成長解析のための補助的手法としての Bayes 型モデルの応用

大西 修平,^{1*} 木曾 克裕²

(2003 年 10 月 29 日受付, 2004 年 3 月 26 日受理)

¹東海大学海洋学部, ²舩水産総合研究センター西海区水産研究所石垣支所

An application of a Bayesian model as an auxiliary method for analyzing growth

SHUHEI OHNISHI^{1*} AND KATSUHIRO KISO²

¹*School of Marine Science and Technology, Tokai University, Shizuoka, Shizuoka 424-8610, ²Ishigaki Tropical Station, Seikai National Fisheries Research Institute, Fisheries Research Agency, Ishigaki, Okinawa 907-0451, Japan*

Before constructing growth formulae with time-varying growth coefficient, the gain in body length should be examined using data. For this purpose, it is necessary to apply an auxiliary analysis model while constraining the fluctuation of body expansion in time. This paper presents an analysis model which has a flexible fitness for the fluctuation in length. In general, the time-varying growth coefficient has been described by a periodic function. On the contrary, we employed the first-order differences of growth coefficient, which follows the probability distribution. Here, we describe the parameter estimation and model selection based on Bayes' Theorem and Marginal Likelihood. In order to compare the result of the Bayesian approach with that given by the conventional way, we introduced the body length data of masu salmon. In a case where only unreliable growth information has been obtained, it is necessary to extract the biological characters from data, which is useful for model exploration and construction. Through the auxiliary utilization of the Bayesian model, the biological features related to the complicated mechanism of growth may be specified.

キーワード：成長式，成長係数，Bayes の定理，事前分布，周辺尤度，サクラマス

水産生物の成長式において，成長速度の変動が考慮される場合がある。Kiso *et al.*¹⁾では，成長係数 k に時間変化を与えるために周期関数が用いられた。そこでは，いくつかの周期関数が候補に挙げられ，モデルの妥当性の比較が統計的に行われた。成長係数を変動させる要因は，生理学的な研究手法によって解明できると思われる。それと同時に，体長データが正確に測定されていれば，データにもその特徴は潜在的に含まれる。したがって，関数の特定の形に拘束されないように k の時間変化を表現し，データ解析を通じて k の時間変化パターンが抽出できれば，潜在的な特徴の把握につながると期待できる。これにより，より詳細で現実的な成長解析が可能となる。本研究ではこのような，モデル比較の前段階として，データに基づいたモデルの探索が必要であるという考え方を前提に展開を進める。成長式のパラメータの時間変化を，特定の関数を使用せずに表した事例として

は，体長と体重の相対成長いわゆる allometry 式のパラメータの時間変化にランダムウォークを導入した例がある。²⁾ 本研究ではランダムウォークにあたる 1 次階差³⁾を用いて， k に時間変化を与えるモデルを展開する。成長係数を固定的な関数で扱う Kiso *et al.*¹⁾ のモデルでは，扱うデータの特性によっては適用上問題が起きる可能性があることを指摘し，特定の関数を用いない方法の必要性について述べる。最適モデルとされた Gompertz 式に基づいて展開をすすめ， k のモデリングの可能性についても考察する。

方 法

モデルの展開 t を一般的な齢とする。齢ごとの体長 l_t について，Gompertz 式の基本形 $l_t = l_\infty \exp[-\exp\{-k(t-t_0)\}]$ を用いて展開する。成長を考えるうえで起点となる齢を $t=\alpha$ において Gompertz 式を変形し，

成長係数 k に時間変化を許して $k(t)$ と一般化すれば,

$$\begin{aligned} & \log \{-\log(l_t/l_\infty)\} - \log(-\mu) \\ &= -\int_{\alpha}^t k(s) ds \quad (t \geq \alpha) \end{aligned} \quad (1)$$

となる。ただし l_∞ は最大体長, $\mu = \log(l_\alpha/l_\infty)$ である。(1)式の導出については **Appendix 1** に示した。ここで(1)式の左辺全体を $g(l_t)$ において, これを $g(l_t) = l'_t$ とする。右辺については齢を離散値で扱い, $\int_{\alpha}^{t+1} k(s) ds = k_t$ とし, さらに誤差 ε_t を考慮して, (1)式を線形モデル

$$l'_t = -\sum_{s=\alpha}^{t-1} k_s + \varepsilon_t \quad (t = \alpha+1, \alpha+2, \dots) \quad (2)$$

に書きなおす。次に, 齢 t における体長の観測値を l_t^* とし, 齢 t での体長のばらつきを誤差 δ_t と比例定数 ρ で, $l_t^* = l_t + \rho\delta_t$ と定義する。 l_t は Gompertz 式の理論体長である。 δ_t は正規分布 $N(0, w_t^2)$ にしたがう絶対値または相対値を仮定する。ここで $g(l_t^*)$ が $g(l_t^*) \cong al_t^* + b$ と直線近似できるような関数の場合を考えれば,

$$g(l_t^*) \cong al_t + b + \rho\delta_t \quad (3)$$

である。(2)式の l'_t は $l'_t = g(l_t) + \varepsilon_t$ であるので, これと(3)式の関係から, 観測値による $g(l_t^*)$ を(2)式の l'_t にあてはめれば,

$$al_t + b + \rho\delta_t \cong g(l_t) + \varepsilon_t$$

の関係が得られ, ここで $g(l_t) \cong al_t + b$ であることから $\varepsilon_t = \rho\delta_t$ となる。以上のことから, 観測値 l_t^* に正規分布 $N(l_t, \rho^2 w_t^2)$ を仮定するとき, $g(l_t^*)$ を l'_t に用いれば, (2)式の誤差項 ε_t も $\varepsilon_t \sim N(0, a^2 \rho^2 w_t^2)$ という正規分布にしたがうことになる。次に全齢に共通の誤差 ε を考え, $\varepsilon \sim N(0, \sigma^2)$ と正規分布を仮定し, $a^2 \rho^2 = \sigma^2$ とすれば, $\varepsilon_t \sim N(0, w_t^2 \sigma^2)$ と書くことができる。これをもとにすれば(2)式は

$$l'_t/w_t = -1/w_t \sum_{s=\alpha}^{t-1} k_s + \varepsilon \quad (t = \alpha+1, \alpha+2, \dots, T) \quad (4)$$

となる。ただし T は齢の最大値である。(4)式は重み付きモデルになっている。次に,

$$\mathbf{Y} = [l'_{\alpha+1}/w_{\alpha+1}, \dots, l'_T/w_T]', \quad \mathbf{K} = [k_\alpha, \dots, k_{T-1}]'$$

$$\mathbf{D}_1 = -\begin{bmatrix} w_{\alpha+1}^{-1} & 0 & \dots & 0 \\ w_{\alpha+2}^{-1} & w_{\alpha+2}^{-1} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ w_T^{-1} & \dots & \dots & w_T^{-1} \end{bmatrix}$$

とおく。太字はベクトルや行列で, これにプライム符号(') がつく場合は転置を意味する。ベクトルと行列を用いて, パラメータ $\mathbf{K}$, σ^2 の尤度関数 $L(\mathbf{K}, \sigma^2)$ は

$$L(\mathbf{K}, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{T-\alpha} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{D}_1 \mathbf{K}\|^2 \right\} \quad (5)$$

と記述できる。 $\|\cdot\|$ はベクトルのノルムである。ここで

$k_\alpha, \dots, k_{T-1}$ の時間変化は, 特定の関数を想定せず, 1次階差値 $v_t = k_t - k_{t-1}$ が確率分布に従う³⁾ という, 束縛の弱いモデルで表す。本研究では最も単純に正規分布により, $v_t \sim N(0, \sigma^2/\phi^2)$ とした。なお分散 σ^2/ϕ^2 は, 新たなパラメータ ϕ によって σ^2 をスケーリングしたもので, k_t の変動の大きさを調節する役割を持つ。さて $v_t (t = \alpha+1, \dots, T)$ が互いに独立ならば, v_t から時間変化傾向が排除されるので, 制約のない k_t の時間変化パターンが描ける。そこで同時分布 $\Pi_{t=\alpha+1}^T v_t$ すなわち

$$\begin{aligned} & \pi(\mathbf{K} | \phi, k_T, \sigma^2) \\ &= \left(\frac{\phi}{\sqrt{2\pi}\sigma} \right)^{T-\alpha} \exp \left\{ -\frac{\phi^2}{2\sigma^2} \|\mathbf{Z} - \mathbf{D}_2 \mathbf{K}\|^2 \right\} \end{aligned} \quad (6)$$

を k_t の時間変化モデルと定義した。なお $\mathbf{Z}$ は要素数 $T - \alpha$ の列ベクトル $\mathbf{Z} = [0, \dots, 0, k_T]'$, $\mathbf{D}_2$ は

$$\mathbf{D}_2 = \begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & 0 \\ & & \dots & \\ & 0 & & 1 & -1 \\ & & & & 1 \end{bmatrix}$$

という形式の $(T - \alpha)$ 次正方形行列である。 $\mathbf{D}_2$ は $-1, 0, 1$ だけで構成され, 対角成分は全て 1 である。

本モデルで推定の対象となる未知パラメータは, $\mathbf{K}$ と σ^2 であり, これらは尤度関数 $L(\mathbf{K}, \sigma^2)$ に含まれている。一方, パラメータ $\mathbf{K}$ の時間変化の特徴は, 特定の関数ではなく, (6)式のような確率分布で表現されている。 $\mathbf{K}$ の変化パターンが確率的に記述されているモデルに基づいて, 与えられた体長の観測値から, $\mathbf{K}$ と σ^2 を推定することが, 本研究で扱う問題の本質である。このような, 尤度関数の未知パラメータが確率分布に従うモデルは, Bayes 型モデル⁴⁾ と総称され, Bayes の定理を応用してパラメータ推定を行うことができる。データが与えられるとき, Bayes の定理に基づき(6)式をパラメータ $\mathbf{K}$ の事前分布として, (5)式の尤度関数との積 $L(\mathbf{K}, \sigma^2) \pi(\mathbf{K} | \phi, k_T, \sigma^2)$ によって, 事後分布 $p(\mathbf{K})$ が

$$p(\mathbf{K}) \propto L(\mathbf{K}, \sigma^2) \pi(\mathbf{K} | \phi, k_T, \sigma^2) \quad (7)$$

と計算できることを利用する。なお事前分布だけを規定する未知パラメータ ϕ, k_T は, 超パラメータと呼ばれる。⁴⁾ 事後分布 $p(\mathbf{K})$ は(6)式で表されるパラメータ $\mathbf{K}$ の挙動を考慮に入れながら, 同時に $\mathbf{K}$ の出現頻度を表現する機能をもっている。 $L(\cdot) \pi(\cdot | \cdot)$ が単峰で十分に尖っていれば, 事後分布のモードを代表値とみなして, $\mathbf{K}$ の推定値とすることが実用的である。⁴⁾ 本研究の尤度関数, 事前分布ともに正規分布を前提としており, このことから $L(\cdot) \pi(\cdot | \cdot)$ も単峰で, モードによる解法が適切と判断できる。(7)式は具体的には

$$p(\mathbf{K}) \propto \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{2(T-\alpha)} \phi^{T-\alpha} \exp \left[-\frac{1}{2\sigma^2} \left\{ \|\mathbf{Y} - \mathbf{D}_1 \mathbf{K}\|^2 + \phi^2 \|\mathbf{Z} - \mathbf{D}_2 \mathbf{K}\|^2 \right\} \right] \quad (8)$$

と書けるので、事後分布のモードに対応する $\mathbf{K}$ を、 $\{\bullet\}$ 内の最小化によって求めればよい。

ベクトルと行列を整理して、

$$\tilde{\mathbf{Y}} = [\mathbf{Y}', 0, \dots, 0]', \quad \tilde{\mathbf{K}} = [\mathbf{K}', k_T]',$$

$$\mathbf{G} = \begin{bmatrix} & 0 \\ & \vdots \\ \mathbf{D}_1 & \vdots \\ & 0 \\ \hline & 0 \\ \phi \mathbf{D}_2 & \vdots \\ & 0 \\ & -\phi \end{bmatrix}$$

と再構成し、これらを用いて(8)式の $\{\bullet\}$ 内を

$$\|\mathbf{Y} - \mathbf{D}_1 \mathbf{K}\|^2 + \phi^2 \|\mathbf{Z} - \mathbf{D}_2 \mathbf{K}\|^2 = \|\tilde{\mathbf{Y}} - \mathbf{G} \tilde{\mathbf{K}}\|^2$$

と簡略化できる。 $\tilde{\mathbf{Y}}$ は $\mathbf{Y}$ にゼロベクトルを結合した、要素数 $2(T-\alpha)$ の列ベクトルである。 $\tilde{\mathbf{K}}$ は $\mathbf{K}$ に超パラメータ k_T を付加したものである。また $\mathbf{G}$ は $2(T-\alpha) \times (T-\alpha+1)$ の行列である。このような書きかえにより、(8)式 $\{\bullet\}$ 内の最小化において、超パラメータ k_T がパラメータ $\mathbf{K}$ と一緒に扱える利点がある。 ϕ を与えたときの $\|\tilde{\mathbf{Y}} - \mathbf{G} \tilde{\mathbf{K}}\|^2$ の最小化と、それに伴う解 $\tilde{\mathbf{K}}^*$ の推定は、Householder 変換によるアルゴリズム⁵⁾を用いて実行できる。本研究の数値計算でも同アルゴリズムを用いた。

数値計算について、残る問題はパラメータ σ^2 と超パラメータ ϕ の決定であるが、これについては周辺尤度 (Marginal Likelihood: ML)⁴⁾の最大化に基づくことができる。周辺尤度 ML は

$$ML = \int \cdots \int L(k_\alpha, \dots, k_{T-1}, \sigma^2) \times \pi(k_\alpha, \dots, k_{T-1} | \phi, k_T, \sigma^2) dk_\alpha, \dots, dk_{T-1} \quad (9)$$

と定義される。実用上は(9)式の数値 $\log ML$ の最大化を考えればよい。本研究では(8)式右辺の積分の対数を最大化するように σ^2 と ϕ を決定する。周辺尤度の具体的な導出と、パラメータ推定手順については **Appendix 2** に示した。

Bayes 型モデルによる解析 適用例として、Kiso *et al.*¹⁾のサクラマス (*Oncorhynchus masou*) の成長解析で扱われたデータに対する解析を試みた。同報告によれば、本州北部太平洋岸、特に三陸地方南部の河川に生息するサクラマス個体群の雌には、河川に終生留まる河川残留型 (以下、残留型とする) と、降海して海洋で生活する降海型と、2 型の生活史が確認されている。同水域

でのサクラマスの産卵盛期は 10 月である。0 歳魚の夏季以降の成長速度の個体差が、2 型への分化に関与している可能性が指摘されてきた。⁶⁾ Kiso *et al.*¹⁾は、降海型の降海が終わる時期までの、若齢魚の体長測定データを、2 つの型に区別したうえで成長解析を行った。河川の水温の変動は周期的な季節性を持ち、成長速度の変化とも関連が強い。⁷⁾ Kiso *et al.*¹⁾では、成長係数の時間変化を三角関数で表したモデルを用いて、最尤法によるパラメータ推定を行っている。AIC によるモデル比較の結果、Gompertz 式を土台に、成長係数に周期的変化を与えたモデルが最適モデルに選択されている。

ここで本研究では、先に理論的展開を示した Bayes 型モデルで、Kiso *et al.*¹⁾の体長測定データを解析した。成長係数を推定し、生物学的知見との整合性について検討する。この解析事例を通じて、Bayes 型モデルについて考察する。データの性質によっては、周期関数で成長係数の変動を表すことに問題がおきる可能性にふれ、Bayes 型モデルの必要性を述べる。

Kiso *et al.*¹⁾のモデルと Bayes 型モデルの違いを比較するために、入力データは全て Kiso *et al.*¹⁾に表示されている値を用いた。 t は月齢に対応し、データは月齢 2 から 19 にわたっており、 $\alpha=2$, $T=19$ である。なお $T=19$ は、降海型になった個体の大部分の降海が終了する月齢にあたる。入力データは l_t^* と w_t^2 であり、 l_t^* は月齢ごとの平均尾叉長、 w_t^2 は月齢ごとの l_t^* の分散推定値である。入力データ $\mathbf{Y}$ は、(1)式左辺にならって、 $g(l_t^*)$ に変換した値と w_t^2 より求める。 l_∞ は一般的には推定の対象であるが、Kiso *et al.*¹⁾では、降海型については観測された後期スモルトの尾叉長最大値 17.70 cm を、残留型については、同水域で過去実施された成長研究での最大体長推定値 24.12 cm⁸⁾を、いずれも l_∞ として所与としている。したがって本研究で推定の対象となるパラメータは、月齢ごとの成長係数 k_t ($t=2, \dots, 18$)、分散推定値 σ^2 、さらに超パラメータ k_{19} と ϕ である。数値計算プログラムは、MATLAB (MathWorks, Inc.) 上でプログラミングしたものを使用した。

結果および考察

$g(l_t)$ が直線近似できる性質について 本研究で扱っている Bayes 型モデルは、(1)式の左辺、すなわち $g(l_t)$ が近似的に直線とみなせる関数であることを前提とする。このことの確認のために、 l_t^* ($t=2, \dots, 19$) を、(1)式左辺に代入したものを Fig. 1 にプロットした。Fig. 1 の点の配置から $g(l_t)$ の大まかな関数形が読み取れ、残留型と降海型とも、 $g(l_t)$ は直線から大きくは逸脱しないと判断した。

ここで $g(l_t)$ の直線性が成立しやすい条件を考える。 $g(l_t)$ を直線で置き換えるためには、広範囲の l_t の値に

Fig. 1 Relationship between fork length l_t and $g(l_t)$ which consists of log transformation. Solid and open circles correspond to the data of fluvial form and sea-run form, respectively.

ついて、 $g(l_t)$ の曲率がゼロになっていれば都合が良い。曲率は $g(l_t)$ の変曲点の近傍において 0 に近い値をとる。(1)式に戻り $l_t/l_\infty = x_t$ として、 $g(x_t) = \log(-\log x_t) - \log(-\mu)$ ($0 < x_t < 1$) に注目する。 $g(x_t)$ の 2 次導関数 $g''(x_t)$ を求めると、 $g''(x_t) = -(\log x_t + 1)/(x_t \log x_t)^2$ となる。変曲点を求めるために $g''(x_t) = 0$ において解けば、変曲点は $x_t = e^{-1}$ であることがわかる。したがって $l_t = l_\infty e^{-1}$ の近傍では $g(l_t)$ は直線に置換えやすい。 $e^{-1} \approx 0.37$ であるので、おおむね l_∞ に対して 30~40% の体長サイズのデータを中央値にもつような標本が、モデルの前提条件をみたすデータといえる。ただし、 $g(l_t)$ の関数の特性からみて、 l_t が 0 や l_∞ に近い部分では関数の変化率が急激に大きくなる。そのため、解析対象となる標本に、このような発生直後の若齢魚や l_∞ に近い高齢魚がなるべく含まれないことも必要である。Fig. 1 では、降海型のデータが残留型のデータに比べて直線近似の程度がやや悪い。2 型それぞれのデータの最大値と最小値により、それぞれの最大体長から比 x_t を求めると、残留型では $0.074 \leq x_t \leq 0.697$ であるのに対して、降海型では $0.101 \leq x_t \leq 0.859$ であった。0.859 のように、1.0 に近い値を含む標本は、 $g(l_t)$ の変化率の大きくなる部分に接近しており、直線への近似が悪かったものと思われる。

成長式に Bertalanffy 式を用いる場合についても簡単にふれる。Bertalanffy 式では、簡単な変形によって、(1)式に相当する $\log(1 - l_t/l_\infty) - \log \mu = -\int_0^t k(s) ds$ を導くことができる。ただし $\mu = 1 - l_0/l_\infty$ である。Gompertz 式の場合にならい、 $l_t/l_\infty = x_t$ において、 $g(x_t) = \log(1 - x_t) - \log \mu$ ($0 < x_t < 1$) に注目する。 $g(x_t)$ の 2 次導関数は、 $g''(x_t) = -(1 - x_t)^{-2}$ となる。 $g''(x_t)$ の定義範囲が $0 < x_t < 1$ であることから、 $g''(x_t) = 0$ となる点は存在せず、したがって $g(x_t)$ は変曲点をもたない。この

Fig. 2 Estimators of time-varying growth coefficient.

Lines with and without marker correspond to the estimator of the Bayesian model and the periodic function model, respectively. Bars indicate $\pm$ SD based on the variance-covariance matrix of the posterior distribution. Except $k_2, \dots, k_{18}$, unknown parameters were k_{19} , ϕ , and σ^2 . Estimators obtained for the two forms were as follows. For the fluvial form, $k_{19} = 0.104$, $\phi = 1.052$, and $\sigma^2 = 4.089 \times 10^{-3}$ were estimated. For the sea-run form, $k_{19} = 0.183$, $\phi = 1.233$, and $\sigma^2 = 6.781 \times 10^{-3}$ were estimated. The maximum value of log ML given by the optimal model for fluvial form and sea-run form was 8.532 and 5.569, respectively.

性質に基づけば、Bertalanffy 式の場合は、Gompertz 式の場合に比べて、 $g(l_t)$ の直線性が成り立つ範囲を明確に定義することができない。したがって、なるべく狭い体長範囲のデータを解析対象とするように注意が必要である。

解析結果における Bayes 型モデルと周期関数モデルの比較 Kiso *et al.*¹⁾ は、成長係数の時間変化を周期関数で表し、モデル比較を行い、残留型、降海型ともに

$$l_t = l_\infty \exp[-\exp\{-k(F(t) - F(t_0))\}]$$

$$F(t) = \frac{t}{12} + \frac{A}{2\pi} \sin \frac{2\pi}{12} (t - t_1) + \frac{C}{6\pi} \sin \frac{6\pi}{12} (t - t_1)$$

と、成長係数が 12 ヶ月と 4 ヶ月の周期変動を持つ Gompertz 式を最適モデルとして選択している。ここで、

k, A, C, t_0, t_1 に推定値を代入して描いた成長係数の周期変化と、本研究の Bayes 型モデルによる k_t 推定値を対比させて Fig. 2 に示した。なお Bayes 型モデルによる推定値には、事後分布の分散 ($\sigma^2(\mathbf{D}_3'\mathbf{D}_3)^{-1}$ の対角成分) から標準偏差を求めて、推定精度の目安として示している。

Fig. 2 の周期関数には、残留型、降海型ともに月齢 6 から 8 付近にかけて成長速度に急激な停滞がみられる。また月齢 12 から 14 付近にかけては、逆に急激な上昇が示されている。この 2 つの変化については、生物学的には説明がつかないことが指摘されている。¹⁾ 2 つの変化が生じた理由として、月齢 2 から 5 の若齢時の尾叉長の分散が特に小さいこと、さらにパラメータ推定が重み付き最小二乗法に基づいていることが挙げられている。¹⁾ すなわち、Fig. 2 の周期関数は、解析対象の月齢 2 から 19 の尾叉長データのうち、分散の値の小さい月齢 2 から 5 のデータへの、あてはまりが特に良い関数と考えることができる。残差平方和に占める、月齢 2 から 5 のデータによる残差平方の影響が強調された結果という意味になる。したがって、先に述べた 2 つの急激な変化 (月齢 6-8, 12-14) は、月齢 2 から 5 のデータを説明するための形式的な変化と考えられる。周期関数では、月齢 8 から 10 付近の値が増加しているが、Bayes 型モデルでは一方的に減少しており、逆の変化となっている。これについても、周期関数の変化は生物学的に説明のつかないもので、月齢 2 から 5 のデータの説明のための、形式的な変化のひとつと考えられる。本来、残留型と降海型の解析結果を比較する目的は、2 型の成長速度の違いを通して、2 型の分化の時期を探ることであった。ところが、若齢部分の成長速度の変化の説明に都合の良い関数形が、全ての齢に共通で使われた。さらにデータの若齢部分は、残留型と降海型で共通である¹⁾ ことも関連して、Fig. 2 の周期関数では振幅に違いがみられるだけで、位相には顕著な違いがみられない。したがって、ここで扱われた材料の場合、固定的な周期関数による結果からは、2 型の分化の時期を探ることは困難であると考えられる。特定の周期関数を規定するモデルでは、このように、データの説明に不可欠な周期と形式的な周期が、分離されず扱われてしまう場合が存在する。モデル評価は情報量規準に基づくことができるが、生物学的に有意義な解析結果が常に与えられる保証はない。

Bayes 型モデルでは、 k_t の挙動は事前分布だけで特徴付けられる。しかも (6) 式のように、各月齢の v_t を独立に扱うため、 k_t の動き方は制約が少ないものである。Bayes 型モデルも、(4) 式のように本質的に重み付きモデルになっているので、 $k_t(t=2, \dots, 5)$ の推定値は、月齢 2 から 5 のデータを再現できる値が選ばれる。しか

しながらそれ以外の月齢については、月齢 2 から 5 のデータの影響は及ばない。Bayes 型モデルによる結果は、上に述べたような周期関数の形状と生物学的知見が整合しにくいこととは逆に、生物学的に説明のつきやすいものとなっている。月齢 6 から 8 付近にかけての変化と、月齢 12 から 14 付近にかけての変化は緩やかで、木曾⁹⁾ が述べているような、成長速度の変化が水温の変化と連動している可能性を示唆するものとなっている。また、降海型が海洋生活期に移行する段階で、大型個体ほど早い時期に降海することが述べられている。⁹⁾ これに対応して、Fig. 2 の降海型の成長速度は、月齢 12 から大部分の個体が降海する月齢 19 (成長速度としては k_{18}) にかけて大きくなっており、この時期の変化は残留型と降海型で異なるパターンを見せている。この点でも Bayes 型モデルによる結果は、生物学的知見と対応がつきやすいものである。当該水域でのサクラマス¹⁰⁾ の生物学的研究では、生殖腺の発達や餌選択性についての知見も蓄積されている。⁹⁾ これら生理学的な動態が成長に与える影響についても定量的に明らかになれば、Bayes 型モデルで得られた成長速度の情報とあわせて吟味することで、2 型の分化のもとになる生理的条件について探ることが可能である。Bayes 型モデルの基本的な考え方は、関数形にとらわれることなく、データを平滑化することであるため、推定結果を様々な生物学的問題に活用することが可能である。

採集される生物の体長のばらつきには、生物自身の行動や生態の影響が無視できない。生物の行動が原因で、採集が特定のサイズの個体に偏ったものとなれば、生息個体の体長が均一であるという誤解を生む。ここでの分散は過小となる。仮にこのような内容のサンプリングが数ヶ月間続いたようなデータに対して、特定の関数で成長係数の変化を表したモデルをあてはめることを考える。この場合も、上で述べた、月齢 2 から 5 のデータにウェイトが置かれる例と同様、分散が過小になる期間のデータの影響を強く受けた関数形が選択される。ひとまとまりの魚群を構成する個体のサイズが揃っている場合はこのような状況が考えられる。また、餌を用いた釣によるサンプリングの場合、成長の段階に応じて餌を選択する生物に対しては、特定のサイズの個体が選択的に釣られやすいことがある。また釣り漁具の選択性も無視できず、やはり同様に、特定のサイズの個体が重点的に採集されることがありうる。このように、採集個体のサイズの偏りから、ある期間データの分散が過小になっても、Bayes 型モデルの場合 k_t の動きが自由であるため、その期間のデータが推定結果全体に与える影響は小さいものと思われる。

パラメータの変化率のモデリング 月齢の間のパラメータ変化の大きさについては、階差の次数を変えるこ

とで調節できる。1次階差は、パラメータの局所的な変化を0に押えるためのものである。一方で2次階差は、パラメータの局所的な変化は直線であればよい。したがって階差の次数の違いによって、扱える変化率は異なる。パラメータの1次階差によるモデルは中村¹⁰⁾の事例などに見ることができるが、石黒、荒畑¹¹⁾や石黒¹²⁾のように、一般的には2次階差が用いられることが多い。本研究では1次階差だけを用いたが、2次階差もモデルの候補に含め、比較検討する必要がある。

パラメータの変化率は、超パラメータ ϕ によっても調節できる。急速な成長をする期間と、ほとんど成長しない期間があり、2者がごく短時間で切り替わるような場合を考える。このような成長様式では、パラメータの急激な変化と0に近い変化を、ひとつのモデルの中で扱うことが求められる。このような問題については、期を区切り、それぞれの期で個別の超パラメータ ϕ を使用することで、対応が可能である。期ごとにパラメータの変化率を変えて表現できるため、急激な変化も扱うことができる。

成長速度は様々な環境要因から、複雑な制御をうけているものと思われる。したがって、成長係数の時間変化を固定的な関数で扱う考え方では、重要な特性の見落としにつながるおそれがある。Bayes型モデルのような、成長係数の動きに制約が少ない柔軟な方法を、成長データ解析の初期の段階で使用し、成長速度の複雑な変化を細部にわたり細かく眺めることができる。そこで得られる情報を環境特性データとあわせて検討することで、成長と環境要因の関係が徐々に明らかになる。さらに解析を積み重ねることで、最終的に生物学的に意義を持った適切な関数形が、成長式として明らかにされると期待できる。Bayes型モデルは、このようにプロセスを実現するための、補助的ツールと位置づけることができる。Bayes型モデルの応用範囲は広いと考えられる。様々な適用と拡張がさらに必要である。

謝 辞

本稿の丁寧な査読を通して、貴重な助言を頂戴した2名の匿名の査読者に、深く感謝いたします。

文 献

- 1) Kiso K, Akamine T, Ohnishi S, Matsumiya Y. Mathematical examinations of the growth of sea-run and fluvial forms of female masu salmon *Oncorhynchus masou* in rivers of the southern Sanriku district, Honshu, Japan. *Nippon Suisan Gakkaishi* 1992; **58**: 1779-1784.
- 2) Szalai EB, Fleischer GW, Bence JR. Modeling time-varying growth using a generalized von Bertalanffy model with application to bloater (*Coregonus hoyi*) growth dynamics in Lake Michigan. *Can. J. Fish. Aquat. Sci.* 2003; **60**: 55-66.

- 3) Akaike H. Likelihood and the Bayes procedure with discussion. In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds) *Bayesian Statistics*. University Press, Valencia. 1980; 143-166.
- 4) 坂元慶行. 「カテゴリカルデータのモデル分析」共立出版, 東京. 1985; 12-15.
- 5) 坂元慶行, 石黒真木夫, 北川源四郎. 「情報量統計学」共立出版, 東京. 1983; 149-154.
- 6) 久保達郎. 北海道のサクラマス我的生活史に関する研究. 北海道さけ・ますふ化場研究報告 1980; **34**: 1-95.
- 7) 赤嶺達郎. 水産資源学における成長式に関する数理的な研究. 中央水研研報 1995; **7**: 189-263.
- 8) Kiso K, Matsumiya Y. Growth of fluvial form masu salmon *Oncorhynchus masou* in southern Sanriku district, Honshu, Japan. *Nippon Suisan Gakkaishi* 1992; **58**: 9-13.
- 9) 木曾克裕. 本州北部太平洋岸の河川を母川とするサクラマスの生活史の研究. 中央水研研報 1995; **7**: 1-188.
- 10) 中村 隆. ベイズ型コウホート・モデル—標準コウホート表への適用—. 統計数理研究所彙報 1982; **29**: 77-97.
- 11) 石黒真木夫, 荒畑恵美子. ベイズ型スプライン回帰. 統計数理研究所彙報 1982; **30**: 29-36.
- 12) 石黒真木夫. ベイズ型季節調整プログラム BAYSEA. 統計数理 1986; **34**: 241-254.

Appendix 1 (1)式の導出について

元の式を

$$l_t = l_\infty \exp [-\exp \{-k(t-t_0)\}] \quad (\text{A1-1})$$

として、 $t = \alpha$ のとき

$$l_\alpha = l_\infty \exp [-\exp \{-k(\alpha-t_0)\}] \quad (\text{A1-2})$$

を得る。(A1-2)式を変形して両辺の対数をとれば

$$\log (l_\alpha / l_\infty) = -\exp \{-k(\alpha-t_0)\} \quad (\text{A1-3})$$

となる。ここで(A1-3)式の左辺を $\log (l_\alpha / l_\infty) = \mu$ とおいて

$$\exp (kt_0) = -\mu \exp (k\alpha) \quad (\text{A1-4})$$

と表し、これを(A1-1)式に代入すれば t_0 が消えるので、

$$l_t = l_\infty \exp [\mu \exp \{-k(t-\alpha)\}] \quad (\text{A1-5})$$

となる。さらに(A1-5)式を変形して対数を2回とれば

$$\log \{-\log (l_t / l_\infty)\} = \log (-\mu) - k(t-\alpha) \quad (\text{A1-6})$$

となる。ここで(A1-6)式の k を時間 t の関数として一般化し、両辺を整理すれば

$$\log \{-\log (l_t / l_\infty)\} - \log (-\mu) = -\int_\alpha^t k(s) ds \quad (\text{A1-7})$$

と(1)式が導かれる。

Appendix 2 周辺尤度の導出とパラメータ推定手順について

周辺尤度 ML の導出を考える。 ML は実用上その対数 $\log ML$ が用いられる。(9)式にしたがえば、 ML の導出には、(8)式を $\mathbf{K}$ について積分する必要がある。最初に、

$$\mathbf{X} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix}, \quad \mathbf{D}_3 = \begin{bmatrix} \mathbf{D}_1 \\ \phi \mathbf{D}_2 \end{bmatrix}$$

という新たな定義を用いて、(8)式の指数関数の部分を

$$\begin{aligned} & \exp \left[-\frac{1}{2\sigma^2} \left\{ \|\mathbf{Y} - \mathbf{D}_1 \mathbf{K}\|^2 + \phi^2 \|\mathbf{Z} - \mathbf{D}_2 \mathbf{K}\|^2 \right\} \right] \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{D}_3 \mathbf{K}\|^2 \right\} \end{aligned} \quad (\text{A2-1})$$

と書きなおす。(A2-1)式右辺の $\|\cdot\|^2$ について、これを最小にするような $\mathbf{K}$ を $\mathbf{K}^*$ とすると、最小二乗解の幾何学的な特性を用いて

$$\|\mathbf{X} - \mathbf{D}_3 \mathbf{K}\|^2 = \|\mathbf{X} - \mathbf{D}_3 \mathbf{K}^*\|^2 + \|\mathbf{D}_3 \mathbf{K} - \mathbf{D}_3 \mathbf{K}^*\|^2 \quad (\text{A2-2})$$

と分解することができる。(A2-2)式の関係を用いたうえ、(A2-1)式右辺を $\mathbf{K}$ で積分すれば

$$\begin{aligned} & \int \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{D}_3 \mathbf{K}\|^2 \right\} d\mathbf{K} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{D}_3 \mathbf{K}^*\|^2 \right\} \\ & \quad \times \int \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{D}_3 \mathbf{K} - \mathbf{D}_3 \mathbf{K}^*\|^2 \right\} d\mathbf{K} \end{aligned} \quad (\text{A2-3})$$

となる。ここで(A2-3)式右辺の積分が $(T-\alpha)$ 次元正規分布 $N(\mathbf{K}^*, \sigma^2(\mathbf{D}_3' \mathbf{D}_3)^{-1})$ の指数部分の積分になっていることを利用して、(A2-3)式は

$$\begin{aligned} & \int \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{D}_3 \mathbf{K}\|^2 \right\} d\mathbf{K} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{D}_3 \mathbf{K}^*\|^2 \right\} \left[\left(\frac{1}{\sqrt{2\pi\sigma}} \right)^{T-\alpha} \right. \\ & \quad \left. \times (\det |\mathbf{D}_3' \mathbf{D}_3|)^{1/2} \right]^{-1} \end{aligned} \quad (\text{A2-4})$$

と書きなおすことができる。 $\det|\cdot|$ は行列式の値である。(A2-4)式は(8)式の指数関数の積分値に相当する。

したがって(8)式右辺と(A2-4)式から、(9)式で定義される周辺尤度 ML が求められ、その対数をとって

$$\begin{aligned} \log ML &= -\frac{T-\alpha}{2} \log(2\pi\sigma^2) + \frac{T-\alpha}{2} \log \phi^2 \\ & \quad - \frac{1}{2\sigma^2} \|\mathbf{X} - \mathbf{D}_3 \mathbf{K}^*\|^2 - \frac{1}{2} \log(\det |\mathbf{D}_3' \mathbf{D}_3|) \end{aligned} \quad (\text{A2-5})$$

と表される。(A2-5)には σ^2 が未知パラメータとして含まれるが、これについては $\log ML$ を最大化する条件

$$\frac{\partial \log ML}{\partial \sigma^2} = -\frac{1}{2} \left\{ \frac{T-\alpha}{\sigma^2} - \frac{\|\mathbf{X} - \mathbf{D}_3 \mathbf{K}^*\|^2}{(\sigma^2)^2} \right\} = 0$$

を考慮して σ^2 について整理すれば、推定値 $\hat{\sigma}^2$ は

$$\hat{\sigma}^2 = \frac{\|\mathbf{X} - \mathbf{D}_3 \mathbf{K}^*\|^2}{T-\alpha} \quad (\text{A2-6})$$

となる。 $\hat{\sigma}^2$ を(A2-5)式に代入すれば、 ϕ を除いて、パラメータの推定値が与えられた $\log ML$ は、

$$\begin{aligned} \log ML &= -\frac{T-\alpha}{2} \left\{ \log(2\pi\hat{\sigma}^2) - \log \phi^2 + 1 \right\} \\ & \quad - \frac{1}{2} \log(\det |\mathbf{D}_3' \mathbf{D}_3|) \end{aligned} \quad (\text{A2-7})$$

となる。

実際のパラメータ推定の手順は、まず ϕ 所与のもとで $\mathbf{K}^*$ を求め、(A2-6)式より $\hat{\sigma}^2$ を計算し、(A2-7)式の $\log ML$ を求める。 ϕ を変化させながら以上の計算を繰り返し、 $\log ML$ の最大値を与える超パラメータ $\hat{\phi}$ を探索する。最後に、 $\phi = \hat{\phi}$ のもとで $\|\mathbf{Y} - \mathbf{G}\mathbf{K}\|^2$ を最小化し、 $\mathbf{K}^*$ を最終的な推定値とする。

(A2-3)式右辺の被積分関数は、事後分布の $\mathbf{K}$ の関数にあたり、事後分布は $N(\mathbf{K}^*, \sigma^2(\mathbf{D}_3' \mathbf{D}_3)^{-1})$ に該当する。パラメータの推定精度はこの分散共分散行列をもとに評価できる。