非線形ダイナミカルシステムの「モデル周辺ゆう度」 重み付き再構成と予測 —階層ベイズ的アプローチ*

齊藤 幹貴[†]・榎本 剛[‡]・松本 隆[‡]

Reconstructions and Predictions of Nonlinear Dynamical Systems Weighted by Model Marginal Likelihoods —A Hierarchical Bayesian Approach*

Motoki SAITO[†], Tsuyoshi ENOMOTO[‡] and Takashi MATSUMOTO[‡]

A hierarchical Bayesian approach is formulated for nonlinear time series prediction problems with neural nets. The proposed scheme consists of several steps:

- (i) Formulae for posterior distributions of parameters, hyper parameters as well as models via Bayes formula.
- (ii) Derivation of predictive distributions of future values taking into account model marginal likelihoods.
- (iii) Using several drastic approximations for computing predictive mean of time series incorporating model marginal likelihoods.

The proposed scheme is tested against two examples; (A) Time series data generated by noisy chaotic dynamical system, and (B) Building air-conditioning load prediction problem. The proposed scheme outperforms the algorithm previously used by the authors.

1. はじめに

時系列データ $\{x_t\}_{t=0}^N \subset \mathbb{R}$ が与えられたとき、その将来値 $\{x_t\}_{t>N}$ を予測する問題は工学のみならず、あらゆる分野に自然に発生する根本問題の一つと見てよいであろう。データには一般に雑音が含まれているから、その予測に統計的手法が用いられるのは自然なことであるが、時系列データが一般の統計的データと異なるのは、時系列の将来の動きが現在および過去の値の影響を受けることによる。したがって、時系列の背後に潜んでいるダイナミカルシステムをとらえる必要がある。非線形ダイナミカルシステムが潜む時系列データに対して、既存の線形予測手法を適用することには限界がある。筆者ら

は「階層ベイズ的」枠組みから、新しい非線形時系列予測手法を提案しており、いくつかの具体的問題に比較的成功裡に適用している [1,2]。

階層ベイズ的枠組みは次のようなものである。適当な“基底関数”を用意し、付随するパラメータを調整して、与えられた学習用時系列データに fit するモデルダイナミカルシステムを構成する。この際、よく知られているように overfitting の問題が起こる。すなわち、雑音を含むデータに fit し過ぎると予測性能が落ちる。ベイズ的枠組みでは、パラメータの先験分布 (prior) を考慮してパラメータの事後分布を計算し、結果的に overfitting の問題を克服する。ここで prior に付随して、いわゆるハイパーパラメータが現れ、階層性が生じる。また、モデルダイナミカルシステムの次数推定も階層ベイズ構造から求まる事後分布によって計算可能である。これらの推定に基づき、目的とする時系列の予測分布を計算する。したがって、(i) “基底関数”のパラメータ、(ii) ハイパーパラメータ、そして、(iii) モデル、の3階層からなるベイズ推定を行うことになり、これが「階層ベイズ的」と呼ばれるゆえんである。

本論文の要点は次のようにまとめられる：

- (i) ニューラルネットのパラメータ、ハイパーパラメータ、そしてモデルを、ベイズ公式によりそれらの事

* 原稿受付 2000 年 11 月 24 日

[†] 早稲田大学 理工学部 電気電子情報工学科 Department of Electrical, Electronics and Computer Engineering, Waseda University; 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, JAPAN

[‡] 早稲田大学 理工学部 電気電子情報工学科, 科学技術振興事業団 Department of Electrical, Electronics and Computer Engineering, Waseda University, CREST, JST.; 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, JAPAN

Key Words: prediction weighted by marginal likelihood, neural net, hierarchical Bayesian approach, nonlinear time series prediction.

後分布から推定する定式化を行う。

- (ii) 予測値の「予測分布」のモデル周辺ゆう度の重みを考慮にいた定式化を行う。
- (iii) いくつかの近似を用いてカオスのダイナミカルシステムと空調機熱負荷予測問題の二つの具体的な時系列予測問題に適用し、モデル事後分布最大のモデルただ一つを用いた予測より、提案手法が優れている場合があることを示す。

なおここでの結果は、参考文献 [3-5] で発表したものを正式にまとめたものである。

2. 定式化

次のような二つのクラスの時系列予測問題を考える：

問題 A: 時系列データ $D := \{x_t\}_{t=0}^N \subset \mathbb{R}$ が与えられたとき、 $\{x_t\}_{t=N+1}^T$ を予測せよ。

問題 B: 時系列データ

$D := (\{x_t\}_{t=0}^N, \{u_t\}_{t=0}^N) \subset \mathbb{R} \times \mathbb{R}^m$ と、入力データ $\{u_t\}_{t=N+1}^{T-1} \subset \mathbb{R}^m$ から出力データ $\{x_t\}_{t=N+1}^T$ を予測せよ。

ここでは、前者に対して自律ダイナミカルシステム

$$x_{t+1} = f(x_t), \quad x_t \in \mathbb{R}^T \quad (1)$$

そして後者には非自律ダイナミカルシステム

$$x_{t+1} = f(x_t; u_t), \quad x_t \in \mathbb{R}^T, u_t \in \mathbb{R}^m \quad (2)$$

を、おのおのモデルとして考える。(1) 式は、たとえばカオスのダイナミカルシステムがあり、(2) 式の例としては、空調機熱負荷予測問題などがある。

2.1 階層ベイズ的定式化

定式化は参考文献 [1] で行われているが、明解さを期すため、改めて正確な定式化を行う。

ここでは、問題 A の定式化を行う。問題 B は適当な修正が必要である [2]。ベイズ統計的枠組みでは、未知パラメータのゆう度 (likelihood) と先験確率分布 (prior) を指定することで問題が定式化される。これらをまとめて $\mathcal{H}$ (Hypothesis または model) と表すことにする。具体的には次の諸項目を意味する。

モデル (Hypothesis) $\mathcal{H}$:

- (i) アーキテクチャ:

用いるモデルの構造を指定する。ここでは中間素子数 h 、出力関数 $1/(1+e^{-u})$ の 3 層パーセプトロンを用いる¹。

- (ii) ゆう度 (likelihood):

以下のような遅れ座標系でガウス過程を仮定し

$$x_{t+1} = f(x_t, \dots, x_{t-\tau+1}) + \mu_t \quad (3)$$

$$\mu_t \sim N(0, 1/\beta) \quad (4)$$

次式でゆう度 (likelihood) を導入する。

¹これ以外の基底関数でも、もちろん可能である。

$$\begin{aligned} & P(\{x_t\}_{t=\tau}^N, \{x_{\tau-1}, \dots, x_0\} | \mathbf{w}, \beta, \mathcal{H}) \\ & := \underbrace{\prod_{t=0}^{N-\tau} \frac{1}{Z_D(\beta)} \exp(-\beta E_D(x_{t+\tau} | x_{t+\tau-1}, \dots, x_t; \mathbf{w}))}_{\text{noisy dynamics}} \\ & \quad \times \underbrace{P(x_{\tau-1}, \dots, x_0 | \mathcal{H})}_{\text{initial state uncertainty}} \end{aligned} \quad (5)$$

$$\begin{aligned} & E_D(x_{t+\tau} | x_{t+\tau-1}, \dots, x_t; \mathbf{w}) \\ & := \frac{1}{2} (x_{t+\tau} - f(x_{t+\tau-1}, \dots, x_t; \mathbf{w}))^2 \end{aligned} \quad (6)$$

ただし $f(\cdot)$ は weight パラメータ $\mathbf{w} \in \mathbb{R}^k$ をもつパーセプトロン出力、 β はノイズの分散の逆数 (未知)、 $Z_D(\beta)$ は正規化定数、そして τ はダイナミカルシステムの次数 (未知) である。

- (iii) $\mathbf{w}$ の prior:

$\mathbf{w}$ をネットワークグループごとに分割し、

$$\mathbf{w} := (\mathbf{w}_1, \dots, \mathbf{w}_C), \mathbf{w}_c \in \mathbb{R}^{k_c} \quad (7)$$

ここでは各 $\mathbf{w}_c$ は大きいものは少なく、小さいものは多く、正負両方の値を取りうるとし、次式で prior を導入する。

$$P(\mathbf{w} | \alpha, \mathcal{H}) := \prod_{c=1}^C \frac{1}{Z_w(\alpha_c)} \exp(-\alpha_c E_{w_c}(\mathbf{w}_c)) \quad (8)$$

ここで

$$E_{w_c}(\mathbf{w}_c) := \frac{1}{2} \|\mathbf{w}_c\|^2 \quad (9)$$

$$\alpha := (\alpha_1, \dots, \alpha_C), \alpha_c \in \mathbb{R} \quad (10)$$

である。

- (iv) (α, β) , $\mathcal{H}$ の prior: 特定のハイパーパラメータ (α, β) や特定のモデル $\mathcal{H}$ に特に重きをおく理由がなければ、 $P(\alpha, \beta | \mathcal{H})$ や $P(\mathcal{H})$ は平坦と考えるのが妥当である。以下この論文では、そのように考える。

以上の設定に基づき、予測分布 $P(\{x_t\}_{t=N+1}^T | D)$ を計算するのが、ここでの目的である。記号を簡単にするために、(5) 式を $P(D | \mathbf{w}, \beta, \mathcal{H})$ と書く。

上に述べた (i)–(iv) より、ベイズ公式から Fig. 1 のような自然な階層構造が帰結する。各レベルの分母にある周辺ゆう度

$$\begin{aligned} & P(D | \alpha, \beta, \mathcal{H}) \\ & = \int_{\mathbb{R}^k} P(D | \mathbf{w}, \beta, \mathcal{H}) P(\mathbf{w} | \alpha, \mathcal{H}) d\mathbf{w} \quad (11) \\ & \quad P(D | \mathcal{H}) \\ & = \int_{\mathbb{R}^C} \int_{\mathbb{R}} P(D | \alpha, \beta, \mathcal{H}) P(\alpha, \beta | \mathcal{H}) d\beta d\alpha \quad (12) \end{aligned}$$

Fig. 1 Hierarchical Bayes structure

が本質的役割を演じる。

具体的な計算は次章以後で行う。

3. 予測

まず, Fig. 1 で与えられた階層的スキームに従って各パラメータの事後分布を求める。

いま, $(\alpha_{MP}, \beta_{MP})$ をハイパーパラメータの事後分布のモードとする:

$$(\alpha_{MP}, \beta_{MP}) = \arg \max_{\alpha, \beta} P(\alpha, \beta | D, \mathcal{H}) \quad (13)$$

つぎに (α, β) を固定した時の $\mathbf{w}$ の事後分布のモード

$$\mathbf{w}_{MP} = \arg \max_{\mathbf{w}} P(\mathbf{w} | D, \alpha, \beta, \mathcal{H}) \quad (14)$$

を考える。 $(\alpha_{MP}, \beta_{MP})$ がひとつ用意されれば, $\mathbf{w}_{MP}$ はペナルティー付きニューラルネット最適化となる。そこで $\mathbf{w}_{MP}, (\alpha_{MP}, \beta_{MP})$ を交互に共役勾配法により更新学習を行う。(13)式の最大化は, $\mathbf{w}$ に関して周辺化された関数(11)式の最適化を含み、非自明であって、それなりの工夫を要する。参考文献[1,2]および、ここでは、次の2次近似を用いている:

$$\begin{aligned} \log P(D | \alpha, \beta, \mathcal{H}) &\approx -M(\mathbf{w}_{MP}; \alpha, \beta) - \frac{1}{2} \log \det \mathbf{A} \\ &+ \frac{N}{2} \log \frac{\beta}{2\pi} + \sum_c \frac{k_c}{2} \log \alpha_c \end{aligned} \quad (15)$$

ここに

$$\begin{aligned} M(\mathbf{w}) &:= \frac{1}{2} \beta \sum_{t=0}^{N-\tau} (x_{t+\tau} - f(x_{t+\tau-1}, \dots, x_t; \mathbf{w}))^2 \\ &+ \frac{1}{2} \sum_{c=1}^{k_c} \alpha_c \|\mathbf{w}_c\|^2 \end{aligned} \quad (16)$$

また $\mathbf{w}_{MP}$ は

$$\mathbf{w}_{MP} = \arg \min_{\mathbf{w}} M(\mathbf{w})$$

と書き直すことができる。

$\mathbf{A}$ は $\mathbf{w}_{MP}$ において評価した $M(\mathbf{w})$ に関する Hessian である。

つぎに予測分布

$$\begin{aligned} P(\{x_t\}_{N+1}^T | D) &= \sum_{\mathcal{H}} \int \int \int P(\{x_t\}_{N+1}^T | \mathbf{w}, \beta, \mathcal{H}) \\ &\times P(\mathbf{w}, \alpha, \beta, \mathcal{H} | D) d\mathbf{w} d\alpha d\beta \end{aligned} \quad (17)$$

を計算する。Fig. 1 の諸公式、特に(17)式も厳密なものである一方、実際の計算には著しい困難が生じる理由は、(11), (12)式そして(17)式がパラメータやハイパーパラメータに関する「周辺化」を含むからである。将来的には、これらの周辺化は「瞬時」に計算可能になる時が来るかも知れないが、少なくとも現在は思い切った近似が必要である。この妥当性については参考文献[6]で考察されている。

$P(\alpha, \beta | D, \mathcal{H})$ が $(\alpha_{MP}, \beta_{MP})$ 近傍で十分鋭いと仮定すると、

$$\begin{aligned} P(\mathbf{w}, \alpha, \beta, \mathcal{H} | D) &= P(\mathbf{w} | D, \alpha, \beta, \mathcal{H}) \\ &\times P(\alpha, \beta | D, \mathcal{H}) P(\mathcal{H} | D) \end{aligned} \quad (18)$$

より予測分布は次で近似される。

$$\begin{aligned} P(\{x_t\}_{N+1}^T | D) &\approx \sum_{\mathcal{H}} P(\mathcal{H} | D) \int P(\{x_t\}_{N+1}^T | \mathbf{w}, \beta_{MP}, \mathcal{H}) \\ &\times P(\mathbf{w} | D, \alpha_{MP}, \beta_{MP}, \mathcal{H}) d\mathbf{w} \end{aligned} \quad (19)$$

さらに, $(x_{t+1} - f(x_t, \dots, x_{t-\tau+1}; \mathbf{w}_{MP}))$ を $\mathbf{w} = \mathbf{w}_{MP}$ でテーラー展開し, 1次の項までとると

$$\begin{aligned} P(\{x_t\}_{N+1}^T | \mathbf{w}, \beta_{MP}, \mathcal{H}) &\approx \prod_t \frac{1}{Z_D(\beta_{MP})} \\ &\times \exp \left\{ -\frac{\beta_{MP}}{2} (x_{t+1} - f(x_t, \dots, x_{t-\tau+1}; \mathbf{w}_{MP}) \right. \\ &\quad \left. - \frac{\partial f}{\partial \mathbf{w}}^T (\mathbf{w} - \mathbf{w}_{MP}))^2 \right\} \end{aligned} \quad (20)$$

学習が終了すれば $\partial f / \partial \mathbf{w} = 0$ となる。

あとは近似(15)式, (20)式を用いて予測分布(19)式の最適化であるがここで確率 $P(D | \mathcal{H})$ で分布

$$\exp \left(-\frac{\beta_{MP}}{2} (x_{t+1} - f(x_t, \dots, x_{t-\tau+1}; \mathbf{w}_{MP}))^2 \right)$$

の最適値を選ぶと考えると予測分布(19)式から $\{x_t\}_{N+1}^T$ の predictive mean は次式で与えられる。

$$x_{t+1, MP} = \sum_{\mathcal{H}} f(x_t, MP, \dots, x_{t-\tau+1, MP}; \mathbf{w}_{MP}) \times P(\mathcal{H} | D), N \leq t \leq T-1 \quad (21)$$

ここで $P(\mathcal{H})$ を一様, すなわち $P(\mathcal{H}) = \text{const}$ とすると

$P(\mathcal{H}|D)$ は Bayes 公式より

$$P(\mathcal{H}|D) = \frac{P(D|\mathcal{H})}{\sum_{\mathcal{H}} P(D|\mathcal{H})} \quad (22)$$

であるので

$$x_{t+1,MP} = \sum_{\mathcal{H}} f(x_{t,MP}, \dots, x_{t-\tau+1,MP}; \mathbf{w}_{MP}) P(D|\mathcal{H}) \times \frac{1}{\sum_{\mathcal{H}} P(D|\mathcal{H})}, N \leq t \leq T-1 \quad (23)$$

これは, *predictive mean* が, ニューラルネットの出力 $f(\cdot)$ にモデル周辺ゆう度 $P(D|\mathcal{H})$ で重み付けして平均をとったものとなることを示している.

本論文の提案予測スキーム (23) 式と参考文献 [1,2] で用いられたそれとの違いを述べる. 参考文献 [1,2] で提案された予測手法は, (23) 式の特殊な場合になっている. すなわち, 参考文献 [1,2] では

$$\mathcal{H}_{MP} = \operatorname{argmax}_{\mathcal{H}} P(\mathcal{H}|D) \quad (24)$$

とすると,

$$x_{t+1,MP} = f(x_{t,MP}, \dots, x_{t-\tau+1,MP}; \mathbf{w}_{MP}) \quad (25)$$

で予測を行っている. 以下, ふたつの具体例を取り上げ, 両者の予測性能比較を行う.

4. 具体的問題への適用

4.1 カオス時系列予測

カオスの連続ダイナミカルシステムの代表的なものの一つである Rössler 系への適用を試みる.

Rössler 系:

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= bx - cz + xz \end{aligned} \quad (26)$$

は, パラメータをたとえば $(a, b, c) = (0.36, 0.4, 4.5)$ としたとき, Fig. 2 のようにカオス的な振舞いを示す.

Fig. 2 Rössler attractor

(26) 式は連続力学系システムである. ここで, 次のようなシステムノイズを含んだ系を考える.

$$\begin{aligned} \dot{x} &= -y - z + \nu_t^1 \\ \dot{y} &= x + ay + \nu_t^2 \\ \dot{z} &= bx - cz + xz + \nu_t^3 \end{aligned} \quad (27)$$

$\nu_t^1, \nu_t^2, \nu_t^3$ はノイズプロセスである. 連続ノイズプロセスには理論的困難が伴うため, ここでは次のように離散系として考える.

$$\begin{aligned} x_{(t+1)\delta} &= f(x_{t\delta}, y_{t\delta}, z_{t\delta}) + \nu_{t\delta}^1 \\ y_{(t+1)\delta} &= g(x_{t\delta}, y_{t\delta}, z_{t\delta}) + \nu_{t\delta}^2 \\ z_{(t+1)\delta} &= h(x_{t\delta}, y_{t\delta}, z_{t\delta}) + \nu_{t\delta}^3 \\ \nu_{t\delta}^i &\sim i.i.d. N(0, \sigma^2), i=1, 2, 3 \end{aligned} \quad (28)$$

ただし, $f(\cdot), g(\cdot), h(\cdot)$ は Runge-Kutta 法 (step size δ) による数値積分から帰結する関数である.

観測データを x 座標系の時系列 $\{x_{t\delta}\}_{t \geq 0}$ とし, サンプリング周期 η にて連続する 500 点を学習データとした. ただし, $\delta = 0.01, \eta = 70, \sigma = 0.02$ である. $(x_{t\delta\eta}, x_{(t-1)\delta\eta}, x_{(t-2)\delta\eta})$ プロットを Fig. 3 に示す.

Fig. 3 Training data: $(x_{t\delta\eta}, x_{(t-1)\delta\eta}, x_{(t-2)\delta\eta})$

このデータを, 前章までのアルゴリズムを用いて学習を行い, 次に示す計 105 個のモデルを考慮した:

- (i) $\tau: 3 \sim 7$;
- (ii) $h: 2 \sim 8$;
- (iii) $\mathbf{w}_{MP}$ の推定のための $\mathbf{w}$ の初期値を 3 種類

Table 1 は, 上記 105 モデルのうち周辺ゆう度の値が大きかった 10 モデルを示したものである. ただし, 表中の model ID とは, (埋め込み次元 τ , 中間素子数 h , $\mathbf{w}$ の初期値) を示している. たとえば 5-5-1 は, 埋め込み次元 $\tau=5$, 中間素子数 $h=5$, 推定時における第 1 番目の初期値を意味する. 経験的に各モデルの周辺ゆう度が大きいほど良い予測であるが必ずしも周辺ゆう度最大の単一モデルを用いた予測が累積二乗誤差最小とはならない

Table 1 Marginal likelihoods of top ten models

model ID	5-5-1	4-6-1
$\log P(D \mid \alpha_{\text{MP}}, \beta_{\text{MP}}, \mathcal{H})$	-133.14954	-133.51516
$P(D \mid \alpha_{\text{MP}}, \beta_{\text{MP}}, \mathcal{H})$	1.4924E-58	1.0353E-58

5-6-1	4-7-1	4-7-3	5-7-1
-134.23158	-137.70883	-137.92675	-147.78589
5.0578E-59	1.5624E-60	1.2565E-60	6.5675E-65

4-8-2	5-8-1	5-8-2	4-8-3
-148.92888	-150.07391	-151.66608	-153.89545
2.0941E-65	6.6639E-66	1.3560E-66	1.4590E-67

Fig. 4 Average of squared errors predicted by 10 kinds of initial conditions

Fig. 5 Comparison of predicted x-trajectory

い場合がある。Table 1 をみると、各モデルの周辺ゆう度の値には大きな開きがあり、(23) 式の計算に効いてくるのはせいぜい上位 3 モデルと考えられる。そこで、この上位 3 モデルを用いて重み付き予測を行い、評価する。

予測性能評価では、初期値として学習データに含まれないデータを 10 用意し、100 点の予測を行って予測誤差をとった。

$$\sum_{t=1}^T \frac{1}{2} (x_{t\delta\eta} - \hat{x}_{t\delta\eta})^2, \quad \sum_{t=1}^T \frac{1}{2} (x_{t\delta\eta} - x_{t\delta\eta}(\mathcal{H}))^2 \quad (29)$$

を、おのおの 10 種の初期値について加え合わせ、予測区間 $T=1, \dots, 100$ で比較したものが Fig. 4 である。ただし、 $\hat{x}_{t\delta\eta}$ は (23) 式で与えられる重み付き予測値、 $x_{t\delta\eta}(\mathcal{H})$ は、各モデルごとの予測値である。重み付き予測が、累積二乗誤差に関して優れていることが見てとれる。

Fig. 5 は、典型的初期値に対する三つの単一モデルと、提案スキームの 100 ステップ予測、そして真の時系列を

比較したものである。

4.2 空調機熱負荷予測

1997 年夏、空調調和・衛生工学会によって主催された、熱負荷予測ベンチマークテスト [7] に筆者らも参加し、結果は [2] に報告した。この際のデータを用い、単一モデルの場合と、提案手法の予測性能比較を行った。

Fig. 6 で示すようなアーキテクチャを使用して、次に示す計 45 個のモデルに対して前章までのアルゴリズムを用いて学習を行った。

(i) $h: 5 \sim 9$;

(ii) w_{MP} の推定のための w の初期値を 9 種類

ただし、Fig. 6 中の $u_{t,time}$ は時間変数を表し、 $u_{t,OT,RT,RH}$ は時刻 t での外気温度、室内温度、室内湿度を表す。また、 x_t は時刻 t での予測すべき熱負荷を表す。8 月 12 日～16 日の連続する平日 5 日間を予測した。

Table 2 は、上記 45 モデルのうち周辺ゆう度の値が大きかった 5 モデルを列挙したものである。ただし、表中

Fig. 7 Weighted prediction and prediction with highest marginal likelihood

Fig. 6 Architecture of air-conditioning load prediction

Table 2 Marginal likelihoods of top five models

model ID	5-0.1-3	6-0.3-2
$\log P(D \alpha_{MP}, \beta_{MP}, \mathcal{H})$	703.33499	702.61230
$P(D \alpha_{MP}, \beta_{MP}, \mathcal{H})$	2.8478E+305	1.3824E+305

7-0.2-1	6-0.2-2	5-0.2-3
697.85489	689.25989	681.04352
1.1872E+303	2.1967E+299	5.9353E+295

の model ID とは、(中間素子数 h , 学習の際の $\mathbf{w}$ の初期値の分散, $\mathbf{w}$ の初期値の種類) を示している. Table 2 より, 前節と同様に周辺ゆう度上位 3 モデルを用いて重み付き予測を行った.

Fig. 7 は, 重み付き予測と周辺ゆう度が最大のモデル (model ID: 5-0.1-3) を用いて行った予測, および実測値をプロットしたものであり, Fig. 8 は, 重み付け予測と, 上位 3 モデルおのおのの予測の実測値との二乗誤差 ((a):1 日ごと, (b):5 日間の合計) をプロットしたものである. この実験の場合, 真の値 (観測値) は一つし

かないため, 前例のように異なる複数の初期値からの総合性能を評価することはできない. 次のことが見てとれる:

- (i) モデル周辺ゆう度上位 3 者のうち, 2 者よりは提案手法 (重み付き予測) が優れているが, 一つ (モデル (6-0.3-2)) よりは重み付き予測による誤差が多い.
- (ii) モデル (6-0.3-2) は周辺ゆう度では第 2 位である.
- (iii) 以上から次のことが帰結する.
 - (a) モデル周辺ゆう度による「モデル選択」は, かなり良い手法ではあるが, ダイナミカルシステムに注目すると常に最良であるとは限らない.
 - (b) したがって, モデル周辺ゆう度上位者の重み付き予測を遂行すれば, “取りこぼし” なく, かなり良い予測が可能である. たとえば, モデル周辺ゆう度最大のモデル (5-0.1-3) よりは, 重み付き予測が優れている.

5. おわりに

筆者らがすでに提案している階層ベイズ的非線形時系列予測手法を改良し, モデル周辺ゆう度による重み付き予測アルゴリズムを提案した. ふたつの具体的問題に適用し, その有効性を示した.

残された問題は,

- (i) 学習の高速化
 - (ii) MCMC 等による周辺ゆう度計算法の改良
- などである.

謝 辞

有益なコメントをいただいた査読者の方々に感謝します.

(a) Cumulative squared prediction error for each of the five days

(b) Sum of cumulative squared prediction error for five days (From left to right, show weighted prediction and three individual predictions with top three marginal likelihoods according to priority)

Fig. 8 Squared prediction errors for the weighted prediction and three individual predictions

参考文献

- [1] 松本, 浜岸, 杉, 斉藤, 長南: データから非線形ダイナミクスへ—階層ベイズの時系列予測—; 信学論 (D-II), Vol. J82-D-II, No. 6, pp. 1059–1071 (1999)
- [2] 中島, 斉藤, 杉, 浜岸, 松本: 階層ベイズの空調機熱負荷予測—ニューラルネットによる非線形ダイナミカルシステムのアプローチ—; 信学論 (D-II), Vol. J82-D-II, No. 11, pp. 2075–2083 (1999)
- [3] 斉藤, 松本: 非線形時系列のモデル周辺ゆう度による重み付き予測; 信学技報, NC98–115, pp. 129–136 (1999)
- [4] M. Saito, J. Sugi and T. Matsumoto: Nonlinear time series prediction weighted by marginal likelihoods: A hierarchical Bayesian approach; *Proc. of 1999 In-*

ternational Joint Conference on Neural Networks, No. 0572, Vol. 4, pp. 2604–2607 (1999)

- [5] T. Matsumoto, M. Saito, J. Sugi and Y. Nakajima: A hierarchical Bayesian scheme for reconstruction and prediction of nonlinear dynamical systems; *Proc. of The International Symposium on Frontiers of Time Series Modeling*, pp. 336–337 (2000)
- [6] Y. Nakajima, M. Asano, Y. Nakada, Y. Satoh and T. Matsumoto: Quadratic approximation vs. MCMC predictive mean computations and model comparisons hierarchical bayes approach with neural nets; *Proc. of IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS2000) Time Series Modeling*, pp. 336–337 (2000)
- [7] URL: <http://www.t3.rim.or.jp/~bmtest>