

## 周期時系列のノンパラメトリックベイズ多重整列 Non-parametric Bayesian Alignment of Periodic Time Series

森 源太<sup>†</sup>  
Genta Mori

末松 伸朗<sup>†</sup>  
Nobuo Suematsu

岩田 一貴<sup>†</sup>  
Kazunori Iwata

林 朗<sup>†</sup>  
Akira Hayashi

### 1 はじめに

時系列とは、時間変化に伴い変動する観測値の系列で、人の成長データ、気象観測データ、GDP、株価の変動データや、音楽データ、DNA配列など様々なものが時系列データとして扱われている。これらの時系列データから、目に見えない特徴や、その内部構造を見つけ出すように分析を行うことでそのデータを扱うことができるようになる。また、分析を行うために時系列データを整列し、時間変化を除去することが有効であり、その整列手法は統計学、音声認識、信号処理など広い分野で研究されている。

時系列整列を行う手法で、幅広く導入されているアルゴリズムの1つに、Dynamic Time Warping(DTW)[1]がある。DTWは2つの時系列を比較し、整列させるためのアルゴリズムで、初期の音声認識で利用されていたが、現在は他の様々な分野でも利用されている。DTWは2つの時系列の要素間の距離が最小になるような経路を見つけることで時系列を整列する。これは経路の数を制限できないくつかの合理的な制約と、動的計画法を用いることで整列を実現される。

関連研究の1つに、Ramsayの曲線整列法[2]がある。誤差と滑らかさを調整する目的関数が最小となるように、ある単調増加関数のクラスに限定して時間変換関数を推定する。この手法は局所最適化を行う手法であり、スプライン関数を用いるセミパラメトリックな手法である。

本研究では、ガウス過程事前分布を利用した時系列整列法[3]をもとに、特に周期的なデータを持つ時系列についての整列手法を提案している。複数の周期時系列データについて、標準時系列に時間伸縮を加えることで観測値が得られると仮定する生成モデルを考え、そのモデルで仮定される標準時系列と時間変換関数をベイズ解析することで関数を推定し、周期時系列データの多重整列を実現する。この手法では、推定される関数の事後分布は、解析的に直接求めることができないため、マルコフ連鎖モンテカルロ法(MCMC)を用いてサンプルを生成し、それらの事後平均を得ることで推定している。

提案手法では、連続的な複数の周期時系列のデータについての整列を高い精度で得ることが期待される。提案手法の有効性は、人工データを用いた実験を行い検証している。

### 2 ガウス過程回帰

確率的に定まる実数上の関数(もしくは、確率過程) $g(x)$ について、任意の自然数 $p$ と任意の $x_1, \dots, x_n \in \mathbf{R}$ に対して、 $\mathbf{g} = (g(x_1), \dots, g(x_n))^T$ が正規分布に従

うとき、 $g(x)$ はガウス過程に従うといわれる。ガウス過程は、

$$m(x) = \mathbb{E}[g(x)] \quad (1)$$

$$k(x, x') = \mathbb{E}[\{g(x) - m(x)\}\{g(x') - m(x')\}] \quad (2)$$

で定義される平均関数 $m(x)$ と共分散関数 $k(x, x')$ により完全に指定される。ここで $\mathbb{E}[\cdot]$ は期待値を表す。そしてこのとき、

$$g(x) \sim \mathcal{GP}(m(x), k(x, x')) \quad (3)$$

と表記される。

事前分布がガウス過程である未知の関数の観測値が得られたとき、事後分布はガウス過程となり、回帰問題をベイズ推論の枠組みで取り扱うことができる。これをガウス過程回帰という。以下で、ガウス過程回帰を利用した事後分布の導出について述べる。

未知の関数 $g(x)$ の事前分布が $\mathcal{GP}(m(x), k(x, x'))$ であるときに、ノイズを含む観測値 $\{(x_i, y_i)\}_{i=1}^n$ が与えられたとする。ここで、 $y_i = f(x_i) + \epsilon_i$ であり、 $\epsilon_i$ は平均0、分散 $\sigma_\epsilon^2$ の正規分布に従う独立なノイズである。ここで、学習入力 $\mathbf{x} = (x_1, \dots, x_n)^T$ に対する学習出力 $\mathbf{y} = (y_1, \dots, y_n)^T$ とテスト入力 $\mathbf{x}^* = (x_1^*, \dots, x_{n^*}^*)^T$ に対するテスト出力 $\mathbf{g}^* = (g(x_1^*), \dots, g(x_{n^*}^*))^T$ の同時分布はガウス過程の定義より、

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{g}^* \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (4)$$

ここで、

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{m}(\mathbf{x}) \\ \mathbf{m}(\mathbf{x}^*) \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I & K(\mathbf{x}, \mathbf{x}^*) \\ K(\mathbf{x}^*, \mathbf{x}) & K(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix}$$

である。また、 $\mathbf{m}(\mathbf{x}) = (m(x_1), \dots, m(x_n))^T$ 、 $\mathbf{m}(\mathbf{x}^*) = (m(x_1^*), \dots, m(x_{n^*}^*))^T$ である。ただし、 $K(\cdot, \cdot)$ は共分散行列を示し、 $I$ は $n \times n$ の単位行列である。この分布から得られる事後予測分布は

$$\begin{aligned} \mathbf{m}^*(x) &= \mathbf{m}(\mathbf{x}^*) \\ &+ K(\mathbf{x}^*, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{x})) \end{aligned}$$

$$\begin{aligned} K^* &= K(\mathbf{x}^*, \mathbf{x}^*) \\ &- K(\mathbf{x}^*, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}K(\mathbf{x}, \mathbf{x}^*) \quad (5) \end{aligned}$$

<sup>†</sup>広島市立大学大学院情報科学研究科

となる。よって、 $g(x)$ の事後分布は

$$g(x) | \mathbf{x}, \mathbf{y} \sim \mathcal{GP}(m^*(x), k^*(x, x')) \quad (6)$$

となる。ここで、

$$m^*(x) = m(x) + K(x, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}(\mathbf{y} - \mathbf{m}(\mathbf{x})),$$

$$k^*(x, x') = k(x, x') - K(x, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I]^{-1}K(\mathbf{x}, x')$$

である。

### 3 生成モデル

本研究で用いる生成モデルについて説明する。周期時系列データが $n$ 本あるとき、 $i$ 番目の周期時系列データ $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,T})$ は、時刻を $\mathbf{t} = (t_1, \dots, t_T)$ として、標準時系列 $f(t)$ に時間変換 $s_i(t)$ が加わり、その振幅を $a_i$ 倍することで得られると仮定する。

すなわち、 $i = 1, \dots, n, j = 1, \dots, T$ に対して、

$$y_{i,j} = a_i f(s_i(t_j)) + \epsilon_{i,j} \quad (7)$$

であるとする。ここで、 $\epsilon_{i,j}$ は平均0、分散 $\sigma_\epsilon^2$ の正規分布に従うノイズである。そして、生成された時系列データの集合 $\{\mathbf{y}_i\}$ が与えられたときに、標準時系列 $f(t)$ と各時間変換関数 $s_i(t)$ を推定する問題を考える。

ただし、この問題には次のような不定性があることに注意する。ある単調な時間変換関数 $\phi(t)$ によって $f'(t) = f(\phi(t))$ と定義される $f'(t)$ を標準時系列だと考えても、時間変換関数を $s'_i(t) = \phi^{-1}(s_i(t))$ と変更すれば式(7)と同様の関係が成立する。従って、標準時系列関数と時間変換関数は一意には定まらず、一つの単調な時間変換関数 $\phi(t)$ だけの不定性が存在している。

本研究では、標準関数 $f(t)$ と各時系列 $i$ の時間変換関数 $s_i(t)$ に対してガウス過程事前分布を仮定する。従って、

$$f(t) \sim \mathcal{GP}(m_f(t), k_f(t, t')) \quad (8)$$

$$s_i(t) \stackrel{\text{iid}}{\sim} \mathcal{GP}(m_s(t), k_s(t, t')) \quad (9)$$

$$a_i \stackrel{\text{iid}}{\sim} \mathcal{N}_{>0}(\mu_a, \sigma_a^2) \quad (10)$$

と書ける。また、 $\mathcal{N}_{>0}(\mu_a, \sigma_a^2)$ は定義域が正の実数である平均 $\mu_a$ 、分散 $\sigma_a^2$ の切断正規分布である。切断正規分布とは、ある範囲が切り出された正規分布のことであり、切り出された範囲において、確率の総和が1となるように高さを取り直す処理を行った分布である。つまり、 $\mathcal{N}_{>0}(\mu_a, \sigma_a^2)$ とは、平均 $\mu_a$ 、分散 $\sigma_a^2$ の正規分布から0以下の部分を切り取り、高さを取り直した分布を表している。

### 4 共分散関数

共分散関数は確率変数は確率変数の組 $g(t)$ と $g(t')$ の間の共分散を定めるものである。本研究では、周期的なデータを対象とするので、 $f(t)$ も $s_i(t)$ も周期性をもたなければならない。よって、ガウス過程事前分布の共分散関数に周期性を持たせることで、これを満たす。本研究では、[4]で提案されている関数を用いる。

すなわち、

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2}\right)$$

という二次元空間での距離を用いる共分散関数で、入力に単位円上の角度を用いて、 $\mathbf{x} = (\sin(t), \cos(t))^T$ とおくことで得られる[5]

$$k(t, t') = \sigma^2 \exp\left(-\frac{2 \sin^2\left(\frac{t-t'}{2}\right)}{l^2}\right) \quad (11)$$

を用いる。ここで $\sigma^2$ はある1点の値 $f(t)$ の周辺分布の分散、 $l$ はスケールパラメータである。スケールパラメータとは、確率変数の間の相関を与える範囲を決めるパラメータである。平均関数と共分散関数を特定することは関数空間上での分布を決めることになり、平均関数と共分散関数が与えられたとき、任意の数の点でサンプリングをすることができる。

### 5 提案するMCMCアルゴリズム

解析したいのは観測値 $\mathbf{y}_i$ と観測時刻 $\mathbf{t}$ が与えられたときの $f(t)$ と $\{s_i(t)\}$ である。ただし、 $\{s_i(t)\}$ は時刻 $\mathbf{t}$ に対して必要となるので、 $\mathbf{s}_i = \{s_i(t_1), \dots, s_i(t_T)\}$ とする。この同時事後分布は直接求めることができないので、MCMCを用いる。MCMCのギブス・サンプラーを考えると、次の3ステップにより実現できる。 $n$ 本の時系列に対して、 $\mathbf{s} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ 、 $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ 、 $\mathbf{s}_{-i} = \{\mathbf{s}_1, \dots, \mathbf{s}_{i-1}, \mathbf{s}_{i+1}, \dots, \mathbf{s}_n\}$ 、 $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$ とすると、

- $\mathbf{s}, \mathbf{y}, \mathbf{a}, \mathbf{t}$ が与えられたときの $f(t)$ の条件付き分布からサンプルを生成する
- $f(t), \mathbf{s}_{-i}, \mathbf{y}, \mathbf{a}, \mathbf{t}$ が与えられたときの $\mathbf{s}_i$ の条件付き分布からサンプルを生成する ( $i = 1, \dots, n$ )
- $f(t), \mathbf{s}, \mathbf{y}, \mathbf{t}$ が与えられたときの $a_i$ の条件付き分布からサンプルを生成する ( $i = 1, \dots, n$ )

以下の節で各のステップについて説明する。

#### 5.1 $f(t)$ のサンプリング

$f(t)$ の事前分布を式(8)について、 $\mathbf{s}, \mathbf{y}, \mathbf{a}, \mathbf{t}$ が与えられたときの条件付き分布を求め、サンプリングを行う。 $f(t)$ の事後分布は、ガウス過程回帰より、

$$f(t) | \mathbf{s}, \mathbf{y} \sim \mathcal{GP}(m_f^*(t), k_f^*(t, t')) \quad (12)$$

となる。ここで、

$$m_f^*(t) = m_f(t) + K_f(t, \mathbf{s})[K_f(\mathbf{s}, \mathbf{s}) + \sigma_n^2 I]^{-1}(\mathbf{y} - \mathbf{m}_f(\mathbf{s})),$$

$$k_f^*(t, t') = k_f(t, t') - K_f(t, \mathbf{s})[K_f(\mathbf{s}, \mathbf{s}) + \sigma_n^2 I]^{-1}K_f(\mathbf{s}, t').$$

この事後分布に従うサンプルをギブス・サンプラーにより生成する。ただし、上でも述べたように実際に関

数  $f(t)$  をサンプリングすることはできない。本研究では、サンプル点数を十分多く取り、スプライン補間を用いて  $f(t)$  のサンプリングを実現している。生成されたサンプルは  $\mathbf{s}_i$  のサンプリングの受理確率を計算する際に用いられる。

## 5.2 $a$ のサンプリング

$a_i$  の事前分布を

$$p(a_i) = \mathcal{N}_{>0}(a_i | \bar{a}, \sigma_a^2) \quad (13)$$

とおく。このように切断正規分布を事前分布とすれば、事後分布は切断正規分布に収まり、

$$p(a_i | \mathbf{y}_i, f, \mathbf{s}_i, \mathbf{t}) = \mathcal{N}_{>0} \left( a_i \left| \frac{\left( \frac{\mathbf{y}_i^T \mathbf{f}_i}{\sigma_\epsilon^2} + \frac{\bar{a}}{\sigma_a^2} \right)}{\left( \frac{\mathbf{f}_i^T \mathbf{f}_i}{\sigma_\epsilon^2} + \frac{1}{\sigma_a^2} \right)}, \left( \frac{\mathbf{f}_i^T \mathbf{f}_i}{\sigma_\epsilon^2} + \frac{1}{\sigma_a^2} \right)^{-1} \right. \right) \quad (14)$$

となる。ここで、 $\mathbf{f}_i = (f(s_i(t_1)) \cdots f(s_i(t_T)))^T$  である。この条件付き分布に従うサンプルを生成する。

## 5.3 $\mathbf{s}_i$ のサンプリング

$\mathbf{s}_i$  の条件付き分布は  $f(t)$  の場合と違って、解析的に得ることができない。これは、 $\mathbf{s}_i$  は非線形な関数  $f(t)$  を通した観測値に依存するためである。そこで、メトロポリス・ヘイスティング法を適用する。 $\mathbf{s}_i$  にはガウス過程事前分布をおいているため、 $\mathbf{s}_i$  の成分は相互に強い相関を持っている。この相関を大きく乱さない提案分布が必要である。そこで、 $\mathbf{s}_i$  を観測値と見なすことで得られる事後ガウス過程を提案分布とする。しかし、 $\mathbf{s}_i$  の全ての要素に一斉に変更を加えると、非常に大きな変動を生じやすくなる。そこで、本研究では時間軸上にウィンドウを定義し、ウィンドウ内の点に変動を限定する手法をとっている。つまり、ウィンドウ外ではノイズなしの観測値、ウィンドウ内ではノイズありの観測値と見なして、候補を生成することで変動を制限し、受理確率を調整するのである。ウィンドウ  $b$  に含まれる  $\mathbf{s}_i$  の添字集合を  $J_b$ 、 $s_{i,j} = s_i(t_j)$  とすると、 $\mathbf{s}_{i,J_b} = (s_{i,j})_{j \in J_b}$  に対する提案分布には、 $\{(t_j, s_{i,j}) | j \notin J_b\}$  をノイズなしの観測値と見なしたときの事後分布を用いる。ここで、 $\mathbf{s}_{i,J_b}$  に対応する時刻を  $\mathbf{t}_{J_b} = (t_j)_{j \in J_b}$  とし、残りの時刻を  $\mathbf{t}_{-J_b} = (t_j)_{j \notin J_b}$  とすると、提案分布は

$$q(\mathbf{s}_{i,J_b}^* | \mathbf{s}_i) \sim \mathcal{N}(\mathbf{s}_{i,J_b}^* | \mathbf{s}_{i,J_b}, K_s^*) \quad (15)$$

となる。ここで、

$$K_s^* = K_s(\mathbf{t}_{J_b}, \mathbf{t}_{J_b}) - K_s(\mathbf{t}_{J_b}, \mathbf{t}_{-J_b}) K_s(\mathbf{t}_{-J_b}, \mathbf{t}_{-J_b})^{-1} K_s(\mathbf{t}_{-J_b}, \mathbf{t}_{J_b}) \quad (16)$$

である。生成された候補を受理、または棄却し、ウィンドウを要素一つ分ずらし、同様に提案分布を生成する。このように、ウィンドウサイズを決めて、時刻を一つずつずらしたものが次のウィンドウとなる。また、受理確率  $\alpha$  は、

$$\alpha(\mathbf{s}_i^*, \mathbf{s}_i) = \min \left\{ 1, \frac{p(\mathbf{s}_i^* | f(t), \mathbf{s}_{-i}^*, \mathbf{y}_{-i}, \mathbf{t}) q(\mathbf{s}_{i,J_b} | \mathbf{s}_i^*)}{p(\mathbf{s}_i | f(t), \mathbf{s}_{-i}, \mathbf{y}_{-i}, \mathbf{t}) q(\mathbf{s}_{i,J_b}^* | \mathbf{s}_i)} \right\}$$

(17)

で得られる。ここで、 $\mathbf{s}_i^*$  は  $s$  中の  $\mathbf{s}_{i,J_b}$  を  $\mathbf{s}_{i,J_b}^*$  で置き換えたものである。

## 6 提案アルゴリズム

提案手法のアルゴリズムをまとめる。3行目で  $f(t)$  のサンプリングを行い、5行目で条件付き分布から  $a_i$  のサンプリングを行い、4~10行目で  $\mathbf{s}_i$  のサンプリングを行う。これらのサンプリングを繰り返し行い、その平均を取ることで未知の関数を推定する。

### Algorithm

---

$\mathbf{s}_1, \dots, \mathbf{s}_n$  の初期値を  $\mathbf{s}_i = \mathbf{t}, a_1, \dots, a_n$  の初期値を  $a_i = 1$  とする。  
**for**  $k = 1 : M$  **do**  
 $f(t) \leftarrow$  式 (12) の条件付き分布からサンプリング。  
**for**  $i = 1 : n$  **do**  
 $a_i \leftarrow$  式 (14) の条件付き分布からサンプリング。  
**for**  $b = 1 : B$  **do**  
式 (15) を提案分布として候補  $\mathbf{s}_{i,J_b}^*$  を生成。  
 $\mathbf{s}_{i,J_b}^*$  を式 (17) の確率に従って受理し、  
 $\mathbf{s}_{i,J_b} \leftarrow \mathbf{s}_{i,J_b}^*$   
**end for**  
**end for**  
**end for**

---

## 7 実験

提案手法を用いて、人工データを用いた多重整列実験を行い、提案手法の有効性を検証する。

整列の精度を比較する上でまず思いつくのは、標準時系列  $f(t)$  の推定精度である。しかし、3節で述べたように、時系列の整列問題に存在する不定性のために、 $f(t)$  の推定精度が必ずしもよい指標とはいえない。

そこで、本実験では2つの時系列データ間の時間の対応を示すワーピングパス [1] の推定精度を測る。ワーピングパスは2つの時系列間の対応を示す経路なので、整列問題の不定性に関わらず、一意に決まる。2つの時系列データ  $\mathbf{y}_1, \mathbf{y}_2$  とすると、その間のワーピングパスは  $s_2^{-1}(s_1(t))$  により得られる。そして、推定されたワーピングパスの真のワーピングパスに対する誤差は、それら2つのワーピングパスにより挟まれる部分の面積で評価することにする。

### 7.1 人工データ実験

人工データを整列する実験を行った。人工データは提案手法の生成モデルに従って生成した1セット5本の図1(a)のような周期時系列データを100セット用意した。提案手法と Ramsay の手法 [2] との比較を行う。本実験では、MCMC を10,000ステップ行い、後半の5,000サンプルより事後平均を求めた。図1(a)の周期時系列データを整列したものを図1(b)に示す。本実験では5本の時系列を1セットとしているので、それぞれの評価の平均を1セットの評価とする。100セットの結果を図2の箱ひげ図にまとめる。評価値を比較す

ると、Ramsay の手法より提案手法の方が、低い値で小さくまとまっており、良い結果となった。

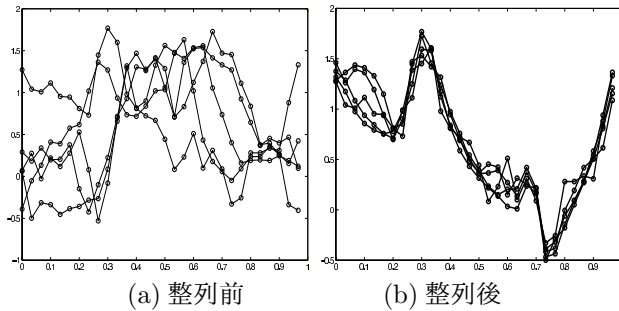


図 1: 周期時系列データの整列例

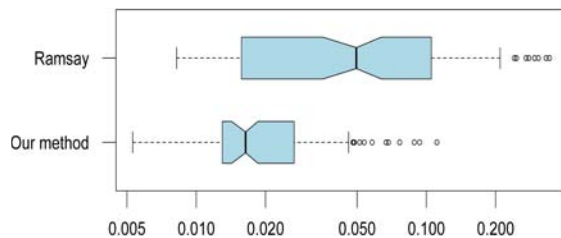


図 2: 評価の箱ひげ図

## 7.2 輪郭の位置合わせ

提案手法を用いて図3の輪郭に対して多重整列を行い、輪郭の位置合わせを行う。実データには図3の猫とカンガルーの輪郭を用いる。本実験では観測値として重心から輪郭までの距離を時系列データとして用いて、整列を行った。データ点数は64として実験を行った。得られた周期時系列データを図4(a)に示す。横軸は時刻  $t$ 、縦軸は観測値  $y$  である。また、猫のデータが実線、カンガルーのデータが破線である。本実験ではMCMCを20,000ステップ行い、後半の10,000サンプルより事後平均を求めた。

この周期時系列データに提案手法を用いて整列した結果が、図4(b)である。さらに、この周期時系列データの対応を示したものが図5である。この図から輪郭の手足や耳、尻尾などの対応がとれていることがわかる。

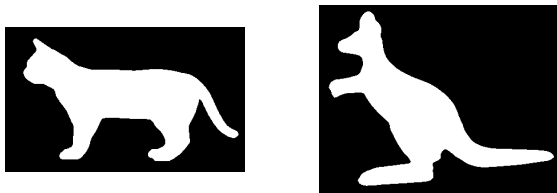


図 3: 猫とカンガルーの輪郭画像

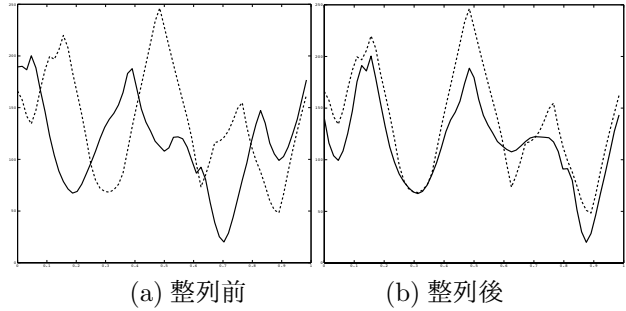


図 4: 輪郭の周期時系列データ

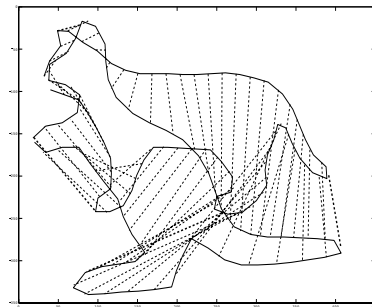


図 5: 位置合わせ後の輪郭

## 8 まとめ

ガウス過程事前分布を用いた時系列の生成モデルを考案し、MCMCを用いてベイズ解析を行う手法を提案した。共通する標準時系列が複数の時間変換関数により変換することで、それぞれの時系列が得られるとする生成モデルに対して、事前分布にガウス過程をおき、MCMCを用いることで標準時系列と時間変換関数を同時に推定する。ここで、共分散関数に周期関数を用いることで、関数に周期性を与え、周期時系列データを推定することができる。周期時系列データの整列実験の結果より、提案手法の有効性が確認できた。

## 参考文献

- [1] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol.26, no.1, pp.43-49, 1978.
- [2] J.O. Ramsay and X. Li, "Curve registration," Journal of the Royal Statistical Society. Series B (Statistical Methodology), vol.60, no.2, pp.351-363, 1998.
- [3] 秋本 真治, 末松 伸朗, 林 朗, 岩田 一貴, 電子情報通信学会論文誌. D, 情報・システム J96-D(3), 587-595, 2013-03-01
- [4] C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
- [5] MacKay, D. J. C. (1998). Introduction to Gaussian Processes. In Bishop, C. M., editor, Neural Networks and Machine Learning. Springer-Verlag. pp. 84, 92