ダイナミカルシステムの階層ベイズ的最小埋め込み次元推定

杉　淳二郎[†]　栗原貴之[†]　松本　隆[†]

システムノイズを含むダイナミカルシステムの最小埋め込み次元を推定する．最小埋め込み次元の推定法は FNN（False Nearest Neighbors）法がよく知られている．本稿は新たな手法として階層ベイズ的推定法を提案して具体例に適用し，FNN 法と比してより客観的な推定が可能であり，データ数が少なく，データにノイズが含まれる場合でも健全に働くことを示す．

Hierarchical Bayesian Estimation of Minimum Embedding Dimension for Dynamical Systems

Junjiro Sugi,[†] Takayuki Kurihara[†] and Takashi Matsumoto[†]

Given one-dimensional time seires data, estimation of the order of the dyamics behind the time series is of vital theoretical as well as practical importance. The problem is far from trivial when (1) the dynamics behind the data is nonlinear and (2) the data is noisy. This paper proposes a scheme to solve this problem via Hiararchical Bayesian framework, where Model Marginal Likelihood is used for estimating the order of the dynamics. The algorithm is tested against noisy chaotic time series data and compared with the conventional algorithm. The proposed scheme outperforms the conventional scheme at least in the examples tested.

1. まえがき

1 次元の時系列データ $\{x_t\}_{t=0}^{N} \subset \mathbb{R}$ からその将来値 $\{x_t\}_{t>N}$ を予測する問題は，多くの科学技術分野に現れる根本問題の 1 つである．時系列の予測問題の困難はある時刻の値が「過去の値に依存する」ことにあり，より具体的には

(A) 時刻 t の値 x_t を考えるとき，過去何ステップの値までの依存性を考慮すればよいか？

を考えなければならない．何らかの推定スキームにより，たとえば τ ステップの過去までさかのぼれば十分と推定されたとき，

(B) そのような τ の最小値は何か？

は，理論的にも具体的問題への適用にも重要な問題である．実データは，その数が有限なので “次元の呪い”（Course of Dimensionality）を避けるため，次元は小さいことが望ましい．実データの有限性に対する考慮は予測手法の基本の 1 つである．

$\{x_t\}_{t=0}^{N}$ の背後に線形ダイナミカルシステムが想定可能な場合は線形ダイナミカルシステムの次数推定に相当し，たとえば AIC[1)] が広く用いられている．

一方，時系列の背後に非線形ダイナミカルシステムが想定される場合，次数推定には大きな困難がともなう．文献 2) および関連する文献 3)〜15) で提案された非線形ダイナミカルシステムの再構成と予測手法は階層ベイズ的なものであり，尤度を階層的に周辺化することにより，いくつかの重要パラメータを推定しようとするものである．本稿の目的は，これらの文献に述べられた手法を明示的に説明したうえ，よく知られた FNN（False Nearest Neighbors）法[16)] と比較検討することである．

本稿の結果は以下のようにまとめられる：

（1）　FNN 法は原理が単純で多くの問題に有効に働くが，近傍であるか否かを決める「閾値」の適切な選択が決定的であり，この選択に任意性が内在している．具体的にはデータに含まれるノイズレベルを考慮しながら閾値を選択しなければならない．一方，提案手法では原理的にはこのような任意性はなく，階層ベイズ構造からノイズレベルの推定も自然に組み込まれている．

（2）　少なくとも本稿で検討した具体例においては，FNN 法による推定よりも安定した結果を示している．

（3）　最小埋め込み次元の推定が最終目的ではなく，それに基づく時系列予測が目的の場合，提案手

† 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

法ではそのまま予測に必要なスキームが含まれているが FNN 法自体にはそのような構造は用意されていない．

（4）　アルゴリズム実装の単純さと計算時間では FNN 法が優れている．

遅延座標系への埋め込み（Delay Coordinate Embedding）や FNN 法はダイナミカルシステムの分野ではよく知られた手法であるが，本稿の明快さを期すため各々 2 章，3 章で簡単に説明する．また，3 章では FNN 法と基本発想は同一であるが，FNN 法では主観的に決められることの多い閾値を用いない AFN（Averaged False Neighbors）法[17),18)] の説明を加える．4 章では本稿で提案する階層ベイズ的手法を説明し，5 章で各手法の比較を行う．6 章は残された問題点と今後の展望である．

2. 遅延座標系への埋め込み（Delay Coordinate Embedding）

非線形ダイナミカルシステム

$$\boldsymbol{y}_{t+1} = F(\boldsymbol{y}_t),\quad \boldsymbol{y}_t \in \mathbb{R}^K \tag{1}$$

とスカラー観測データ

$$x_t = G(\boldsymbol{y}_t),\ x_t \in \mathbb{R} \tag{2}$$

を考えるとき，次が知られている[19)]：

ダイナミカルシステム (1) の不変集合 Y は，$\mathbb{R}^K$ の開部分集合 U のコンパクト部分集合とし，Y の box counting 次元を d とする．このとき，

$$\tau > 2d \tag{3}$$

であれば，遅延座標系

$$\boldsymbol{x}_t := (x_t, x_{t+1}, \ldots, x_{t+(\tau-1)}) \in \mathbb{R}^\tau \tag{4}$$

と，$\boldsymbol{y}_t \in Y$ との関係

$$\boldsymbol{x}_t = \Phi(\boldsymbol{y}_t) \tag{5}$$

は，ほとんどすべての滑らかな観測関数 G に対して次を満たす．

（1）　Y 上で，1 : 1

（2）　Y 内の滑らかな多様体のコンパクト部分集合上で，immersion☆．

図 1 に模式図を示す．

この結果から，少なくとも 2 つの大切な事実が帰結する．

(a)　Φ が Y 上で 1 : 1 なので，適当なアルゴリズムを準備すれば $\boldsymbol{x}_t$ を用いて将来の値を予測できる．

(b)　Y がカオス的アトラクタのとき一般に不安定多様体方向については滑らかなので，上記 (2) から微分構造が保たれ，したがって正の Lyapnov 指数[20)] を $\boldsymbol{x}_t$ から計算することができる．安定多様体方向は，典型的にはカントール構造を持つので負の Lyapnov 指数の計算は困難である．

☆ immersion は局所的な微分構造が保たれることを意味する．ただし，(2) については，周期点に関する若干の技術的条件が要請されるが，省略する．

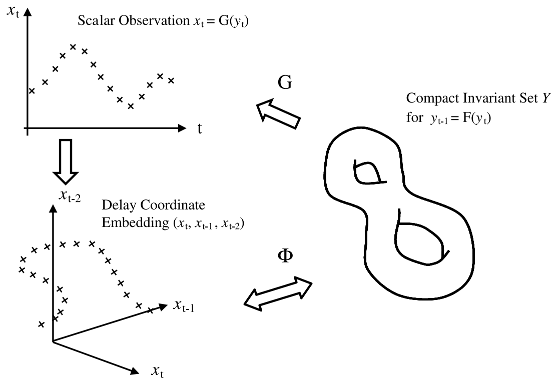


図 1　遅延座標系への埋め込みの模式図
［文献 22) より許可を得て転載］
Fig. 1　Delay coordinate embedding.

上述の結果は離散ダイナミカルシステムに関するものであるが，連続ダイナミカルシステムの場合はサンプリング間隔を η とし，

$$\boldsymbol{x}_t := (x_t, x_{t+\eta}, \ldots, x_{t+(\tau-1)\eta}) \tag{6}$$

を考えると，同様の結果が成立する．

上述の結果は，スカラー時系列データから非線形ダイナミカルシステムを再構成し，それに基づいた時系列予測を遂行する際の，最重要基本原理の 1 つといってよいであろう．ただし次の点は注意するべきである：

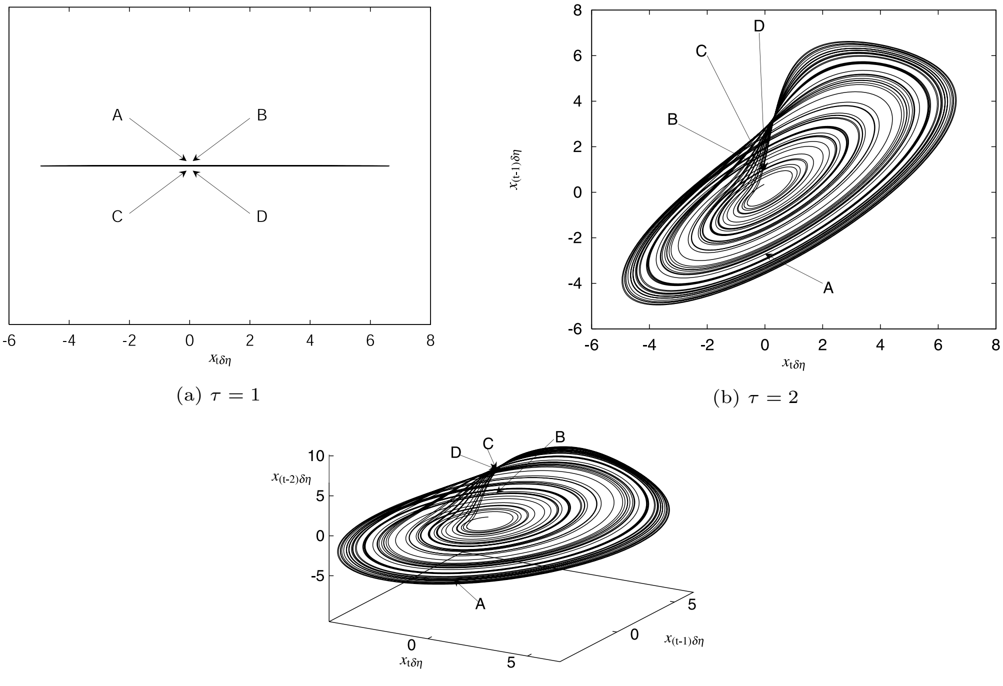
（1）　条件 $\tau > 2d$ は十分条件であって，必要条件ではない．$\tau \leq 2d$ でも埋め込み可能な場合は多数ある．ここに最小値の推定問題が生じている．

（2）　この結果はダイナミクスおよび観測系にノイズがない場合に対するものであり，少なくとも著者らが知る限りではノイズが含まれる場合の結果は得られておらず，最先端研究課題である[21)]．

（3）　与えられた $\{x_t\}$ からどのような間隔でサンプルし，学習・予測に用いるか，はもう 1 つの問題である．これは離散系の場合にも，連続系の場合にも付随する問題であり，後者に関していえば，式 (6) における η である．本稿の主たる目的は τ の推定であり，5.2.5 項の例以外は τ の推定のみを考察する．

3. False Nearest Neighbors（FNN）法

この手法[16)] の考え方は単純である．**図 2** はノイズを含まない Rössler 系（後述）の x 座標から得られ



(a) $\tau = 1$

(b) $\tau = 2$

(c) $\tau = 3$

図 2 ノイズを含まない Rössler 系の遅延座標系への埋め込み
Fig. 2 Delay coordinate embedding for noiseless Rössler system.

た時系列を埋め込み次元 $\tau = 1$〜3 の空間にプロットしたものである．$\tau = 2$ の空間では，$\tau = 1$ で近傍にあった 2 点 A，B 間の距離が極端に増大している．また，$\tau = 3$ の空間では，$\tau = 1$，2 で近傍にあった 2 点 B，C 間の距離が極端に増大している．このような関係を「偽近傍」（False Neighbor）と呼ぶ．これに対し，2 点 C，D 間の距離は $\tau = 1$〜3 を通してあまり増大していない．このような関係を「真近傍」（True Neighbor）と呼ぶ．

「偽近傍」か「真近傍」かは埋め込み次元を上げたときの 2 点間の距離の増加率が，あらかじめ決めてある閾値 R_{tol} を超えるか否かによって判定する．前章の埋め込み定理から，埋め込み次元 τ を徐々に上げていけばいつかは「偽近傍」が消失するはずであり，その最小値を最小埋め込み次元と考える．

データを τ 次元遅延座標空間に埋め込んだとき，$\boldsymbol{x}_t$ と距離基準 $D_\tau(t,i)$ に関して最も近い点（最近傍点，Nearest Neighbor）を $\boldsymbol{x}_{n_\tau(t)}$ と書くことにする：

$$D_\tau(t,i) = \max_{0 \le m \le \tau-1} |x_{t+m} - x_{i+m}| \tag{7}$$

$$\boldsymbol{x}_{n_\tau(t)} = \left(x_{n_\tau(t)}, x_{n_\tau(t)+1}, \ldots, x_{n_\tau(t)+(\tau-1)}\right) \tag{8}$$

$$n_\tau(t) = \arg\min_i D_\tau(t,i) \tag{9}$$

$\boldsymbol{x}_t$，$\boldsymbol{x}_{n_\tau(t)}$ の 2 点を $\tau+1$ 次元遅延座標空間に埋め込んだときの距離は，

$$D_{\tau+1}(t, n_\tau(t)) = \max_{0 \le m \le \tau} \left| x_{t+m} - x_{n_\tau(t)+m} \right| \tag{10}$$

となる．$D_\tau(t, n_\tau(t))$ に比して $D_{\tau+1}(t, n_\tau(t))$ が極端に大きいとき，$\boldsymbol{x}_t$ と $\boldsymbol{x}_{n_\tau(t)}$ は「偽最近傍点」（False Nearest Neighbor）であると考える☆：

☆ 文献 16) の FNN 法では，$D_\tau(t,i)$ を

$$D_\tau^2(t,i) = \sum_{m=0}^{\tau-1} [x_{t+m} - x_{i+m}]^2 \tag{11}$$

と定義し，基準 1 を

$$\left[\frac{D_{\tau+1}^2(t, n_\tau(t)) - D_\tau^2(t, n_\tau(t))}{D_\tau^2(t, n_\tau(t))}\right]^{\frac{1}{2}} > R_{tol} \tag{12}$$

としているが，著者らの実験の限りでは式 (13) と式 (12) のどちらを用いても結果は同等であったので，記述を AFN 法と統一するために本稿では式 (13) を用いる．

［基準 1］

$$a(t,\tau) = \frac{D_{\tau+1}(t, n_\tau(t))}{D_\tau(t, n_\tau(t))} > R_{tol} \qquad (13)$$

基準 1 でデータが「完全に stochastic」か「高次元の力学系から生成された」かを識別するのは困難である．データが stochastic なときには deterministic なときと比してアトラクタが比較的大きいため，相当数のデータがない場合には遅延座標空間でデータの密度が小さくなって，$D_\tau(t, n_\tau(t))$ が大きな値となりやすい．この場合，式 (13) の左辺の値が小さくなって，基準 1 では比較的小さな τ で遅延座標空間に埋め込まれているかのような様相を呈することがある．

$\boldsymbol{x}_t$，$\boldsymbol{x}_{n_\tau(t)}$ の 2 点は τ 次元遅延座標空間で最近傍であるから，データがダイナミカルシステムから生成されていて，τ が十分に大きければその 1 ステップ先の $x_{t+\tau}$ と $x_{n_\tau(t)+\tau}$ も十分に近い値となるはずである．これに対してデータが stochastic な要素が強いときは $x_{t+\tau}$ と $x_{n_\tau(t)+\tau}$ は独立な値をとることが予想される．この議論をもとに，新たな基準を用意する：

［基準 2］

$$b(t,\tau) = \frac{\left| x_{t+\tau} - x_{n_\tau(t)+\tau} \right|}{R_A} > A_{tol} \quad (14)$$

ここで，R_A はアトラクタの大きさを表す量であり，たとえばデータの標準偏差を用いる．A_{tol} はあらかじめ決めてある閾値である．データが stochastic な要素が強いのであれば基準 2 を満たす点が相当数存在するはずである．基準 1，2 のいずれかを満たす点が偽最近傍点（False Nearest Neighbor）であり，その数が 0 になるような最小の次元が最小埋め込み次元であろう，というのがこの手法の考え方である．

データ中の偽最近傍点の割合 FNNP（False Nearest Neighbors Percentage）は埋め込みがどの程度うまくいっているかを示すと同時に，埋め込み次元を必要以下にしたときのペナルティの情報も含んでいると考えられる[16]．

この手法は考え方も具体的アルゴリズムも確かに単純であって，多くの例題においてうまく機能する．しかし，それは「R_{tol}，A_{tol} のふさわしい値が選ばれたとき」であって，「R_{tol}，A_{tol} をどのように推定するか」に関しては特にアルゴリズムは提案されておらず，ad hoc といわざるをえない．特にデータにノイズが含まれる場合はそのノイズレベルを考慮して閾値を選択する必要が生じる．

一方，AFN（Averaged False Neighbors）法[17),18)]の，基本発想は FNN 法と同一であるが，「平均化」によって閾値 R_{tol}，A_{tol} を用いずに判定を行う手法である．

まず，式 (13) に対し，次式のように $a(t,\tau)$ の平均値 $E^a(\tau)$ を定義する：

$$E^a(\tau) = \frac{1}{N-\tau+1} \sum_{t=0}^{N-\tau} a(t,\tau) \qquad (15)$$

τ の値を上げていき，埋め込みが成立すると $E^a(\tau)$ の値はほとんど変化しなくなるであろう．そこで，基準値 $E1(\tau)$ を次のように定義する：

［基準 A］

$$E1(\tau) = \frac{E^a(\tau+1)}{E^a(\tau)} \qquad (16)$$

埋め込み次元を上げていき，$E1(\tau)$ が 1 に飽和したとき，埋め込みが成立していると考える．

同様の考えに従って，式 (14) に対しても $b(t,\tau)$ の平均値 $E^b(\tau)$，基準値 $E2(\tau)$ を次のように定義する：

$$E^b(\tau) = \frac{1}{N-\tau+1} \sum_{t=0}^{N-\tau} b(t,\tau) \qquad (17)$$

［基準 B］

$$E2(\tau) = \frac{E^b(\tau+1)}{E^b(\tau)} \qquad (18)$$

データがたとえば白色雑音であれば，τ によらずに $E2(\tau) \equiv 1$ になるであろう．

AFN 法は上記の基準 A，B をあわせて埋め込みが成立しているか否かを判定する．この手法は FNN 法で値の選択が問題となる閾値を用いずに判定を行い，その意味でかなり客観的な評価基準であるといえる．しかし noisy なデータに対しては，「埋め込みが成立しているときは τ を変化させても $E^a(\tau)$ がほぼ一定である」というこの手法の基本原理が成立するとは限らないであろう．

次章で説明する提案手法は「階層ベイズ的」枠組み[2)~15),22)]から「モデル周辺尤度」を評価基準として最小埋め込み次元を推定しようとするものである．

4. 階層ベイズ的手法

与えられた時系列 $\{x_t\}_{t=0}^N$ からその将来値を予測することが目的の場合，何らかの関数系，たとえばパーセプトロン，RBF（Radial Basis Function）などにより data fitting を行う手法と，必ずしも関数系を用いない手法がある．将来値の予測が目的ではなく，データに内在する何らかの特徴の抽出のみが目的の場合は，関数系で data fitting を行う必要はないであろう．

以下に略述する「階層ベイズ的」手法の最終目標は予測であり，その重要ステップの 1 つが最小埋め込み

次元の推定である．

ベイズ統計的枠組みでは，未知パラメータの尤度（likelihood）と先験分布（prior）が与えられることで問題が定式化される．本稿では，中間層素子数 h，出力関数 $1/(1+e^{-u})$ の 3 層パーセプトロンで学習・予測を行うことを考える．入力数を τ とすると，これが埋め込み次元に相当する．パーセプトロン以外の関数族でも同様に機能する．これらをモデル，あるいは Hypothesis$\mathcal{H}$ と表現する．

（1） 尤度（likelihood）：

$$P\left(\{x_t\}_{t=\tau}^{N}, \{x_{\tau-1}, \ldots, x_0\} \mid \boldsymbol{w}, \beta, \mathcal{H}\right) := \prod_{t=0}^{N-\tau} \frac{1}{Z_D(\beta)} \underbrace{\exp\left(-\beta E_D(x_{t+\tau} \mid x_{t+\tau-1}, \ldots, x_t; \boldsymbol{w})\right)}_{\text{noisy dynamics}} \times \underbrace{P(x_{\tau-1}, \ldots, x_0 \mid \mathcal{H})}_{\text{initial state uncertainty}} \quad (19)$$

$$E_D(x_{t+\tau} \mid x_{t+\tau-1}, \ldots, x_t; \boldsymbol{w}) := \frac{1}{2}(x_{t+\tau} - f(x_{t+\tau-1}, \ldots, x_t; \boldsymbol{w}))^2 \quad (20)$$

ただし $f(\cdot)$ は weight パラメータ $\boldsymbol{w}$ を持つパーセプトロン出力，β は不確定性レベル（未知），$Z_D(\beta)$ は正規化定数である．

（2） $\boldsymbol{w}$ の prior：

$$P(\boldsymbol{w} \mid \boldsymbol{\alpha}, \mathcal{H}) := \prod_{c=1}^{C} \frac{1}{Z_W(\alpha_c)} \exp(-\alpha_c E_{W_c}(\boldsymbol{w}_c))$$

$$E_{w_c}(\boldsymbol{w}_c) := \frac{1}{2} \| \boldsymbol{w}_c \|^2 \quad (21)$$

$\boldsymbol{w}$ は部分ベクトルに分割され，

$$\boldsymbol{w} := (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_C) \quad (22)$$

対応する $\boldsymbol{\alpha}$ はベクトル

$$\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_C)\ ,\ \alpha_c \in \mathbb{R} \quad (23)$$

である．各 $\exp(-\alpha_c E_{W_c}(\boldsymbol{w}_c))/Z_W(\alpha_c)$ は，各 $\boldsymbol{w}_c$ がいかに分布しているかの先験情報である．

（3） $(\boldsymbol{\alpha}, \beta)$ そして $\mathcal{H}$ の prior：$P(\boldsymbol{\alpha}, \beta \mid \mathcal{H})$，$P(\mathcal{H})$

説明

(i) 時系列からダイナミカルシステムの再構成に基づいて予測を行う手法では，3 つの不確定性が考えられる．第 1 はシステム雑音，第 2 はシステムの初期状態の不確定性，そして第 3 は観測雑音である．再構成・予測手法により，困難さは異なるのは当然であるが，カオス的ダイナミカルシステムが関与する場合，どの不確定性も困難な問題と考えられる[23]．本稿ではシステム雑音のみを考慮する．したがって，状態変数は誤差なしで観測可能，そして，初期状態も観測可能と仮定している．第 2，第 3 の不確定性は今後の検討課題である．

(ii) 階層ベイズ的手法は単一の事前分布ではなく，ハイパーパラメータによってパラメータ付けされた事前分布の族を考え，データにハイパーパラメータの分布を語らせようとする手法である．パラメータ推定，ハイパーパラメータ推定，モデル比較に関して，各階層の分母である「周辺尤度」$P(D \mid \boldsymbol{\alpha}, \beta, \mathcal{H})$，$P(D \mid \mathcal{H})$ により統一的に扱うのが要点の 1 つである（**図 3**）．ここで $D = \{x_t\}_{t=0}^{N}$ である．原理としては周辺化を 2 回遂行すれば $P(D \mid \mathcal{H})$ を得る．これを最大化する $\mathcal{H}$ を選び，$P(\boldsymbol{\alpha}, \beta \mid D, \mathcal{H})$ を最大化する $(\boldsymbol{\alpha}, \beta)$ を計算する．これを基に $P(\boldsymbol{w} \mid D, \boldsymbol{\alpha}, \beta, \mathcal{H})$ を最大化して weight の事後モードを計算する．周辺化は一般に解析表示不可能なので文字どおり思いきった近似が必要である．

近似の第 1 は，$\log P(\boldsymbol{w} \mid D, \boldsymbol{\alpha}, \beta, \mathcal{H})$ の 2 次近似である．すなわち，事後モード $\boldsymbol{w}_{MP}$ において 2 次関数で近似して，周辺尤度 $P(D \mid \boldsymbol{\alpha}, \beta, \mathcal{H})$ を計算する．

近似の第 2 は，ハイパーパラメータ事後分布 $P(\boldsymbol{\alpha}, \beta \mid D, \mathcal{H})$ に関するものである．事後モード $(\boldsymbol{\alpha}_{MP}, \beta_{MP})$ で鋭いピークを持つと仮定し，また，ハイパーパラメータの事前分布を一様と仮定し，

$$P(\mathcal{H} \mid D) \propto P(D \mid \boldsymbol{\alpha}_{MP}, \beta_{MP}, \mathcal{H}) \quad (24)$$

を考える．

(iii) 具体的な計算の実装は次のようになる．まず $\mathcal{H}$ を複数用意する．各々の $\mathcal{H}$ に対して，$\boldsymbol{w}$ に関する $P(\boldsymbol{w} \mid D, \boldsymbol{\alpha}, \beta, \mathcal{H})$ の最大化と $\boldsymbol{\alpha}$，β に関する $P(D \mid \boldsymbol{\alpha}, \beta, \mathcal{H})$ の最大化を交互に繰り返して事後モード $\boldsymbol{w}_{MP}$，$\boldsymbol{\alpha}_{MP}$，β_{MP} を計算する．そして最終的に，$P(D \mid \mathcal{H})$ が最大となる $\mathcal{H}$，つまり $P(D \mid \boldsymbol{\alpha}_{MP}, \beta_{MP}, \mathcal{H})$ が最大となる $\mathcal{H}$ を選択し，

Level 1：パラメータの事後分布

$$P(\boldsymbol{w} \mid D, \boldsymbol{\alpha}, \beta, \mathcal{H}) = \frac{P(D \mid \boldsymbol{w}, \beta, \mathcal{H})\ P(\boldsymbol{w} \mid \boldsymbol{\alpha}, \mathcal{H})}{P(D \mid \boldsymbol{\alpha}, \beta, \mathcal{H})}$$

Level 2：ハイパーパラメータの事後分布

$$P(\boldsymbol{\alpha}, \beta \mid D, \mathcal{H}) = \frac{P(D \mid \boldsymbol{\alpha}, \beta, \mathcal{H})\ P(\boldsymbol{\alpha}, \beta \mid \mathcal{H})}{P(D \mid \mathcal{H})}$$

Level 3：モデルの事後分布

$$P(\mathcal{H} \mid D) = \frac{P(D \mid \mathcal{H})\ P(\mathcal{H})}{P(D)}$$

（$\mathcal{H}$：モデル，$\boldsymbol{w}$：荷重，$\boldsymbol{\alpha}$，β：ハイパーパラメータ，D：データ）

図 3 階層ベイズ構造

Fig. 3 Hierarchical Bayesian structure.

それが $P(\mathcal{H} \mid D)$ の事後モード $\mathcal{H}_{MP}$ と考える．その埋め込み次元を最小埋め込み次元の推定値とする．$P(\boldsymbol{w} \mid D, \boldsymbol{\alpha}, \beta, \mathcal{H})$ および $P(D \mid \boldsymbol{\alpha}, \beta, \mathcal{H})$ の最大化は，適当な適化手法，たとえば，共役勾配法を用いて行う．計算は比較的重くなるが，この手法は後述する問題でも，それ以外の問題[12)~15)]でも機能する場合は少なくない．

(iv) weight パラメータの次元，分割などについて述べる．遅延座標値 $\{x_t, x_{t-1}, \cdots, x_{t-\tau+1}\}$ がパーセプトロン各入力層素子ごとに入力され，最初の座標 x_t から中間層素子への weight をまとめて $\boldsymbol{w}_1$ とする（次元 h）．同様に，第 2 座標 x_{t-1} から中間層素子への weight をまとめて $\boldsymbol{w}_2$ とする（次元 h）．以下同様とすれば，τ 個の weight グループができる．また，中間層素子のバイアス項を 1 つのまとまりとして考える．さらに，中間層素子から出力素子につながる weight とそのバイアスを 1 つにまとめると，あわせて $C = \tau + 2$ 個の weight グループができる．全体としての weight $\boldsymbol{w}$ の次元は，$h(\tau+2)+1$ となる．各 weight のまとまりごとに，先験分布として平均 0，分散 $1/\alpha_c$，$\{c = 1, \cdots, \tau+2\}$ の正規分布を仮定する．たとえば，ある weight のまとまりに関する α_c の事後分布が狭い領域に大きな値で集中していれば，これはこの weight のまとまりが原点近くに集中しており，対応する遅れ座標は予測に対する貢献度が低い，と考える．実際，後述する数値実験ではこのような考えに基づき，良好な結果が導かれる．このような考え方は他の問題でも良好に機能している[24)]．

(v) 具体的関数系（パーセプトロン）が指定されているので，モデル $\mathcal{H}$ の自由度は埋め込み次元 τ と中間層素子数 h である．モデルの prior が一様の場合，最小埋め込み次元はモデル周辺尤度 $P(D \mid \mathcal{H})$ が最大になるようなモデルを選択することにより推定される．

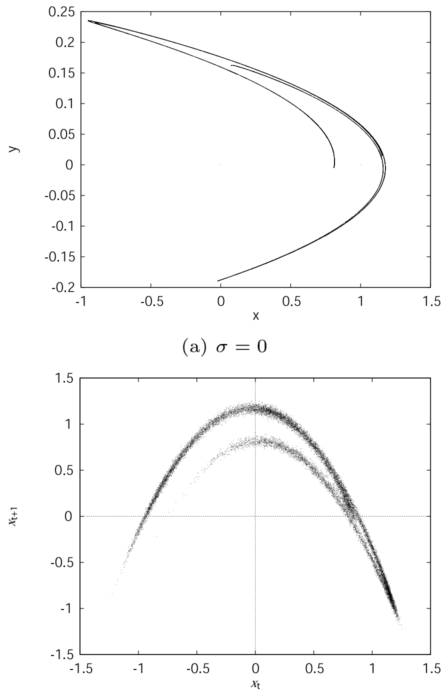
5. 階層ベイズ的手法と FNN 法の比較

代表的なカオス系である Henon 系，Rössler 系にシステムノイズを加えた系から得られた 1 次元時系列データ，および正規白色雑音 $\{x_t\}_{t=0}^{N}$ に対して FNN 法，AFN 法，階層ベイズ的手法を適用し，最小埋め込み次元を推定する．具体的には次のようにして得られるデータを用いる：

(1) Henon 系

システムノイズを含む Henon 系

$$\begin{cases} x_{t+1} = y_t + 1 - a x_t^2 + \mu_t^1 \\ y_{t+1} = b x_t + \mu_t^2 \end{cases} \tag{25}$$



(a) $\sigma = 0$

(b) $\sigma = 0.02$

図 4 Henon アトラクタ
Fig. 4 Henon attractor.

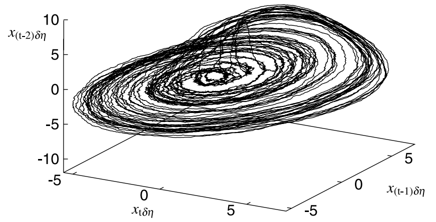


図 5 Rössler 系の遅延座標系への埋め込み（$\sigma = 0.02$）
Fig. 5 Delay coordinate embedding for Rössler system ($\sigma = 0.02$).

$a = 1.4$，$b = 0.2$，$\mu_t^i \sim i.i.d.\ N(0, \sigma^2)$，$i = 1, 2$ を考える．**図 4** (a) は $\sigma = 0$（ノイズがない場合），同図 (b) は $\sigma = 0.02$ の場合の (x_t, y_t)–プロットである．この系から得られた x 座標 $\{x_t\}_{t=0}^{N}$ を用いる．

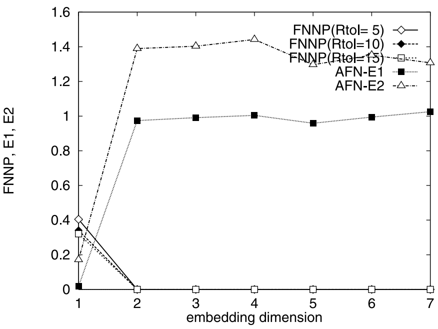
(2) Rössler 系

図 5 はシステムノイズ ν_t^i（$i = 1$〜3）を含む Rössler 系

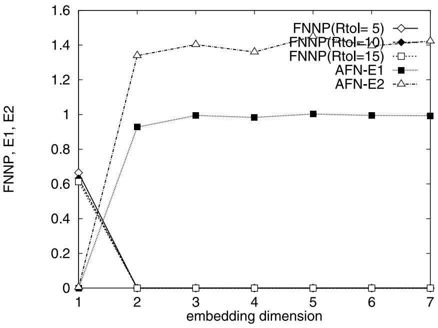
$$\begin{cases} \dot{x} &= -y - z + \nu_t^1 \\ \dot{y} &= x + ay + \nu_t^2 \\ \dot{z} &= bx - cz + xz + \nu_t^3 \end{cases} \tag{26}$$

$$(a, b, c) = (0.36, 0.4, 4.5)$$

をステップ幅 δ の Runge-Kutta で離散化して得られ



(a) データ数 $N = 300$



(b) データ数 $N = 30000$

図 6 ノイズのない Henon 系（$\sigma = 0$）に対する FNNP および $E1(\tau)$, $E2(\tau)$

Fig. 6 FNNP and $E1(\tau)$, $E2(\tau)$ for noiseless Henon system ($\sigma = 0$).

るダイナミカルシステム

$$\begin{cases} x_{(t+1)\delta} &=& f(x_{t\delta}, y_{t\delta}, z_{t\delta}) + \nu^1_{t\delta} \\ y_{(t+1)\delta} &=& g(x_{t\delta}, y_{t\delta}, z_{t\delta}) + \nu^2_{t\delta} \\ z_{(t+1)\delta} &=& h(x_{t\delta}, y_{t\delta}, z_{t\delta}) + \nu^3_{t\delta} \end{cases} \quad (27)$$

$$\nu^i_{t\delta} \sim i.i.d.\ N(0, \sigma^2),\ i = 1, 2, 3$$

の時系列 $\{x_{t\delta\eta}\}$ を $\tau = 3$ の遅延座標空間に埋め込んだものである．ここに，$\delta = 0.01$ は Runge-Kutta ステップサイズ，したがって $\delta\eta$ はサンプリング間隔を意味している（$\eta = 70$, $\sigma = 0.02$）．既出の図 2 は $\sigma = 0$ の場合である．この系から得られた x 座標 $\{x_t\}_{t=0}^N$ を用いる．

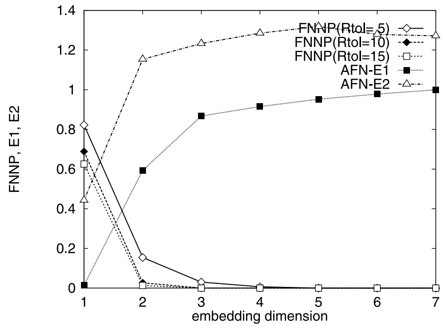
(3) 正規白色雑音

$$x_t \sim i.i.d. N(0, 1) \quad (28)$$

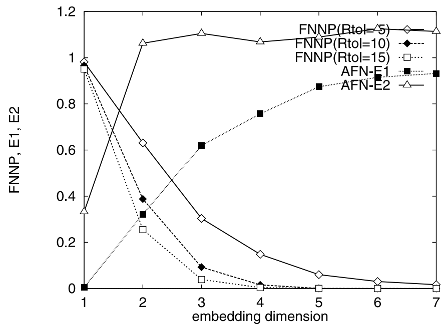
から得られたデータ $\{x_t\}_{t=0}^N$ を用いる．

5.1 FNN 法，AFN 法

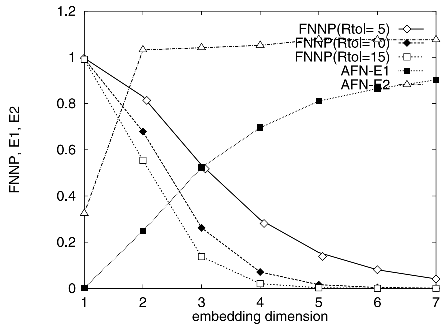
FNN 法では 2 つの閾値 R_{tol}, A_{tol} をあらかじめ選択する必要があるが，16) では経験的に適切な値として $A_{tol} = 2.0$ を用いており，著者らが行った実験の限りでも A_{tol} の選択はデータの種類にはほぼ依存せず，$A_{tol} = 1.5$〜2.5 程度で基準 2（式 (14)）は良好に機能した．したがって本稿では $A_{tol} = 2.0$ に固定して R_{tol} のみを変化させ，R_{tol} の選択が結果に与える影響を調べた．また，同じデータに対して AFN 法も適用した．



(a) データ数 $N = 300$



(b) データ数 $N = 5000$



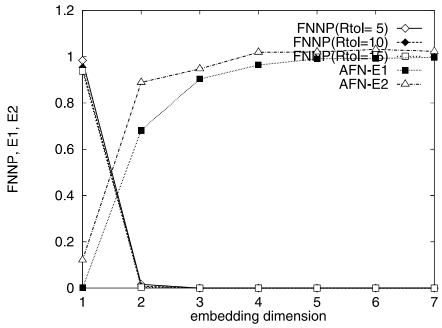
(c) データ数 $N = 30000$

図 7 Noisy な Henon 系（$\sigma = 0.02$）に対する FNNP および $E1(\tau)$, $E2(\tau)$

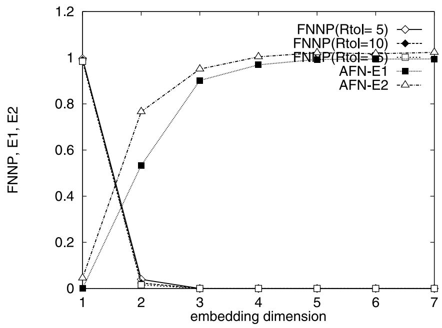
Fig. 7 FNNP and $E1(\tau)$, $E2(\tau)$ for noisy Henon system ($\sigma = 0.02$).

5.1.1 Henon 系

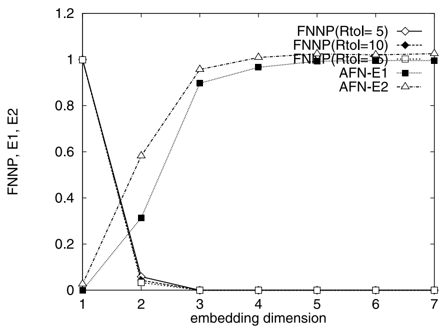
図 6, **図 7** はそれぞれ $\sigma = 0$, 0.02 の Henon 系から得られたデータ $\{x_t\}_{t=0}^N$ に対して FNN 法，AFN 法を適用したものである．ノイズを含まない場合（$\sigma = 0$, 図 6），FNN 法ではデータ数，閾値 R_{tol} の値によらずに FNNP が埋め込み次元 $\tau = 2$ で 0 に飽和しており，最小埋め込み次元は 2 と推定される．



(a) データ数 $N = 500$



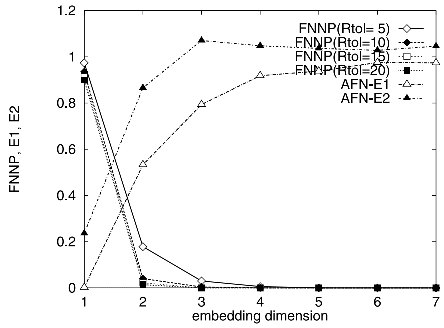
(b) データ数 $N = 5000$



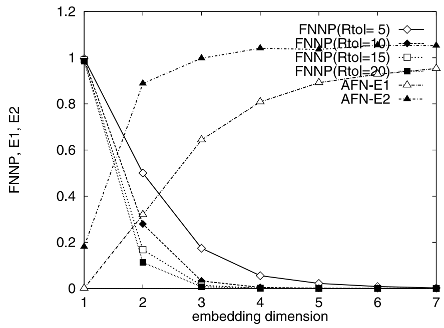
(c) データ数 $N = 30000$

図 8　ノイズのない Rössler 系（$\sigma = 0$）に対する FNNP および $E1(\tau)$，$E2(\tau)$

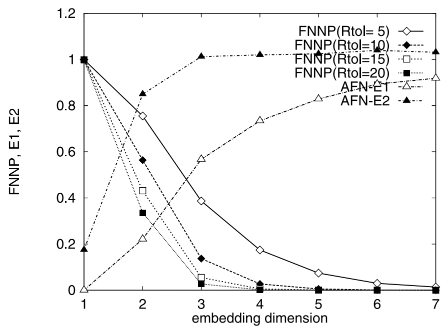
Fig. 8　FNNP and $E1(\tau)$, $E2(\tau)$ for noiseless Rössler system ($\sigma = 0$).



(a) データ数 $N = 500$



(b) データ数 $N = 5000$



(c) データ数 $N = 30000$

図 9　Noisy な Rössler 系（$\sigma = 0.02$）に対する FNNP および $E1(\tau)$，$E2(\tau)$

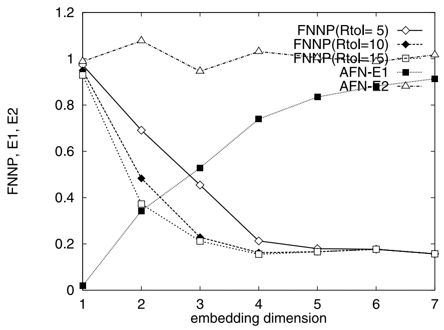
Fig. 9　FNNP and $E1(\tau)$, $E2(\tau)$ for noisy Rössler system ($\sigma = 0.02$).

AFN でもデータ数によらずに $E1$ は $\tau = 2$ で 1 に飽和しており，最小埋め込み次元は 2 と推定される．$\sigma = 0$ のときの最小埋め込み次元は式 (25) より解析的に 2 であることが示せ，FNN 法，AFN 法による推定結果と一致している．

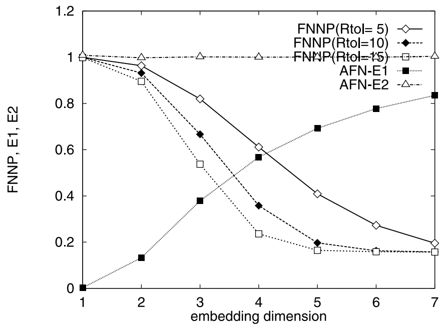
noisy な信号の場合（$\sigma = 0.02$，図 7），FNN 法は FNNP がデータ数，閾値 R_{tol} の値に大きく依存しており，R_{tol} の適切な値の選択は困難で，かなり主観的に R_{tol} を選択しないと機能しないように思われる．また，AFN 法ではノイズの影響で埋め込み次元 $\tau \leq 7$ では $E1$ が 1 に飽和していない．

5.1.2　Rössler 系

図 8，**図 9** はそれぞれ $\sigma = 0$，0.02 の Rössler 系から得られたデータ $\{x_t\}_{t=0}^{N}$ に対して FNN 法，AFN 法を適用したものである．ノイズを含まない場合（$\sigma = 0$，図 8），FNN 法はデータが少ないとき（$N = 500$）は $\tau = 2$ で FNNP がほぼ 0 となっているが，データが多いとき（$N = 30000$）は $\tau = 3$ で FNNP が 0 となっている．図 2 (b) より Rössler 系は実際には $\tau = 2$ では埋め込みが成立しないことは明らかであり，FNN 法はかなり多くのデータが必要であると推察される．また，データ数を考慮して閾値を選択する必要がある．



(a) データ数 $N = 300$



(b) データ数 $N = 30000$

図 10　正規白色雑音 $i.i.d.N(0,1)$ に対する FNNP および $E1(\tau)$，$E2(\tau)$

Fig. 10　FNNP and $E1(\tau)$，$E2(\tau)$ for gaussian white noise.

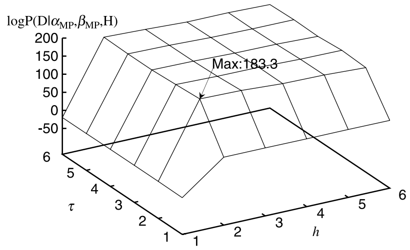


図 11　ハイパーパラメータ対数周辺尤度（$\sigma = 0.02$ の Henon 系，データ数 $N = 100$）

Fig. 11　$\log P(D \mid \boldsymbol{\alpha}_{\mathrm{MP}}, \beta_{\mathrm{MP}}, \mathcal{H})$ against τ and h for Henon system with $\sigma = 0.02$, $N = 100$.

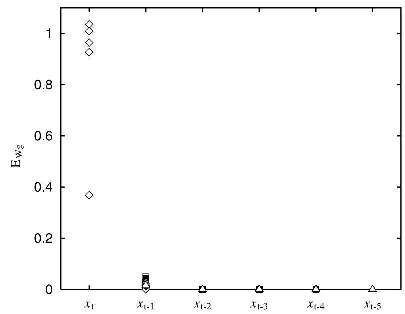


図 12　各遅延座標に対応する E_{W_c}（$\sigma = 0.02$ の Henon 系，データ数 $N = 100$）

Fig. 12　E_{W_c} against each delay coordinate for Henon system with $\sigma = 0.02$, $N = 100$.

AFN 法では $E1$ が 1 に飽和したときに埋め込みが成立していると考えるが，図 8 から最小埋め込み次元を 3〜5 のどの次元と考えるかは主観的判断が必要である．

noisy な信号の場合（$\sigma = 0.02$，図 9）は Henon 系の場合と同様に FNN 法，AFN 法ともに最小埋め込み次元の推定が困難である．

5.1.3　正規白色雑音

図 10 は式 (28) の正規白色雑音 $\{x_t\}_{t=0}^{N}$ に対して FNN 法，AFN 法を適用したものである．FNN 法では τ を上げても FNNP は 0 にはならない．AFN 法では τ によらずに $E2(\tau) = 1$ になっている．したがって，FNN 法，AFN 法ともに白色雑音と deterministic な信号を識別できているといえる．

5.2　階層ベイズ的手法

5.2.1　noisy な Henon 系（$\sigma = 0.02$）

図 11 は $\sigma = 0.02$ の Henon 系から得られたデータ $\{x_t\}_{t=0}^{100}$ に対して 4 章で述べたアルゴリズムを適用し，埋め込み次元 τ，中間層素子数 h に対してハイパーパラメータ対数周辺尤度 $\log P(D \mid \boldsymbol{\alpha}_{\mathrm{MP}}, \beta_{\mathrm{MP}}, \mathcal{H})$ をプロットしたものである．$(\tau, h) = (2, 2)$ で $\log P(D \mid \boldsymbol{\alpha}_{\mathrm{MP}}, \beta_{\mathrm{MP}}, \mathcal{H})$ は最大となり以後は飽和しており，最小埋め込み次元は 2 と推定される．

図 12 は各モデルに関して各遅延座標につながる事後モード $\boldsymbol{w}_{MP}$ における二乗和 $E_{W_c} = \frac{1}{2}||\boldsymbol{w}_{cMP}||^2$ をプロットしたものである．事後モード $\boldsymbol{w}_{MP}$ の計算は前述のように，共役勾配法を用いている．Monte Carlo も可能であるが，これについては別の論文で報告する予定である．図 12 はスケールの都合上，x_t に対応するプロットの大部分がグラフの上にはみ出している．複数プロットはモデル（遅延座標の数と中間層素子数）が複数あることによる．$x_{t-i}, i \geq 2$ につながる荷重はほぼ完全に 0 になっており，埋め込み次元は 2 で十分であるという推定結果を裏付けている．

また，データを作成する際の初期値，つまり式 (25) の (x_0, y_0) を 20 種類用意し，その各々に対して同様の手法で最小埋め込み次元の推定を行った．すべてのデータセットに対して図 11 と同様に $(\tau, h) = (2, 2)$ でハイパーパラメータ対数周辺尤度がほぼ飽和していることが確認された．ハイパーパラメータ対数周辺尤度が最大となる埋め込み次元のヒストグラムを**図 13** に示す．3 つのデータセットに対しては埋め込み次元が 2 よりも大きい場合に周辺尤度が最大となったが，

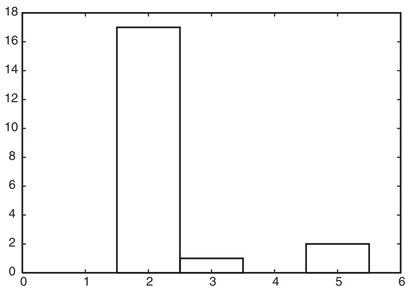


図 13 Henon 系のデータセットを 20 セット用意して階層ベイズ的手法を適用したときの，ハイパーパラメータ対数周辺尤度が最大となる埋め込み次元のヒストグラム

Fig. 13 Histogram of embedding dimensions with maximum hyperparameter marginal likelihood for 20 different data sets of the Henon system.

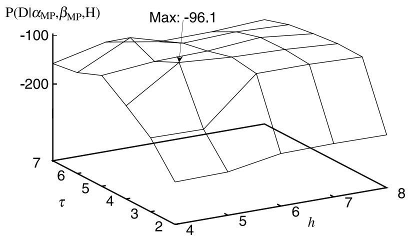


図 14 ハイパーパラメータ対数周辺尤度（$\sigma = 0.02$ の Rössler 系，データ数 $N = 500$）

Fig. 14 $\log P(D \mid \boldsymbol{\alpha}_{\mathrm{MP}}, \beta_{\mathrm{MP}}, \mathcal{H})$ against τ and h for Rössler system with $\sigma = 0.02$, $N = 500$.

そのモデルでは $x_{t-i}, i \geq 2$ の入力につながる荷重はほぼ 0 であることが確認できた．したがって，すべてのデータセットに対して最小埋め込み次元が 2 と推定されたと考えることができ，周辺尤度の評価と荷重値の評価を組み合わせて推定を行うことの有効性を示している．

5.2.2 noisy な Rössler 系（$\sigma = 0.02$）

図 14 は $\sigma = 0.02$ の Rössler 系から得られたデータ $\{x_t\}_{t=0}^{500}$ に対して 4 章で述べたアルゴリズムを適用し，埋め込み次元 τ，中間層素子数 h に対してハイパーパラメータ対数周辺尤度をプロットしたものである．$(\tau, h) = (4, 5)$ で周辺尤度は最大となりそれ以後は各々ほぼ飽和していると考えられ，最小埋め込み次元は 4 と推定される．

図 15 は各遅延座標系につながる荷重の二乗和 E_{W_c} をプロットしたものである．$x_{(t-i)\delta\eta}, i \geq 4$ につながる荷重は極端に小さく推定されており，埋め込み次元は 4 で十分であるという推定結果を裏付けている．

また，上述の Henon 系と同様に複数のデータセットを準備した．式 (27) の (x_0, y_0) を 20 種類用意し，その各々に対して同様の手法で最小埋め込み次元の推

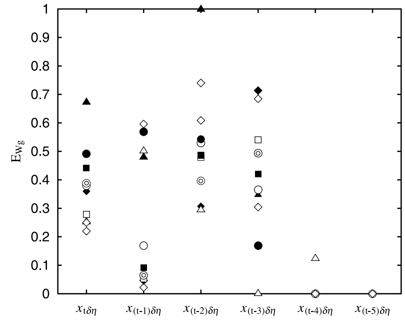


図 15 各遅延座標に対応する E_{W_c}（$\sigma = 0.02$ の Rössler 系，データ数 $N = 500$）

Fig. 15 E_{W_c} against each delay coordinate for Rössler system with $\sigma = 0.02$, $N = 500$.

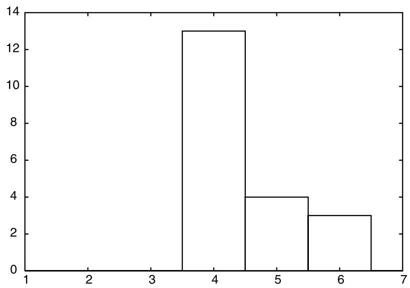


図 16 Rössler 系のデータセットを 20 セット用意して階層ベイズ的手法を適用したときの，ハイパーパラメータ対数周辺尤度が最大となる埋め込み次元のヒストグラム

Fig. 16 Histogram of embedding dimensions with maximum hyperparameter marginal likelihood for 20 different data sets of the Rössler system.

定を行った．ハイパーパラメータ対数周辺尤度が最大となる埋め込み次元のヒストグラムを**図 16** に示す．

階層ベイズ的手法は計算時間は長いが（FNN 法，AFN 法の 100 倍程度），データがノイズを含み，データ数が少ない場合でも最小埋め込み次元が良好に推定されていると思われる．

5.2.3 正規白色雑音

図 17 は式 (28) の正規白色雑音 $\{x_t\}_{t=0}^{500}$ に対して 4 章のアルゴリズムを適用し，各遅延座標系につながる荷重の二乗和 E_{W_c} をプロットしたものである．すべての遅延座標につながる荷重が 0 になっており，x_{t+1} は $x_t, x_{t-1}, ..., x_{t-\tau+1}$ とは独立である可能性が高いことが分かる．したがって，階層ベイズ的手法では deterministic な信号と白色雑音の識別が可能であるといえる．

5.2.4 階層ベイズ的手法の限界

データ数が少なくなっていった場合，提案手法がどの程度うまく働くかを調べるのは興味深いことであろう．**図 18** は，5.2.1 項で行った Henon 系の実験に対

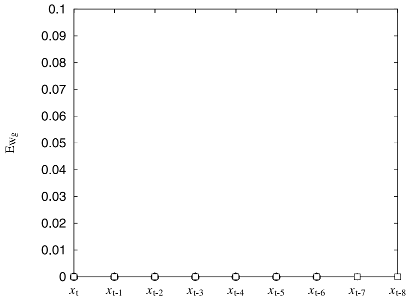


図 17 各遅延座標に対応する E_{W_c}（正規白色雑音）

Fig. 17 E_{W_c} against each delay coordinate (gaussian white noise).

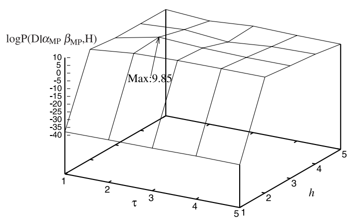


図 18 ハイパーパラメータ対数周辺尤度（$\sigma = 0.08$ の Henon 系，データ数 $N = 50$）

Fig. 18 $\log P(D \mid \boldsymbol{\alpha}_{\mathrm{MP}}, \beta_{\mathrm{MP}}, \mathcal{H})$ against τ and h for Henon system with $\sigma = 0.08$, $N = 50$.

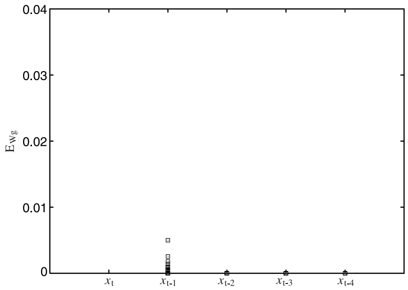


図 19 各遅延座標に対応する E_{W_c}（$\sigma = 0.08$ の Henon 系，データ数 $N = 50$）

Fig. 19 E_{W_c} against each delay coordinate for Henon system with $\sigma = 0.08$, $N = 50$.

してデータ数をさらに少なくし，$\sigma = 0.08$ の場合，Henon 系から得られたデータ $\{x_t\}_{t=0}^{50}$ に対して 4 章のアルゴリズムを適用してモデル周辺尤度をプロットしたものである．この場合でも推定アルゴリズムは機能しているように思われる．**図 19** は各遅延座標に接続している荷重の二乗和をプロットしたものである．

次に，$\sigma = 0.08$ は同様とし，データ数を 20 点とした場合を調べた．周辺尤度は計算不能となり，重み二乗和は，**図 20** で示されたようになり，すべての遅延座標につながる荷重が 0 となった．これは 5.2.3 項の正規白色雑音に対するものと同様の結果となっている．アルゴリズムの限界を示すものと思われる．

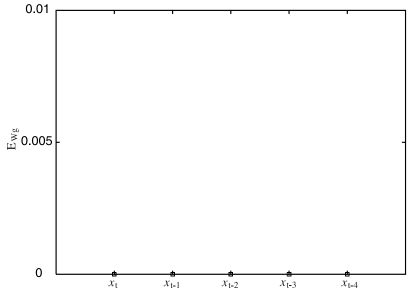


図 20 各遅延座標に対応する E_{W_c}（$\sigma = 0.08$ の Henon 系，データ数 $N = 20$）

Fig. 20 E_{W_c} against each delay coordinate for Henon system with $\sigma = 0.08$, $N = 20$.

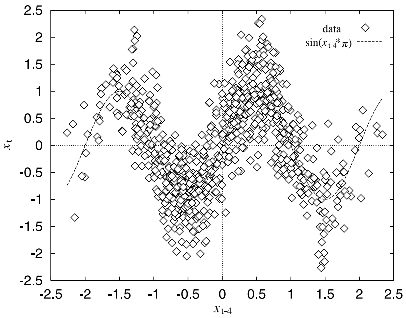


図 21 $x_t = sin(x_{t-4}\pi) + \mu_t$ から得られるデータの (x_t, x_{t-4}) の遅延座標空間でのプロット

Fig. 21 Delay coordinate plot (x_t, x_{t-4}) from $x_t = sin(x_{t-4}\pi) + \mu_t$.

5.2.5 サンプリング間隔の考察

上述の例ではサンプリング間隔が既知とし，最小埋め込み次元の推定のみを行った．次に述べる例は単純ではあるが，サンプリング間隔の考察も必要とする問題である．提案手法が機能することを述べる．

離散ダイナミカルシステム

$$x_t = sin(x_{t-4}\pi) + \mu_t \qquad (29)$$
$$\mu_t \sim i.i.d.N(0, 0.5^2)$$

から得られる時系列データ $\{x_t\}_{t=0}^{N}$ を (x_t, x_{t-4}) 空間にプロットすると**図 21** のようになる．

図 22 はデータ数 $N = 200$，時間遅れの大きさを 1，埋め込み次元 $\tau = 4$ として 4 節で説明した階層ベイズ的手法を適用し，各遅延座標につながる荷重の二乗和をプロットしたものである．x_t，x_{t-1}，x_{t-2} の座標につながる荷重はほぼ 0 になっており，時間遅れの大きさを 4 とすれば最小埋め込み次元は 1 になることを意味している．

各遅延座標につながる荷重の値を評価することで埋め込み次元と時間遅れの大きさの選択が可能であり，階層ベイズ的手法には図 3 の階層構造の中に自動的に埋め込み次元と時間遅れの大きさの選択を行うような

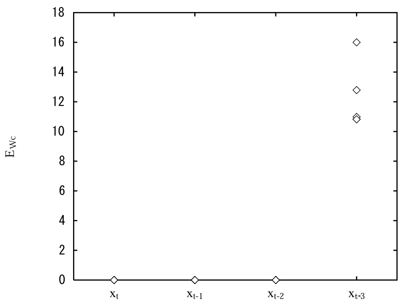


図 22 各遅延座標に対応する E_{W_c}（$x_t = sin(x_{t-4}\pi) + \mu_t$）
Fig. 22 Squared weight norm associated with each delay coordinate for $x_t = sin(x_{t-4}\pi) + \mu_t$.

構造が用意されていると考えられる．

6. まとめ

Henon 系，Rössler 系の x–時系列から階層ベイズ的手法，FNN 法，AFN 法の各手法で最小埋め込み次元を推定した．FNN 法は原理の簡易さ，計算時間において優れているが，特にデータがノイズを含む場合「閾値をどのようにして決定するのか」という根本問題があり，客観的な決定は困難である．AFN 法では閾値を用いずに最小埋め込み次元の推定を行い，その意味で客観的な基準であるといえるが，noisy なデータの最小埋め込み次元を推定することは困難である．階層ベイズ的手法では現れるパラメータはすべてデータに語らせることを基本原理としており，FNN 法での問題が克服されているといえる．また，FNN 法，AFN 法と比してノイズに対して頑健であると推察される．文献 16) に指摘されているように，大きなノイズを含むデータに FNN 法を用いるときは何らかのフィルタを用いてノイズを軽減させる必要があり，これは AFN 法でも同様である．

FNN 法，AFN 法は計算時間が非常に短く，最小埋め込み次元の目安を推定するには強力な手法である．階層ベイズ的手法は FNN 法，AFN 法と比して計算時間は長いが，データ数が少ない場合でも安定して埋め込み次元を推定できると同時に，ダイナミカルシステム再構成・時系列予測までを一貫して行うことが可能である．したがって FNN 法，あるいは AFN 法で最小埋め込み次元の目安を推定し，その結果をふまえて階層ベイズ的手法で最小埋め込み次元の推定・時系列予測などを行うことは有効な手法であろう．

推定モデルの複雑さに関して，MDL 原理がある．この枠組みでは事象の生起確率の代わりに，メッセージを誤りなく転送するために必要なビット長を考察する．単純化していえば，確率モデルの事象 X に関して，

$$L(X) = -\log_2 P(X) \qquad (30)$$

で定義される量に注目する．モデル選択について考えるため，どのモデル $\mathcal{H}$ を用いるかを宣言するメッセージと，このモデル $\mathcal{H}$ を用いてデータ D を転送することを考える．これは，次のようなメッセージ長を定義する：

$$L(D, \mathcal{H}) = L(\mathcal{H}) + L(D \mid \mathcal{H}) \qquad (31)$$

異なるモデル $\mathcal{H}$ に対するメッセージ長 $L(\mathcal{H})$ はモデルの prior（先験情報）を定義し，$L(D \mid \mathcal{H})$ はモデル周辺尤度 $P(D \mid \mathcal{H})$ に対応する．パラメータ数の少ないモデル $\mathcal{H}$ はモデルに関するメッセージ長は短いが，データフィットの度合いが低いため，データ長が増加する．一方，データ長が短くてすむようにすると，パラメータ数が多いモデルを用いる必要が生じる．両者の兼ね合いから，最もふさわしい複雑度を持つモデルを選択することができるであろう．このように，本稿で述べたベイズ的手法は，基本的に MDL 原理との対応付けが可能である．

メッセージを送る際の精度の問題があるが，これに関してはここではこれ以上触れないほうが賢明と思われる．

残された問題としては，次のようなものがある：

(1) モデル周辺尤度 $P(D \mid \mathcal{H})$ の計算に対する近似式 (24) が働かない場合，いかにして計算するか．

(2) 4 章の定式化はシステムノイズのみを考慮している．観測ノイズを含めた定式化はカオスの初期値依存性のために技術的にきわめて困難であるが，非常に興味深い問題である．

(3) もともと Takens 埋めこみ定理の理論的背景は，ダイナミカルシステム全体を記述する関数空間における埋めこみ写像の，適当な位相に関する genericity（開，かつ稠密な集合）に基づくものである．文献 21) は確率的ダイナミカルシステムに対応する埋め込み定理を，このような考え方の延長線上でとらえている．当然であるが確率過程全体を記述する空間に何らかの位相的概念を持ちこみ，そこで genericity を考える必要がある．本稿で扱っているモデルや関連する多くの論文で使用されているモデルに対して厳密に文献 21) の定理 6 あるいは定理 7 が適用されるわけではない．理由は，ノイズ過程の標本を 1 つ固定するごとに埋め込み写像が存在することが保証されているのであって，ノイズへの滑らかな依存性は保証されていないからである．本稿の結果を含め，いくつかの数値実験結

果は理論家に対して興味深いチャレンジを提起しているといってよいであろう．

謝辞 2人の査読者から建設的コメントを多数いただき，論文の趣旨をより明確にすることができた．

参 考 文 献

1) Akaike, H.: Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory*, Petrov, B.N. and Csaki, F. (Eds.), pp.267–281 (1973).
2) 松本　隆，浜岸広明，杉淳二郎，斉藤幹貴，長南吉正：データから非線形ダイナミックスへ—階層ベイズ的時系列予測，電子情報通信学会論文誌 D-II，Vol.J82-D-II, No.6, pp.1059–1071 (1999).
3) Matsumoto, T., Hamagishi, H. and Chonan, Y.: A Hierarchical Bayes Approach to Nonlinear Time Series Prediction with Neural Nets, *Proc. 1997 International Conference on Neural Networks*, Houston, pp.2028–2033 (1997).
4) Chonan, Y., Nishida, K. and Matsumoto, T.: A Bayesian Nonlinear Regression with Multiple Hyperparameters on the ASHRAE II Time Series Data, Vol.102 PART 2, *Trans. ASHRAE*, pp.405–411 (1996).
5) Matsumoto, T., Hamagishi, H., Sugi, J. and Saito, M.: From Data-to Dynamics: Predicting Chaotic Time Series by Hierarchical Bayesian Neural Nets, *Proc. 1998 IEEE World Congress on Computational Intelligence*, Anchorage, pp.2535–2540 (1998).
6) Nakajima, Y., Sugi, J., Saito, M., Hamagishi, H., Hattori, D. and Matsumoto, T.: Hierarchical Bayesian neural nets for air-conditioning load prediction: Nonlinear Dynamics Approach, *Proc. 1998 IEEE World Congress on Computational Intelligence*, Anchorage, pp.1948–1953 (1998).
7) Matsumoto, T., Nakajima, Y., Hamagishi, H., Sugi, J. and Saito, M.: From Data to Nonlinear Dynamics : A Hierarchical Bayes Approach with Neural Nets, *Proc.1998 IEEE Neural Networks for Signal Processing*, pp.333–342 (1998).
8) Matsumoto, T., Hamagishi, H., Sugi, J. and Saito, M.: Chaotic Time Series Prediction via Hierarchical Bayesian Neural Net, *Proc. 5th International Conference on Neural Information Processing*, Kitakyushu, pp.1020–1023 (1998).
9) Nakajima, Y., Sugi, J., Saito, M., Hamagishi, H. and Matsumoto, T.: A Hierarchical Bayes Algorithm for Air-conditioning Load Prediction: Nonlinear Dynamics Approach, *Proc. 5th International Conference on Neural Information Processing*, Kitakyushu, pp.1347–1350 (1998).
10) Matsumoto, T., Saito, M., Nakajima, Y., Sugi, J. and Hamagishi, H.: Reconstruction and prediction of nonlinear dynamical systems: a hirearchical Bayes approach with neural nets, *1999 IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, USA, pp.1057–1060 (1999).
11) Matsumoto, T., Saito, M., Nakajima, Y., Sugi, J. and Hamagishi, H.: A Hierarchical Bayes Approach To Reconstruction and Prediction Of Nonlinear Dynamical Systems, *Proc. IEEE EURASIP Workshop on Nonlinear Signal and Image Processing*, Antalya, Turkey, pp.114–118 (1999).
12) Matsumoto, T., Saito, M. and Sugi, J.: Nonlinear Time Series Prediction Weighted by Marginal Likelihoods: A Hierarchical Bayesian Approach, *Proc.International Joint Conference on Neural Networks*, Washington, DC (1999).
13) Nakajima, Y., Saito, M., Sugi, J. and Matsumoto, T.: Nonlinear Dynamical Systems Approach to Building Energy Prediction Problems, *Proc. Building Simulation'99*, Kyoto, Japan, pp.901–907 (1999).
14) Matsumoto, T., Nakajima, Y., Saito, M. and Sugi, J.: A Hierarchical Bayesian Scheme for Nonlinear Dynamical System Reconstruction and Prediction with Neural Nets, *Proc. 1999 IEEE International Conference on Systems, Man, and Cybernetics*, Tokyo, Japan, Vol.IV, pp.1119–1124 (1999).
15) 中島芳徳，斉藤幹貴，杉淳二郎，浜岸広明，松本　隆：階層ベイズ的空調機熱負荷予測—ニューラルネットによる非線形ダイナミカルシステム的アプローチ，電子情報通信学会論文誌 D-II，Vol.J82-D-II, No.11, pp.2075–2083 (1999).
16) Kennel, M.B. and Abarbanel, R.B.H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Phy. Review*, Vol.A 45, pp.3403–3411 (1992).
17) Cao, L.: Practical method for determining the minimum embedding dimension of a scalar time series, *Physica*, Vol.D 110, pp.43–50 (1997).
18) Cao, L., Mees, A., Judd, K. and Froyland, G.: Determining the Embedding Dimension of Input-output Time Series Data, *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, Vol.8, pp.1491–1504 (1998).
19) Sauer, T., Yorke, J. and Casdagli, M.: Em-

beddology, *J. Stat. Phys.*, Vol.65, pp.579–616 (1991).
20) Matsumoto, T., Komuro, M., Kokubu, H. and Tokunaga, R.: *Bifurcations*, Springer-Verlag (1993).
21) Stark, J., Broomhead, D.S., Davies, M.E., and Huke, J.: Takens embedding theorems for forced and stochastic systems, *Nonlinear Analysis, Theory Methods and Applications*, Vol.30, No.8, pp.5303–5314 (1997).
22) Matsumoto, T., Nakajima, Y., Saito, M., Sugi, J. and Hamagishi, H.: Reconstructions and Predictions of Nonlinear Dynamical Systems: A Hierarchical Bayesian Approach, *IEEE Trans. Signal Processing*, Vol.49, pp.2138–2155 (2001).
23) Judd, K.: Chaotic-time-series reconstruction by the Bayesian paradigm: Right results by wrong methods, *Phys. Rev. E*, Vol.67, 026212 (2003).
24) MacKay, D.J.C.: Bayesian non-linear modeling for the prediction competition, *Trans. ASHRAE*, Vol.100, Pt. 2, pp.1053–1062 (1994)

(平成 15 年 5 月 6 日受付)
(平成 15 年 10 月 16 日採録)

杉　淳二郎

平成 10 年早稲田大学理工学部電気電子情報工学科卒業．平成 12 年同大学大学院修士課程修了（電気工学）．在学中は階層ベイズ的アプローチによる時系列予測等の研究に従事．

栗原　貴之

平成 13 年早稲田大学理工学部電気電子情報工学科卒業．現在，同大学大学院理工学研究科電気工学専攻修士課程 1 年在学．階層ベイズ的アプローチによる時系列予測や on-line 学習の研究に従事．

松本　隆（正会員）

昭和 41 年早稲田大学理工学部電気工学科卒業．昭和 45 年ハーバード大学大学院・応用数学修士．昭和 47 年工学博士（早稲田大学）．昭和 52 年～54 年カリフォルニア大学バークレー校電気工学–計算機科学研究員．現在，早稲田大学理工学部電気・情報生命工学科教授．ベイズ学習と予測，MCMC，バイオメトリック個人認証，信号処理用集積回路設計等に興味を持つ．Circuit，Systems and Signal Processing 編集委員．テレビジョン学会次世代画像入力研究会委員（平成 7 年より）．電子情報通信学会非線形問題調査専門委員会委員長（平成 2，3 年度）．平成 5 年度日本神経回路学会論文賞．著書「Bifurcations」（Springer-Verlag）等．IEEE（Fellow），電気学会，テレビジョン学会，日本神経回路学会，計測自動制御学会等各会員．