

まぐろはえ縄データを用いた標準化 CPUE の推定

市野川 桃子（水産総合研究センター 遠洋水産研究所）

1. はじめに

日本のはえ縄漁業が提供する漁獲成績報告書は、漁場のカバー率が広く、諸外国に比べて精度が良く、50年以上の長期間にわたって整備されている。そのため、日本延縄漁業から得られる CPUE（漁獲尾数を単位努力量で割った指数）は、はえ縄漁業によって漁獲される多くの種の資源評価において、最も重要なインプットデータとして用いられている。しかし、延縄データのように長期間・広範囲にわたる漁業データから資源量指数としての CPUE を計算する場合、漁場や季節によって対象種の釣れやすさが異なる効果を統計的に取り除く（CPUE の標準化）必要がある。

本発表では、延縄漁業データでよく行われている CPUE の標準化の概念を簡単に説明し、標準化 CPUE を実際に推定する際に考慮すべきいくつかの点を紹介する。

2. 標準化 CPUE とは

一般に、CPUE は、資源豊度の相対的な増減を示す指標と考えられている。それは、「ある年(y)の漁獲量(C_y)は、海の中の資源(N_y)が多ければ多いほど、また、漁獲のために費やす努力量 (E_y) が多ければ多いほど、多くなる（式 1）」という単純な仮説に基づいている。

$$C_y = q E_y N_y \quad (1)$$

上式の両辺を努力量 (E_y) で除すると左辺が CPUE となり、つまり、CPUE は資源量と比例関係にある、という関係が得られる。

$$C_y / E_y (= \text{CPUE}_y) = q N_y \quad (2)$$

ここで、 N に乘じられている数 q は漁獲能率(catchability)と呼ばれ、魚の釣れやすさを示す係数である。

しかし、実際の漁業を考えてみると、魚の釣れやすさ(q)というのは、海域や時期、用いる漁具によって異なることが容易に想像できる。それに従って式 2 を書き直すと、CPUE または $\log(\text{CPUE})$ は、例えば

$$C_y / E_y (= \text{CPUE}_y) = q_a q_s q_g N_y \quad (3),$$

$$\log(C_y / E_y = \text{CPUE}_y) = \log(q_a) + \log(q_s) + \log(q_g) + \log(N_y) \quad (4)$$

のようになる。式 3 は、「海の中の資源が多ければ多いほど、漁獲のために費やす努力が多ければ多いほど、よい漁場 (q_a) や季節 (q_s)、漁具 (q_g) を選べば選ぶほど、漁獲量は多くなる」という仮説を表している。資源の相対的な豊度(N_y)だけでなく、漁場や季節、漁具といった様々な効果が混ざった生の CPUE（ノミナル CPUE、3 式の左辺）から、それらの要素を除いた相対的な資源量のみを示す指数（標準化 CPUE、 N_y)を推定する作業が CPUE の標準化である。標準化された CPUE は、ノミナル CPUE よりも、実際の資源の動向をより反映すると考えられている。

3. 標準化 CPUE の推定

式 4 は、左辺の観察データ（生の CPUE、応答変数）が、右辺の様々な説明変数の線形結合によって説明されていることを示している。但し、実際の観察データは様々な原因によってモデルによる予測値から外れることが考えられる。そこで、式 4 に適切な誤差項を加え（式 5）、観測値とモデルによる予測値の残差が仮定された誤差分布に最も尤もらしく近づくようなパラメータを統計的に推定（最尤法）

することで、観測された CPUE データから標準化 CPUE(N_y)を推定することができる。

$$\log(C/E = \text{CPUE}) = \log(q_a) + \log(q_s) + \log(q_e) + \log(N_y) + \text{誤差項} \quad (5)$$

式 5 のような単純な線形モデル (GLM) を用いた CPUE の標準化は、様々な統計パッケージを使って簡単に実施できる。例えば、フリーの統計ソフト'R'を用いる場合、以下のコマンドを 1 行打てばよい。

```
res <- glm(logcpue~effect1+effect2+effect3+as.factor(year), data=fishery.data)
```

ここでは、ある漁業データ (fishery.data) 内の応答変数 logcpue を、年の効果 year と 3 つの主効果 effect1, effect2, effect3 が説明している。

4. glm のコマンドの以前と以後

glm のコマンドを 1 行打つのは簡単である。しかし、glm コマンドの実行後は、推定された標準化 CPUE が、実際の資源量により近い適切な値として推定されているかどうかを、検討することが大切である。真の資源量が不明である以上、標準化 CPUE の確からしさは、GLM で推定されたパラメータが、いかに統計学的・漁業学的・生物学的に妥当であるかといった検討を積み上げることでしか判断する方法がないからである。モデルの統計的な妥当性は、各説明変数の有意性の検定や、情報量基準を用いたモデル選択 (cf. Burnham & Anderson 2002) により客観的に判断することができる。一方で、推定されたモデルが生物学的、漁業学的に妥当であるかという検討は、客観的に行うことが困難である分、しばしば忘れられがちである。

年以外の効果は適切に導入・推定されているか？ GLMには、資源量のトレンドを示す「年」以外の説明変数として、漁獲能率に影響を与えそうな説明変数 (海域や季節、漁具) が導入される。特に、まぐろやかじきのような高度回遊性魚類の場合、その CPUE の空間的分布 (図 1 のノミナル CPUE) は季節によって異なる。そのような場合は、季節による海域の CPUE の違いを、海域と季節の交互作用によって説明する必要がある。図 1 は、季節による CPUE の空間分布の違いが、濃淡で示された海域と季節の交互作用のパラメータで説明されていることを示している。

GLM に導入すべき説明変数は、AIC 等の情報量基準を用いたモデル選択によって機械的に取捨選択することも可能である。しかし、膨大な観測データからなる漁業データを用いる場合、情報量基準を用いたモデル選択では、解釈が難しい複雑すぎるモデルが選択される傾向が強い。そのため、GLM に導入する説明変数はできるだけ生物学的・漁業学的に解釈が可能なものだけに限り、推定されたパラメータが妥当であるかどうかを、既存の生態的知見やノミナル CPUE の分布等と照らし合わせ、常にチェックすることが理想的である。生物学的・漁業学的な解釈が難しい効果や交互作用については、固定効果でなくランダム効果としてモデルに取り入れる方法もある (Venables & Ripley 2002)。

残差 前述したように、GLMによる最尤推定では、予測値と観測値の誤差が仮定した誤差分布に似るようパラメータが推定される。そこで、残差の分布が仮定した誤差分布に類似しているか、誤差の偏りがないか等を診断する必要がある。例えば、図 2 のような偏った残差分布は、多くの観測値がモデルによって十分に説明できていないことを示唆している。偏った残差分布の原因と解決方法は様々であるが、ゼロキャッチデータが原因の場合はそれを回避するための方法がいくつか提案されている (Shono et al. 2008)。

5. おわりに

本発表では、はえ縄データを用いた CPUE の標準化の例を紹介したが、はえ縄以外の様々な漁業データにおいても標準化 CPUE は推定・利用されている。さらに、本文中でもいくつか紹介したが、CPUE の標準化の手法として、GLM だけでない様々な統計モデルも提案されている (cf. Maunder & Punt 2004, 庄野 2008)。一方で、標準化 CPUE が本当に真の資源量を反映しているのか、操業分布や漁獲対象種の歴史的なシフトによる潜在的なバイアス等、未だ解決されていない問題も多い。それでも、CPUE の標準化といった一連の統計的手法を漁業データに適用し、CPUE データがどのように生物学的・漁業学的に説明されるのかを考えることは、用いている漁業データ、ひいては、対象資源の動向に対する理解をより深く知る上で非常に重要かつ有用なプロセスであると言える。

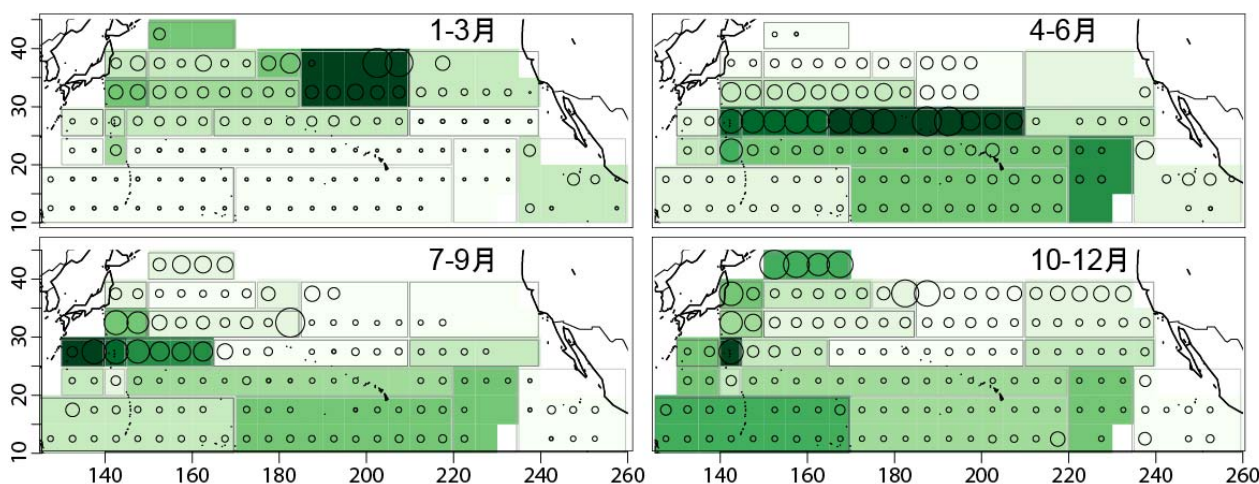


図 1. 北太平洋のメカジキの CPUE 分布。図中の丸の大きさは、緯度・経度ごとのノミナル CPUE (平均値) の大きさと比例し、背景の濃淡は季節と海域の交互作用で推定されたパラメータの大きさ (CPUE の予測値) と比例している。

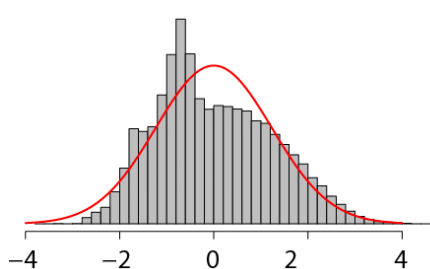


図 2. 偏った残差分布の例。メカジキ CPUE の GLM モデルより。図中の線は仮定した正規分布。この例における残差の偏りは、CPUE を対数変換する際に、メカジキの漁獲尾数がゼロであるデータ (ゼロキャッチデータ) に微小な数を足し込んだことによって生じている。

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. *Fish. Res.* **70**: 141-159.

Shono, H. 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. *Fish. Res.* **93**(1-2): 154-162.

Venables, W.N., and Dichmont, C.M. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fish. Res.* **70**(2-3): 319-337.

庄野宏. 2008. 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. 水研センター研報 **22**: 1-85.