MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Time Series Decomposition into Oscillation Components and Phase Estimation

Takeru MATSUDA and Fumiyasu KOMAKI

METR 2016–8 May 2016

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Time Series Decomposition into Oscillation Components and Phase Estimation

Takeru MATSUDA and Fumiyasu KOMAKI

Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo
{Takeru_Matsuda,komaki}@mist.i.u-tokyo.ac.jp

May 11th, 2016

Abstract

Many time series are naturally considered as a superposition of several oscillation components. For example, electroencephalogram time series (EEG) include oscillation components such as alpha, beta, and gamma. We propose a method for decomposing time series into such oscillation components using state space models. Based on the concept of random frequency modulation, Gaussian linear state space models for oscillation components are developed. In this model, the frequency of an oscillator is fluctuated by noise. Time series decomposition is accomplished by this model like the Bayesian seasonal adjustment method. Since the model parameters are estimated from data by the empirical Bayes method, the frequencies of oscillation components are determined in a data-driven manner. Also, appropriate number of oscillation components is determined with the Akaike Information Criterion (AIC). In this way, the proposed method provides a natural decomposition of the given time series into oscillation components. In neuroscience, the phase of neural time series plays an important role in neural information processing. The proposed method can be used to estimate the phase of each oscillation component and has several advantages over conventional method based on the Hilbert transform. Thus, the proposed method enables an investigation of the phase dynamics of time series. We apply the proposed method to real data from various fields such as astronomy, ecology, tidology and neuroscience.

1 Introduction

Many time series are naturally considered as a superposition of several oscillation components. For example, EEG time series are composed of several oscillation components such as alpha, beta and gamma. Each oscillation

component has its own amplitude, frequency and phase. In general, they vary over time and these dynamics characterize the process underlying the oscillation component. In neuroscience, such phase of neural time series plays an important role in neural information processing (Buzsaki, 2011). O'Keefe and Recce (1993) found that the theta phase of the hippocampal local field potential (LFP, extracellular recording of voltage) at spike timing of the place cells encodes the place of a rat. This study initiated important follow-up research on the place cells and grid cells in the hippocampus. In addition, phase synchronization between distant brain areas has been deemed important (Siapas et al., 2005), the phase reset phenomenon has been reported in human LFP (Makeig et al., 2002), and the LFP phase has been found to encode some sensory information (Lopour et al., 2013). Time series from other fields such as astronomy, ecology, and tidology, are considered to have their own oscillation components. An investigation of such time series from the viewpoint of phase dynamics would provide interesting insights.

In neuroscience, the phase of a neural time series is computed by the Hilbert transform (Cohen, 2014). After applying a band-pass filter to the original time series, the analytical signal is computed by the Hilbert transform and the phase is determined from this analytical signal. Although this method is widely used and easy to implement, it has several shortcomings (Siapas et al., 2005). First, this method does not account for measurement noise, which is inevitable in real time series data. Second, band-pass filtering can distort the waveform, which leads to inaccurate estimate of the phase. Third, the selection of band-pass filter is subjective and it may control the final estimate of phase. For example, the definition of alpha band, which is around 8-13 Hz, seems to vary among literatures. Since it is natural that there are individual differences in the frequency band of alpha, it would be useful if the frequency band is determined in a data-driven manner.

In this study, we propose a method for decomposing time series into oscillation components with state space models and also estimating the phase of each oscillation component. Based on the concept of random frequency modulation by (Wiener, 1966), Gaussian linear state space models for oscillation components are developed. In this model, the frequency of an oscillator is fluctuated by noise. Time series decomposition is accomplished by this model like the Bayesian seasonal adjustment method of Kitagawa and Gersch (1984) and Akaike and Ishiguro (1980). Since parameters of the state space model are estimated from data by empirical Bayes method, the amplitude and frequency of each oscillation component are determined in a data-driven manner. Also, the number of oscillation components is determined with the Akaike Information Criterion (AIC) from Akaike (1980). In this way, the proposed method accomplishes a natural decomposition of the given time series into oscillation components. Following the decomposition, the phase of each oscillation component can be estimated. This phase es-

timator has several advantages over the conventional phase estimator using the Hilbert transform as discussed in Section 4.3.

We first explain about the existing method for phase estimation with the Hilbert transform in Section 2. Then, we introduce random frequency modulation models in Section 3. Based on this model, details of the proposed method are explained in Section 4. Results of numerical experiments are presented in Section 5. The proposed method is applied to real time series from various fields in Section 6.

2 Phase estimation with the Hilbert transform

We explain about the existing method for phase estimation with the Hilbert transform (Cohen, 2014). Given a time series $y_0(t)$, its phase in the frequency band $[f_1, f_2]$ is estimated as follows. First, a band-pass filter with the passband $[f_1, f_2]$ is selected, applied to $y_0(t)$ and the band-pass filtered signal $y(t)$ is obtained. To prevent the band-pass filtering from causing phase delay, the band-pass filter $b = (b_0, b_1, \cdots, b_L)$ is applied twice in opposite directions as follows:

$$\widetilde{y}(t) = \sum_{l=0}^{L} b_l y_0(t-l), \quad y(t) = \sum_{l=0}^{L} b_l \widetilde{y}(t+l).$$

The MATLAB function *filtfilt* performs this operation. Next, the band-pass filtered signal $y(t)$ is transformed into another signal $y_H(t)$ by the Hilbert transform:

$$y_H(t) = \int_{-\infty}^{\infty} \frac{y(\tau)}{t-\tau} \mathrm{d}\tau. \tag{1}$$

Taking $y(t)$ as the real part and $y_H(t)$ as the imaginary part, the analytical signal $z(t)$ is determined as follows:

$$z(t) = y(t) + i y_H(t), \tag{2}$$

where i denotes the imaginary unit $\sqrt{-1}$. Finally, we define the phase $\phi(t)$ of $y_0(t)$ in the frequency band $[f_1, f_2]$ as the angle of the analytical signal $z(t)$:

$$\phi(t) = \arg z(t). \tag{3}$$

Throughout this paper, we take the phase to be $-\pi \leq \phi(t) < \pi$.

Note that (1) and (2) can be rewritten as

$$Z(\omega) = \begin{cases} 2Y(\omega) & (\omega > 0) \\ 0 & (\omega < 0) \end{cases}, \tag{4}$$

where $X(\omega)$ denotes the Fourier transform of $x(t)$. The MATLAB function *hilbert* computes the Hilbert transform using the Fast Fourier Transform and (4).

Although this method is widely used and easy to implement, it has several shortcomings that have recently been recognized in the field of neuroscience (Siapas et al., 2005). First, this method does not account for measurement noise in $y(t)$, which is inevitable in real time series data. Second, band-pass filtering can distort the waveform of $y(t)$, which leads to inaccurate estimation of the phase. Third, the selection of a band-pass filter is subjective and may affect the final estimate of the phase. For example, the definition of an alpha band seems to vary among studies. Since it is natural that there are individual differences in the frequency band of alpha, it would be useful if the frequency band is determined in a data-driven manner.

3 Random frequency modulation model

3.1 Motivation

In general, a state space model is a pair of the state model and the observation model:

$$x_{t+1} = f(x_t, v_t), \quad v_t \sim p(v),$$
$$y_t = g(x_t, w_t), \quad w_t \sim p(w).$$

In particular, a state space model of the following form is called a Gaussian linear state space model:

$$x_{t+1} = Fx_t + v_t, \quad v_t \sim \mathrm{N}(0, Q),$$

$$y_t = Hx_t + w_t, \quad w_t \sim \mathrm{N}(0, R).$$

The central problem in state space models is to estimate the state x from the observation y.

We note that (2) can be interpreted as an observation model in a Gaussian linear state space model:

$$y(t) = x_1(t),$$

where

$$x(t) = (x_1(t), x_2(t)) = (\mathrm{Re}z(t), \mathrm{Im}z(t))^\top.$$

Here, no observation noise is assumed. Therefore, the Hilbert transform can be interpreted as estimating the two-dimensional state vector $x(t)$ from the one-dimensional observation $y(t)$. Based on this idea, we propose state space models that describes oscillation components in time series in the following.

3.2 Definition

Wiener (1966) noted that the spectrum of EEG has a characteristic peak around 10 Hz, corresponding to the alpha wave. In order to understand how this peak is established, he developed a stochastic process model of brain wave that represents random fluctuation in the frequency of an oscillator. He showed that such random frequency modulation gives rise to the alpha peak in the spectrum of brain wave.

Based on Wiener's "random frequency modulation" concept, we develop Gaussian linear state space models for oscillation components as follows:

$$\begin{pmatrix} x_{t+1,1} \\ x_{t+1,2} \end{pmatrix} = a \begin{pmatrix} \cos(2\pi f \Delta t) & -\sin(2\pi f \Delta t) \\ \sin(2\pi f \Delta t) & \cos(2\pi f \Delta t) \end{pmatrix} \begin{pmatrix} x_{t,1} \\ x_{t,2} \end{pmatrix} + \begin{pmatrix} v_{t,1} \\ v_{t,2} \end{pmatrix}, \tag{5}$$

$$y_t = x_{t,1} + w_t, \tag{6}$$

where

$$\begin{pmatrix} v_{t,1} \\ v_{t,2} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right),$$

$$w_t \sim \mathrm{N}(0, \tau^2).$$

Here, $0 < a < 1$, $0 \leq 2\pi f \Delta t \leq \pi$ and Δt is the sampling period of y_t. In Section 3.5, we discuss the extension of this model by adding moving average (MA) terms to (5). From the system model (5), x_{t+1} is determined by rotating x_t through an angle $2\pi f \Delta t$ about the origin, multiplying by a, and adding system noise v_t. These operations represent the random frequency modulation. We added the constant a to make the model stationary. From the observation model (6), y_t is the first coordinate of x_t plus observation noise w_t. Like (3), the phase of the oscillator is naturally defined as

$$\phi_t = \arg\left(x_{t,1} + i x_{t,2}\right). \tag{7}$$

Note that we take the phase to be $-\pi \leq \phi(t) < \pi$ throughout this paper.

In practice, a given time series is often a superposition of several oscillation components. Thus, we model a time series with K oscillation components as follows:

$$\begin{pmatrix} x^{(k)}_{t+1,1} \\ x^{(k)}_{t+1,2} \end{pmatrix} = a_k \begin{pmatrix} \cos(2\pi f_k \Delta t) & -\sin(2\pi f_k \Delta t) \\ \sin(2\pi f_k \Delta t) & \cos(2\pi f_k \Delta t) \end{pmatrix} \begin{pmatrix} x^{(k)}_{t,1} \\ x^{(k)}_{t,2} \end{pmatrix} + \begin{pmatrix} v^{(k)}_{t,1} \\ v^{(k)}_{t,2} \end{pmatrix}, \quad (k = 1, \cdots, K), \tag{8}$$

$$y_t = \sum_{k=1}^{K} x^{(k)}_{t,1} + w_t, \tag{9}$$

where

$$\begin{pmatrix} v^{(k)}_{t,1} \\ v^{(k)}_{t,2} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_k^2 & 0 \\ 0 & \sigma_k^2 \end{pmatrix} \right),$$

$$w_t \sim \mathrm{N}(0, \tau^2).$$

Here, $0 < a_k < 1$, $0 \leq 2\pi f_k \Delta t \leq \pi$ and Δt is the sampling period of y_t. The frequencies of K oscillation components underlying the given time series are $f_1, \cdots, f_K$. The phase of the k-th oscillator is naturally defined as

$$\phi_t^{(k)} = \arg\left(x_{t,1}^{(k)} + i x_{t,2}^{(k)}\right). \tag{10}$$

In the following, we denote the state vector and the observation matrix of the model (8) and (9) by $x_t = (x_{t,1}^{(1)}, x_{t,2}^{(1)}, \cdots, x_{t,1}^{(K)}, x_{t,2}^{(K)})^\top$ and $H = (1, 0, 1, 0, \cdots, 1, 0)$.

3.3 Power spectrum

Since the model (5) describes a stationary Gaussian time series, we discuss its second-order spectrum (Brockwell and Davis, 2009). The autocovariance function of $x_{t,1}$ is

$$\mathrm{E}\left[x_{t,1} x_{t+h,1}\right] = \frac{\sigma^2}{1-a^2} a^h \cos\left(2\pi h f \Delta t\right).$$

This coincides with the autocovariance function of the ARMA(2,1) model

$$X_t - 2a\cos(2\pi f \Delta t) X_{t-1} + a^2 X_{t-2} = Z_t + bZ_{t-1}, \tag{11}$$

where

$$b = \frac{1}{2}\left(A - 2a\cos(2\pi f \Delta t) + \sqrt{\left(A - 2a\cos(2\pi f \Delta t)\right)^2 - 4}\right),$$

$$A = \frac{1 - 2a^2\cos^2(2\pi f \Delta t) + a^4 \cos(4\pi f \Delta t)}{a(a^2-1)\cos(2\pi f \Delta t)},$$

and $Z_t \sim \mathrm{N}(0, V)$ are independent random variables with variance

$$V = \frac{-\sigma^2 a \cos(2\pi f \Delta t)}{b} \ (> 0). \tag{12}$$

The power spectrum of the ARMA(2,1) model (11) is

$$p(\lambda) = \frac{V}{2\pi} \frac{|1 + b\exp(-i\lambda)|^2}{|1 - 2a\cos(2\pi f \Delta t)\exp(-i\lambda) + a^2\exp(-2i\lambda)|^2} \quad (-\pi < \lambda \leq \pi), \tag{13}$$

and it has one peak around $\lambda = 2\pi f \Delta t$. We note that (11) specifies a submodel of the full ARMA(2,1) model.

In the model (8) and (9), the spectrum of y_t is the sum of the spectrums of $x_{t,1}^{(1)}, \cdots, x_{t,1}^{(K)}$ and white noise. Therefore, the model (8) and (9) can be viewed as representing the spectrum of a given time series as a sum of the spectrums of the ARMA(2,1) models (11).

3.4 Oscillator interpretation of the AR model

The submodel (11) of the ARMA(2,1) models was introduced naturally as the description of an oscillator with random frequency modulation On the other hand, the AR(2) models are commonly used to describe time series with single characteristic frequencies (Quinn and Hannan, 2013). Here, we explain how the AR process of arbitrary order can also be interpreted as a superposition of several oscillation components (Huerta and West, 1999a,b).

For the AR(p) process

$$x_t = \sum_{k=1}^{p} a_k x_{t-k} + \varepsilon_t, \tag{14}$$

the characteristic roots $\xi_1, \cdots, \xi_p$ are defined as the roots of the algebraic equation

$$z^p - \sum_{k=1}^{p} a_k z^{p-k} = 0. \tag{15}$$

In the following, we assume that (15) has no multiple roots, $\xi_1, \cdots, \xi_{2m_p}$ are imaginary numbers satisfying $\xi_2 = \bar{\xi}_1, \cdots, \xi_{2m_p} = \bar{\xi}_{2m_p-1}$ and $\xi_{2m_p+1}, \cdots, \xi_p$ are real numbers. Let V be a Vandermonde matrix

$$V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \xi_1^{-1} & \xi_2^{-1} & \cdots & \xi_p^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1^{1-p} & \xi_2^{1-p} & \cdots & \xi_p^{1-p} \end{pmatrix}. \tag{16}$$

Since $\xi_1, \cdots, \xi_p$ are all different, V is nonsingular. Then, the state space form of the AR model

$$\begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{p-1} & a_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is reduced to the following canonical form:

$$\begin{pmatrix} u_t^{(1)} \\ u_t^{(2)} \\ \vdots \\ u_t^{(p)} \end{pmatrix} = \begin{pmatrix} \xi_1 & 0 & \cdots & 0 \\ 0 & \xi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \xi_p \end{pmatrix} \begin{pmatrix} u_{t-1}^{(1)} \\ u_{t-1}^{(2)} \\ \vdots \\ u_{t-1}^{(p)} \end{pmatrix} + V^{-1} \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{17}$$

where

$$\begin{pmatrix} u_t^{(1)} \\ u_t^{(2)} \\ \vdots \\ u_t^{(p)} \end{pmatrix} = V^{-1} \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix}. \tag{18}$$

Now, since the multiplication by a complex number ξ_k corresponds to the composition of the scalar multiplication by $|\xi_k|$ and the rotation by $\arg \xi_k$ on the complex plane, each $u_t^{(k)}$ in (17) represents an oscillator with random frequency modulation. However. from $\xi_2 = \bar{\xi}_1, \cdots, \xi_{2m_p} = \bar{\xi}_{2m_p-1}$, we can show that $u_t^{(2)} = \bar{u}_t^{(1)}, \cdots, u_t^{(2m_p)} = \bar{u}_t^{(2m_p-1)}$. Therefore, each pair $(u_t^{(1)}, u_t^{(2)}), \cdots, (u_t^{(2m_p-1)}, u_t^{(2m_p)})$ essentially represents one oscillator. Since

$$x_t = \sum_{k=1}^{K} u_t^{(k)}$$

from (18), the AR process (14) is interpreted as the superposition of $p - m_p$ oscillators.

The phase of each oscillator in the AR process can now be defined as

$$\phi_t^{(k)} = \arg u_t^{(k)} \quad (k = 1, 3, \cdots, 2m_p + 1, 2m_p + 2, \cdots, p). \tag{19}$$

However, as described in Section 5, this phase definition did not perform well in numerical experiments. Therefore, we use the model (8) in the proposed method. Just as AR(2) models have been extended so that frequencies vary with time (Nguyen et al., 2009), the model (8) can also be made to have a frequency that varies with time.

Whereas the system noise is degenerated in (17), the system noise is rotation invariant in (8). In other words, the modulation does not depend on direction in (8). Therefore, the model (8) describes the random frequency modulation more naturally than the model (17). Also, whereas the system noises for different oscillators are strongly correlated in (17), the system noises for different oscillators are independent in (8). Therefore, the model (8) is preferable to the AR model (17) when one decomposes a given time series into independent oscillation components.

3.5 Addition of MA term

We found some real time series for which the fit of the model (8) and (9) is not very good in the high frequency range. This problem was overcome by

adding the following MA terms to the model (8) and (9):

$$\begin{pmatrix} x_{t+1,1}^{(k)} \\ x_{t+1,2}^{(k)} \end{pmatrix} = a_k \begin{pmatrix} \cos(2\pi f_k \Delta t) & -\sin(2\pi f_k \Delta t) \\ \sin(2\pi f_k \Delta t) & \cos(2\pi f_k \Delta t) \end{pmatrix} \begin{pmatrix} x_{t,1}^{(k)} \\ x_{t,2}^{(k)} \end{pmatrix} + \begin{pmatrix} v_{t,1}^{(k)} \\ v_{t,2}^{(k)} \end{pmatrix}$$
$$+ b_{k,1} \begin{pmatrix} v_{t-1,1}^{(k)} \\ v_{t-1,2}^{(k)} \end{pmatrix} + \cdots + b_{k,q} \begin{pmatrix} v_{t-q,1}^{(k)} \\ v_{t-q,2}^{(k)} \end{pmatrix} \quad (k = 1, \cdots, K), \quad (20)$$

$$y_t = \sum_{k=1}^{K} x_{t,1}^{(k)} + w_t, \quad (21)$$

where

$$\begin{pmatrix} v_{t,1}^{(k)} \\ v_{t,2}^{(k)} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_k^2 & 0 \\ 0 & \sigma_k^2 \end{pmatrix} \right),$$

$$w_t \sim \mathrm{N}(0, \tau^2).$$

This model corresponds to a submodel of the ARMA(2,1 + q) model. Compared with the original model (8) and (9), the spectrum of the model (20) and (21) decay faster in the high frequency range. Therefore, for time series whose spectrum decay fast in the high frequency range, the model (20) and (21) provides better fit than the original model (8) and (9). This phenomenon is demonstrated using tidal data in Section 6.3.

4 Proposed method

4.1 Decomposition into oscillation components and phase estimation

Based on the random frequency modulation model (8) and (9), the proposed method decomposes the given time series into oscillation components and estimates the phase of each component. Decomposition is accomplished using the same approach as in the Bayesian seasonal component adjustment method of Kitagawa and Gersch (1984) and Akaike and Ishiguro (1980). The methods of selecting K and estimating the parameters $f_1, \cdots, f_K$, $a_1, \cdots, a_K$, $\sigma_1^2, \cdots, \sigma_K^2$, and τ^2 are explained in Section 4.2.

For Gaussian linear state space models, filtering and smoothing are realized by the Kalman filter and smoother. The Kalman filter algorithm calculates the filtering distribution $p(x_t \mid y_1, \cdots, y_t)$ and the one-step ahead predictive distribution $p(x_{t+1} \mid y_1, \cdots, y_t)$. The Kalman smoother algorithm calculates the smoothing distribution $p(x_s \mid y_1, \cdots, y_t)$ where $s < t$. These algorithms form the basis of the proposed method. Since all of the conditional distributions $p(x_s \mid y_1, \cdots, y_t)$ are Gaussian, they are determined by the mean and the covariance matrix. We define the conditional mean as

$$x_{s|t} = E[x_s \mid y_1, \cdots, y_t]$$

and the conditional covariance matrix as

$$\Sigma_{s|t} = E[(x_s - x_{s|t})(x_s - x_{s|t})^\top \mid y_1, \cdots, y_t].$$

Kitagawa and Gersch (1984) and Akaike and Ishiguro (1980) developed the Bayesian seasonal adjustment method. They assumed that a given time series y_t consists of a trend component T_t, a seasonal component S_t and observation noise W_t, as follows:

$$y_t = T_t + S_t + W_t.$$

The trend component T_t and the seasonal component S_t are described by Gaussian linear state space models. By applying the Kalman smoother, the given time series y_t is decomposed into these components:

$$y_t = T_{t|N} + S_{t|N} + W_{t|N},$$

where N denotes the length of the time series. West (1997) also proposed a method for decomposing a given time series into trend and AR components.

The proposed method decomposes a given time series into oscillation components based on the model (8) and (9). By applying the Kalman smoother, y_t is represented as the superposition of K oscillation components as follows:

$$y_t = \sum_{k=1}^{K} x^{(k)}_{t,1|N} + w_{t|N}, \tag{22}$$

where w_t represents the observation noise. The phase $\phi^{(k)}_t(t)$ of the k-th oscillator (10) is then estimated as

$$\hat{\phi}^{(k)}_t = \arg\left(x^{(k)}_{t,1|N} + i x^{(k)}_{t,2|N}\right). \tag{23}$$

We can construct the credible intervals of each oscillation component $x^{(k)}_{t,1}$ using the conditional covariance $\left(\Sigma_{t|N}\right)_{2k-1,2k-1}$. The $(2\Phi(1) - 1 =)$ 68 % credible interval is

$$\left[x^{(k)}_{t,1|N} - \sqrt{\left(\Sigma_{t|N}\right)_{2k-1,2k-1}}, x^{(k)}_{t,1|N} + \sqrt{\left(\Sigma_{t|N}\right)_{2k-1,2k-1}}\right],$$

where Φ is the cumulative distribution function of the standard normal distribution $\mathrm{N}(0,1)$. Also, we can construct the credible intervals of the phase $\phi^{(k)}_t$ defined as (10) using the conditional covariance $\left(\Sigma_{t|N}\right)_{2k-1,2k-1}$, $\left(\Sigma_{t|N}\right)_{2k-1,2k}$ and $\left(\Sigma_{t|N}\right)_{2k,2k}$. Let $\widetilde{x}_1 = (\widetilde{x}_{1,1}, \widetilde{x}_{1,2}), \cdots, \widetilde{x}_M = (\widetilde{x}_{M,1}, \widetilde{x}_{M,2})$ $(M = 10^3)$ be independent samples from the conditional distribution of $(x^{(k)}_{t,1}, x^{(k)}_{t,2})^\top$

$$\mathrm{N}\left(\begin{pmatrix} x^{(k)}_{t,1|N} \\ x^{(k)}_{t,2|N} \end{pmatrix}, \begin{pmatrix} \left(\Sigma_{t|N}\right)_{2k-1,2k-1} & \left(\Sigma_{t|N}\right)_{2k-1,2k} \\ \left(\Sigma_{t|N}\right)_{2k,2k-1} & \left(\Sigma_{t|N}\right)_{2k,2k} \end{pmatrix}\right).$$

The phase of each sample is defined as

$$\widetilde{\phi}_k = \arctan \frac{\widetilde{x}_{k,2}}{\widetilde{x}_{k,1}}.$$

We sort $\widetilde{\phi}_k - \phi_{t|N}^{(k)}$ and put them $\widetilde{\phi}_{(1)}, \cdots, \widetilde{\phi}_{(M)}$. Then, we construct the 100α % credible interval of the phase $\phi_t^{(k)}$ as $[\phi_{t|N}^{(k)} + \widetilde{\phi}_{(l)}, \phi_{t|N}^{(k)} + \widetilde{\phi}_{(u)}]$, where $\widetilde{\phi}_{(l)}, \cdots, \widetilde{\phi}_{(u)}$ have $\lceil M\alpha \rceil$ smallest absolute values among $\widetilde{\phi}_{(1)}, \cdots, \widetilde{\phi}_{(M)}$.

Since the proposed method is based on the state space model (8) and (9), it can be applied to time series with missing values; the filtering step is skipped for time points without observation (Kitagawa, 2010). In this way, the state vectors can be estimated even for time points without observation. This capability is demonstrated using tidal data with missing values in Section 6.4.

4.2 Parameter estimation and model selection

Before decomposing a given time series, we need to determine the parameters of the model (8) and (9) and the number of oscillation components K. We estimate the parameters by the empirical Bayes method and select K with the Akaike Information Criterion (AIC) of Akaike (1980).

First, we explain the method for parameter estimation. We denote the parameters in the model (8) and (9) as $\theta_K = (f_1, \cdots, f_K, a_1, \cdots, a_K, \widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2, \tau^2)$, where

$$\widetilde{\sigma}_k^2 = \tau^{-2}\sigma_k^2.$$

We estimate θ_K by maximizing the marginal likelihood $L(\theta_K) = p(y_1, \cdots, y_N \mid \theta_K)$ of the model (8) and (9) with numerical optimization.

As noted in Kitagawa (2010), the maximization of $L(\theta_K)$ with respect to τ^2 is solved in closed form. The log marginal likelihood of the model (8) and (9) is computed using the Kalman filter as follows:

$$\log L(\theta_K) = \sum_{t=1}^{T} \log p(y_t \mid y_1, \cdots, y_{t-1}; \theta_K), \tag{24}$$

$$\log p(y_t \mid y_1, \cdots, y_{t-1}; \theta_K) = -\frac{1}{2}\log\Big(2\pi(H\Sigma_{t|t-1}H^\top + \tau^2)\Big) - \frac{(y_t - Hx_{t|t-1})^2}{2(H\Sigma_{t|t-1}H^\top + \tau^2)},$$

where $x_{1|0}$ and $\Sigma_{1|0}$ are the mean and the covariance matrix of the stationary distribution of the model (8):

$$x_{1|0} = 0,$$

$$(\Sigma_{1|0})_{2k-1,2k-1} = (\Sigma_{1|0})_{2k,2k} = \frac{\tau^2\widetilde{\sigma}_k^2}{1 - a_k^2} \quad (k = 1, \cdots, K),$$

$$(\Sigma_{1|0})_{ij} = 0 \quad (i \neq j).$$

When $(f_1, \cdots, f_K, a_1, \cdots, a_K, \widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2)$ are fixed, $x_{t|t-1}$ and $x_{t|t}$ do not depend on τ^2, and $\Sigma_{t|t-1}$ and $\Sigma_{t|t}$ are proportional to τ^2. Therefore, the log-likelihood function (24) is rewritten as

$$\log L(\theta_K) = \sum_{t=1}^{T} \left(-\frac{1}{2} \log \left(2\pi\tau^2 (H\widetilde{\Sigma}_{t|t-1} H^\top + 1) \right) - \frac{(y_t - H\widetilde{x}_{t|t-1})^2}{2\tau^2 (H\widetilde{\Sigma}_{t|t-1} H^\top + 1)} \right), \tag{25}$$

where $\widetilde{\Sigma}_{t|t-1}$ and $\widetilde{\Sigma}_{t|t}$ are conditional covariance matrices with $\tau^2 = 1$ and fixed parameters $f_1, \cdots, f_K$, $a_1, \cdots, a_K$, $\widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2$. The function (25) is maximized as a function of τ^2 by

$$\hat{\tau}^2 = \frac{1}{T} \sum_{t=1}^{T} \frac{(y_t - H\widetilde{x}_{t|t-1})^2}{H\widetilde{\Sigma}_{t|t-1} H^\top + 1},$$

and the maximum value is

$$\max_{\tau^2} \log L(\theta_K) = -\frac{T}{2} \log \left(2\pi e \hat{\tau}^2 \right) - \frac{1}{2} \sum_{t=1}^{T} \log (H\widetilde{\Sigma}_{t|t-1} H^\top + 1). \tag{26}$$

Thus, we have reduced the number of parameters to be estimated by numerical optimization. We only need to maximize (26) with respect to $f_1, \cdots, f_K$, $a_1, \cdots, a_K$, and $\widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2$.

We maximize (26) with respect to $f_1, \cdots, f_K, a_1, \cdots, a_K$, and $\widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2$ using the quasi-Newton method. Since (26) is not concave in general, the initial values for $f_1, \cdots, f_K$, $a_1, \cdots, a_K$, and $\widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2$ must be carefully determined. The method of setting the initial values, based on the oscillator interpretation of AR models discussed in Section 3.4, is detailed in the Appendix.

We can also use the expectation maximization (EM) algorithm to estimate the parameters (Shumway and Stoffer, 1982). For the model (8) and (9), M-step can be written down in a closed form. However, this algorithm was slower than the quasi-Newton method in the numerical experiments. Therefore, here, we use the quasi-Newton method for the parameter estimation.

For the model (20) and (21), we take $\theta_{q,K} = (f_1, \cdots, f_K, a_1, \cdots, a_K, \widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2, b_1^{(1)}, \cdots, b_q^{(K)}, \tau^2)$ and $\theta_{q,K}$ is estimated in the same way as θ_K.

Next, we explain the method for model selection. We select the number K of oscillation components based on the Akaike Information Criterion (AIC) from Akaike (1980), Since $3K+1$ parameters $(f_1, \cdots, f_K, a_1, \cdots, a_K, \sigma_1^2, \cdots, \sigma_K^2, \tau^2)$ are estimated from the data, the AIC of the model (8) and (9) with K oscillation components is defined as

$$\mathrm{AIC}_K = -2 \log L(\hat{\theta}_K) + 2(3K + 1),$$

where $L(\theta_K) = p(y_1, \cdots, y_N \mid \theta_K)$ is the marginal likelihood and $\hat{\theta}_K$ is the estimate of θ_K. Similarly, when we add MA terms, the AIC of the model (20) and (21) with K oscillation components is defined as

$$\mathrm{AIC}_{q,K} = -2 \log L(\hat{\theta}_{q,K}) + 2\{(3+q)K + 1\}.$$

We select the model with the minimum AIC to determine K.

4.3 Comparison with the Hilbert transform method

In Section 2, we discussed several shortcomings of the existing phase estimation method using the Hilbert transform. The proposed method overcomes these shortcomings. First, since the proposed method is based on the state space model (8) and (9), measurement noise is naturally accounted for in the decomposition (22). Next, by virtue of parameter estimation with the empirical Bayes method and model selection with information criterion AIC, the proposed method decomposes a given time series into several oscillation components in a data-driven manner. The frequency, amplitude and the degree of random frequency modulation of each component, together with the number of components, are determined from data objectively. Therefore, the proposed method avoids arbitrary selection of the band-pass filter. Finally, the proposed method is free from waveform distortion caused by band-pass filtering.

Furthermore, the proposed method can evaluate the error of the phase estimate (23) using the conditional covariance $\left(\Sigma_{t|T}\right)_{2k-1,2k-1}$, $\left(\Sigma_{t|T}\right)_{2k-1,2k}$ and $\left(\Sigma_{t|T}\right)_{2k,2k}$, and can also be applied to time series with missing values since it is based on the state space model (8) and (9). The Hilbert transform method is not capable of performing these tasks.

5 Numerical experiments

In this Section, we compare the proposed method to the Hilbert transform method with simulated data. We generated simulated data from the following state space model:

$$\log r_{t+1}^{(k)} \sim \mathrm{N}(a_k \log r_t^{(k)}, \sigma_k^2) \quad (k = 1, \cdots, K),$$

$$\phi_{t+1}^{(k)} \sim \mathrm{VM}(\phi_t^{(k)} + 2\pi f_k \Delta t, \kappa_k) \quad (k = 1, \cdots, K),$$

$$y_t \sim \mathrm{N}\left(\sum_{k=1}^{K} r_t^{(k)} \cos \phi_t^{(k)}, \tau^2\right),$$

where $\mathrm{VM}(\mu, \kappa)$ represents the von Mises distribution with a mean parameter μ and a concentration parameter κ.This model represents a time series

Table 1: AIC values of the models (8) and (9) with K oscillation components

K	1	2	3	4	5
AIC	1793	1659	1442	1448	1453

with K oscillation components. In this model, the amplitude r_t and the phase ϕ_t of each oscillator develop independently.

In the simulation, the parameters were set with $K = 3$ as follows:

$$a_1 = 0.5, \sigma_1^2 = 0.01, f_1 = 25, \kappa_1 = 20,$$

$$a_2 = 0.4, \sigma_2^2 = 0.01, f_2 = 50, \kappa_2 = 30,$$

$$a_3 = 0.3, \sigma_3^2 = 0.01, f_3 = 75, \kappa_3 = 40,$$

and $\tau^2 = 1$. Therefore, the simulated data consists of three oscillation components with frequencies of 25 Hz, 50 Hz, and 75 Hz. We used the MATLAB toolbox CircStat (Berens, 2009) to generate random variables from the von Mises distribution. The sampling frequency was set to 200 Hz, and the data length was set to 1000.

Table 1 shows the AIC values of the models (8) and (9) with K oscillation components. The AIC attains minimum at $K = 3$. With $K = 3$, the estimated frequencies are $\hat{f}_1 = 25.22$, $\hat{f}_2 = 50.40$ and $\hat{f}_3 = 75.10$. Therefore, the proposed method detects the correct number and frequencies of oscillators in a data-driven manner.

Fig. 1 shows the decomposition of the simulated data obtained by the proposed method. The proposed method succeeds in denoising the given time series and decomposing it into three oscillation components.

Table 2 shows the mean squared errors of phase estimators. The proposed method has smaller mean squared errors than the Hilbert transform method. Here, the Hilbert transform method adopted band-pass filters with a passband of 20-30 Hz (for $\phi^{(1)}$), 40-60 Hz (for $\phi^{(2)}$) and 60-90 Hz (for $\phi^{(3)}$). We used the MATLAB function *filtfilt* and *hilbert* to prevent phase delay during band-pass filtering and to compute the Hilbert transform, respectively. Also, we examined the phase estimator (19) defined from the canonical form (17) of the AR models. Among the AR models with observation noise (27) and (28), the AR(7) model had the minimum AIC. As described in Section 3.4, the fitted AR(7) model had four oscillation components with frequencies of 0 Hz, 25.17 Hz, 50.41 Hz and 75.13 Hz. Therefore, the phases of three oscillation components appear estimable using the estimator (19). However, as shown in Table 2, this estimator has a mean squared error that is significantly larger than the other estimators.

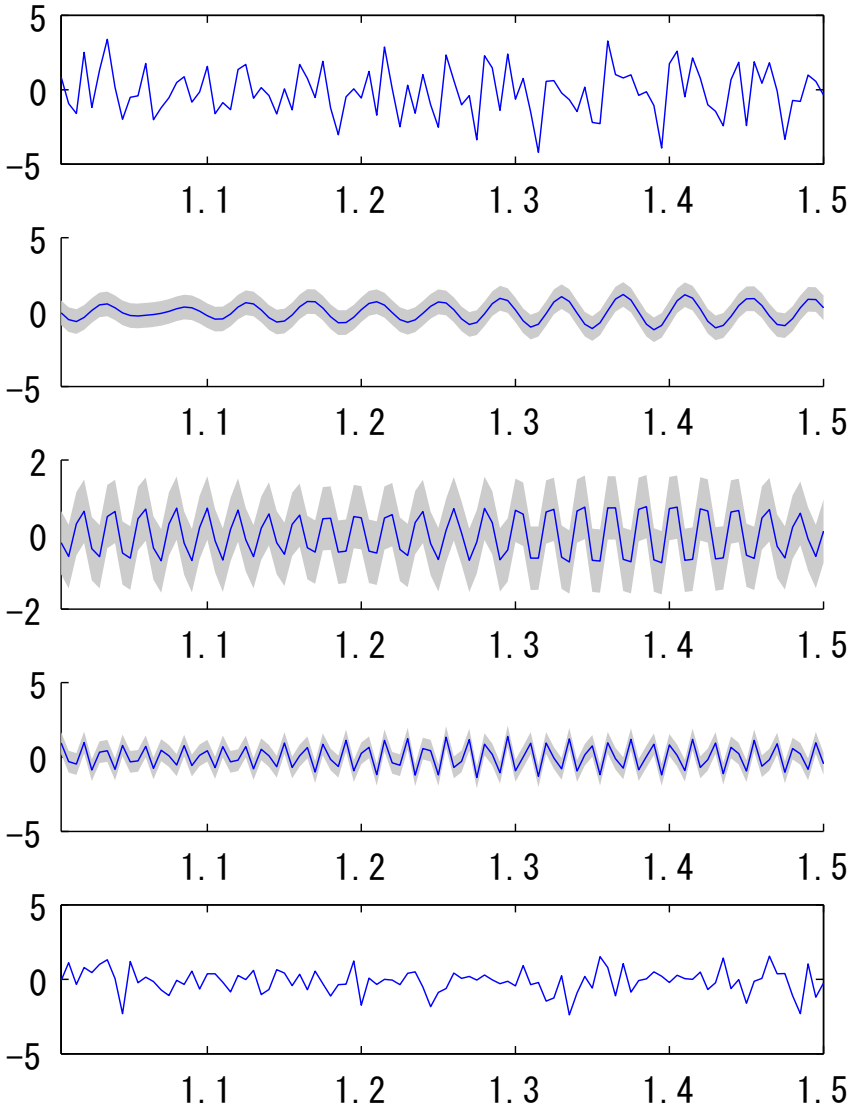

Figure 1: Decomposition of simulated data. First panel: raw data. Second-fourth panel: oscillation components with 68% credible intervals. Fifth panel: observation noise.

Table 2: Mean squared errors of phase estimators

	$\phi^{(1)}$	$\phi^{(2)}$	$\phi^{(3)}$
Hilbert transform	2.28	2.16	2.63
AR(7)	7.01	9.71	7.19
Proposed method	2.33	1.59	1.70

6 Applications to real data

6.1 Canadian Lynx data

The Canadian Lynx data is a famous time series that appears in many textbooks of time series analysis. Here, we use data from Brockwell and Davis (2009) that is the annual number of the Canadian lynxes trapped in the Mackenzie River District of the Northwestern Canada for the period 1821-1934. Many statistical analyses of this data have been performed (Campbell and Walker, 1977; Tong, 1977). We log-transformed the original data.

Table 3 shows the AIC values of several models for Canadian Lynx data. As noted in Tong (1977), the AR(11) model is the minimum AIC model among AR models without observation noise. Among AR models with observation noise (27) and (28), the AR(13) model is the minimum AIC model. Our model (8) and (9) attains the minimum AIC when $K = 6$ and provides better fit than the AR models. The estimated frequencies of the six oscillation components are

$$\hat{f}_1 = 0, \hat{f}_2 = 0.03, \hat{f}_3 = 0.10, \hat{f}_4 = 0.20, \hat{f}_5 = 0.30, \hat{f}_6 = 0.41,$$

where the frequency units are *per year*.

Fig. 2 presents the decomposition into oscillation components obtained by the proposed method. Here, observation noise is omitted since it is in the order 10^-5. Since $\hat{f}_1 = 0$, the first component is considered to be the trend component. As we can see from the first panel in Fig. 2, the Canadian Lynx data has a characteristic period of approximately 10 years. The third component, which has a 9.78 years period, corresponds well to this value.

Fig. 3 shows the phases of the six oscillation components. Because the first component represents the trend, its phase remains largely unchanged. For the third component, the phase develops regularly and the credible interval of the phase is very narrow. For the fourth component, which has a 5.02 years period, a phase reset phenomenon occurs around 1850 and the credible interval of the phase widens around this phenomenon. This observed phase reset may in fact represent a real-world event involving the Canadian Lynx. For example, it would be interesting to apply the proposed method to the annual number of the Lepus americanus, on which the Canadian lynx feeds, and investigate the phase relationships between them. In this way, the proposed method enables investigation of time series from a phase

Table 3: AIC values of several models for Canadian Lynx data. The AR order p and number of components K are optimized.

our model (8) and (9) with $K = 6$	-27.36
AR(11)	-25.79
AR(13) with observation noise	-26.47

Table 4: AIC values of several models for Wolfer sunspot data. The AR order p and number of components K are optimized.

our model (8) and (9) with $K = 3$	-19.71
AR(9)	-19.54
AR(7) with observation noise	-21.44

dynamics perspective and may give rise to an interesting approach in time series analysis.

6.2 Wolfer sunspot data

Similar to the Canadian Lynx data, the Wolfer sunspot data is also a famous time series that appears in many textbooks. Here, we use data from Tong (1990) that is the annual number of recorded sunspots on the sun's surface for the period 1770-1869. We log-transformed the original data.

Table 4 shows the AIC values of several models for Wolfer sunspot data. The proposed method decomposes this data into three oscillation components. The estimated frequencies of the three oscillators are

$$\hat{f}_1 = 0, \hat{f}_2 = 0.086, \hat{f}_3 = 0.180,$$

where the frequency units are *per year*.

Fig. 4 presents the decomposition into oscillation components accomplished by the proposed method. Here, observation noise is omitted since it is in the order $10^{-}6$. Since $\hat{f}_1 = 0$, the first component is considered to be the trend component. The number of sunspots is known to oscillate with a period of approximately 11 years. The second component, which has 11.6 years period, corresponds well to this known value.

Fig. 5 shows the phases of the three oscillation components. Because the first component represents the trend, its phase remains largely unchanged. Comparatively, the second and third component phases change with time. For the third component, the phase fluctuates randomly and the credible interval of the phase is consistently wide.

6.3 Tidal data

Tidal data also show characteristic oscillatory behavior (Tamura et al., 1991). Here, we apply the proposed method to tidal data in Tokyo obtained

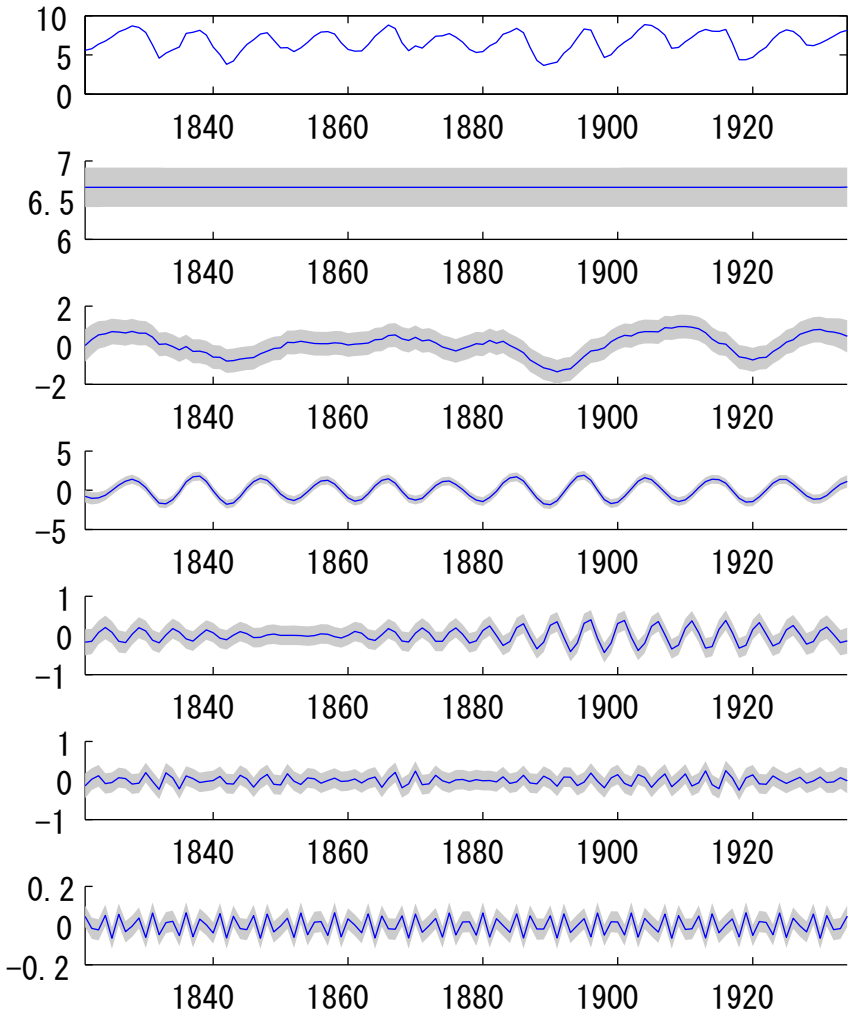

Figure 2: Decomposition of Canadian Lynx data. First panel: raw data (log-transformed Lynx data). Second-seventh panel: oscillation components with 68% credible intervals.

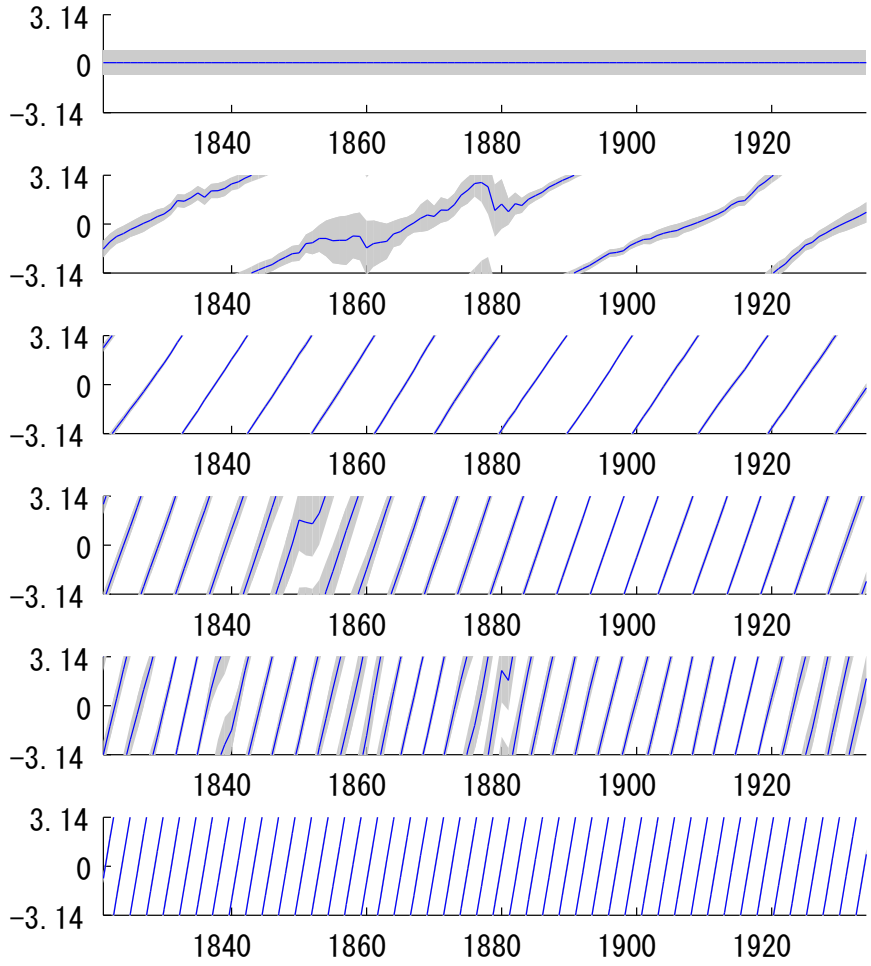

Figure 3: Estimated phases of six oscillation components in Canadian Lynx data with 68% credible intervals.

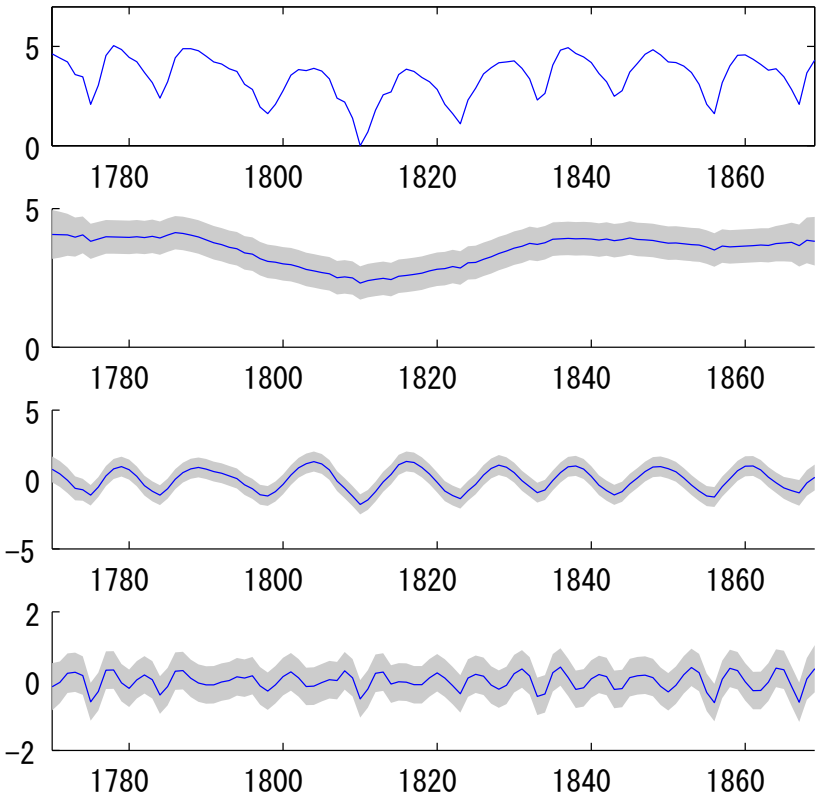

Figure 4: Decomposition of Wolfer sunspot data. First panel: raw data (log-transformed sunspot data). Second-fourth panel: oscillation components with 68% credible intervals.

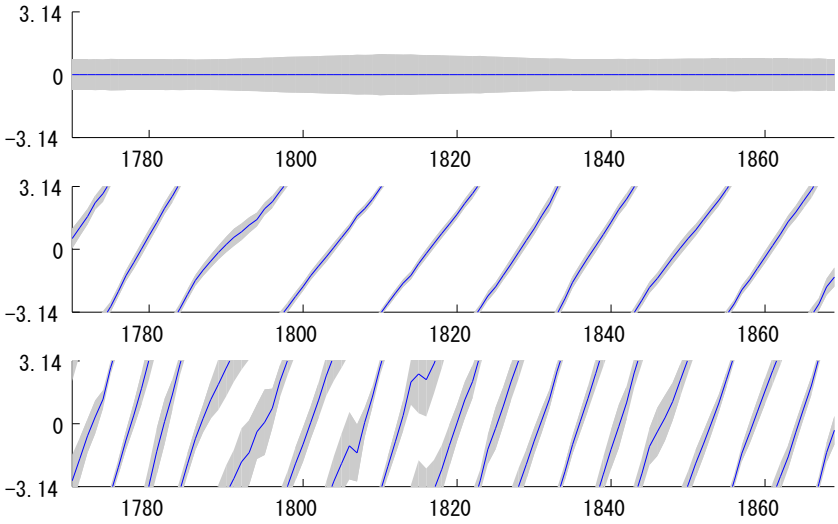

Figure 5: Estimated phases of three oscillation components in Wolfer sunspot data with 68% credible intervals

from the homepage of the Japan Meteorological Agency (`http://www.data.jma.go.jp/kaiyou/db/tide/suisan/`). We use data from January-May 2010, which are sampled every hour with no missing values.

First, we down-sampled the original data to every four hours and then applied the proposed method. The data length is 750. Fig. 6 plots the AIC values of our models (8) and (9) and the AR models with observation noise (27) and (28). Our model attains a minimum AIC at $K = 11$ and it is lower than that of the AR models. Fig. 7 shows the decomposition for the 201st-300th data points (17 days) and Fig. 8 shows the phase of each component for the 201st-250th data points (8 days). Frequencies of large-power components are

$$\hat{f}_1 = 0, \hat{f}_2 = 0.17, \hat{f}_3 = 0.39, \hat{f}_4 = 0.93, \hat{f}_7 = 1.93,$$

where the frequency units are *per day*. The phase of the fourth, fifth, sixth and seventh components develops regularly, and their credible intervals are narrow. Fig. 9 presents the fourth and seventh component for fifty days. Interestingly, the amplitudes of these components oscillate with a period of approximately 15 days.

Next, we applied the proposed method to the original tidal data, which is sampled every hour. The data length is 3000. Fig. 10 plots the AIC of our models (8) and (9) and the AR models with observation noise (27) and (28). Here, we also present the AIC of our models with MA term added (20) and (21) where $q = 2$.We can see that addition of MA terms improves

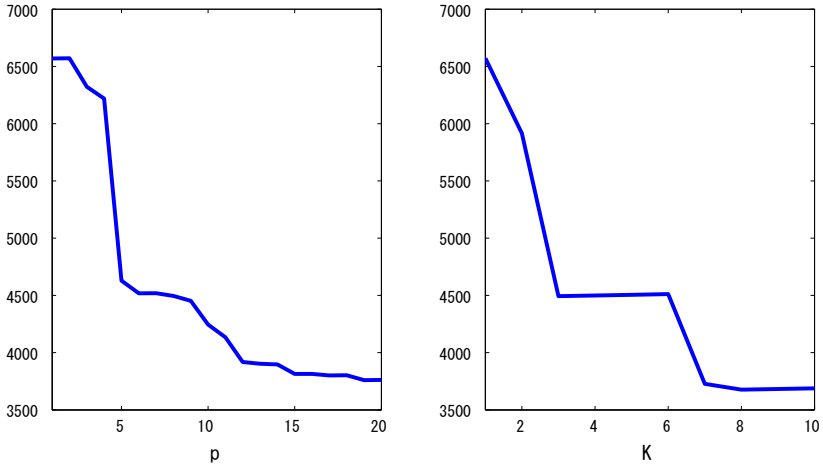

Figure 6: AIC values of several models for tidal data. Left: AR(p) models with observation noise (27) and (28). Right: our models (8) and (9) with K oscillation components.

the model fit.

6.4 Tidal data with missing values

In tidal data, measurement equipment problems sometimes result in missing values in practice. The proposed method can be applied to such data since it is based on state space models. Here, in order to assess the effects of missing values, we artificially dropped the 221st-250th data points from the down-sampled tidal data described above. We applied the proposed method to this time series.

Fig. 7 shows the decomposition for the 201st-300th data points (17 days). Here, we used the same parameter as the non-missing case. The decomposition is qualitatively the same with the non-missing case. For several components like the first and second, the conditional variance becomes large in the interval without observation. However, regularly oscillating components like the fourth and seventh are not affected much.Thus, the proposed method detects the oscillation components underlying the tidal data robustly.

6.5 Hippocampal local field potential

Finally, we apply the proposed method to rat hippocampal LFP data from Mizuseki (2009a). The sampling frequency is 125 Hz, and the data length is 250. The original paper (Mizuseki et al., 2009b) focused on the LFP theta band (5-10 Hz) of this LFP. Fig. 12 depicts this data; periodic activity in this frequency band is apparent.

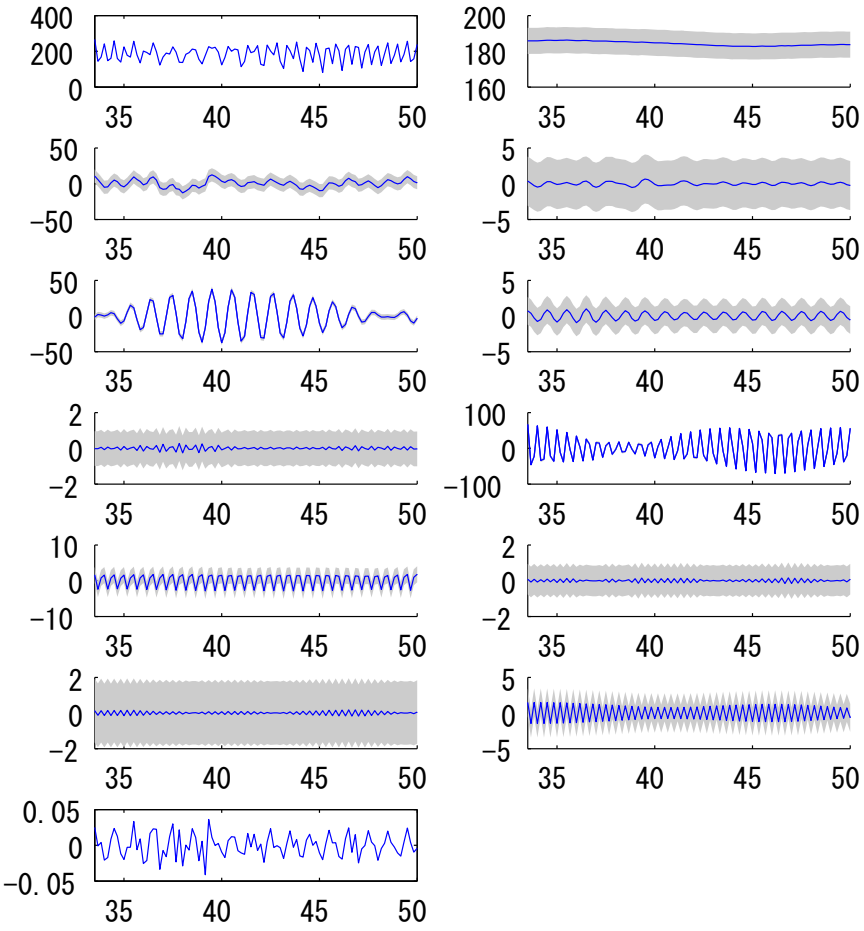

Figure 7: Decomposition of tidal data. First panel: raw data. Second-12th panel: oscillation components with 68% credible intervals. 13th panel: observation noise.

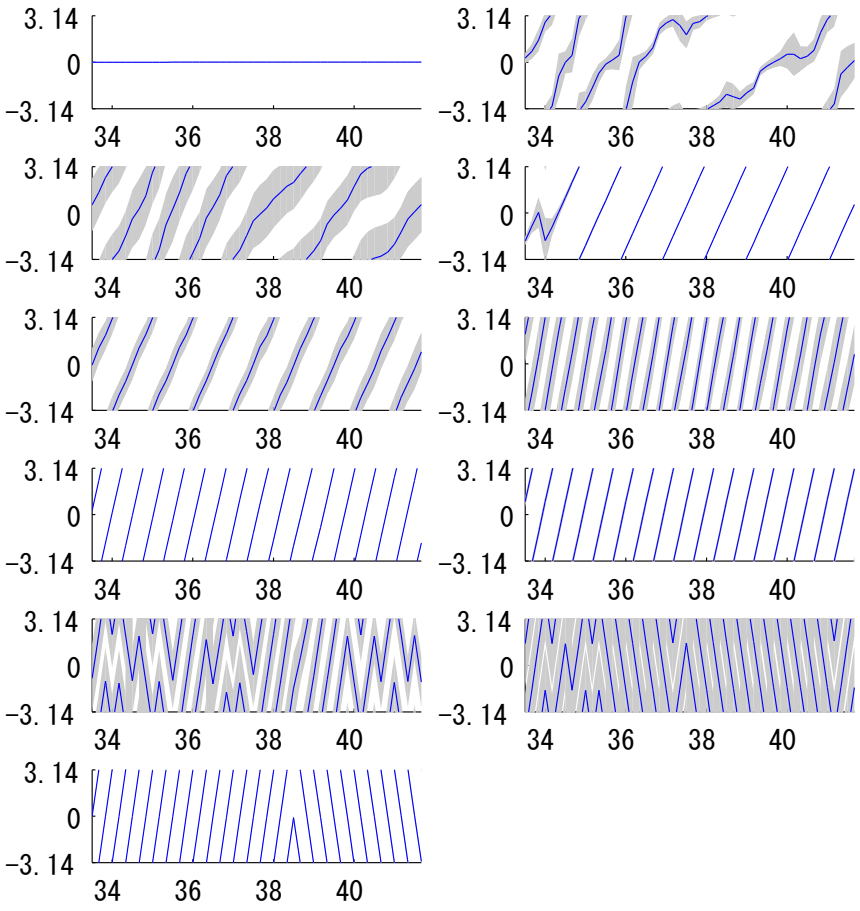

Figure 8: Estimated phases of eleven components in tidal data with 68% credible intervals.

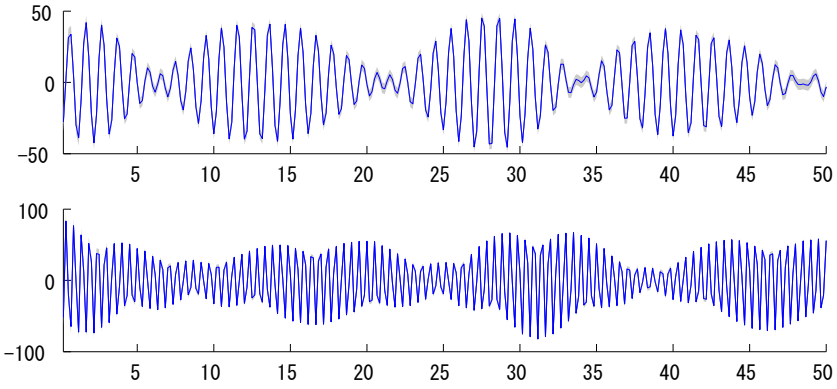

Figure 9: The fourth and seventh component in tidal data with 68% credible intervals.

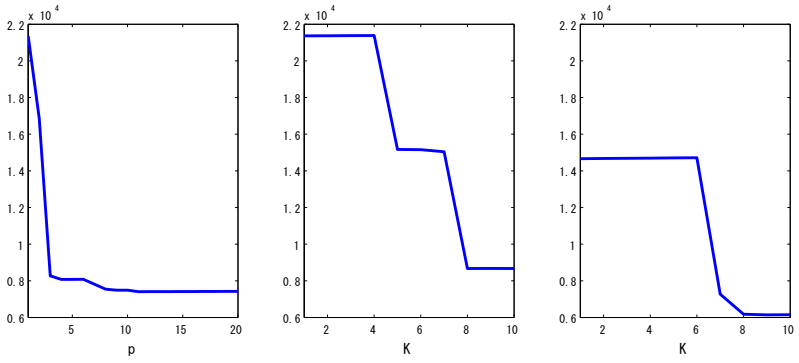

Figure 10: AIC values of several models for tidal data. Left: AR(p) models with observation noise (27) and (28). Middle: our models (8) and (9). Right: our models (20) and (21) with $q = 2$.

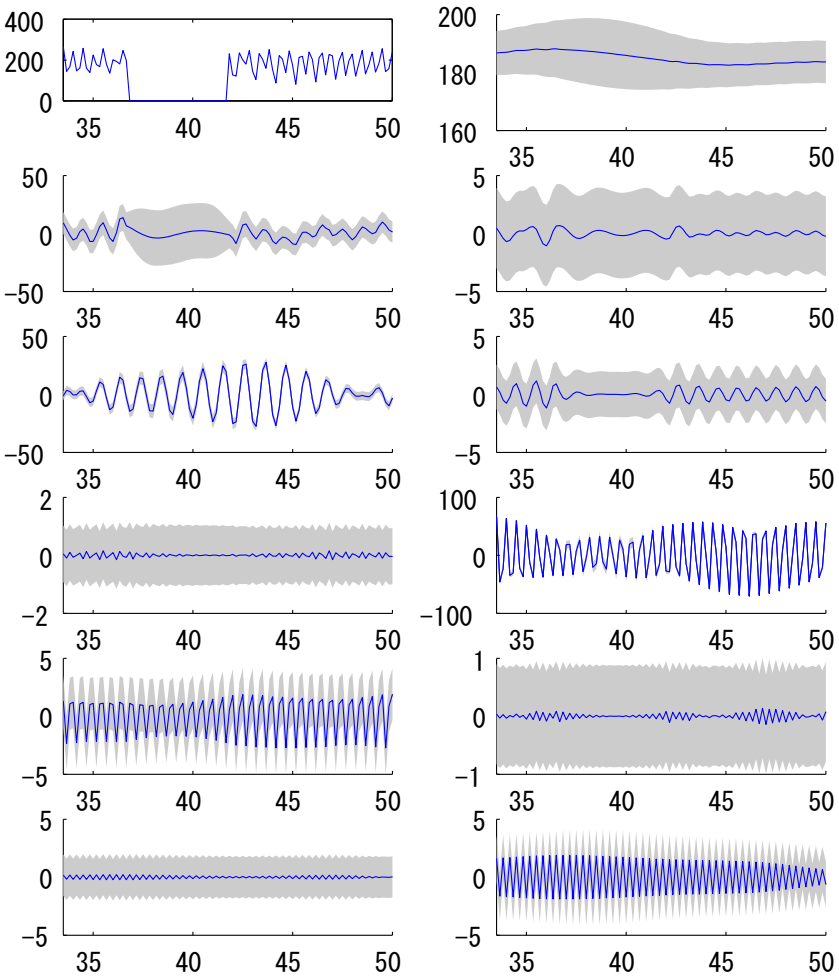

Figure 11: Decomposition of tidal data with missing values. First panel: raw data (zero value corresponds to missing). Second-12th panel: oscillation components with 68% credible intervals.

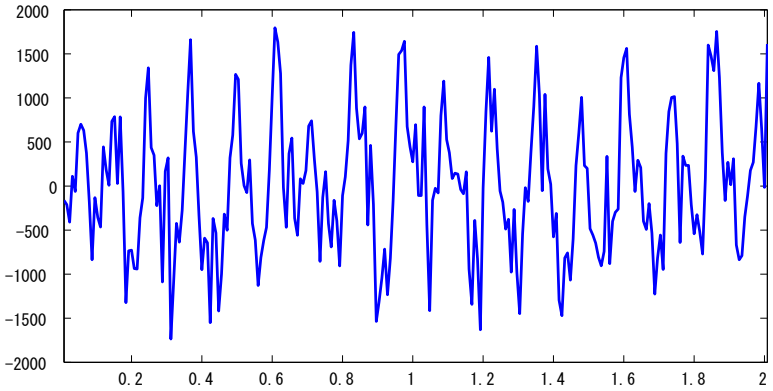

Figure 12: hippocampal LFP data

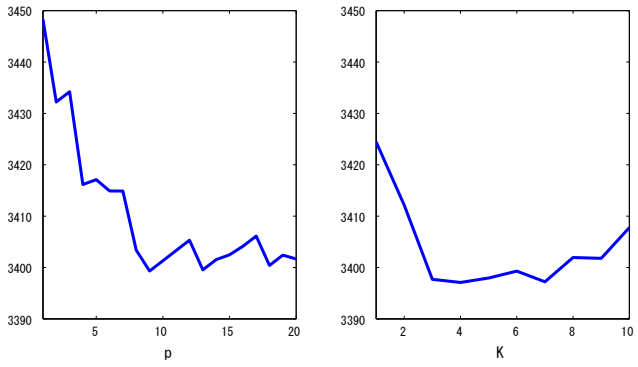

Figure 13: AIC values of several models for hippocampal LFP data. Left: AR(p) models with observation noise (27) and (28). Right: our models (8) and (9) with K oscillation components.

Fig. 13 plots the AIC values of our models (8) and (9) and the AR models with observation noise (27) and (28). Our model attains minimum AIC at $K = 4$ and it is less than the AIC of AR models. When $K = 4$, the estimated frequencies are

$$\hat{f}_1 = 6.00, \hat{f}_2 = 8.02, \hat{f}_3 = 15.96, \hat{f}_4 = 35.60,$$

where the frequency units are *per second* (Hz). Fig. 14 shows the decomposition and Fig. 15 shows the phase of each component. For the second, third, and fourth component, the phases develop regularly.

These results imply that one oscillator corresponds to the theta rhythm with a frequency $\hat{f}_2 = 8.02$ Hz. A second oscillator corresponding to the alpha rhythm likely exists with a frequency $\hat{f}_3 = 15.96$ Hz. Higher frequency oscillators are also evident.

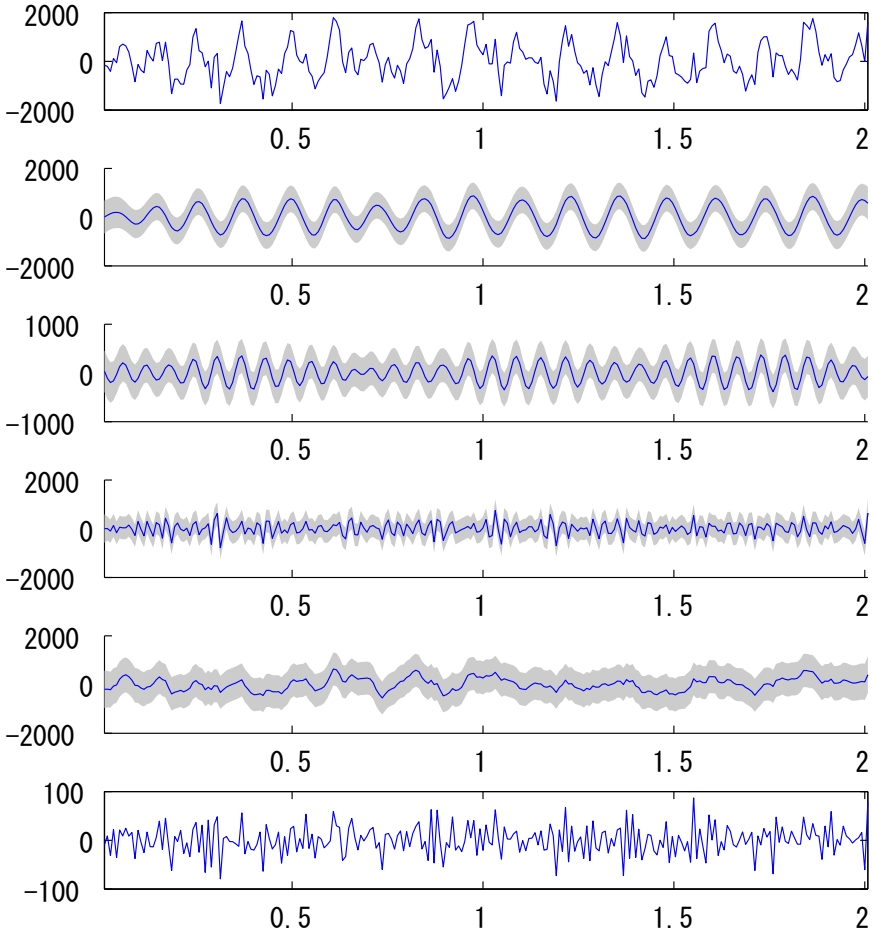

Figure 14: Decomposition of hippocampal LFP data. First panel: raw data. Second-fifth panel: oscillation components with 68% credible intervals. Sixth panel: observation noise.

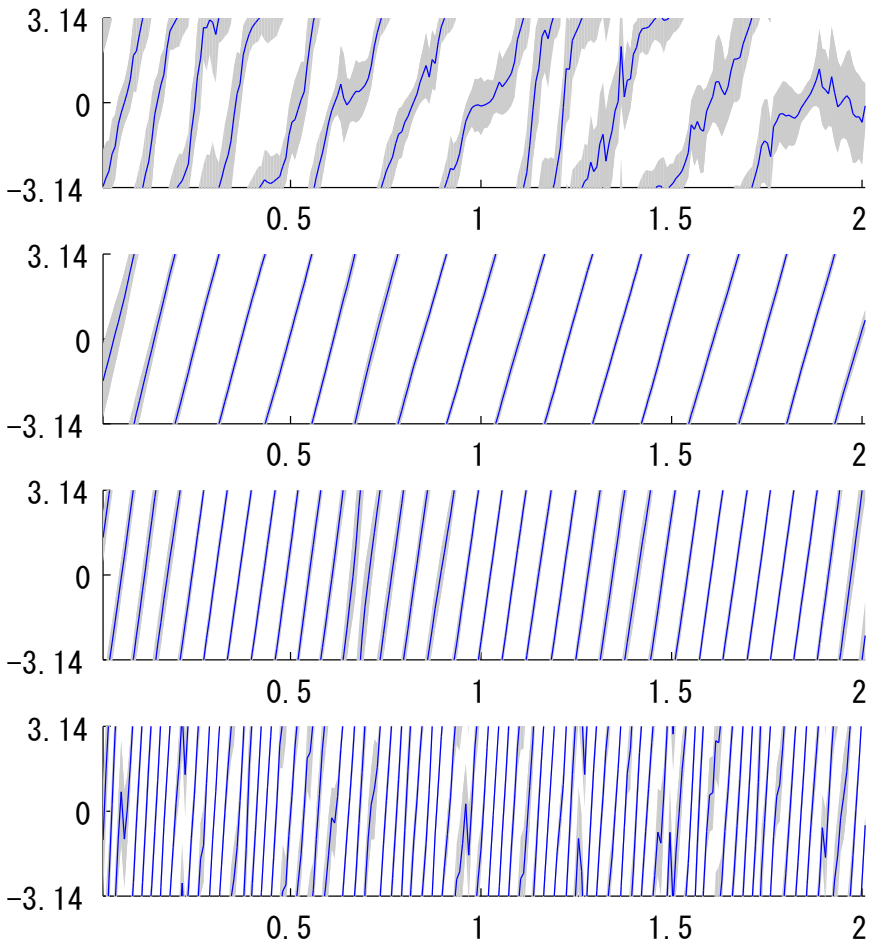

Figure 15: Estimated phases of the four oscillation components in hippocampal LFP data with 68% credible intervals.

Conventional analysis of neural time series such as LFP focuses on frequencies below 50 Hz, since the signal-to-noise ratio is lower for higher frequencies. However, the proposed method denoises the time series and decomposes it into oscillation components. Therefore, the observed high frequency oscillators may have a significant role in neural information processing. Future research may investigate the role of these oscillators, including the potential for their phases to encode internal information.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 26280005 and Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science.

References

Akaike, H. (1980) Likelihood and the Bayes procedure. In *Bayesian Statistics*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, Valencia, Spain, 1–13.

Akaike, H. & Ishiguro, M. (1980) BAYSEA: A Bayesian seasonal adjustment program. *The Institute of Statistical Mathematics: Computer Science Monographs* **13**.

Berens, P. (2009) CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software* **31**, 10.

Brockwell, P. J. & Davis, R. A. (2009) *Time Series: Theory and Methods*. New York: Springer.

Buzsáki, G. (2011) *Rhythms of the Brain*. Oxford: Oxford University Press.

Campbell, M. J. & Walker, A. M. (1977) A survey of Statistical Work on the Mackenzie River Series of Annual Canadian Lynx Trappings for the Years 1821-1934 and a New Analysis. *Journal of the Royal Statistical Society A* **140**, 411–431.

Cohen, M. X. (2014) *Analyzing neural time series data*. Cambridge: MIT Press.

Huerta, G. & West, M. (1999) CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of the Royal Statistical Society B* **61**, 881–899.

Huerta, G. & West, M. (1999) Priors and component structures in autoregressive time series models. *Journal of Time Series Analysis* **20**, 401–416.

Kitagawa, G. (2010) *Introduction to Time Series Modeling.* London: Chapman & Hall.

Kitagawa, G. & Gersch, W. (1984) A smoothness priors-state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association* **79**, 378–389.

Lopour, B. A., Tavassoli, A., Fried, I. & Ringach, D. L.. (2013) Coding of Information in the Phase of Local Field Potentials within Human Medial Temporal Lobe. *Neuron* **79**, 594–606.

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E. & Sejnowski, T. J. (2002) Dynamic brain sources of visual evoked responses. *Science* **295**, 690–694.

Mizuseki, K., Sirota, A., Pastalkova, E. & Buzsaki, G. (2009a) Multi-unit recordings from the rat hippocampus made during open field foraging. http://dx.doi.org/10.6080/K0Z60KZ9

Mizuseki, K., Sirota, A., Pastalkova, E. & Buzsaki, G. (2009b) Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron* **64**, 267–280.

Nguyen, D. P., Wilsona, M. A., Brown, E. N. & Barbieri, R. (2009) Measuring instantaneous frequency of local field potential oscillations using the Kalman smoother. *Journal of Neuroscience Methods* **184**, 365–374.

O'Keefe, J. & Recce, M. L. (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330.

Quinn, B. G. & Hannan, E. J. (2013) *The Estimation and Tracking of Frequency.* Cambridge: Cambridge University Press.

Shumway, R. H. & Stoffer, D. S. (1982) An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* **3**, 253–264.

Siapas, A. G., Lubenov, E. V. & Wilson, M. A. (2005) Prefrontal phase locking to hippocampal theta oscillations. *Neuron* **46**, 141–151.

Tamura, Y., Sato, T., Ooe, M. & Ishiguro, M.. (1991) A procedure for tidal analysis with a Bayesian information criterion. *Geophysical Journal International* **104**, 507–516.

Tong, H. (1977) Some comments on the Canadian Lynx Data. *Journal of the Royal Statistical Society A* **140**, 432–436.

Tong, H. (1990) *Nonlinear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

West, M. (1995) Bayesian inference in cyclical component dynamic linear models. *Journal of the American Statistical Association* **90**, 1301–1312.

West, M. (1997) Time series decomposition. *Biometrika* **84**, 489–494.

Whittle, P. (1953) Estimation and information in stationary time series. *Arkiv for Matematik* **2**, 423–434.

Wiener, N. (1966) *Nonlinear problems in random theory*. Cambridge: MIT Press.

A Estimation of AR models with observation noise

Before fitting our model (8) and (9) to given time series, we need to fit the AR models with observation noise

$$x_t = \sum_{k=1}^{p} a_k x_{t-k} + v_t, v_t \sim \mathrm{N}(0, Q), \tag{27}$$

$$y_t = x_t + w_t, w_t \sim \mathrm{N}(0, R) \tag{28}$$

to given time series, in order to determine the initial values of the parameters in (8) and (9) for numerical optimization as described in Appendix B. Here, we explain the method for fitting the AR models with observation noise (27) and (28). Let

$$\hat{C}_k = \frac{1}{N} \sum_{t=k+1}^{N} y_t y_{t-k}$$

be the sample autocovariance function of the given time series y_t.

First, we calculate the minimum eigenvalue $\overline{R} > 0$ of the Toeplitz matrix

$$\begin{pmatrix} \hat{C}_0 & \hat{C}_1 & \hat{C}_2 & \cdots & \hat{C}_p \\ \hat{C}_1 & \hat{C}_0 & \hat{C}_1 & \cdots & \hat{C}_{p-1} \\ \hat{C}_2 & \hat{C}_1 & \hat{C}_0 & \cdots & \hat{C}_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{C}_p & \hat{C}_{p-1} & \hat{C}_{p-2} & \cdots & \hat{C}_0 \end{pmatrix}.$$

Then, we define a function $g : (0, \overline{R}) \to \mathbb{R}$ as follows. For $\widetilde{R} \in (0, \overline{R})$, we

solve the Yule-Walker type equation

$$\begin{pmatrix} \hat{C}_0 - \widetilde{R} & \hat{C}_1 & \hat{C}_2 & \cdots & \hat{C}_{p-1} \\ \hat{C}_1 & \hat{C}_0 - \widetilde{R} & \hat{C}_1 & \cdots & \hat{C}_{p-2} \\ \hat{C}_2 & \hat{C}_1 & \hat{C}_0 - \widetilde{R} & \cdots & \hat{C}_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{C}_{p-1} & \hat{C}_{p-2} & \hat{C}_{p-3} & \cdots & \hat{C}_0 - \widetilde{R} \end{pmatrix} \begin{pmatrix} a_1(R) \\ a_2(R) \\ a_3(R) \\ \vdots \\ a_p(R) \end{pmatrix} = \begin{pmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \hat{C}_3 \\ \vdots \\ \hat{C}_p \end{pmatrix},$$

calculate the log-likelihood of the model (27) and (28) with parameter

$$a_1 = a_1(\widetilde{R}), \cdots, a_p = a_p(\widetilde{R}), Q = \hat{C}_0 - \sum_{k=1}^{p} a_k(\widetilde{R})\hat{C}_k, R = \widetilde{R}$$

using Kalman filter (Kitagawa, 2010) and denote it as $g(\widetilde{R})$. Note that the stationarity of the AR process with $a_1 = a_1(\widetilde{R}), \cdots, a_p = a_p(\widetilde{R})$ and the positivity of $\hat{C}_0 - \sum_{k=1}^{p} a_k(\widetilde{R})$ is guaranteed since $\widetilde{R} \in (0, \overline{R})$. We maximize $g(\widetilde{R})$ by numerical optimization. Let R_0 be the solution.

Next, we maximize the log-likelihood of the model (27) and (28) with respect to $a_1, \cdots, a_p, Q, R$ by numerical optimization. The initial value is set to

$$a_1 = a_1(R_0), \cdots, a_p = a_p(R_0), Q = \hat{C}_0 - \sum_{k=1}^{p} a_k(R_0)\hat{C}_k, R = R_0.$$

In numerical optimization, we transform the AR coefficients $a_1, \cdots, a_p$ to the partial autocorrelation coefficients (PARCOR) $c_1, \cdots, c_p$ to enforce the stationarity (Kitagawa, 2010).

Thus, we fit the AR models with observation noise (27) and (28) to the given time series.

B Initial values of parameters in model fitting

Since the log-likelihood (26) is not concave in general, in the numerical optimization, initial values for $f_1, \cdots, f_K$, $a_1, \cdots, a_K$, $\widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2$ must be carefully chosen. We utilize the interpretation of AR process as a superposition of oscillation components discussed in Section 3.4.

Suppose we want to fit the model (8) and (9) with K oscillation components to given time series y_t. First, for each $p = K, \cdots, 2K$, we fit the AR(p) model with observation noise (27) and (28) to y_t using the method described in Appendix A. Let $\xi_1^{(p)}, \cdots, \xi_p^{(p)}$ be the characteristic roots of the fitted AR(p) model (27). Here, $\xi_2^{(p)} = \bar{\xi}_1^{(p)}, \cdots, \xi_{2m_p}^{(p)} = \bar{\xi}_{2m_p-1}^{(p)}$ are imaginary numbers and $\xi_{2m_p+1}^{(p)}, \cdots, \xi_p^{(p)}$ are real numbers. Therefore, from Section 3.4,

the fitted AR(p) model has $p-m_p$ oscillators. Note that $|\xi_1^{(p)}|, \cdots, |\xi_p^{(p)}|$ are all smaller than one since the fitted AR(p) model is stationary. Among the fitted AR(p) models with $p - m_p = K$, let AR(p_*) be the model with the minimum AIC. We put $\xi_1 = \xi_1^{(p_*)}, \xi_2 = \xi_3^{(p_*)}, \cdots, \xi_{m_p} = \xi_{2m_{p_*}-1}^{(p_*)}, \xi_{m_{p_*}+1} = \xi_{2m_{p_*}+1}^{(p_*)}, \cdots, \xi_K = \xi_{p_*}^{(p_*)}$. If there are no model with $p - m_p = K$, then let AR(p_*) be the model with the minimum AIC among those with $p-m_p > K$. In this case, among $\xi_k^{(p_*)}$ with nonnegative imaginary parts, we set ξ_i to be that with the i-th maximum absolute value. In both cases, let the estimated variance R of the observation noise in the model (27) and (28) be R_*.

Then, the initial values for $f_1, \cdots, f_K$ and $a_1, \cdots, a_K$ are set to

$$2\pi f_1 \Delta t = \arg \xi_1, \cdots, 2\pi f_K \Delta t = \arg \xi_K,$$

$$a_1 = |\xi_1|, \cdots, a_K = |\xi_K|.$$

Thus, the location and degree of sharpness of peaks in the spectrum are set to be the same as the best AR model. Also, the initial values for $\widetilde{\sigma}_1^2, \cdots, \widetilde{\sigma}_K^2$ are set to

$$\widetilde{\sigma}_1^2 = \frac{\sigma_1^2}{R_*}, \cdots, \widetilde{\sigma}_K^2 = \frac{\sigma_K^2}{R_*},$$

where $\sigma_1^2, \cdots, \sigma_K^2$ are calculated from the linear equation

$$\begin{pmatrix} \widetilde{p}_1(\arg \xi_1) & \widetilde{p}_2(\arg \xi_1) & \cdots & \widetilde{p}_K(\arg \xi_1) \\ \widetilde{p}_1(\arg \xi_2) & \widetilde{p}_2(\arg \xi_2) & \cdots & \widetilde{p}_K(\arg \xi_2) \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{p}_1(\arg \xi_K) & \widetilde{p}_2(\arg \xi_K) & \cdots & \widetilde{p}_K(\arg \xi_K) \end{pmatrix} \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_K^2 \end{pmatrix} = \begin{pmatrix} I(\arg \xi_1) \\ I(\arg \xi_2) \\ \vdots \\ I(\arg \xi_K) \end{pmatrix}.$$

Here,

$$\widetilde{p}_k(\lambda) = \frac{-a_k \cos(2\pi f_k \Delta t)}{2\pi b_k} \frac{|1 + b_k \exp(-i\lambda)|^2}{|1 - 2a_k \cos(2\pi f_k \Delta t) \exp(-i\lambda) + a_k^2 \exp(-2i\lambda)|^2}$$

is the spectrum (13) of the ARMA(2,1) model (11) with $\sigma^2 = 1$ and

$$I(\lambda) = \frac{1}{N} \left| \sum_{t=1}^{N} y_t \exp(-it\lambda) \right|^2$$

is the sample spectrum. Thus, values of the power spectrum on $f_1, \cdots, f_K$ are made to match the sample spectrum on $f_1, \cdots, f_K$. This method is motivated from the theory of Whittle likelihood for Gaussian time series (Whittle, 1953).

For the model with MA term added (20) and (21), the initial values for $f_1, \cdots, f_K$ and $a_1, \cdots, a_K$ are set in the above way and the initial values for $b_{k,1}, \cdots, b_{k,q}$ $(k = 1, \cdots, K)$ are set to zero.