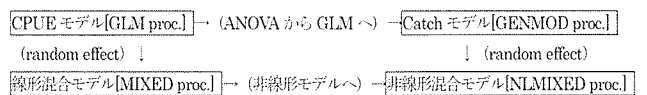


図3. CPUE標準化に使用されるモデル間の関係と対応するSAS procedureについて (ボックス内の[]はSAS/STATを用いた解析の際に使用するprocedureを表している。)

なお、固定効果モデルと混合効果モデルの関係を図示すると、図3のようになる。

2-6. その他の問題

CPUE標準化への応用が考えられるその他の統計手法としては、構造方程式モデリング（共分散構造分析）(Bollen, 1989; 狩野・三浦, 2002; 豊田, 1998a, 2000, 2003)が挙げられるが、現在のところこれら手法の水産分野への適用例はほとんど存在しない。著者らは現在構造方程式モデリングを利用して漁船に装備されている装置類の効果推定に取り組んでおり、(操業機器・探索機器・漁船情報の3つの潜在因子を設定して)どの装置類がCPUEに影響を与えているかを測定している。なお、Bayes流アプローチを利用したGLMのCPUE標準化への適用事例も報告されている(Badcook and Mcallister, 2001)。

また、著者らのCPUE標準化では漁船に装備されているソナーなどの探索機器や低温畜養装置などの操業機器、漁船や漁具の規模、表面水温や塩分濃度などの環境要因などがCPUEに与える影響について検討を行ってきた(Shono and Ogura, 1999, 2000; Shono *et al.*, 2000)。その結果、多くは過去の知見と一致したが、一部は機器を装備しない場合のCPUEが高くなるという負の効果も認められ、不自然な部分も残ってしまった。これは、ある意味で一般化線形モデルの限界を示していると考えられないこともない。そこで、このような過去の知見と矛盾する負の効果の問題を解決する手段として、漁船の番号を(カテゴリカル変数として)要因に取り込むことや一般化線形モデルの代替としてデータマイニング手法が考えられており、後者については第3章で取り上げる。

3. データマイニング手法

これまでの一般化線形モデルに代表される統計手法に加えて、近年樹形モデルやニューラルネットワークなどのデータマイニング手法(Berry and Linoff, 1997; Hastie *et al.*, 2001; Witten and Frank, 2000; 内田, 2002)が、Watters and Deriso (2000)をはじめとしてCPUE標準化に使用されるようになってきている。その理由として、数学・物理学・情報科学などから水産資源解析に転向した研究者によってこれらの手法が紹介されたことが挙げられるが、統計モデルによる標準化された年トレンドなどが現状にそぐわないことも多く、観測誤差などのバイアスの大きいデータに対応する斬新かつ有効な手法の登場が期待されていたという背景も一因であると考えられる。

なお、本報告でのデータマイニング手法の定義として、与えられた大量のデータの中から有用な規則や仮説を発見するために行われる探索的な方法を表すこととし(寺野, 2002)、統計モデルを構築してからデータをモデルに当てはめるという検証的なアプローチと区別することとする。データマイニング的なアプローチを大別すると教師付き学

習と教師無し学習⁴に分けられるが、本報告では教師付き学習についてのみ取り上げる。

本章では、まずCPUE標準化への応用例が存在するデータマイニング手法である樹形モデル(tree-regression models, TRM)とニューラルネットワークについて取り上げて、CPUE標準化における適用の現状と問題点について説明する。また、一般化線形モデルとの比較も含めて議論されることの多い一般化加法モデル(generalized additive models, GAM)について、先行研究などの紹介を行う。なお、今後の使用が考えられる可能性のあるデータマイニング手法についても、本章の最後で簡単に取り上げる。

3-1. 樹形モデル

樹形モデルは、ある基準に従ってデータセットを複数のサブセットに次々と分割していく手法であり、応答変数が単一である場合の分類問題(判別問題)や回帰問題⁵に有用である(Kass, 1980)。樹形モデル分析によって得られたルールは、一般的な高級言語やSQLのようなデータベース言語によって簡単に表すことが出来る。

樹形モデルの代表的な計算アルゴリズムとしては、CART(Breiman *et al.*, 1983; 大滝ら, 1998)、C5.0(Quinlan, 1993)、CHAID(Hartigan, 1975)などが挙げられるが、大別するとCART系(CART, C5.0など)とCHAID系(CHAIDなど)に分けられる。これらの主な違いとしては、CART系では量的変数を連続のまま処理を行って最終的な分岐後に枝刈りを行うのに対し、CHAID系では量的変数を内部でカテゴリカル変数に変換してから計算を行って分岐する際にカイ二乗検定を用いてその必要性を判断することが挙げられる。CART系アルゴリズムでは分岐の際の多重比較の問題が未解決のままであるため、統計研究者は一般にCHAID系アルゴリズムを好む傾向にある。

この樹形モデルを一般化線形モデルに代表される統計モデルの代替としてCPUE標準化に用いることに対しては、次のようなメリットがあると考えられる。

- ・いかなる統計分布も仮定することなく、柔軟なモデリングが可能である。
- ・重要な要因(説明変数)を自動的に抽出して、グループ化を行う。
- ・欠損データに対してロバストである(まぐろ類の広範な

⁴教師付き学習とは、原因(要因効果)を表す入力変数(説明変数)に対応する出力変数(応答変数)が存在する場合における学習方法であり、データの因果関係を調べることを主な目的とする。教師無し学習とは、結果系の変数が存在しない場合における学習方法であり、データの相関関係を調べることを主な目的とする。統計モデルとの関連で言うと、前者に対応する手法として重回帰分析や分散分析、一般化線形モデルなどが、後者に対応する手法として因子分析や主成分分析、クラスター分析などが挙げられる。

⁵一般に、分類問題(判別問題)とは原因に対する結果系変数が離散である場合の推定問題を指し、回帰問題とは結果系変数が連続である場合の推定問題を指す。

時空間的な分布パターンの変化ゆえに極めて有意であると考えられる (Year)*(Area) の交互作用を、GLMでは欠損データのために検出できないことが多く見受けられるが、樹形モデルでは原則として交互作用について考慮する必要がない)。

- ・ GLMが持つ技術的な問題 (ゼロ・キャッチの問題・モデル比較の問題 (CPUE-LogNormal vs. Catch-Poisson など)) が解決される。
- ・ 統計分析を実行する前の探索的段階で使用した場合に、データの非線形性や交互作用が検出可能である。具体的には、GLM解析を行う前にエリア分けに対して樹形モデルが適用可能であり、その場合にはCPUEでなく努力量を出力変数に設定することも可能である (Shono *et al.*, 2001)。

そのため、Watters and Deriso (2000) に始まり、樹形モデルを用いたCPUE標準化は、これまでにいくつかの解析例が紹介されている (Shono *et al.*, 2001; Venables and Toscas, 2002)。しかし、上のような長所が存在する一方、使用するアルゴリズム (CHAID vs. CART) によるCPUE年トレンドの違い (Shono *et al.*, 2001)、どのようにしてCPUEの年トレンドを抽出するか (例2)、量的な応答変数に対して推定値が離散になってしまうなどの問題点も残っている。

離散推定値の問題点を克服するために、樹形モデルに対するバギングの使用 (櫻井, 1998) やファジーメンバーシップ関数を用いた分類 (竹澤, 1999) など新しい試みも見られる。また、Venables and Toscas (2002) はバギングを用いたミナマガロCPUEの標準化を行っており、今後は要因分析の計算方法も含めて、より良いモデルの開発ならびに改良が望まれる。

3-2. ニューラルネットワーク

ニューラルネットワーク (麻生, 1988; Haykin, 1994; Smith, 1996; 吉富, 2000) とは、複雑な環境をモデル化する際に良く使用されている脳や神経系の仕組み (生物の神経細胞の回路で行われている情報処理システム) をモデル化したものであり、現在では様々な社会問題や工学上の問題などに適用されている。具体的には、多数の素子が互いに適当な重みを持って結合されたモデルであり、結合の種類によって相互結合型ニューラルネットワーク (各素子が全て相互に結合されたもの) と階層型ニューラルネットワーク (素子が幾つかの層をなして配列されており、各層間の素子は互いに結合されているが、層内の素子は結合されていないもの) に大別出来る。

ユーザーの立場からすれば、教師無し学習アルゴリズムと教師付き学習アルゴリズムとで分類する方が理解し易く、前者の代表的な例としてSOM (self-organization maps, 自己組織化マップ: Kohonen, 1989) が有名であり、後者の代表的な例としてはRBFN (radial basis function network, 動径基底関数ネットワーク: Broomhead and Lowe, 1988) や

MLEBPN (multi layer error back propagation network, 多層誤差逆伝搬ネットワーク: Rumelhart *et al.*, 1986) が挙げられる。SOMは低次元の格子によって構成された出力層が存在しないニューラルネットワークであり、クラスター分析の代わりに使用されることも多い。RBFNはデータが似たような性質を持つ幾つかのクラスターに分割出来るような場合に有効な方法であり、多変量正規混合分布モデルや多変量正規密度をカーネル関数とする確率密度関数推定とも密接な関係がある。しかし、ソフトウェアなどの対応状況などもあり、教師付き学習アルゴリズムにおいて一般に広く用いられているのはMLEBPNモデルである。

MLEBPNとは入力層と出力層の間に任意の個数の中間層 (隠れ層) を設ける方法であり、入力に対する出力と望ましい出力 (教師信号) との誤差を減らすようにニューラルネットワークの各素子の結合加重を修正していく。その際に情報が出力層から入力層に向かって逆方向に伝わっていくことが特徴的である。MLEBPNでは、中間層の数を1つに設定することが多く、その中間素子数を変更させることによって非線形関数に対する高い関数近似能力を持つ。その上、入出力データに対する制限が少なく (入力・出力ともに離散変数と連続変数の両方が扱える) 回帰問題と分類問題の両方が計算可能であり、多くのソフトウェアが開発されていて実行も容易である。後述する水産分野への応用のほとんどがMLEBPNを使用しており、本研究ではこのアルゴリズムのみを取り上げることとする。なお、ニューラルネットワークと統計モデルとの関係については、佐藤 (1996) や豊田 (1998b) などに詳しく述べられている。

ニューラルネットワークの水産分野への応用としては、加量量・資源量予測や魚種判別の適用例が幾つか見受けられるが (青木・小松, 1992; Haralabous and Georgakarakos, 1996; Chen and Ware, 1999など)、CPUE標準化への適用例はそれほど多くない。Shono (2002) は、ミナマガロ資源における過去に漁業が存在して現在は存在しない部分のCPUE予測を3層のエラーバックプロパゲーションを仮定した教師付き学習の典型的なニューラルネットワーク (MLEBPN) を用いて行っている。

ニューラルネットワーク (MLEBPN) においては、入力 (原因) と出力 (結果) の関係を表す式を記述する必要がなく、非線形かつ柔軟なモデリングが可能であるが、一方では要因分析や (中間層の数の設定をも含めた) モデル選択の難しさなどの問題があり、ブラックボックス的な側面も併せ持っている。また、学習速度が遅く、過学習 (overfitting) が起こりやすいことも指摘されている (Repley, 1994)。要因分析については、説明変数の数が少なくかつカテゴリカル変数の水準数が少ない場合 (カテゴリカル変数を想定しているが、連続変数の場合にはカテゴリに置き換える必要がある) には、添字を固定して平均的な予測

値を比較することが可能であるが、一般には難しい。ルール抽出に基づく方法も一部のソフトウェアで実装されているが、実用化の域には達していない感もある (Tsukimoto, 2000; 月本・森田, 2000)。そのため、CPUE 標準化の主目的である年トレンドの抽出に対する定まった方法が存在しないという問題は残されたままである。

なお、中間層の素子数の推定に関して、認定可能性の問題ゆえに最尤推定量の漸近的な性質が成立しないため、情報量規準を用いたモデル選択が理論的には不適切である。そのため、クロスバリデーションなどに基づく有効な推定方法の開発が望まれる。

3-3. 一般化加法モデル

CPUE 標準化での統計モデルの使用は、1990年代中頃まで一般化線形モデル (GLM) 一辺倒であった。しかし、最近になって平滑化スプライン (smoothing-spline: Whittaker, 1923) などの一般化加法モデル (GAM) (Hastie and Tibshirani, 1990; Simonoff, 1998) も CPUE 標準化に使用されるようになってきている (Wise *et al.*, 2002)。一般化加法モデルは、一般化線形モデルにおける各々の要因効果と応答変数の線形な関係を非線形関数で置き換えたものと解釈することが可能であり、smoothing-spline や局所重み付き多項式回帰 (locally weighed polynomial regression: Cleveland and Grosse, 1991) など多くの計算方法が提案されている。

一般化加法モデルを使用する利点としては、何と云っても柔軟なモデリングが可能となることが挙げられる。一般化線形モデルでは、CPUE と要因 (説明変数) の間に直線関係あるいは指数関係 (応答変数を $\log(\text{CPUE})$ と考えることが多い) が仮定されるのに対して、一般化加法モデルでは非線形な関係を記述することが可能である。そのため、表面水温などに対して平滑化スプラインを用いることが有効である場合が多い。例えば、ある一定の水温帯で CPUE が高くなっておりその周りの水温帯で CPUE 低くなっている場合には、その関係を一般化線形モデルで記述することは難しいが、表面水温に対してスプライン関数を使用することによって一般化加法モデルによる定式化が可能となる。一般に環境要因などに対して平滑化スプラインなどの手法を用いることが有効である場合が多い。その一方で、一般化加法モデルにおいては自由度の指定方法などの問題も残っている (竹澤, 2001)。SAS や S-Plus などの統計パッケージによる計算も一般化線形モデルと比較して複雑であることから、理論と応用の両面においてこれらの問題点の改良が望まれる。

なお、一般化線形モデルにおいて従来カテゴリカル変数として組み込んでいた要因に対して非線形な仮定を取り入れること (例えば漁業海区を便宜的に区分したカテゴリカル変数の代わりに緯度・経度のスプライン関数を用いることなど) が場合によっては有効であると思われる。しかし、水産資源分野での実際例はほとんど存在しない。

3-4. その他の手法

その他に CPUE 標準化への応用が考えられるデータマイニング手法としては、グラフィカルモデリング、ベイジアンネットワークなどが挙げられる (Jensen, 2001; Edwards, 2000; 宮川, 1997)。また、2値分類問題での有効な手法であり、多値判別問題や回帰問題に拡張されている SVM (Support Vector Machine, サポートベクターマシーン: Vapnik, 1998; 麻生ほか, 2003) も将来的には CPUE 標準化に対する有効な方法の一つになるのではないかと著者は考えている。現在のところこれらの手法の水産資源解析への適用例は少ないが、因果関係や相関関係を検出するためのデータマイニング手法の今後の発展が期待される。

例3. ニューラルネットワークによる要因分析と CPUE 予測の例

Shono (2002) では、月別/5×5度で集計された日本のはえ縄漁船によるミナミマグロの年齢別漁獲尾数データに対し、ニューラルネットワークを用いて CPUE 標準化を行い、要因分析 (CPUE の年トレンド抽出) と操業がない部分の CPUE 予測を試みた。

その概略は以下の通りであるが、この方法が現実の CPUE 年トレンドをより良く反映しているかは不明である。今後はデータマイニング的なアプローチによる要因分析や教師無し部分の CPUE 予測に関する優れた方法の開発ならびに改良が期待される。

出力変数 (応答変数)—CPUE: 4歳以上のミナミマグロ

CPUE(number/1000hooks)

入力変数 (説明変数)—Year: 年 (1969–2000)

Month: 月 (4–9の6ヶ月間)

Latitude: 緯度 (30S–50S,
5度刻み)

Longitude: 経度 (20W–90E–
180E–175W, 5度刻み)

注) 入力変数はすべてカテゴリカル変数とおく。

Step-1) 操業があるセル (5×5度) のデータ (全7,267レコード) を用いて、ニューラルネットワークによるルールを作成し、対応する CPUE の予測値を算出する。この計算においては、学習用データと検証用データをランダムに 85% と 15% に分割した⁶ (比較のために全て学習用データに設定したものも合わせて使用した)。

Step-2) Step-1 で作成したルールを用いて、操業のないセル (全33,047レコード) の CPUE 予測値を推定する。

Step-3) Step-1 と Step-2 で計算された全てのセル (全40,320

⁶ 統計分析では通常は全てのデータを使用するため、データを分割することはさほど多くないが、データマイニングにおいては (教師付きの) データを学習用データと検証用データにランダムに分割して、学習用データを用いて作成したルールの良さを検証用データによりチェックすることが一般的である。この方法はクロスバリデーションそのものである。

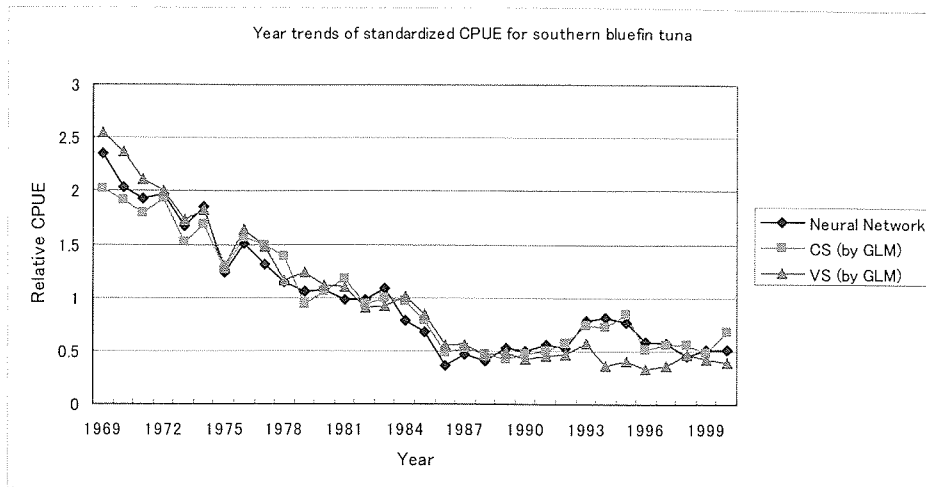


図4. ニューラルネットワークを用いた要因分析により推定されたミナミマグロ標準化CPUEの年トレンド (Step-2でMonthによる平均を使用) 並びにGLM (一般化線形モデル) との比較 (Shono, 2002)

レコード)の予測値からCPUEの年トレンドを求める。具体的には、最初に月(Month)による平均(もしくは総和)を計算し、次にエリア(Latitude & Longitude)による総和を計算した。

Step-1では教師付きデータを用いて、ニューラルネットワークによるルールを作成している。このステップでは個々のCPUE観測値(教師付きデータ)に対応する推定値が計算される。Step-2では、作成されたルールを用いて教師無し部分のCPUE予測をニューラルネットワークにより行っており、得られた予測値を用いることによって教師無しデータが教師付きデータに変換される。Step-3ではCPUE年トレンドの抽出(要因分析)を行っているが、その際にはStep-1で得られた各々の教師付きデータに対応するCPUEの推定値とStep-2で得られた個々の教師無しデータに対応するCPUEの予測値(いずれもニューラルネットワークを用いて作成されたルールに基づいている)を使用している。

この手順で得られたニューラルネットワークによるCPUE予測と要因分析の結果を図4に示す。ミナミマグロの標準化されたCPUE年トレンドは、CS仮説(CS仮説とVS仮説の違いについては(4.2)節で詳しく説明している)に基づくGLM(一般化線形モデル)によるCPUE年トレンドと比較的良く似ている。

4. 水産資源に特有の問題

本章では、CPUE標準化における漁業資源特有の問題について論じる。主に観測値がゼロであるデータが含まれている場合の取り扱いとエリアサイズによる重み付けの問題であり、その他としてまき網漁業での努力量の定義や説明変数として標準化に含める要因、ハビタットモデルなどについて取り上げる。

4-1. ゼロ・キャッチの取り扱い

CPUE標準化において広く用いられてきたCPUE-LogNormalモデルについて、応答変数であるCPUEに対して自然対数を取っていることから、CPUEがゼロとなるデータは $\text{Log}(\text{CPUE})=-\infty$ となってしまうのでそのままでは取り扱うことが出来ない。このことをゼロ・キャッチ(CPUE=0とCatch=0は同じ意味であることからこのような名前が付いたと考えられる)の問題と呼んでおり、標準化の計算を行うために、大きく分けて以下の2つの方法が使用されている。

- 1) 全てのデータに対して一定量(定数項)を足し込む方法
- 2) CPUEがゼロか否かを分けてからゼロ・キャッチ率をlogitモデル(4.1式)によって推定し、ゼロでない部分のみに対して通常のモデル(CPUEモデルやCatchモデルなど)を適用する方法(Delta型2段階法)

$$E[\text{Log}\{R/(1-R)\}] = (\text{Intercept}) + (\text{Year}) + (\text{Area}) + \dots + (\text{Interactions}) + (\text{Log}(\text{Effort})) \quad (4.1)$$

(但しR(ゼロ・キャッチ率)~Binomial(p)とする)

1)はユーザーにとって扱い易い反面、区間推定における偏りの原因になってしまう(点推定に関しては、推定値からこの定数項を差し引くことによって偏りを防ぐことが出来る)。また、一定量としてどのような値を取れば良いのか、という問題もある。現状では1)の方法が多く使用されており、国際委員会のICCATなどではCPUEに足し込む一定量として平均CPUEの10%が用いられている(ICCAT, 1997)。しかし根拠は不明であり、長所も感じられない。一般には微量の方が区間推定への影響が少ないと感じられるが、データに依存しない形でこのことを示すことは困難である。

Shono (2001) では、CPUE-LogNormalモデルに対する(全てのCPUEデータに一律に足し込むための)値を、複数の候補の中から情報量規準を用いて決定出来ることを示した。また、尤度関数の形に着目してこの定数項の数値的最適化が可能であることを指摘した。この計算は、例えば統計パッケージのSASにおいてはORと呼ばれるprocedureの中のNLPステートメントを使用することによって実行可能となる。

$$E[\text{Log}(\text{CPUE} + \text{const.})] = (\text{Intercept}) + (\text{Year}) + (\text{Area}) + \dots + (\text{EMT}) + (\text{Interactions}) \quad (4.2)$$

(但し $\text{Log}(\text{CPUE} + \text{const.}) \sim N(\mu, \sigma^2)$ とする)

$$E[\text{Log}(\text{Catch} + \text{const.} * \text{Effort})] = \log(\text{Effort}) + (\text{Intercept}) + (\text{Year}) + \dots + (\text{Interactions}) \quad (4.3)$$

$$E[\text{Log}(\text{Catch} + \text{const.})] = \log(\text{Effort}) + (\text{Intercept}) + (\text{Year}) + \dots + (\text{Interactions}) \quad (4.4)$$

ただし、ゼロ・キャッチが含まれる場合のCPUEモデルとCatchモデルの比較(2.4節)に際しては、応答変数を揃えるに当たって注意が必要である。すなわち、上の(4.2)式においてCPUE (= Catch/Effort)の定義を代入すると(4.3)式で表されるが、このモデルではCPUE年トレンドの抽出が難しい面も持っている。そのため、(4.2)式のCPUEをダイレクトにCatchに置き換えたモデル((4.4)式)を用いることも1つの手段であるが、このモデルの仮定は(4.2)式で表されるモデルと異なっている。その上、Catchモデル((2.4)式)では応答変数に定数項が含まれていないのに対し、CPUEモデルの変形版((4.3)式または(4.4)式)では含まれており、モデルの微妙な違いも認められる。

そのため、このゼロ・キャッチの問題に対して、出来る限り(次に説明する)2)のDelta型2段階法を用いるべきと著者は考えており、2)の方法を用いることが定数項を含まない形でのCPUEモデルとCatchモデルとの比較を可能にする。

2)の方法では、最初のlogitモデル((4.1)式)と次の通常モデル(CPUEモデルまたはCatchモデル)を別々に計算することがCPUE解析において多く行われており、Delta型モデル(あるいは2段階型モデル)と呼ばれている(Lo, 1992; Stefansson, 1996)。なお、logitモデルのみを用いてCPUE解析を行った例としては、Miyashita *et al.* (2000)が挙げられる。

このDelta型アプローチは実務家にとって理解し易い部分があり、SASなどの統計パッケージによる解析も比較的容易に行われる。その反面、区間推定が難しい面があり、2つのモデルにおいて有意と認められた要因(説明変数)が異なる場合には、CPUE年トレンドの推定がかなり複雑になる欠点も併せ持つ。2つのモデル(logitモデルと通常

モデル)の尤度を1つに書き下して同時に推定することも理論的には可能であり、Zero-Inflatモデルと呼ばれている(Lambert, 1992; Ridoud *et al.*, 2001)。しかし、ソフトウェアによる計算手順が複雑なこともあり、水産資源分野における適用例はほとんど見受けられない。

4-2. 面積指数での重み付け

CPUE標準化の第1の目的は年効果の推定であり、抽出された年トレンドは相対的な資源の増減傾向を表している。なぜなら、CPUEは資源量に比例すると考えられているからである。(CPUEと資源量の比例関係が線形か非線形か、あるいはどのような関数形で表すべきかという問題は古くから議論されているが、CPUE標準化とは直接関係がないこともあり、本報告では割愛したい。)

相対資源量という立場でCPUEを捉えた場合、エリア分けが全て等しいサイズの場合にはCPUEと相対資源量は同じことを表すが、個々の海区の大きさが異なり、なおかつ年とエリアの交互作用が認められる場合には、エリアサイズを考慮した補正が必要である。すなわち、推定されたCPUEの年効果に相対的なエリアサイズを掛け合わせたものを資源量指数(abundance index: AI)と呼んでおり、通常はCPUE推定値に基づいた資源量指数が相対資源量に対応すると考えられている(能勢ほか, 1988; 山田・田中, 1999)。

$$AI_{ij} = w_j \text{CPUE}_{ij} \quad (4.5)$$

(但し $\sum_j w_j = 1$ (i : Yearの添字, j : Areaの添字, w_j : Area(j)の相対面積指数) とする)

この相対面積指数は、式(4.5)のように年に依存しないと考えることが多い。しかし、一般にまぐる類は広範囲の時空間的な移動を行い、それに合わせて漁場が変化する場合も多い。そのため、年が経つにつれて漁場が縮小しているような場合に(4.5)式を用いて資源量指数を計算すると、資源の過大評価につながる恐れもある。

この問題に対して、ミナミマグロ漁業を取り扱っている国際委員会CCSBTでは、(4.5)式の方法の他に1990年代初めから(4.6)式のような相対面積指数が年に依存するという考え方を取り入れている(CCSBT, 1998)。すなわち漁業の変化に対してエリアサイズも変化させるという方法である。この考え方は(4.5)式のCS仮説(constant square仮説)に対してVS仮説(variable square仮説)と呼ばれている。

$$AI_{ij} = w_{ij} \text{CPUE}_{ij} \quad (4.6)$$

(但し $\sum_j w_{ij} = 1$ (w_{ij} : Year(i)かつArea(j)の相対面積指数) とする)

現実問題として、ミナミマグロ資源のCPUE標準化に際していずれの仮説を用いるかにより、1990年以降の相対資源量の年トレンドのみならずこれらをチューニングインデックスとして使用している資源評価モデルの結果が全く異なるものになってしまった。すなわち、CS仮説によると資源状態は上向きであり、VS仮説に基づく資源は下方に向かっているという結論になる (Takahashi *et al.*, 2001)。

いずれの仮説が現実の状況に良く当てはまっているか、という問題については多く議論が行われているが、本質は漁場の縮小が乱獲によって魚の存在するエリアが狭くなったのか、あるいはその他の原因によるものなのか (規制により操業エリアや漁船の数が減少したからなのか、(漁獲効率のみならず気象条件なども含めて) 良い条件の場所での操業が多くなった結果として操業海域が狭くなったのか)、いずれに多く起因する問題なのか、ということに尽きる。

この問題に対して、Toscas and Thomas (1998) は混合効果モデルの一種である repeated measure の考え方に基づいて、過去に漁獲があって現在は漁獲が無いエリアのCPUEを推定しており (論文では空間統計学的手法を用いて補間しているが、repeated measureによる推定と読み替えることが可能である)、Shono (2002) はニューラルネットワークを用いて操業がないセルのCPUE予測を行った (例3)。

4-3. その他の話題

本節では、これまでに述べられなかった水産資源特有のCPUE標準化におけるその他の問題について、簡単に整理する。

努力量の定義

まぐろ漁業における努力量として、はえ縄漁業や竿釣り漁業では針数 (1000 hook を1単位とする場合が多い) や竿数を使用することが多く、国際委員会等においてもこれらを努力量の定義とすることに対しておおよそのコンセンサスが得られている。

しかし、まき網漁業については、探索時間をどのようにカウントするか、あるいはFADsと呼ばれる魚を集めるための人工浮き漁礁の影響をどのように測定するか、など多くの問題点が指摘されており、努力量の定義自体が難しい側面を併せ持つ。実際問題として、まき網漁業のCPUEにおいては操業日数や操業回数を使用すること (O'Brien *et al.*, 1997; Shono *et al.*, 2000; Soto *et al.*, 2000; Soto *et al.*, 2002) が多く見受けられるが、これらの努力量の定義が漁業の現状にマッチしているかどうかについては、注意深く検討していく必要がある。

努力量の不均一性

一般に努力の絶対量が多い部分と少ない部分でその推定精度に差が出るのは、統計学的に考えれば当然のことである。操業の少ない部分を解析から除くことも一案であるが、こ

の努力量の不均一性を解決するための1つのオプションとして、LSMEANを計算して要因効果を抽出する際にデータ数に応じて重み付けすること⁷が考えられる ((2.5)式)。

説明変数として取り入れる要因の検討

一般化線形モデルなどを用いたCPUE解析においては、多くの説明変数が要因として組み込まれている。現状では、年 (Year)、海区 (Area)、季節 (Season) (四半期もしくは月を単位と考える場合が多い) などに加えて、表面水温などに代表される環境要因、漁船に装備されている操業機器や探索機器などの装備類や漁船の規模などの要素、(はえ縄船における) 枝縄数などの情報 (Yokawa and Shono, 2000)、あるいはターゲット種以外の漁獲の有無などを使用することもある (Shono and Ogura, 2000; Shono *et al.*, 2000)。ややもすると多重共線性などの問題が生じるため、モデルの初期設定 (変数選択を行う前に仮定したモデル) の際には、説明変数間の相関関係について考慮すべきである。また、最初に仮定する要因効果の取捨選択に関しては、主効果のみならず交互作用の条件設定についても十分な注意が必要である。なお、統計モデルにおける要因分析に際して、特に単調性が想定される効果に関しては説明変数としての組み込み方、すなわち (連続値を有限個のカテゴリーに分解して) カテゴリカル変数と設定するか、それとも連続変数と指定するかについて、十分な検討が必要である。

CPUE解析に使用されるデータの種類の

まぐろ類の商業船による操業データのCPUEは、通常は漁法・魚種・海域 (系群) 別に標準化されることが多い。これらの漁業データは大きく分けて shot-by-shot と呼ばれる操業毎のデータと (5×5度/月別などに) 集計を施したデータに分類されるが、データの与える重みなどの問題もあり、一般には操業毎のデータが好ましいと考えられている。しかし、膨大な件数になることもあり、場合によってはデータの入力や加工段階で集計されることも多いため、操業毎のデータと集計されたデータがともに広く使用されている。通常は操業毎のデータが存在する場合にこれを集計する必要はないが、1つの入力セット (同じ添字を持つ説明変数の組) に対して複数の異なる出力 (CPUE) が存在する場合もあるため、ニューラルネットワークなどにおいては注意が必要である。

ハビタットモデル

1990年代後半から、ハビタットモデル (Hinton and Nakano, 1996) と呼ばれるはえ縄漁具の鉛直分布パターンと漁獲対象種の鉛直分布パターンを用いて、(水域別時期別の魚の

⁷ 各セルの重みを同じとした通常の平方和 (Type III) と並んで、データ数に応じた重み付けを行った平方和 (Type II) も SAS や S-Plus などの統計ソフトウェアに装備されている。しかし、データ数の不均一性も含めた様々な要因効果の除去というCPUE標準化の目的からすると Type II 平方和の多用は問題があると著者は考えており、まぐろ類のCPUE解析においても一般に Type III 平方和が広く使用されている。

鉛直分布確率と漁具のそれとの積の形で) 有効努力量を直接推定するCPUE標準化の方法が、遠洋はえ縄漁業における一部の魚種・海域(南西太平洋や東部太平洋のまぐろ・かじき類など)で使用されている。具体的には、漁具の鉛直分布パターンははえ縄に装着した小型水深水温計などのデータを基に、漁獲対象種の鉛直分布パターンは魚に装着したポップアップアーカイバルタグなどのデータを基にして推定されることが多い。この方法はかじき類など魚の鉛直方向の分布が非常に浅いところに偏っている場合などに有効であると考えられており、生物データを直接取り扱えるという長所がある。その一方、漁獲対象種の鉛直分布パターンのモデル化が難しいという欠点や、ハビタットモデルにより推定されたCPUEの年トレンドが一般化線形モデルによるそれと異なるケースも存在する(Yokawa *et al.*, 2002)。そのため、今後ハビタットモデルと一般化線形モデルとの比較を含めた詳細な検討が必要である。

謝 辞

原稿を読んで有益なコメントを下された筑波大学教授の椿広計博士ならびに遠洋水産研究所室長の平松一彦博士に深く感謝申し上げる。また、提出原稿の不備な部分を指摘して下さった匿名の2名の査読者の方々に厚く御礼申し上げます。

引用文献

- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory. Petrov, B. N., and Csaki, F. (eds.), Akademiai Kiado, Budapest, 267-281.
- 青木一郎・小松輝久(1992)ニューラルネットワークによるマイワシ成魚漁獲量の予測。水産海洋研究, **56**, 113-120.
- 麻生英樹(1988)ニューラルネットワーク情報処理。産業図書, 198 pp.
- 麻生英樹・津田宏治・村田昇(2003)パターン認識と学習の統計学。岩波書店, 225 pp.
- Badcock, E. A. and M. M. Mcallister (2001) Bayesian generalized linear models for standardization catch rate indices of abundance. ICCAT/SCRS/01/43, 29 pp.
- Bayler, P. B. (1993) Quasi-likelihood estimation of marked fish recapture. Can. J. Fish. Aquat. Sci., **50**, 2077-2085.
- Berry, M. J. A. and G. Linoff (1997) Data mining techniques: for marketing, sales, and customer support. John Wiley and Sons, New York, 454 pp.
- Bollen, K. A. (1989) Structural equations with latent variables. John Wiley and Sons, New York, 528 pp.
- Bozdogan, H. (1987) Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, **52**(3), 345-370.
- Breiman, L. J., R. A. Friedman, R. A. Olshen and C. J. Stone (1983) Classification and regression trees. Wadsworth International Group, Belmont, California, 368 pp.
- Broomhead, D. S. and D. Lowe (1988) Multivariable functional interpolation and adaptive networks. Complex Systems, **2**, 321-355.
- Burnham, K. P. and D. R. Anderson (1998) Model selection and inference: A practical information—Theoretic approach—Springer, New York, 353 pp.
- CCSBT (1998) Report of the first meeting of the scientific assessment group (SAG), 41 pp.
- Chen, D. G. and D. M. Ware (1999) A neural network model for forecasting fish stock recruitment. Can. J. Fish. Aquat. Sci., **56**, 2385-2396.
- Cleveland, W. S. and E. Grosse (1991) Computational methods for local regression. Statistics and Computing, **1**, 47-62.
- Dobson, A. J. (1990) An introduction to generalized linear models. Chapman and Hall, London, 174 pp.
- Edwards, D. (2000) Introduction to graphical modeling (2nd edition). Springer, New York, 333 pp.
- Fahrmeir, L., G. Tutz and L. Fahrmeir (2001) Multivariate statistical modeling based on generalized linear models. 2nd edition. Springer, New York, 517 pp.
- Gavaris, S. (1980) Use of a multiplicative model to estimated catch rate and effort from commercial data. Can. J. Fish. Aquat. Sci., **37**, 2272-2275.
- Gavaris, S. (1988) An adaptive framework for the estimation of population size. CAFSAC Research Document, 88/12, 12 pp.
- Hannan, E. J. and B. G. Quinn (1979) The determination of the order of autoregression. J. Royal Statist. Soc. Ser. B, **41**, 190-195.
- Haralabous, J. and S. Georgakarakos (1996) Artificial neural networks as a tool for species identification of fish schools. ICES J. Mar. Sci., **53**, 173-180.
- Hartigan, J. A. (1975) Clustering algorithms. John Wiley and Sons, New York, 352 pp.
- Hastie, T. and R. Tibshirani (1990) Generalized Additive Models. Chapman and Hall, London, 335 pp.
- Hastie, T., R. Tibshirani and J. Friedman (2001) The elements of statistical learning: Data mining, inference and prediction. Springer, New York, 533 pp.
- Haykin, S. (1994) Neural networks: A comprehensive foundation. Macmillan, New York, 750 pp.
- Hilborn, R. and C. L. Walters (1992) Quantitative fisheries stock assessment. Chapman and Hall, 570 pp.
- 平松一彦(1995)統計モデルによるCPUE標準化。漁業資源研究会議北日本底魚部会報, **28**, 87-97.
- Hinton, M. G. and H. Nakano (1996) Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (*Makaira nigricans*) catch and effort data from Japanese longline fisheries in the Pacific. IATTC Bull., **21**(4), 171-200.
- ICCAT (1997) Report of the bluefin tuna methodology session. ICCAT, Coll. Vo. Sci. Pap. Vol. XL VI(1), 187-201.
- Jensen, F. V. (2001) Bayesian networks and decision graphs. Springer, New York, 268 pp.
- 狩野裕・三浦麻子(2002)AMOS, EQS, CALISによるグラフィカル多変量解析(増補版):目で見える共分散構造分析。現代数学社, 293 pp.
- Kass, G. V. (1980) An exploratory technique for investigating large quantities of categorical data. Appl. Stat., **29**, 119-127.
- Kohonen, T. (1989) Self-Organization and Associative Memory. 3rd edition. Springer, Berlin, 368 pp.
- Lambert, D. (1992) Zero-Inflated Poisson regression, with an application to defects in manufacturing. Technometrics, **34**(1), 1-14.
- Large, P. A. (1992) Use of a multiplicative model to estimate relative abundance from commercial CPUE data. ICES J. Mar. Sci., **49**, 253-261.
- Little, R., C., G. A. Milliken, W. W. Stroup and R. D. Wolfinger (1996) SAS system for mixed models. SAS Institute Inc., USA, 633 pp.
- Lo, N. C. L. D., L. D. Jacobson and J. L. Squire (1992) Indices of relative abundance from fish spotter data based on Delta-Lognormal models. Can. J. Fish. Aquat. Sci., **49**, 2515-2526.

- Matsunaga, H., H. Shono, and K. Yokawa (2002) Standardization of CPUE of Pacific bluefin tuna caught by Japanese distant-water and offshore longliners in the spawning ground from 1953–2000. ISC-BFT-WG/02/Doc.11, 7 pp.
- McCullagh, P. and J. A. Nelder (1989) Generalized linear models. 2nd edition Chapman and Hall, London, 511 pp.
- 宮川雅巳 (1997) グラフィカルモデリング. 朝倉書店, 177 pp.
- Miyashita, T., Shono, H. and Okamura, H. (2001) GLM analysis of the JSV data for the Antarctic minke whale. IWC-SC/53/IA15, 11 pp.
- 能勢幸雄・石井丈夫・清水 誠 (1988) 水産資源学. 東京大学出版会, 217 pp.
- O'Brien, C. M. B., L. T. Kell, J. Santiago and V. Ortiz de Zarate, V. (1997) The use of generalized linear models for the modeling of catch-effort series. II: Application to north Atlantic albacore surface fishery. ICCAT/SCRS/97/49, 14 pp.
- Okamoto, H., K. Satoh, H. Shono, and N. Miyabe (2003) Standardized Japanese longline CPUE for yellowfin tuna in the Atlantic Ocean up to 2001. ICCAT/SCRS/2003/056, 24 pp.
- 大滝 厚・堀江有治・Dan Steinberg (1998) 応用2進木解析法. 日科技連, 273 pp.
- Pella, J. J. and P. K. Tomlinson (1969) A generalize stock production model. IATTC Bull., **13**, 421–496.
- Quinlan, J. R. (1993) C4.5: Programs for machine learning. Morgan Kaufmann. San Mateo, 302 pp.
- Quinn, T. J. and R. B. Deriso (1999) Quantitative fish dynamics. Oxford, New York, 542 pp.
- Reed, W. J. (1996) Analyzing catch-effort data allowing for randomness in the catching process. Can. J. Fish. Aquat. Sci., **43**, 174–186.
- Repley, B. D. (1994) Neural networks and flexible regression and discrimination. Advances in Applied Statistics, J. Appl. Stat., Suppl. **21**, 39–57.
- Ridout, M., J. Hinde, and C. G. B. Demetrio (2001) A score test for testing a zero-inflated Poisson regression model against zero inflated negative binomial alternatives. Biomaterics, **57**, 219–223.
- Rissanen, J. (1983) A universal prior for integers and estimation by minimum description length. Ann. Statist., **11**, 416–431.
- Robson, D. S. (1966) Estimation of the relative fishing power of individual ships. Research Bull., International Commission for the North-west Atlantic Fisheries, 3, 5–14.
- Rumelhart, D. E., J. L. McClelland and PDP Research Group (1986) Learning internal representation by error propagation. Parallel Distributed Processing: Explorations in the Microstructures of Cognition, 1, MIT Press, Cambridge, 318–362.
- Russell, E. S. (1931) Some theoretical consideration on the “overfishing” problem. Rapp. P.-v. Reun. Cons. int. Explor. Mer., **6**(1), 3–20.
- 櫻井茂明 (1998) ファジィ帰納学習による数値予測規則の学習. 電気学会論文誌, **C118**(9), 1369–1375.
- 佐藤整尚 (1996) 統計モデルとしてのニューラルネットワーク. 統計数理, **44**(1), 85–98.
- Schwarz, G. (1978) Estimating the dimension of a model. Ann. Statist., **6**, 461–464.
- Searle, S., R., G. Casella and C. E. McCulloch (1992) Variance components. John Wiley and Sons, New York, 501 pp.
- Shono, H. (2000) Efficiency of finite correction of Akaike's information criteria. Fish. Sci., **66**(3), 608–610.
- Shono, H. (2001) Comparison of statistical models for CPUE standardization by information criteria —Poisson model vs. Log-normal model— IOTC-WPM/01/1, 12 pp.
- Shono, H. (2002) Attempts for estimation of standardized CPUE by tree-regression modes and neural networks. CCSBT-ESC/0209/38, 18 pp.
- Shono, H. and M. Ogura (1999) The standardized skipjack CPUE including the effect of searching devices, of the Japanese distant water pole and line fishery in the Western Central Pacific Ocean. ICCAT-SCRS/99/59, 18 pp.
- Shono, H. and M. Ogura (2000) The standardized albacore CPUE of the Japanese distant water pole and line fishery, including the effect of searching devices. NPALB/00/8, 9 pp.
- Shono, H., H. Okamoto, and T. Nishida (2002) Standardized CPUE for yellowfin tuna (*Thunnus albacares*) resources in the Indian Ocean by generalized linear models (GLM) (1960–2000). IOTC/WPTT/02/12, 12 pp.
- Shono, H., T. Matsumoto, M. Ogura, and N. Miyabe (2000) Preliminary analysis of effect of fishing gears on catch rate for the Japanese purse seine fishery. SPC-SCTB13/WP/RG-3, 13 pp.
- Shono, H., S. Tsuji, N. Takahashi and T. Itoh (2001) Preliminary analysis for CPUE standardization and area stratification by tree-regression models. CCSBT-SC/0108/30, 17 pp.
- 庄野 宏 (2000) 情報量規準とステップワイズ検定の比較と水産資源解析への応用. 遠洋水研報, **37**, 1–8.
- 庄野 宏 (2001) 情報量規準 TIC と c-AIC によるモデル選択の有効性. 遠洋水研報, **38**, 21–28.
- Simonoff, J. S. (1998) Smoothing methods in statistics. Springer, New York, 338 pp.
- Smith, M. (1996) Neural networks for statistical modeling. International Thomson Computer Press, Boston, 256 pp.
- Soto, M., J. Moron and P. Pallares (2000) Standardized catch rates for yellowfin (*Thunnus albacares*) from the Spanish purse seine fleet (1984–1995). IOTC-WPTT/00/04, 15 pp.
- Soto, M., P. Pallares, D. Gaertner, Delgado de Molina, A. Fonteneau and Y. J. Ariz (2002) Standardization of tropical purse seine fishing effort by generalized linear model (GLM). IOTC-WPTT/02/26, 15 pp.
- Stefansson, G. (1996) Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J. Mar. Sci., **53**, 577–588.
- Sugiura, N. (1978) Further analysis of the data by Akaike's information criterion and the finite corrections. Commun. Statist.-Theory. Meth., **7**(1), 13–26.
- Takahashi, N., S. Tsuji, T. Itoh, and H. Shono (2001) Abundance indices of southern bluefin tuna based on the Japanese longline fisheries data, 1969–2000, along the interim approach agreed for the 2001 stock assessment. CCSBT-SC/0108/28, 39 pp.
- 高橋正雄・大橋靖雄・芳賀敏郎 (1989) SASによる実験データの解析. 367 pp.
- 竹内 啓 (1976) 情報統計量の分布とモデルの適切さの規準. 数理科学, **153**, 12–18.
- 竹澤邦夫 (1999) バギング樹形モデルを用いたモデル合成. システム農学, **15**(1), 1–8.
- 竹澤邦夫 (2001) みんなのためのノンパラメトリック回帰. 吉岡書店, 560 pp.
- 寺野隆雄 (2002) データマイニングの展望. 計測と制御, **41**(5), 315–324.
- Toscas, P. and M. Thomas (1998). Spatial analysis of southern bluefin tuna catch per unit effort data: A best linear unbiased predictor approach. CCSBT-SC/9807/10, 33 pp.
- 豊田秀樹 (1998a) 共分散構造分析 [入門編]. 朝倉書店, 319 pp.
- 豊田秀樹 (1998b) 非線形多変量解析. 朝倉書店, 325 pp.
- 豊田秀樹 (2000) 共分散構造分析 [応用編]. 朝倉書店, 303 pp.
- 豊田秀樹 (2003) 共分散構造分析 [技術編]. 朝倉書店, 238 pp.
- 椿広計 (1988) 一般化線形模型の問題点と疑似尤度の一般化. 応用統計学, **17**(1), 1–12.
- Tsukimoto, H. (2000) Extracting rules from trained neural networks. IEEE Transactions on Neural networks, **11**(2), 377–389.
- 月本 洋・森田千絵 (2000) 予測モデルからのルール抽出. 発見科学とデータマイニング. 共立出版, 24–33.

- 内田 治 (2002) 例解データマイニング入門. 日本経済新聞社, 214 pp.
- Vapnik, V. (1998) Statistical learning theory. John Wiley & Sons, New York, 736 pp.
- Verbeke, G. and G. Molenberghs (1997) Linear mixed models in practice —A SAS oriented approach— Springer, New York, 306 pp.
- Venables, W. N. and P. J. Tascas (2002) Estimating a CPUE series for southern bluefin tuna using enhanced tree-based modeling methods. CCSBT-ESC/0209/31, 24 pp.
- Watters, G. and R. Deriso (2000) Catch per unit of effort of bigeye tuna: a new analysis with regression tree and simulated annealing. IATTC Bull., **21**(8), 531–571.
- Wedderburn, R. W. M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biomatrika*, **61**, 439–447.
- Whittaker, E. (1923) On a new method of graduation. Proceedings of the Edinburgh Mathematics Society, **41**, 63–75.
- Wise, B., A. Bugg, H. Shono, S. Barry, T. Nishida, D. Barratt and J. Kalish (2002) Standardization of Japanese longline catch rates for yellowfin tuna in the Indian Ocean using GAM analyses. IOTC-WPTT/02/11, 15 pp.
- Witten, I. H. and E. Frank (2000) Data mining: Practical machine learning tools and techniques with JAVA implementations. Morgan Kaufman, 371 pp.
- 山田作太郎・田中栄次 (1999) 水産資源解析学. 成山堂書店, 151 pp.
- Yokawa, K. and H. Shono (2000) Preliminary stock assessment of swordfish (*Xiphias gladius*) in the Indian Ocean. IOTC-WPB-00-02, 5 pp.
- Yokawa, K., M. Okazaki, H. Okamura, T. Matsumoto, Y. Uozumi and H. Saito (2002) An estimation of effective fishing effort of Japanese longliners on the Atlantic blue marlin, *Makaira nigricans*, in the Atlantic Ocean. Handbook and Abstract of Third International Billfish Symposium, 25.
- 吉富康成 (2002) ニューラルネットワーク. 朝倉書店, 174 pp.