



特集 2 生態学におけるモデル選択

予測、実証、モデル構築

箱山 洋

水産総合研究センター

Prediction, evidence, and model building

Hiroshi Hakoyama

Fisheries Research Agency

要旨：特集：生態学におけるモデル選択の総括として、3つの論点を議論した。第一に、予測において問題とする自然現象（確率変数）を明確にすることの重要性について論じた。第二に、自然の構造の科学的証拠を得るためのモデル選択の規準を論じた。サンプルサイズが小さい場合、AICは真のモデルの構造から遠いモデルを選ぶ傾向があり、科学的証拠としては弱点がある。近似による不一致は、サンプルサイズによらない真の分布と近似分布の乖離であることから、その推定量を科学的証拠のための規準として提案した。第三に、豊かなモデル構築の方法論と適切なモデル選択の規準が、データから有益な情報を引き出すことを、まとめとして論じた。

キーワード：モデル選択、科学的証拠、AIC、Linhart-Zucchini Criterion

はじめに

「特集：生態学におけるモデル選択」の4つの各論文（箱山 2015；粕谷 2015；巖佐 2015；岸野 2015）の総括として整理すべき主要な論点は3つある。

第一の論点は、予測において問題とする自然現象（確率変数）を明確にすることの重要性である。異なる現象（例えば、年齢別個体数と総個体数）の確率分布を扱うモデル選択規準を比較することはできない。また、同じモデルの比較でも異なる現象を扱えば、最善のモデルも変わり得る。例えば、個体群動態モデルでは、現象として個体数も絶滅確率も扱えるが、予測したい現象が個体数なのか、絶滅確率なのかで最善のモデルが変わり得る。Iwasa et al. (1987, 1989)；巖佐 (2015) のモデルの変数を束ねる一連のモデル・アグリゲーションの研究は、この問題に関連している。

第二の論点は、AIC (Akaike's 'A' information criterion ; Akaike 1973, 1974) など予測のためのモデル選択規準が科学的証拠 (scientific evidence) となりうるのかである。赤

池自身は、AICは統計的モデル同定のための多目的な方法を提供するものであり、その目的とは予測・シグナル検出・パターン認識・最適化であると述べている (Akaike 1974)。本特集では箱山 (2015) が、AICは予測のためのモデルを選ぶ規準であることを解説した。粕谷 (2015) はAICを検定として用いることの問題点について議論したが、ここでは、AICが自然の構造に関する、検定とは異なるある種の科学的証拠となりうるのかという問題について議論する。そのうえで、AICよりも優れた、科学的証拠となりうるモデル選択規準を提案する。

第三の論点は、対象の本質を考えた適切なモデル構築の重要性である。巖佐 (2015) は決定論の個体群・群集動態モデルについて、モデル・アグリゲーションの重要性を議論し、岸野 (2015) は、対比的に、因果の関係式だけではなく、確率モデルにおける誤差、すなわち確率分布の構造のモデリングの重要性を議論した。これらを踏まえて、予測、もしくは自然の理解のために、よいモデルとは何かについて簡単に議論する。

予測の為のモデル選択

AIC

ある単変量の確率変数 X について、真の分布 F とデータに当てはめる近似分布 G_θ の乖離を表す不一致 (discrepancy) の一つであるカルバック・ライブラー不一致 (Kullback-Leibler discrepancy; Linhart and Zucchini 1986, p-18) は、カルバック・ライブラー情報量 (Kullback-Leibler divergence; Kullback and Leibler 1951) のうち真の分布のみに依存する項を除いた量として、次のように定義される:

$$\Delta(\theta) = - \int f(x) \ln g_\theta(x) dx, \quad (1)$$

ただし、 $\theta=(\theta_1, \theta_2, \dots, \theta_p)'$ はデータに当てはめる近似モデル G_θ のパラメータ、 p はパラメーター数、 $f(x)$ は真の分布 F の確率密度関数、 $g_\theta(x)$ は近似モデルの確率密度関数を表す。 θ を最尤推定量で置き換えた不一致 $\Delta(\hat{\theta})$ は確率変数であり、その真の分布 F に関する期待値 $E_F \Delta(\hat{\theta})$ はサンプルサイズごとには異なるが、データには依存しない定数である。期待不一致 $E_F \Delta(\hat{\theta})$ が小さいモデルは予測の観点でよいと考えられるから、データから期待不一致 $E_F \Delta(\hat{\theta})$ を推定してモデルを選択するのが、予測のためのモデル選択である。期待不一致 (もしくは何らかの不一致の統計量) の推定量を規準と呼ぶ。

カルバック・ライブラー期待不一致を $2n$ 倍した定数の推定量である AIC は次のように定義される:

$$\text{AIC} = -2L(\hat{\theta}; \mathbf{x}) + 2p, \quad (2)$$

ただし、 $\mathbf{x}=(x_1, x_2, \dots, x_n)'$ はデータ、 $\hat{\theta}$ は最尤推定値、 $L(\hat{\theta}; \mathbf{x})$ は最大対数尤度である。着目した確率変数に関して、複数の近似確率分布モデルが考えられるとき、同一のデータに対して、近似モデルごとに AIC を計算し、AIC が最も小さいモデルを予測の観点で優れたモデルとして選択する。

問題とする確率分布を明確にする

基本的な AIC によるモデル選択のスキームは上で述べた通りであるが、巖佐 (2015) のモデル・アグリゲーションがモデリングの重要な本質の一つであることから、大規模なシステムの単純化などでは確率モデルの確率変数を束ねることも多いだろう。このとき、確率変数を束ねたモデルは違った現象を扱うモデルとなっており、単

純にそれぞれの対数尤度とパラメータ数を AIC の公式に代入して比較することはできない。

例えば、年齢構成モデルを考えてみよう。個体群に年齢構成 (最高年齢 m) があるとし、時刻 $t+1$ の個体数は時刻 t の個体数の条件付きで決まる離散時間のマルコフ過程を第一の近似モデル G_1 とすると、時刻 $t+1$ の各年齢の個体数 $\{N_{t+1,0}, N_{t+1,1}, \dots, N_{t+1,m}\}$ の条件付き同時確率密度関数は次のように表すことができる:

$$g_1(n_{t+1,0}, n_{t+1,1}, \dots, n_{t+1,m} \mid n_{t,0}, n_{t,1}, \dots, n_{t,m}). \quad (3)$$

一方、関心のある現象として総個体数 Z_{t+1} を考えよう:

$$Z_{t+1} = N_{t+1,0} + N_{t+1,1} + \dots + N_{t+1,m}.$$

この新たな確率変数 Z_{t+1} の真の分布に従う条件付き確率密度関数は、次のように表される:

$$\tilde{g}_1(z_{t+1} \mid n_{t,0}, n_{t,1}, \dots, n_{t,m}). \quad (4)$$

式 4 の確率分布は、条件に時刻 t の年齢別個体数が必要である。

ここで、確率変数が総個体数 Z だけで閉じた近似モデル G_2 を構築しよう。この近似モデルでは対象とする現象が総個体数であり、各年齢の個体数を知ることなく、総個体数の動態を扱うことができる。その条件付き確率密度関数は次のように表される:

$$g_2(z_{t+1} \mid z_t). \quad (5)$$

近似分布 $g_2(z_{t+1} \mid z_t)$ は一般には完全アグリゲーション (Iwasa et al. 1987) ではなく、年齢を考慮した近似モデル G_1 の分布から計算される Z_{t+1} の条件付き分布 $\tilde{g}_1(z_{t+1} \mid n_{t,0}, n_{t,1}, \dots, n_{t,m})$ とは異なっている。

年齢別個体数 (もしくは個体数の推定値) の時系列データが得られたとき、年齢構成モデル G_1 の尤度は年齢別個体数 $N_{t+1,0}, N_{t+1,1}, \dots, N_{t+1,m}$ の同時分布 (式 3) に対して構築される。一方、総個体数モデル G_2 の尤度は総個体数 Z_{t+1} の確率分布 (式 5) に対して構築される。両者は違った現象 (年齢別個体数と総個体数) を扱っていることから、それぞれ AIC を計算しても比較にならない。この場合に比較すべきは、着目した確率変数 Z_{t+1} のモデル G_1 に従う確率分布 (式 4) と G_2 に従う分布 (式 5) に基づいた規

準であり、単なる AIC での比較ではない個々のモデルに即したモデル選択の枠組みを考える必要がある。Seo and Kishino (2008, 2009) はこのような問題に対する一つの回答である。

決定論の線形動態モデルにおける短期予測と長期予測で最善の近似モデルが異なるという巖佐(2015)の結果は、個体群動態の確率モデルにおいても短期予測と長期予測で最善の近似モデルが異なる可能性を示唆している。1年後の個体数 N_{t+1} と L 年後の個体数 N_{t+L} は異なる確率分布を持っており、どのタイムスケールでの個体数の分布に着目してモデル選択するかが重要になる。また、個体群の絶滅リスクを考える場合では、個体数の分布の乖離でモデル選択するのではなく、真のモデルの絶滅確率の分布と近似モデルの絶滅確率の分布の乖離をもってモデルを選択する枠組みを作る必要がある。Focused information criterion (Claeskens and Hjort 2003) がそのような問題を扱う規準の一つであるが、個々の研究はこれからの課題である。

科学的証拠の為のモデル選択

科学的証拠に基づいて自然を理解する実証研究 (empirical research) は、科学の重要な部分を占める。ここでは自然の法則・構造を理解することが目的であり、将来予測は第一の目的ではない。これまで科学における実証研究では、主として、実験と仮説検定を基礎とした還元論的なアプローチから科学的証拠を得て、理論を進展させてきた。還元論的方法は大成功し、現在でも有効な科学的なアプローチであることに変わりはない。

しかしながら、大規模で複雑な要因が絡み合った自然のシステム (生態学的過程もその一つであるが) を理解しようとする、還元論的な方法だけでは十分ではない。なぜなら、生態学的過程について得られるデータは基本的には野外観察に基づいており、多くの場合、我々には野外観察のデータにモデルを当てはめることしかできない。また、仮に要素ごとに実験・仮説検定ができたとしても、それだけでは全体の複雑な因果の構造をうまく推論できないからである。

また、科学的研究では複数の仮説を考えることが重要である (Chamberlin 1890)。自然のシステムが複雑であるほど、現象を説明しうる仮説は増加する。したがって、複雑なシステムでは帰無仮説と対立仮説の二者択一の仮説検定ではなく、複数の仮説の確からしさを比較することが重要になる。実験のデータでも、複雑な問題では複

数仮説の比較が重要であり、岸野 (2015) の殺虫剤処理とタバコズメガの幼虫の個体数や、バイトウヒ幼木の成長へのオゾン暴露の影響はその例である。

これらの制約を背景に複雑な自然のシステムに関する科学的証拠を得る新しい方法論が求められている。この要求に対する一つの考え方として、概念的には次のようなスキームが考えられる：(1) 現象を確率変数で表現する。(2) 自然 = 真の構造・法則 (確率分布関数) が存在すると考える。(3) 真の構造に関する異なる仮説から複数の確率モデルを構築し、野外 (もしくは実験) データに当てはめる。(4) 真の構造と確率モデルの構造の近さを何らかの基準で計り、モデル間で比較する。(5) その構造に関する基準の近さが、真のモデルの構造に関する科学的証拠と考える。

カルバック・ライブラー期待不一致はこのようなスキームにおける基準として用いることができ、その推定量である AIC は複雑な自然の構造を推論するのにも役に立つという考え方がある (例えば、Burnham and Anderson 2002)。統計的検定の意味での科学的証拠とは異なるが、AIC は自然の構造に関する科学的証拠のひとつであると考えているのである。しかしながら、本特集で箱山 (2015) が解説したように、本来 AIC は予測のためのモデルの選択規準である。AIC でのモデル選択の結果を、対象とするシステムの構造についての科学的証拠と考えることは適切であろうか。

AICを自然の構造の科学的証拠と見なすことへの議論

粕谷 (2015) が指摘した生態学分野や日本生態学会での発表における数々の AIC の誤用は、AIC を検定の代用と考えるものである。このような理解不足に基づく誤用が問題なのではなく、AIC でのモデル選択の結果を、対象とするシステムの構造についての科学的証拠と考えることが妥当かが問題であり、この点を議論する必要がある。

箱山 (2015) が解説したように、AIC によるモデル選択の枠組みでは、真のモデル・近似モデル・サンプルサイズによって決定される最善のモデルが存在する。ここでの最善とは、モデルの期待不一致が最小という意味である。候補となる近似モデルの中でどのモデルが最善かは、サンプルサイズに応じて変わりうる。サンプルサイズが小さいとき、単純なモデルが最善となる傾向がある。

それぞれの近似モデルはその構造において、真の構造への近さが違う。例えば、箱山 (2015) の重回帰モデルの例では、単回帰モデル G_1 は切片しかないモデル G_2 よりも、真の重回帰モデル F に構造に近い。厳密には、モ

デルの構造が近いということの定義を与えなければならぬが (Iwasa et al. 1989, を見よ)、この場合は自明だろう。AIC が常に単回帰モデル G_1 を選ぶ傾向を持っていれば、AIC でのモデル選択の結果を科学的証拠と見なす考え方を支持する。しかしながら、AIC によるモデル選択は、必ずしも真のモデルの構造に近い近似モデルを選択する傾向を持つとは限らない。サンプルサイズが小さいときには、真のモデルの構造に遠い単純なモデル (箱山 (2015) の例では G_2) が選ばれる可能性が高い。どの程度のサンプルサイズがあれば、サンプルサイズが大きと言えるのかは対象のシステムによって異なる。即ち、我々のデータのサンプルサイズが小さいことは十分に考えられ、そのとき、AIC でのモデル選択ではモデルの構造について誤った推論をする可能性が高いのである。このことは、AIC が自然の構造に関する科学的証拠のひとつであるという主張の最大の弱点である。統計的検定と比べてみると、サンプルサイズが小さい場合、統計的検定では帰無仮説を捨てる可能性が小さくなり、判断を保留することが多くなるが、AIC では高い確率で積極的に真の構造から遠いモデルを選ぶという誤った推論をしやすい。少ないデータから言えることが制限されるのは当然であるから、前者は許容できるが、後者には問題がある。

では、サンプルサイズ n が十分に大きいときはどうであろうか。ある近似モデル G_0 のなかで、もっとも不一致が小さいパラメータを θ_0 とする。サンプルサイズが大きいつき、 $E_n \Delta(\hat{\theta})$ は「近似による不一致」 $\Delta(\theta_0)$ に近づくことから、AIC は $\Delta(\theta_0)$ を $2n$ 倍した値の不偏推定量でもある (詳しくは竹内 1976; Linhart and Zucchini 1986; 箱山 2015, を見よ)。

近似による不一致 $\Delta(\theta_0)$ は、ある近似モデルに不一致が最小となるベストのパラメータ θ_0 を与えた場合の不一致であり、その近似モデルのなかでの最善のモデルと真の分布との乖離である。この量は、データには依存せず、真のモデルと近似モデルによって決定される。前節で述べたように、我々に必要なのは、「真の構造と確率モデルの構造の近さを何らかの基準で計り、モデル間で比較する。」ことであったが、近似による不一致 $\Delta(\theta_0)$ はその要求を満たしているかもしれない。ある基準の大きさとモデルの構造の近さが一対一対応していれば、その基準はこの要求を満たしている。モデルが真の構造の本質を捉えていて、真のモデルの構造と確率モデルの構造が近いとき、近似による不一致 $\Delta(\theta_0)$ は小さくなる。逆に、近似による不一致 $\Delta(\theta_0)$ が小さいとき、真の構造と確率モデルの構造が近いと言えるだろうか? もし言えるとすれ

ば、近似による不一致 $\Delta(\theta_0)$ は真の構造と確率モデルの構造の近さを計る基準として適切である。異なる構造のモデルが、同一の $\Delta(\theta_0)$ を与えるという反例はすぐ見つかる。データからは仮説が見分けられず、別のモデルが同一の確率分布の形を与える場合がある (例としては、体長分布を年齢で分解する混合正規分布モデルの特別な場合が挙げられる)。また、あるモデルに構造として偽の冗長な部分を付け加えたモデルは同一の $\Delta(\theta_0)$ を持つ (例としては、重回帰モデルで真の係数が $\mathbf{0}$ の説明変数を加えることが挙げられる)。不一致はモデルの構造の結果であり、逆は必ずしも真ではない。しかし、おそらく多くの場合、適切な仮説を立てることで、近似による不一致 $\Delta(\theta_0)$ は構造の近さを計るよい基準となっているだろう。この点に関しては、構造が近いことの定義から考えて、さらに検討が必要であるが、もしそうだとすると、サンプルサイズが十分に大きいときは、AIC の比較はモデルの構造が真のモデルの構造に近いことのひとつの科学的証拠となっている。

自然の構造の科学的証拠となりうるモデル選択規準

そう考えると、近似による不一致 $\Delta(\theta_0)$ の推定量を考えれば、AIC より優れた自然の構造の科学的証拠となりうるモデル選択規準を構築できることに気がつく。

箱山 (2015) の式 9 によれば、 $\Delta(\theta_0)$ は次のように展開できる:

$$\Delta(\theta_0) \approx \Delta_n(\hat{\theta}) + \frac{1}{2}(\hat{\theta} - \theta_0)' J(\theta_0)(\hat{\theta} - \theta_0) + M,$$

ただし、 $\Delta_n(\hat{\theta}) = L(\hat{\theta}; \mathbf{x})/n$ は経験不一致、

$J(\theta_0) = -E_F[\partial^2 \ln g_0(\mathbf{x}) / (\partial \theta \partial \theta') |_{\theta=\theta_0}]$ 、 $M = \Delta(\theta_0) - \Delta_n(\theta_0)$ である。この式の両辺の期待値をとると、

$$\Delta(\theta_0) \approx E_F \left[\Delta_n(\hat{\theta}) \right] + \frac{1}{2n} \text{trace} \left[J(\theta_0)^{-1} I(\theta_0) \right],$$

ただし、 $I(\theta_0) = E_F[(\partial \ln g_0(\mathbf{x}) / \partial \theta)(\partial \ln g_0(\mathbf{x}) / \partial \theta') |_{\theta=\theta_0}]$ である。したがって、 $\Delta(\theta_0)$ の推定量は、

$$\widehat{\Delta(\theta_0)} = \Delta_n(\hat{\theta}) + \frac{1}{2n} \text{trace} \left[J(\hat{\theta})^{-1} I(\hat{\theta}) \right], \quad (6)$$

AIC の導出と同様に式 6 の右辺第二項のトレース項を p で近似し、両辺を $2n$ 倍すると:

$$\text{LZC} = -2L(\hat{\theta}; \mathbf{x}) + p. \quad (7)$$

Linhart and Zucchini (1986) が、この近似による不一致 $\Delta(\theta_0)$ の推定量を導いたことから、この推定量を Linhart-Zucchini Criterion (LZC) と呼ぼう。彼らはこの規準に名前を与えておらず、また、科学的証拠としての意味について深い考察は行っていない。Linhart and Zucchini (1986) は、最善のあてはまりを潜在的に与えるモデルを選びたいときに LZC での選択もあり得ると議論している。それは、与えられた標本によってその潜在性が達成できるかにかかわりなくである。彼らは LZC でのモデル選択の例として、より大規模な実験への予備的研究を挙げている。

LZC と AIC の違いは式 7 の右辺第二項が p か $2p$ だけの違いである。しかし、その意味は大きく違う。AIC がサンプルサイズで変化する期待不一致の推定量であったの対し、LZC はサンプルサイズによらない定数である近似による不一致 $\Delta(\theta_0)$ の推定量である。したがって、LZC でのモデルの比較ではサンプルサイズが小さいときに、バイアスの小さい範囲において、真の構造から遠いモデルを選択するという傾向はない。LZC での比較は、サンプルサイズに関わらず、近似による不一致 $\Delta(\theta_0)$ の推定値が小さなモデルを選ぶ傾向を持つ。これは、最善のパラメータ θ_0 を与えた場合に真の分布と近似分布の乖離が小さいと推定されるモデルを選ぶということである。従って、近似による不一致 $\Delta(\theta_0)$ が、近似モデルの構造と真のモデルの構造の近さを与えると考えられるならば、LZC は自然の構造の科学的証拠となるモデル選択規準として、AIC より優れている。一例として、箱山 (2015) の重回帰モデルの例を考えると、サンプルサイズが小さいとき、AIC では真の構造から遠い切片しかないモデル G_2 が高い確率で選択されるが、LZC では真の構造に近い単回帰モデル G_1 が選択される確率が比較的高く、望ましい性質を持っている。

サンプルサイズが小さいときは、近似による不一致 $\Delta(\theta_0)$ の推定値には大きなばらつきがある。このため、サンプルサイズが小さいとき、LZC での比較では、どのモデルが真の構造に近いのか見分けるのが難しく、LZC の大きさの単純な比較だけでモデルの順位付けをするのでは不十分である。従って、 m 個のモデルを比較するとき、LZC でのモデル選択では、近似による不一致がモデル間で等しいとする帰無仮説 H_0 :

$$H_0 : \Delta(\theta_{1,0}) = \Delta(\theta_{2,0}) = \dots = \Delta(\theta_{m,0})$$

をある種の方法で検定するという手順が必要となる。方向性のない対立仮説を持つ場合、個々のモデルの差を議

論するには事後検定を行う。具体的には、岸野 (2015) が AIC の信頼性評価について指摘したようなデータのサンプリング (ブートストラップ法) や何らかの漸近近似 (例えば、Kishino and Hasegawa 1989) で検定法を構築することになる。

$\Delta(\theta_0)$ の推定量として LZC には改善すべき点もある。第一に、AIC 同様に漸近近似から導いたため、サンプルサイズが小さいときバイアスがある。第二に、サンプルサイズが無限大でも、LZC でのモデル選択では、真の構造に最も近いモデルが確率 1 で選ばれないことがある。AIC 同様に (Shibata 1976)、最善のモデルに冗長な項が付け加わったモデルも一定確率で選ばれてしまう。最善のモデルと冗長なモデルで $\Delta(\theta_0)$ が等しいのだから当然のようであるが、 $\Delta(\theta_0)$ の推定量を工夫すれば、十分大きなサンプルサイズで最善のモデルが高い確率で選ばれるようにすることができるはずである。前者に関しては AICc (Sugiura 1978; Hurvich and Tsai 1989) のような補正、後者に関しては何らかのサンプルサイズ n の増加関数で式 7 の p の項の補正を行うことで (Bozdogan 1987, を見よ)、よりよい $\Delta(\theta_0)$ の推定量 (規準) を構築することができるだろう。

まとめ：自然の理解と予測のためのモデル構築

我々は対象を抽象化したモデルを通してのみ自然を理解できる。その意味では、いかにしてモデルの抽象化を行うかというモデル・アグリゲーション (巖佐 2015) は自然の理解における本質的なテーマである。また、よい予測が目的の場合でも、自然の構造の本質を捉えたモデルは予測精度は高いことから、複雑な対象をいかにうまく抽象化するかが重要である。

モデル選択の目的が予測であれ、自然の構造の科学的証拠を得ることであれ、モデルの候補を作るのは具体的に柔軟なモデル構築の方法論があつてのことである。岸野 (2015) が指摘したように、近年の統計学理論の発展や統計ソフト R の普及は、統計モデル構築における柔軟性と実現性を著しく向上させた。R. A. フィッシャーの分散分析の混合効果モデルは、現在、より柔軟に対象の構造をモデル化できる線形混合モデルや一般化線形混合モデルに発展し、計算機で最尤法や REML に基づいた推定が可能なのである (例えば、混合モデルの R での実装とその適用については Pinheiro and Bates 2000, を見よ)。

よい予測を目的とした AIC や、今回新たに議論した科学的証拠を目的とした LZC によるモデル選択は、このよ

うな膨大なモデル構築の可能性のなかで、それぞれの目的に近づく羅針盤となり、有益な情報をデータから引き出してくれる。

謝 辞

生態学会シンポジウム「生態学におけるモデル選択」および、本特集で、ご講演・ご執筆頂き、全体を通して有意義なコメントを頂いた岸野洋久先生、巖佐 庸先生、粕谷英一先生に感謝致します。また、お二人の査読者に感謝致します。この研究は、水産総合研究センターの運営費交付金の支援を受けた。

引 用 文 献

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Proceedings of the Second International Symposium on Information theory, 267-281, Akademinai Kiado
- Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19:716-723
- Bozdogan H (1987) Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52:345-370
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer Verlag, New York
- Chamberlin TC (1890) The method of multiple working hypotheses. Science, 15:92-96
- Claeskens G, Hjort NL (2003) The focused information criterion. Journal of the American Statistical Association, 98:900-916
- 箱山 洋 (2015) モデル選択と予測：その考え方と方法. 日本生態学会誌, 65:157-167
- Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. Biometrika, 76:297-307
- 巖佐 庸 (2015) 動態モデルにおける完全アグリゲーション：変数を束ねてもモデルの予測に誤りが生じないのはどのようなときか？ 日本生態学会誌, 65:169-177
- Iwasa Y, Andreasen V, Levin SA (1987) Aggregation in model ecosystems. I. Perfect aggregation. Ecological Modelling, 37:287-302
- Iwasa Y, Levin SA, Andreasen V (1989) Aggregation in model ecosystems II. Approximate aggregation. Ima Journal of Mathematics Applied In Medicine and Biology, 6:1-23
- 粕谷 英一 (2015) 生態学におけるAICの誤用-AICは正しいモデルを選ぶものではないので正しいモデルを選ばない. 日本生態学会誌, 65:179-185
- 岸野 洋久 (2015) 関係式のモデリングと誤差のモデリング. 日本生態学会誌, 65:187-196
- Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution, 29:170-179
- Kullback S, Leibler RA (1951) On information and sufficiency. The Annals of Mathematical Statistics, 22:79-86
- Linhart H, Zucchini W (1986) Model selection. Wiley, New York
- Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer-Verlag, New York
- Seo TK, Kishino H (2008) Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. Systematic Biology, 57:367-377
- Seo TK, Kishino H (2009) Statistical Comparison of Nucleotide, Amino Acid, and Codon Substitution Models for Evolutionary Analysis of Protein-Coding Sequences. Systematic Biology, 58:199-210
- Shibata R (1976) Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika, 63:117-126
- Sugiura N (1978) Further analysts of the data by Akaike's information criterion and the finite corrections. Communications in Statistics-Theory and Methods, 7:13-26
- 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