

フーリエ変換とベイズモデルおよびトレンドモデルによる時系列データのノイズ除去

小野寺 光永^a, 井須 芳美^a, 長嶋 雲兵^{a,c*}, 吉田 裕亮^a,
細矢 治夫^a, 永川 祐三^{b,d}

^a お茶の水女子大学理学部情報科学科 〒 112-8610 文京区大塚 2-1-1

^b お茶の水女子大学保健管理センター 〒 112-8610 文京区大塚 2-1-1

^c 現所属: 物質工学工業技術研究所 〒 305-8565 つくば市東 1-1

*e-mail: umpei@nair.go.jp

^d 現所属: (株)三菱電機

(Received: December 4, 1997 ; Accepted for publication: March 3, 1998 ; Published on Web: June 21, 1999)

生体の断層イメージを測定する核磁気共鳴 (NMR) のシグナルや、Belousov-Zhabotinsky 反応 (BZ) などの振動反応の酸化還元ポテンシャルには、測定環境により SN 比が急激に悪くなる時系列データが見られ、そのノイズ除去が問題となっている。また、生体における時系列データの一つである心電図においても、ノイズによる心電図解析の妨げが問題となっている。そこで本研究では、フーリエ変換、ベイズモデル、トレンドモデルの3手法を用いたノイズ除去を行い、それぞれのモデルによる結果の比較、検討を行った。数値実験には、NMR のシグナルに代表的な Multiple nonstationary frequencies の近似関数と心電図の2種類のデータを用いた。

その結果、Multiple nonstationary frequencies の近似関数では、ベイズモデルが最も有効であった。心電図では、フーリエ変換により、基線の揺れを抑えることができ、ベイズモデルやトレンドモデルを用いて平滑化を行った後に、フーリエ変換によってノイズを除去する方法が有効であることが示唆された。

キーワード: Noise filtering, FFT, Bayesian model, Trend model, Time series

1 はじめに

時間の経過とともに不規則に変動する現象の時系列データは、自然現象をはじめ、医学、経済等様々な分野で見ることができる。その解析により我々は多くの情報を得ることができるが、一般に時系列データの解析では、観測された時系列データからノイズを除去して真のデータを引き出すことが重要な問題となる。特に SN 比が悪くなるような大きなノイズがのる場合、それが解析の著しい障害となる。例えば生体の断層イメージを測定する核磁気共鳴 (NMR) のシグナルや、Belousov-Zhabotinsky 反応 (BZ) などの振動反応の酸化還元ポテンシャルには、測

定環境により SN 比が急激に悪くなる時系列データが見られ、そのノイズ除去が問題となっている。また生体における時系列データの一つである心電図においても、携帯心電計による測定では SN 比が急激に悪くなるようなノイズがのるケースがあり、それが狭心症などを診断する心電図解析の一つである RR 間隔測定の際に大きな障害となる。

ノイズ除去の方法として、これまでにフーリエ変換 [1]、ベイズモデル [2, 3]、トレンドモデル [7] 等が提案されてきた。フーリエ変換による方法は、時系列データを正弦波の和として表すことにより周波数分解し、その高周波成分や低周波成分の一部を除くことによりノイズの除去を行う。ベイズモデルでは、観測値に対する確率分布を導くことにより真のモデルの推定を行う。トレンドモデルは、パラメータの確率的变化を表現したモデルであり、時系列の長期的変動であるトレンドを推定する。これらの手法のうちベイズモデルやトレンドモデルは、精度の良い結果を得るために統計的知識を伴った多くの試行錯誤を必要とする。一方フーリエ変換による方法は、統計的な予備知識を必要としない反面、特定の周波数成分を除くことによるトレンドへの影響を見る必要がある。

そこで本研究では、フーリエ変換、ベイズモデル、トレンドモデルの 3 手法を用いて、SN 比が急激に悪くなるデータを含めた 2 種類のサンプルデータのノイズ除去を行い、それぞれのモデルによる結果の比較、検討を行った。

2 サンプルデータ

サンプルデータとして以下の 2 種類のデータを用いた。それぞれを簡単に説明する。

2.1 Multiple nonstationary frequencies

NMR や BZ 反応のシグナルは、Multiple nonstationary frequencies で表されることが知られている。そこで本研究では、Multiple nonstationary frequencies の近似データを式 1 により作成し、この時系列データに SN 比がそれぞれ 20, 10, 5 となるようノイズを加えた 3 種類の時系列データをサンプルデータとして用いた。その中の SN 比が 5 のデータを Figure 1 に示す。ここで、SN 比 sn はデータ $f(t)$ の標準偏差 \sqrt{vs} をノイズの標準偏差 pn で割った値で、式 2 のように表すことができる。

$$data[t] = 100 \cos(0.3(t + 11) + 1) \exp(-0.005(t + 11)) + 25 \cos(0.31(t + 11) + 3) \exp(-0.003(t + 11)) \quad (t = 0 \sim 1023) \quad (1)$$

$$sn = \frac{\sqrt{vs}}{pn} \quad (2)$$

$$vs = \frac{1}{n} \sum_{t=0}^{n-1} |f(t) - m|^2 \quad (3)$$

$$m = \frac{1}{n} \sum_{t=0}^{n-1} f(t) \quad (4)$$

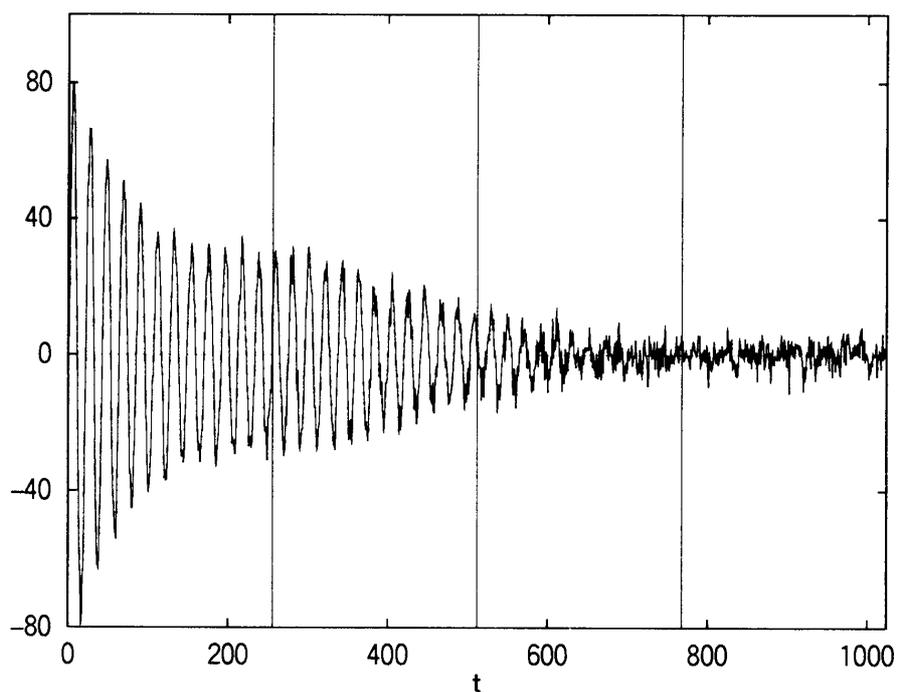


Figure 1. Multiple nonstationary frequencies(SN 比=5)

式 1 により作成されたデータは、Figure 1 からわかる通り、時間における振幅の変化が大きく、データ中の時間変化による SN 比の差が大きい。そこで Figure 1 の縦線で示したように、データを $[0,255]$, $[256,511]$, $[512,767]$, $[768,1023]$ の 4 つの時間区間に分け、データ中の SN 比の大きさの変化による結果の違いを見ることにした。SN 比が 20, 10, 5 であるそれぞれのノイズデータの時間区間ごとの SN 比を Table 1 に示す。 $[0,255]$ は区間 (0)、 $[256,511]$ は区間 (1)、 $[512,767]$ は区間 (2)、 $[768,1023]$ は区間 (3) とした。SN 比が 5 のノイズデータは、区間 [3] で SN 比がかなり悪くなっており、シグナルよりノイズの方が大きくなっている。

Table 1. 各区間の SN 比

全体の SN 比	区間の SN 比			
	(0)	(1)	(2)	(3)
20	60.15	30.37	6.57	2.69
10	30.07	15.18	3.29	1.35
5	15.04	7.59	1.64	0.67

2.2 心電図

実際の時系列データへの応用として、携帯心電計により測定された心電図のデータを使うことにした (Figure 2)。一般の心電図検査では仰臥位安静時に測定を行うが、携帯心電計では、日常生活の運動負荷の状態変化をみることができ、こうして心電図の監視を継続的に行うことで、高血圧障害などの患者の在宅医療が進むものと期待されている。しかし、携帯心電計により測定されたデータには、携帯心電計から記録装置への電波の伝送ミスのため、Figure 2 中の $t=100, 360$ 付近で見られるように SN 比が急激に悪くなる大きなノイズが含まれる。このようなノイズは、RR 間隔 (心電図内の各 QRS 波の間隔 (Figure 3)) 測定の障害となる。この間隔は運動により変化するが、心臓病との関係も大きく、期外収縮ではその短縮、房室ブロックではその延長が見られる。そのため RR 間隔時系列は、医療における重要な診断材料となっている [4]。従って、その正確な測定はたいへん重要であり [5]、そのためのノイズ除去が必要となる。

またこのような測定時のノイズは、BZ 反応などにおいても見ることができるので、心電図以外のデータのノイズ除去への応用分野も広い。

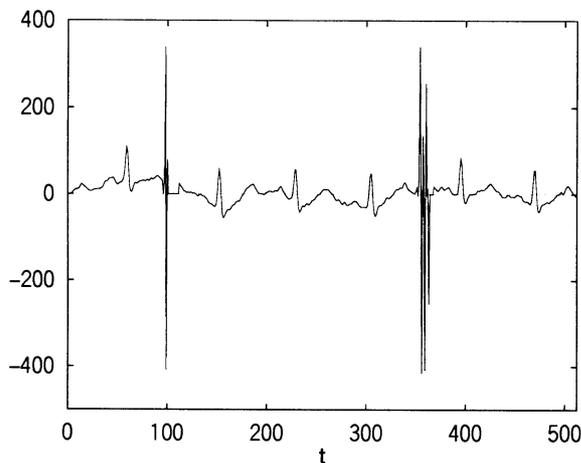


Figure 2. 携帯心電計により測定された心電図

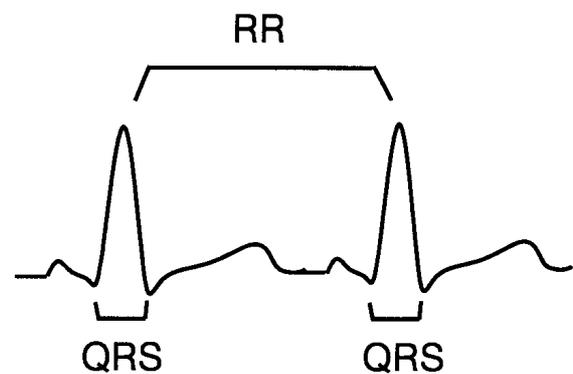


Figure 3. 心電図の RR 間隔

3 ノイズ除去の方法

3.1 フーリエ変換によるノイズ除去

フーリエ変換では、時間関数 ($g(t), t=0, \dots, N-1$) から周波数関数 ($G(f)$) を求める (式 5)。逆フーリエ変換では、反対に周波数関数から時間関数を求める (式 6)。本研究では、フーリエ変換を行い周波数関数を求め、その高周波成分を除くことによりノイズを除去した後、その結果に逆フーリエ変換を行い、時間関数に戻すことを試みた。

$$G(t/N) = \frac{1}{N} \sum_{k=0}^{N-1} g(k) \exp(-i2\pi tk/N) \quad (5)$$

$$g(t) = \sum_{k=0}^{N-1} G(k/N) \exp(i2\pi tk/N) \quad (6)$$

本方法は、ロー/ハイパスフィルタリングとして知られている。

3.2 ベイズモデルによるノイズ除去

ベイズモデルによるノイズ除去には田辺らの方法 [6] を用いた。以下にその方法を説明する。

推定すべき関数 $f(x)$ を等間隔な離散点 x_j 上の値として $f_j = f(x_j)$ と表現し、 f を式 7 のように表す (t は転置)。以下では関数とベクトル f を同一視する。また今、 N 個のデータ $y_i (1 \leq i \leq N)$ が得られているとし、 y_i を式 8 のように表す。 n は N より十分大きくとる。離散点 x_j 上に常に y_i がある必要はない。

$$f = (f_1, f_2, \dots, f_j, \dots, f_n)^t \quad (7)$$

$$y = (y_1, y_2, \dots, y_i, \dots, y_N)^t \quad (8)$$

データ y_i の誤差 ε_i を $\varepsilon_i \sim i.i.d.N(0, \sigma^2)$ とすると、 f_j 上のデータである y_i の分布は式 9 で表すことができ、この時 y の分布は式 10 で表すことができる。ここで *i.i.d.* は、independently and identically distributed (定常独立) の意味であり、 $\|\cdot\|$ は、ユークリッドノルムを表す。また E は y と f の対応を表す ($N \times n$) 行列で、fit-point x_j 上にデータ y_i があれば、その i 行 j 列は 1、その他は 0 となるような行列である。

$$p_i(y_i | f_j, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{1/2} \exp \left(-\frac{1}{2\sigma^2} (y_i - f_j)^2 \right) \quad (9)$$

$$p(y | f, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{N/2} \exp \left(-\frac{1}{2\sigma^2} \|y - Ef\|^2 \right) \quad (10)$$

次に、 f の滑らかさを二階差分の大きさで表現する。各点での二階差分の大きさを式 11 のような量で測り、 f の滑らかさの程度を $\|dDf\|^2$ で測る。 D は式 12 のようにとる。

$$|f_{i-1} - 2f_i + f_{i+1}|, \quad 2 \leq i \leq n_{i-1} \quad (11)$$

$$D = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & & \ddots & & \\ & & & & 1 & -2 & 1 \end{pmatrix} \quad (12)$$

ここで、超パラメータ d を持つ improper な事前分布式 13 を導入し、 f の滑らかさをこの事前分布に含まれる超パラメータ d で制御する。 $l = n - 2$ 、 ψ は $D^t D$ の非零固有値である。

$$\pi(f | d) = \left(\frac{1}{2\pi} \right)^{l/2} d^l \psi \exp \left(-\frac{1}{2} \|dDf\|^2 \right) \quad (13)$$

従って、データ y_i が与えられた時の周辺尤度 $L(\sigma^2, d)$ は式 14 のように定義される。

$$L(\sigma^2, d) = \int p(y|f, \sigma^2) \pi(f|d) df \quad (14)$$

赤池はベイズ型情報規準量 $\text{ABIC} = -2 \log L(\sigma^2, d)$ を最小にする σ^2, d を f の推定に用いる方法を提案している。今、 ABIC を最小にする σ^2, d を $\hat{\sigma}^2, \hat{d}$ (最尤推定値) とすると、ベイズの定理により f の事後分布は式 15 で表されるので、この式を最大にする f を最適近似関数 \hat{f} として選ぶことにする。

$$p(f|y) = \frac{p(y|f, \hat{\sigma}^2) \pi(f|\hat{d})}{\int p(y|f, \hat{\sigma}^2) \pi(f|\hat{d}) df} \quad (15)$$

ここで $p(f|y)$ は、 y という事象が起こった後に f が起こる確率を意味する。 $L(\sigma^2, d)$ は、 $\alpha = d\sigma$ と超パラメータの変数変換を行うと式 16 のように表すことができる。ここで、 $\|b - Z_\alpha f_*\|^2$ は最小二乗問題 $\min_f \|b - Z_\alpha f\|^2$ の最小二乗誤差で、 f_* はその時の最小二乗解である。つまり、 \hat{f} は $\alpha = \hat{\alpha}, \sigma^2 = \hat{\sigma}^2$ のときの f_* に他ならず、問題は式 16 を最大にする σ^2, α を求めることになる。

$$L(\sigma^2, \alpha) = \left(\frac{1}{2\pi}\right)^{\frac{N+l-n}{2}} \left(\frac{1}{\sigma}\right)^{N+l-n} \alpha^l \psi |\det(Z_\alpha^t Z_\alpha)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \|b - Z_\alpha f_*\|^2\right), \quad (16)$$

$$Z_\alpha = \begin{bmatrix} E \\ \alpha D \end{bmatrix}, \quad b = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

$\frac{\delta}{\delta \sigma^2} L(\sigma^2, \alpha)|_{\sigma^2 = \hat{\sigma}^2 = 0}$ より式 17 を得ることから、式 16 は式 18 となる。従って、 α についてその尤度方程式を解けば良いが、これは解析的に解くことが困難である。そこで α を数値的に求める。いろいろな α に対して式 19 を計算し、その ABIC が最小になるような α を求め、その α に対応する f_* を推定関数にする。

$$\hat{\sigma}^2 = \frac{1}{N+l-n} \|b - Z_\alpha f_*\|^2 \quad (17)$$

$$L(\alpha) = \left(\frac{1}{2\pi}\right)^{\frac{N+l-n}{2}} \psi \exp\left(-\frac{N+l-n}{2}\right) \alpha^l |\det(Z_\alpha^t Z_\alpha)|^{-\frac{1}{2}} \left(\frac{1}{N+l-n} \|b - Z_\alpha f_*\|^2\right)^{-\frac{N+l-n}{2}} \quad (18)$$

$$\text{ABIC}(\alpha) = (N+l-n) \log(\|b - Z_\alpha f_*\|^2) - 2(n-2) \log \alpha + \log |\det(Z_\alpha^t Z_\alpha)| + C \quad (19)$$

3.3 トレンドモデルによるノイズ除去

トレンドは時系列のおおよその傾向を表す。逆にいえば、実際に観測される時系列はトレンドに様々な変動成分が加わったものである。その中で最も単純な場合は、 w_n を平均 0、分散 σ^2 の正規白色雑音として、式 20 のように表現できる。

$$y_n = t_n + w_n \quad (20)$$

トレンド t_n は説明変数 x_n の多項式で、式 21 のように表されるものとする。

$$t_n = a_0 + a_1 x_n + \dots + a_k x_n^k \quad (21)$$

ここで、式 21 の多項式をより柔軟な関数に拡張する。一般に、 $k-1$ 次多項式は k 階の差分方程式 (式 22) の解とみなすことができるが、差分方程式の代わりに ${}^k t_n \approx 0$ が成り立つように k 階の確率差分方程式 (式 23) を解くことにする。ただし、 $1-B$ は、 $t_n = t_n - t_{n-1}$ によって定義される時間差分オペレータとし、 v_n は、平均 0、分散 τ^2 の正規白色雑音とする。式 23 をトレンド成分モデルと呼ぶ。

$${}^k t_n = 0 \quad (22)$$

$${}^k t_n = v_n \quad (23)$$

従って、時系列 y_n からトレンド t_n を推定するためには、式 23 と式 20 を組み合わせた次のようなモデルを考えれば良い。これをトレンドモデル (trend model) と呼ぶ。

$${}^k t_n = v_n \quad (24)$$

$$y_n = t_n + w_n \quad (25)$$

式 25 は、時系列 y_n をトレンドに独立なノイズが加わったものを観測した結果とみなした観測モデルである。一方式 24 は、そのトレンドの変化の仕方をモデル化したトレンド成分モデルである。

一般に、トレンド成分モデル式 24 は、式 26 となるので、 $c_i = (-1)^{i+1} {}_k C_i$ によって係数 c_i を定義すると、式 27 のように表現できる。

$$\Delta^k = (1-B)^k = \sum_{i=0}^k {}_k C_i (-B)^i \quad (26)$$

$$t_n = \sum_{i=1}^k c_i t_{n-i} + v_n \quad (27)$$

このモデルは定常性を満たさないが形式的には k 次の AR モデルとみなせるので、

$$x_n = \begin{pmatrix} t_n \\ t_{n-1} \\ \vdots \\ t_{n-k+1} \end{pmatrix}, \quad F = \begin{pmatrix} c_1 & c_2 & \dots & c_k \\ 1 & & & \\ & \ddots & & \\ & & 1 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$H = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \quad (28)$$

とおくことによりトレンド成分モデルの状態空間表現

$$\begin{aligned}x_n &= Fx_{n-1} + Gv_n \\t_n &= Hx_n\end{aligned}\tag{29}$$

が得られる。これを利用することにより、トレンドモデルの状態空間表現 (式 30) が得られる。

$$\begin{aligned}x_n &= Fx_{n-1} + Gv_n \\y_n &= Hx_n + w_n\end{aligned}\tag{30}$$

従って、トレンドモデルの次数 k 及び分散 τ^2 , σ^2 が定まると、カルマンフィルタ及び平滑化のアルゴリズムにより状態ベクトルの平滑値 $x_{1|N}, \dots, x_{N|N}$ を求めることができる [7]。また、 σ^2 はカルマンフィルタを用いることによりその最尤推定値が得られることから、本研究では、次数 k と分散 τ^2 のみをパラメータとして選択した。

4 結果

4.1 Multiple nonstationary frequencies

2.1 節で説明した Multiple nonstationary frequencies の近似関数のノイズ除去をフーリエ変換、ベイズモデル、トレンドモデルそれぞれにより行った。各手法のパラメータの数値は、それぞれの SN 比のノイズデータで区間 (0) の相対誤差が最小になるように選択した。つまりフーリエ変換では、サンプル数として 2048 点を用い、式 5, 6 の $k = 0 \sim 410$ の $G(k)$ を 0 と置き、高周波数成分を除いた。ベイズモデルでは、3.2 節で説明した滑らかさを表すパラメータ α をデータの SN 比が 20 のとき $\alpha=7$ 、10 のとき $\alpha=12$ 、5 のとき $\alpha=16$ とした。トレンドモデルでは、3.3 節で説明したモデルの次数を 1、トレンドのノイズの分散 τ^2 をデータの SN 比が 20 のとき $\tau^2=1.9$ 、10 のとき $\tau^2=1.6$ 、5 のとき $\tau^2=0.9$ とした。

結果を Table 2 に示す。Table 2 は左列から順に、全体の SN 比、(0) ~ (3) の各区間とその SN 比、式 1 に示したノイズを加える前のデータとノイズを加えた後のデータとの各区間における相対誤差の平均値 ($|e-s|/e$)、ノイズを加える前のデータとフーリエ変換を用いてノイズを除去したデータとの各区間における相対誤差の平均値 ($|e-f|/e$)、同様にベイズモデルによる結果の相対誤差の平均値 ($|e-b|/e$) とトレンドモデルによる結果の相対誤差の平均値 ($|e-t|/e$) をそれぞれ表す。また、Figures 4-6 に SN 比が 5 のデータをフーリエ変換、ベイズモデル、トレンドモデルそれぞれによりノイズ除去した図を示す。Figure 7 は、区間 (2) における拡大図である。

Table 2 の右から 4 列目に示したノイズを加える前のデータと加えた後のデータとの相対誤差 ($|e-s|/e$) と、各手法による相対誤差を比較することで、各手法によるノイズ除去の程度を見ることができる。 $|e-s|/e$ と比較して相対誤差が小さいほど、ノイズが除去されたことを表し、逆に大きくなっていればノイズの除去に失敗したことを表す。

Table 2. フーリエ変換 (f)、ベイズモデル (b)、トレンドモデル (t) の各区間における実データとの相対誤差

全体の SN 比	区間	区間の SN 比	$ e-s /e$	$ e-f /e$	$ e-b /e$	$ e-t /e$
20	(0)	60.15	0.10	0.04	0.04	0.08
	(1)	30.37	0.12	0.04	0.05	0.08
	(2)	6.57	4.17	3.17	3.05	3.34
	(3)	2.69	1.51	1.73	1.00	1.18
10	(0)	30.07	0.19	0.07	0.08	0.13
	(1)	15.18	0.25	0.08	0.08	0.13
	(2)	3.29	8.34	6.17	5.55	6.55
	(3)	1.35	3.01	2.57	1.84	2.30
5	(0)	15.04	0.38	0.13	0.14	0.22
	(1)	7.59	0.49	0.15	0.15	0.22
	(2)	1.64	16.69	12.17	10.40	12.01
	(3)	0.67	6.03	4.29	3.54	4.22

Table 3. フーリエ変換 (f)、ベイズモデル (b)、トレンドモデル (t) の全区間における実データとの相対誤差

全体の SN 比	$ e-f /e$	$ e-b /e$	$ e-t /e$
20	1.24	1.04	1.17
10	2.22	1.89	2.28
5	4.19	3.56	4.17

まずフーリエ変換の相対誤差は、区間 (0), (1) では SN 比=20, 10, 5 全てのデータで $|e-s|/e$ の 1/2 以下になっている。区間 (2), (3) では区間 (0), (1) ほどではないが、SN 比=20 の区間 (3) を除いて相対誤差が減少している。

ベイズモデルの相対誤差は、区間 (0), (1) では、フーリエ変換の結果と同様に全てのデータで $|e-s|/e$ の 1/2 以下になっている。また区間 (2), (3) では、区間 (0), (1) ほどではないが、全てのデータで相対誤差が減少している。

次にトレンドモデルの相対誤差を見ると、全ての区間で相対誤差が減少しているが、区間 (0), (1) の相対誤差がフーリエ変換やベイズモデルと比較すると大きい。

各手法の相対誤差を比較すると、区間 (0), (1) ではフーリエ変換とベイズモデルがほぼ同じ誤差であり、トレンドモデルと比較して誤差が小さい。区間 (2), (3) ではベイズモデルの誤差が最も小さい。また、Table 3 の全区間における誤差を比較してもベイズモデルの相対誤差が最も小さい。

Figures 4–6 の時刻 512 以降 (区間 (2),(3)) をみると、Figure 6 のトレンドモデルによる結果は Figures 4, 5 に示した他の手法による結果と比較して滑らかさに欠けている。これは区間 (2) の拡大図である Figure 7 をみると一層よくわかる。 τ^2 を小さい値にすると滑らかな結果が得られ

るが、本研究では、 τ^2 を区間 (0) の相対誤差により選んだのでこのような結果が得られた。一方、同様にパラメータ (σ^2) を選択したベイズモデルは、Figure 7 に示すように、3 手法の中で最も滑らかな結果が得られていることがわかる。

これらの結果から、このサンプルデータによる実験では、ベイズモデルが最も SN 比によらず精度の良いノイズ除去を行うことができるという結論を得た。

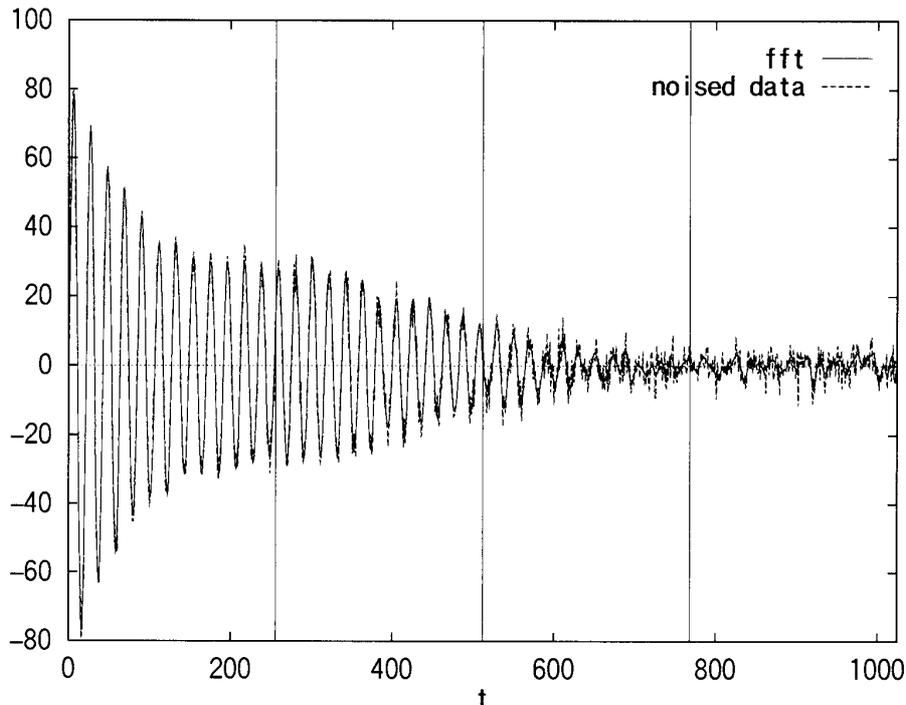


Figure 4. ノイズデータとフーリエ変換 (fft) によるノイズ除去 (SN 比=5)

4.2 心電図

Figure 2 に示した心電図を用いて数値実験を行った。この心電図は、健康な人の階段昇降時の心拍状態を測定したものであり、データ数は 512 である。

まず $t=100, 360$ 付近の電波の伝送ミスによるノイズを除くため、フーリエ変換によるノイズ除去を行った。その結果を Figure 8 に示す。ここでは、サンプル数 512 で、式 5, 6 の $k=0 \sim 170$ の $G(k)$ を 0 と置き、高周波数成分を除いた。QRS 波のピークが Figure 2 と比較してほとんど変化がないのに対し、 $t=100, 360$ 付近で見られた大きな振れは抑えられていることがわかる。

また、Figure 2 では運動負荷による基線の揺れが見られるが、このことも RR 間隔時系列を取り出す際に障害となる。そこで、基線を一定にするため、高周波成分に加え、低周波成分を除くことを試みた。その結果を Figure 9 に示す。ここでは、式 5, 6 の $k=508 \sim 512$ の $G(k)$ を 0 と置き、低周波数成分を除いた。Figure 9 を見ると、Figures 2, 8 と比較して基線の揺れが抑えられ、一定となっていることがわかる。

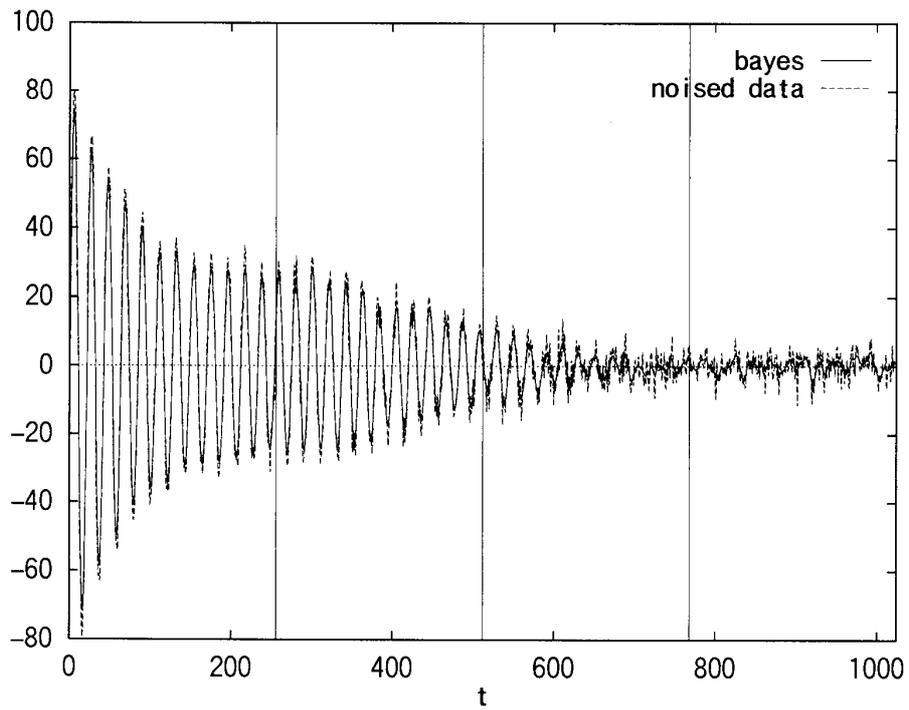


Figure 5. ノイズデータとベイズモデル (bayes) によるノイズ除去 (SN 比=5)

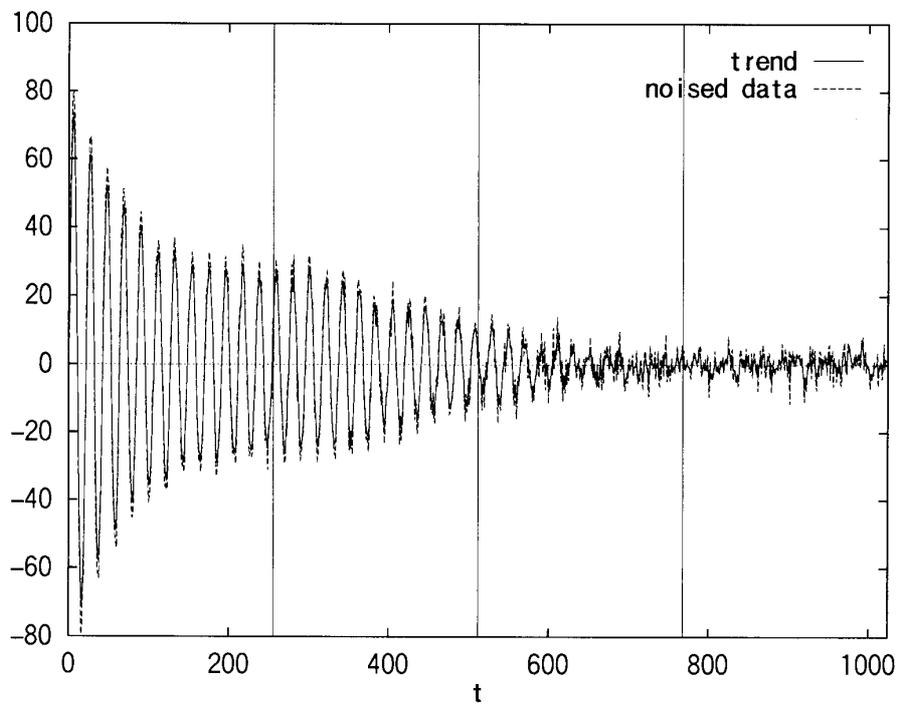


Figure 6. ノイズデータとトレンドモデル (trend) によるノイズ除去 (SN 比=5)

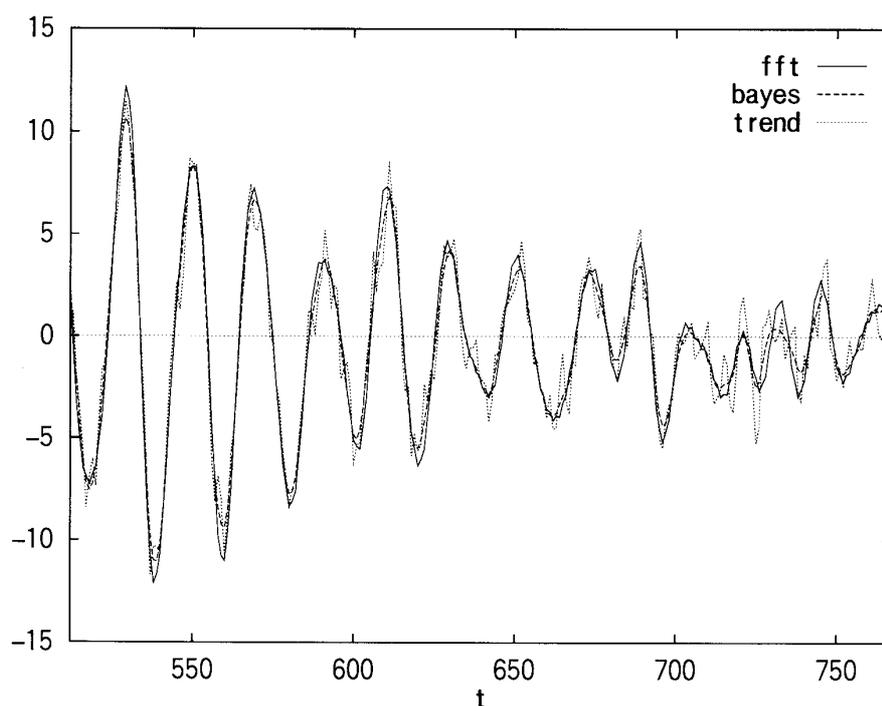


Figure 7. フーリエ変換 (fft)、ベイズモデル (bayes)、トレンドモデル (trend) によるノイズ除去 (SN 比=5)

次に、Figure 2 をベイズモデルにより平滑化した結果を Figures 10, 11 に示す。Figure 10 は $\alpha=1.0$ 、Figure 11 は $\alpha=5.0$ である。 α の値が大きいほど滑らかな曲線が得られるため、Figure 10 では $t=100, 360$ 付近の大きな振れが残っているが、Figure 11 ではほぼ完全に除かれている。また Figure 10 と Figure 11 の QRS 波を比較すると、Figure 11 の方が多少波が低くなっているが、QRS 波を捉えるには十分である。

次に、Figure 2 をトレンドモデルにより平滑化した結果を Figures 12, 13 に示す。Figure 12 は $\tau^2=1.4$ 、Figure 13 は $\tau^2=0.5$ である。 τ^2 の値が小さいほど滑らかな曲線が得られるため、Figure 12 では $t=100, 360$ 付近に振れが残っているが、Figure 13 ではほぼ完全に除かれている。

Figures 9, 11, 13 を比較すると、Figure 9 は全体に細かい波がのっており、そのため滑らかさに欠けるが、このような細かい波は、RR 間隔測定の障害にはならない。また Figure 9 で示した通り、フーリエ変換による方法では、基線の揺れも抑えられるため、Figure 9 にのっている細かい波が気になるならば、ベイズモデルやトレンドモデルにより平滑化した後、フーリエ変換を用いて基線を一定にする方法が有効であると考えられる。

5 まとめと今後の課題

本研究では、2種類のサンプルデータのノイズ除去をフーリエ変換、ベイズモデル、トレンドモデルの3手法により行い、各手法による結果の違いを検討した。

まず NMR や BZ 反応のシグナルに見られる Multiple nonstationary frequencies の近似関数に SN 比が 20, 10, 5 となるようノイズを加え、そのノイズ除去を行った。このデータは時間に

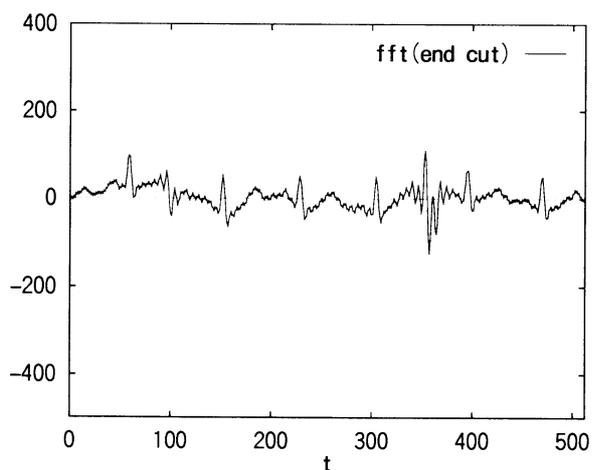


Figure 8. フーリエ変換によるノイズ除去 (高周波成分を除去)

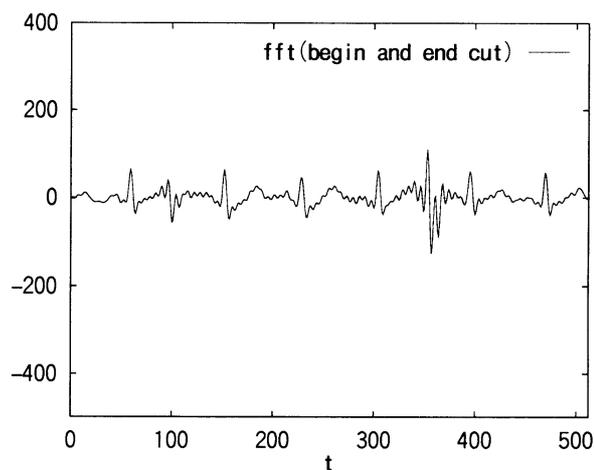


Figure 9. フーリエ変換によるノイズ除去 (高、低周波成分を除去)

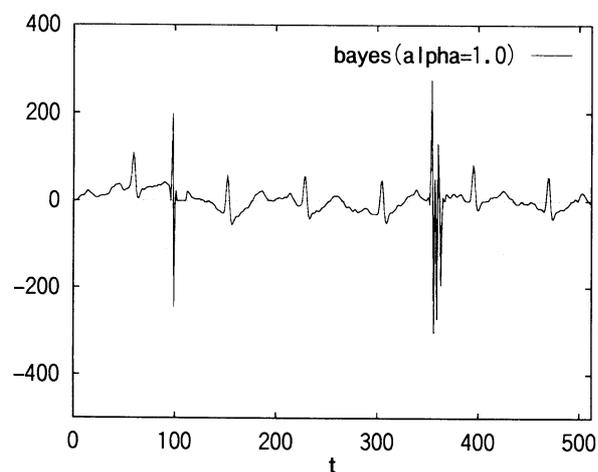


Figure 10. ベイズモデルによるノイズ除去 ($\alpha=1.0$)

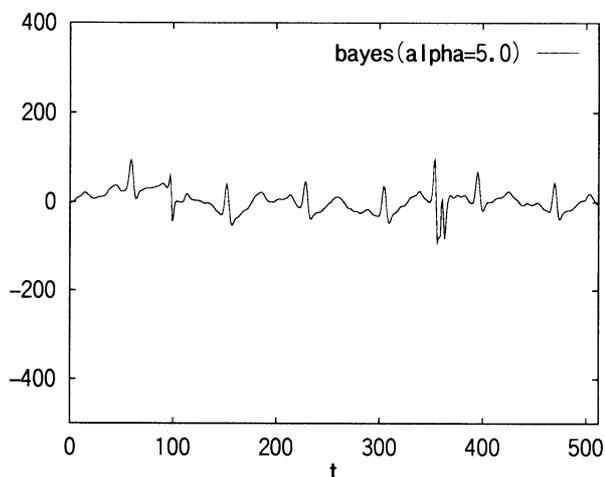


Figure 11. ベイズモデルによるノイズ除去 ($\alpha=5.0$)

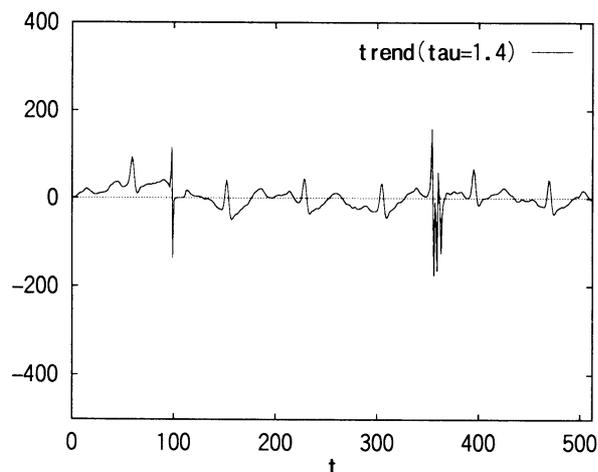


Figure 12. トレンドモデルによるノイズ除去 ($\tau^2=1.4$)

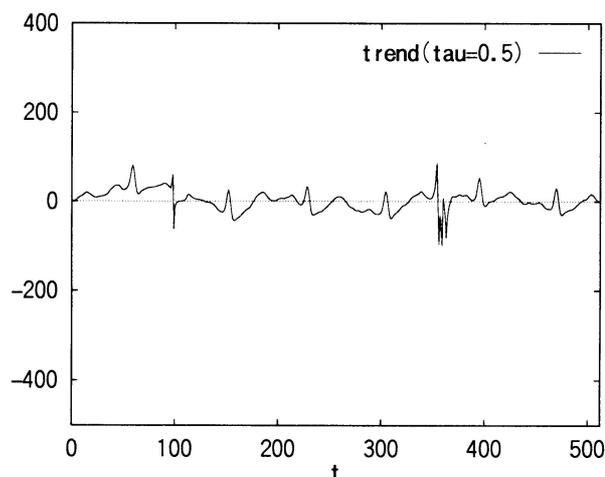


Figure 13. トレンドモデルによるノイズ除去 ($\tau^2=0.5$)

よる SN 比の変化が大きいため、データを 4 つの時間区間に分け、それぞれの区間の誤差を見た。その結果、SN 比が大きいはじめの 2 つの区間では、フーリエ変換とベイズモデルにより、ノイズを除去した後の相対誤差をノイズを除去する以前の相対誤差の 1/2 以下にすることができた。トレンドモデルは、フーリエ変換及びベイズモデルと比較して誤差が大きかった。SN 比が小さい残りの 2 区間では、ベイズモデルが最も誤差が小さかった。また、全区間の誤差を比較してもベイズモデルによる結果が最も誤差が小さかった。これらの結果から、このサンプルデータによる実験では、ベイズモデルが最も SN 比によらず精度の良いノイズ除去を行うことができるという結論を得た。

次に、携帯心電計により測定された心電図を用い、心電計と記録装置間の電波の伝送ミスによるノイズ除去を行った。その結果、フーリエ変換、ベイズモデル、トレンドモデル、どの手法を用いても、電波の伝送ミスによる、SN 比が大きいノイズを除くことができた。さらにフーリエ変換では、運動負荷による基線の揺れも抑えることができた。しかし、フーリエ変換による方法では、全体に細かい波がのり滑らかさに欠けるため、ベイズモデルやトレンドモデルにより平滑化をした後、フーリエ変換を用いて基線を一定にする方法が有効であると考えられる。これらの手法を用いたノイズ除去により、医療における重要な診断材料である心電図の RR 間隔の測定が容易になることが期待される。

ノイズ除去の手法ごとの長所、短所が、Multiple nonstationary frequencies の近似関数や、本来の心電図などの「関数群」によるものなのか、それともその上にかぶせた、ランダムな雑音や、電波伝送ミスにより起こる大きなノイズ等の「ノイズの種類」によるものなのか、という疑問が残るが、この究明を今後の課題として挙げたい。また、心電図以外の時系列データにおけるノイズ除去を行い、その効果を確かめていきたい。

参考文献

- [1] 安居院猛, 中嶋正之, *FFT の使い方*, 秋葉書店, 東京 (1994).
- [2] 鈴木雪夫, 国友直人編, *ベイズ統計学とその応用*, 東京大学出版会, 東京 (1992).
- [3] 繁榊算男, *ベイズ統計入門*, 東京大学出版会, 東京 (1994).
- [4] 五島雄一郎, 大林完二監修, *心電図の ABC*, 日本医師会, 東京 (1997).
- [5] 早野順一郎, 心拍変動による自律神経機能解析, 循環器疾患と自律神経機能, 医学書院, 東京 (1996), p58.
- [6] 田辺國士, 田中輝雄, ベイズモデルによる曲線・曲面のあてはめ, 月刊地球, 5, 179-186 (1983).
- [7] 北川源四郎, *時系列解析プログラミング*, 岩波書店, 東京 (1993).

Noise Filtering Using FFT, Bayesian Model and Trend Model for Time Series Data

Mitsue ONODERA^a, Yoshimi ISU^a, Umpei NAGASHIMA^{a,c*},
Hiroaki YOSHIDA^a, Haruo HOSOYA^a and Yuuzou NAGAKAWA^{b,d}

^aFaculty of Science, Ochanomizu University
Bunkyo-ku, Tokyo, 112-8610 JAPAN

^bHealth Care Center, Ochanomizu University
Bunkyo-ku, Tokyo, 112-8610 JAPAN

^cPresent address: National Institute of Materials and Chemical Research, MITI
1-1 Azuma, Tsukuba 305-8565, JAPAN

**e-mail: umpei@nair.go.jp*

^dPresent address: Mitsubishi Electric Co. Ltd

The applicability of noise filtering methods, FFT, Bayesian model and trend model was evaluated using multiple nonstationary frequencies of noise and electrocardiogram with large signal-to-noise ratio. The results show that the Bayesian model is the most applicable to filtering of multiple nonstationary frequencies with large noise. In the case the best way for filtering out the noise in electrocardiogram suggested that the background rolling noise should be eliminated by FFT after removing the high frequency component by Bayesian or trend model.

Keywords: Noise filtering, FFT, Bayesian model, Trend model, Time series