International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:8, No:2, 2014

Gaussian Process Model Identification Using Artificial Bee Colony Algorithm and Its Application to Modeling of Power Systems

Tomohiro Hachino, Hitoshi Takata, Shigeru Nakayama, Ichiro Iimura, Seiji Fukushima, Yasutaka Igarashi

Abstract—This paper presents a nonparametric identification of continuous-time nonlinear systems by using a Gaussian process (GP) model. The GP prior model is trained by artificial bee colony algorithm. The nonlinear function of the objective system is estimated as the predictive mean function of the GP, and the confidence measure of the estimated nonlinear function is given by the predictive covariance of the GP. The proposed identification method is applied to modeling of a simplified electric power system. Simulation results are shown to demonstrate the effectiveness of the proposed method.

Keywords—Artificial bee colony algorithm, Gaussian process model, identification, nonlinear system, electric power system.

I. Introduction

PRACTICAL systems such as electric power systems are essentially continuous-time nonlinear systems. Development of accurate identification algorithm for such systems is indispensable for precise analysis or control design. Identification based on the continuous-time model has received a little attention owing to difficulty of handling the higher-order derivatives of input and output data. For this approach, identification methods based on neural network model [1], orthogonal least-squares (LS) estimator [2], radial basis function model [3], [4], and automatic choosing function model [5] have been reported. Since these methods are categorized into the parametric identification, one needs many weighting parameters of any basis functions to describe the nonlinearity. Moreover, any confidence measures for the estimated nonlinear functions are not given in such identification methods.

In recent years, the Gaussian process (GP) model has been introduced for the modeling of the nonlinear dynamic systems [6], [7] and the prediction in time series analysis [8], [9]. The GP model was originally utilized for the regression problem by O' Hagan [10] and has recently much attention for regression or classification problem [11]–[13]. Some applications using the GP model have been reported for human motion modeling [14], or predictive control of gas-liquid separation plant [15]. The GP model is a non-parametric model and fits naturally into Bayesian framework. Since it has fewer parameters than parametric models such as the neural network model, we can describe the nonlinearity of the objective system in a few parameters. Moreover, the GP gives us not only the mean function but also the covariance function. Therefore, in this paper, we propose a nonparametric identification of continuous-time nonlinear systems using the GP model.

The hyperparameters included in the GP prior model have to be appropriately determined based on the identification data. Generally this training becomes nonlinear optimization problem. In this paper, the separable LS approach combining the linear LS method with artificial bee colony (ABC) algorithm is presented for this training. ABC algorithm is an optimization algorithm inspired by an intelligent behavior of honeybee swarms and has high potential for both global and local optimizations [16]. This algorithm consists of search by the three types of bees; the employed bees, the onlooker bees, and the scout bees. ABC algorithm consists of only the basic arithmetic operations and does not require complicated coding and genetic operations such as crossovers and mutations for the genetic algorithm. Moreover, the performance of ABC algorithm is better than or similar to those of other population-based algorithms in spite of a few setting parameters [16], [17]. These advantages suggest that the use of ABC algorithm increases efficiency when the GP prior model for identification is trained.

This paper is organized as follows. In Section II the problem is formulated. In Section III the GP prior model for the identification is derived. In Section IV the training method of the GP prior model is presented using ABC algorithm, and the nonlinear function with the confidence measure is estimated from the GP posterior distribution. In Section V numerical simulation for a simplified electric power system is carried out to illustrate the effectiveness of the proposed identification method. Finally some conclusions are given in Section VI.

Tomohiro Hachino is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (phone and fax: +81-99-285-8392; e-mail: hachino@eee.kagoshima-u.ac.jp).
Hitoshi Takata is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: takata@eee.kagoshima-u.ac.jp).
Shigeru Nakayama is with the Department of Information Science and Biomedical Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: shignaka@ibe.kagoshima-u.ac.jp).
Ichiro Iimura is with the Faculty of Administration, Prefectural University of Kumamoto, Kumamoto, 862-8502 Japan (e-mail: iiimura@pu-kumamoto.ac.jp).
Seiji Fukushima is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: fukushima@eee.kagoshima-u.ac.jp).
Yasutaka Igarashi is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: igarashi@eee.kagoshima-u.ac.jp).

II. Statement of the Problem

Consider a single-input, single-output, continuous-time nonlinear system described by

$$\sum_{\substack{i=0\\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i}x(t) = f(\boldsymbol{z}(t)) + \sum_{\substack{j=0\\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j}u(t) \tag{1}$$
$$(a_0 = 1, \quad n \geq m)$$
$$\boldsymbol{z}(t) = \left[p^{n-n_1}x(t), p^{n-n_2}x(t), \cdots, p^{n-n_\alpha}x(t), p^{m-m_1}u(t), p^{m-m_2}u(t), \cdots, p^{m-m_\beta}u(t)\right]^{\mathrm{T}}$$
$$y(t) = x(t) + e(t)$$

where $u(t)$ and $x(t)$ are the true input and output signals, respectively. $y(t)$ is the noisy output that is corrupted by the measurement noise $e(t)$. $f(\cdot)$ is an unknown nonlinear function, which is assumed to be stationary and smooth. p denotes the differential operator. n, n_i $(i = 1, 2, \cdots, \alpha)$, m and m_j $(j = 1, 2, \cdots, \beta)$ are assumed to be known. The purpose of this paper is to identify the parameters $\{a_i\}$ and $\{b_j\}$ of the linear terms and the nonlinear function $f(\cdot)$ with the confidence measure, from the true input and noisy output data in the GP framework.

III. GP Prior model for Identification

Equation (1) can be rewritten as

$$p^n y(t) = f(\boldsymbol{w}(t)) - \sum_{\substack{i=1\\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i}y(t) + \sum_{\substack{j=0\\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j}u(t) + \varepsilon(t)$$
$$\boldsymbol{w}(t) = [p^{n-n_1}y(t), p^{n-n_2}y(t), \cdots, p^{n-n_\alpha}y(t), p^{m-m_1}u(t), p^{m-m_2}u(t), \cdots, p^{m-m_\beta}u(t)]^{\mathrm{T}} \tag{2}$$

where $\varepsilon(t)$ is an error caused by the measurement noise $e(t)$.

The following state variable filter $F(p)$ is introduced in order to evaluate higher-order derivatives of the signals:

$$F(p) = \frac{1}{p^q + \gamma_1 p^{q-1} + \cdots\cdots + \gamma_q} \quad (q > n) \tag{3}$$

Multiplying both sides of (2) by $F(p)$ yields

$$p^n y^f(t) = F(p)f(\boldsymbol{w}(t)) - \sum_{\substack{i=1\\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i}y^f(t) + \sum_{\substack{j=0\\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j}u^f(t) + \varepsilon^f(t) \tag{4}$$

where $u^f(t) = F(p)u(t)$ and $y^f(t) = F(p)y(t)$ are the filtered input and output signals, respectively. When $F(p)$ has a transport lag characteristic, the filter $F(p)$ and the nonlinear function $f(\cdot)$ are interchangeable [2], and it follows that $F(p)f(\boldsymbol{w}(t)) = f(F(p)\boldsymbol{w}(t)) = f(\boldsymbol{w}^f(t))$. Therefore, (4) becomes

$$p^n y^f(t) = f(\boldsymbol{w}^f(t)) - \sum_{\substack{i=1\\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i}y^f(t) + \sum_{\substack{j=0\\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j}u^f(t) + \varepsilon^f(t) \tag{5}$$

where $\varepsilon^f(t)$ is assumed to be zero mean Gaussian noise with variance σ_n^2.

Putting $t = t_1, t_2, \cdots, t_N$ into (5) yields

$$\boldsymbol{y} = \boldsymbol{v} + \boldsymbol{G}\boldsymbol{\theta}_l \tag{6}$$

where

$$\boldsymbol{y} = [p^n y^f(t_1), p^n y^f(t_2), \cdots, p^n y^f(t_N)]^{\mathrm{T}}$$
$$\boldsymbol{v} = [f(\boldsymbol{w}^f(t_1)) + \varepsilon^f(t_1), f(\boldsymbol{w}^f(t_2)) + \varepsilon^f(t_2), \cdots, f(\boldsymbol{w}^f(t_N)) + \varepsilon^f(t_N)]^{\mathrm{T}}$$
$$\boldsymbol{\theta}_l = [a_1, \cdots, a_i, \cdots, a_n, b_0, \cdots, b_j, \cdots, b_m]^{\mathrm{T}} \tag{7}$$
$$\boldsymbol{G} = [\boldsymbol{g}(t_1), \boldsymbol{g}(t_2), \cdots, \boldsymbol{g}(t_N)]^{\mathrm{T}}$$
$$\boldsymbol{g}(t) = [-p^{n-1}y^f(t), \cdots, -p^{n-i}y^f(t), \cdots, -y^f(t), p^m u^f(t), \cdots, p^{m-j}u^f(t), \cdots, u^f(t)]^{\mathrm{T}}$$

A GP is a Gaussian random function and is completely described by its mean function and covariance function. We can regard it as a collection of random variables with a joint multivariable Gaussian distribution. Therefore, the function values $\boldsymbol{f}$ can be represented by the GP:

$$\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}), \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w})) \tag{8}$$

where

$$\boldsymbol{f} = [f(\boldsymbol{w}^f(t_1)), f(\boldsymbol{w}^f(t_2)), \cdots, f(\boldsymbol{w}^f(t_N))]^{\mathrm{T}}$$
$$\boldsymbol{w} = [\boldsymbol{w}^f(t_1), \boldsymbol{w}^f(t_2), \cdots, \boldsymbol{w}^f(t_N)] \tag{9}$$

$\boldsymbol{w}$ is the input of the function $\boldsymbol{f}$, $\boldsymbol{m}(\boldsymbol{w})$ is the mean function vector, and $\boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w})$ is the covariance matrix. The mean function is often represented by a polynomial regression [13]. In this paper, the mean function is expressed by the first order polynomial, i.e., a linear combination of the input variable:

$$m(\boldsymbol{w}^f(t)) = (\boldsymbol{w}^f(t))^{\mathrm{T}}\boldsymbol{\theta}_m$$
$$\boldsymbol{\theta}_m = \left[\theta_{n_1}, \theta_{n_2}, \cdots, \theta_{n_\alpha}, \theta_{m_1}, \theta_{m_2}, \cdots, \theta_{m_\beta}\right]^{\mathrm{T}} \tag{10}$$

where $\boldsymbol{\theta}_m$ is the unknown parameter vector for the mean function. Thus, the mean function vector $\boldsymbol{m}(\boldsymbol{w})$ is described as follows:

$$\boldsymbol{m}(\boldsymbol{w}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_m \tag{11}$$

The covariance $\Sigma_{pq} = s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q))$ is an element of the covariance matrix $\boldsymbol{\Sigma}$, which is a function of $\boldsymbol{w}^f(t_p)$ and $\boldsymbol{w}^f(t_q)$. Under the assumption that the nonlinear function is stationary and smooth, the following Gaussian kernel is utilized in this paper:

$$\Sigma_{pq} = s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = \sigma_y^2 \exp\left(-\frac{||\boldsymbol{w}^f(t_p) - \boldsymbol{w}^f(t_q)||^2}{2\ell^2}\right) \tag{12}$$

where $||\cdot||$ denotes the Euclidean norm. Equation (12) means that the covariance of the outputs of the nonlinear function depends only on the distance between the inputs $\boldsymbol{w}^f(t_p)$ and $\boldsymbol{w}^f(t_q)$. A high correlation between the outputs of the nonlinear function occurs for the inputs that are close to each other. The overall variance of the random function can be controlled by σ_y, and the characteristic length scale of the process can be changed by ℓ.

From (8), the vector $\boldsymbol{v}$ of the noisy function values in (6) can be written as

$$\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}), \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})) \tag{13}$$

where

$$\begin{aligned} \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w}) &= \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}) + \sigma_n^2 \boldsymbol{I}_N \\ \boldsymbol{I}_N &: N \times N \ \ \text{identity} \ \ \text{matrix} \end{aligned} \tag{14}$$

and $\boldsymbol{\theta}_c = [\sigma_y, \ell, \sigma_n]^{\mathrm{T}}$ is called the *hyperparameter* vector. From (6) and (13), the GP model for the identification is derived as

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}) + \boldsymbol{G}\boldsymbol{\theta}_l, \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})) \tag{15}$$

In the following, $\boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})$ is written as $\boldsymbol{K}$ for simplicity.

IV. Identification

A. Training of GP Prior Model by ABC Algorithm

At the first stage of the identification, the GP prior model is trained by optimizing the unknown parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_m^{\mathrm{T}}, \boldsymbol{\theta}_l^{\mathrm{T}}, \boldsymbol{\theta}_c^{\mathrm{T}}]^{\mathrm{T}}$. This training is carried out by minimizing the negative log marginal likelihood of the identification data:

$$\begin{aligned} J &= -\log p(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{G}, \boldsymbol{\theta}) \\ &= \frac{1}{2}\log|\boldsymbol{K}| + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml})^{\mathrm{T}}\boldsymbol{K}^{-1}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml}) \\ &\quad + \frac{N}{2}\log(2\pi) \end{aligned} \tag{16}$$

where

$$\begin{aligned} \boldsymbol{Z} &= [\ \ \boldsymbol{w}^{\mathrm{T}} \ \vdots \ \boldsymbol{G} \ \ ] \\ \boldsymbol{\theta}_{ml} &= [\boldsymbol{\theta}_m^{\mathrm{T}}, \boldsymbol{\theta}_l^{\mathrm{T}}]^{\mathrm{T}} \end{aligned} \tag{17}$$

Although this problem is a nonlinear optimization one, we can separate the linear optimization part and the nonlinear optimization part. The partial derivative of (16) with respect to the parameter vector $\boldsymbol{\theta}_{ml}$ is as follows:

$$\frac{\partial J}{\partial \boldsymbol{\theta}_{ml}} = -\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{y} + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{Z}\boldsymbol{\theta}_{ml} \tag{18}$$

Note that if the candidates of the hyperparameter vector $\boldsymbol{\theta}_c$ of the covariance function are given, the candidates of the covariance matrix $\boldsymbol{K}$ can be constructed. Therefore, the parameter vector $\boldsymbol{\theta}_{ml}$ can be estimated by the linear LS method from (18):

$$\boldsymbol{\theta}_{ml} = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{y} \tag{19}$$

However even if the parameter vector $\boldsymbol{\theta}_{ml}$ is known, the optimization with respect to $\boldsymbol{\theta}_c$ is a complicated nonlinear problem and might suffer from the local minima problem. Therefore, in this paper, we propose a method that combines the linear LS method with ABC algorithm. Only $\boldsymbol{\Omega} = \boldsymbol{\theta}_c = [\sigma_y, \ell, \sigma_n]^{\mathrm{T}}$ is represented with the positions of the food sources and searched by ABC algorithm. The detailed training algorithm is as follows:

Step 1: Initialization

(1-1) Generate an initial population of N_s bees with random positions of the food sources $\boldsymbol{\Omega}_{[i]}$ $(i = 1, 2, \cdots, N_s)$ from (20):

$$\begin{aligned} \Omega_{ij} = \Omega_{min,j} + rand[0,1] \cdot (\Omega_{max,j} - \Omega_{min,j}) \\ (j = 1, 2, 3) \end{aligned} \tag{20}$$

where N_s denotes the size of the employed bees or onlooker bees and Ω_{ij} is the jth element of the vector $\boldsymbol{\Omega}_{[i]}$. $\Omega_{min,j}$ and $\Omega_{max,j}$ are the minimum and maximum values for Ω_{ij}, respectively. $rand[0,1]$ is uniformly distributed random number with amplitude in the range $[0, 1]$.

(1-2) Set the iteration counter l to 1.

(1-3) Set the counter for abandonment $trial_i$ to 0. The counter $trial_i$ shows the number of times that the solution $\boldsymbol{\Omega}_{[i]}$ is not improved by the employed and onlooker bees.

Step 2: Construction of the covariance matrix

Construct N_s candidates of the covariance matrix $\boldsymbol{K}_{[i]}$ using $\boldsymbol{\Omega}_{[i]}$ $(i = 1, 2, \cdots, N_s)$.

Step 3: Estimation of $\boldsymbol{\theta}_{ml}$

Estimate N_s candidates for $\boldsymbol{\theta}_{ml[i]}$ $(i = 1, 2, \cdots, N_s)$ from (19):

$$\boldsymbol{\theta}_{ml[i]} = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}_{[i]}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}_{[i]}^{-1}\boldsymbol{y} \tag{21}$$

Step 4: Fitness value calculation

Calculate the negative log marginal likelihood of the identification data:

$$\begin{aligned} J_i(\boldsymbol{\Omega}_{[i]}) &= \frac{1}{2}\log|\boldsymbol{K}_{[i]}| + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml[i]})^{\mathrm{T}}\boldsymbol{K}_{[i]}^{-1} \\ &\quad \times (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml[i]}) + \frac{N}{2}\log(2\pi) \end{aligned} \tag{22}$$

and the fitness value $F_i(\boldsymbol{\Omega}_{[i]}) = \exp(-J_i(\boldsymbol{\Omega}_{[i]}))$.

Step 5: Search by the employed bees

(5-1) Determine the new positions of the food sources $\boldsymbol{V}_{[i]} = \boldsymbol{\vartheta}_{c[i]}$ around $\boldsymbol{\Omega}_{[i]}$ for the employed bees from (23):

$$\begin{aligned} V_{ij} = \Omega_{ij} + rand[-1,1] \cdot (\Omega_{ij} - \Omega_{kj}) \\ (j = 1, 2, 3) \end{aligned} \tag{23}$$

where V_{ij} is the jth element of the vector $\boldsymbol{V}_{[i]}$ and k is a random integer selected from $\{1, 2, \cdots, N_s\}$, where $k \neq i$.

(5-2) Construct N_s candidates of the covariance matrix $\boldsymbol{\mathcal{K}}_{[i]}$ using $\boldsymbol{V}_{[i]}$ $(i = 1, 2, \cdots, N_s)$.

(5-3) Estimate N_s candidates for $\boldsymbol{\vartheta}_{ml[i]}$ $(i = 1, 2, \cdots, N_s)$ from (19).

(5-4) Calculate the objective function value:

$$\begin{aligned} J_i(\boldsymbol{V}_{[i]}) &= \frac{1}{2}\log|\boldsymbol{\mathcal{K}}_{[i]}| + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\vartheta}_{ml[i]})^{\mathrm{T}}\boldsymbol{\mathcal{K}}_{[i]}^{-1} \\ &\quad \times (\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\vartheta}_{ml[i]}) + \frac{N}{2}\log(2\pi) \end{aligned} \tag{24}$$

and the fitness value $F_i(\boldsymbol{V}_{[i]}) = \exp(-J_i(\boldsymbol{V}_{[i]}))$.

(5-5) If $F_i(\boldsymbol{\Omega}_{[i]}) < F_i(\boldsymbol{V}_{[i]})$, update $\boldsymbol{\Omega}_{[i]}$, $\boldsymbol{\theta}_{ml[i]}$ and $F_i(\boldsymbol{\Omega}_{[i]})$ by $\boldsymbol{V}_{[i]}$, $\boldsymbol{\vartheta}_{ml[i]}$ and $F_i(\boldsymbol{V}_{[i]})$, respectively, and set $trial_i = 0$.

Otherwise set $trial_i = trial_i + 1$. This procedure is called "greedy selection".

Step 6: Search by the onlooker bees

(6-1) Choose one position of the food source for each onlooker bee from $\boldsymbol{\Omega}_{[i]}$ $(i = 1, 2, \cdots, N_s)$ through "roulette-wheel" slots weighted in proportion to the fitness value of the employed bee. Namely each onlooker bee selects one position of the food source with probability of $F_i(\boldsymbol{\Omega}_{[i]})/\sum_{p=1}^{N_s} F_p(\boldsymbol{\Omega}_{[p]})$.

(6-2) Calculate the new positions of the food sources $\boldsymbol{V}_{[i]}$ corresponding to the selected positions $\boldsymbol{\Omega}_i$ from (23).

(6-3) Construct N_s candidates of the covariance matrix $\boldsymbol{\mathcal{K}}_{[i]}$ using $\boldsymbol{V}_{[i]}$ $(i = 1, 2, \cdots, N_s)$.

(6-4) Estimate N_s candidates for $\boldsymbol{\vartheta}_{ml[i]}$ $(i = 1, 2, \cdots, N_s)$ from (19).

(6-5) Calculate the fitness value $F_i(\boldsymbol{V}_{[i]}) = \exp(-J_i(\boldsymbol{V}_{[i]})$ from (24).

(6-6) Carry out the greedy selection with the same way of step 5 (5-5).

Step 7: Search by the scout bees

If the counter for abandonment $trial_i$ is greater or equal to the prespecified number *limit*, carry out the following procedure.

(7-1) Differentiate the corresponding employed bee into the scout bee and generate the new position of the food source $\boldsymbol{\Omega}_{[i]}$ for the scout bee randomly from (20).

(7-2) Construct the covariance matrix $\boldsymbol{K}_{[i]}$ using the corresponding $\boldsymbol{\Omega}_{[i]}$.

(7-3) Estimate $\boldsymbol{\theta}_{ml[i]}$ from (19).

(7-4) Calculate the fitness value $F_i(\boldsymbol{\Omega}_{[i]}) = \exp(-J_i(\boldsymbol{\Omega}_{[i]})$ from (22).

This step means that if the solution is not improved *limit* times through search by the employed and onlooker bees, the corresponding employed bee gives up to search around his food source and transforms himself to the scout bee to search around randomly selected food source. Since the number *limit* is usually set to be the product of the employed bee size and the dimension of the search space [16], this number is taken to be $limit = N_s \times 3$ in this paper.

Step 8: Repetition

Set the iteration counter to $l = l + 1$ and go to *step 5* until the prespecified iteration number l_{max}.

Step 9: Determination of the GP prior model

Determine the vector $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\theta}}_c = [\hat{\sigma}_y, \hat{\ell}, \hat{\sigma}_n]^{\mathrm{T}}$ and the corresponding parameter vector $\hat{\boldsymbol{\theta}}_{ml} = [\hat{\boldsymbol{\theta}}_m^{\mathrm{T}}, \hat{\boldsymbol{\theta}}_l^{\mathrm{T}}]^{\mathrm{T}}$ using the best position of the food source. Construct the suboptimal prior mean function and prior covariance function:

$$m(\boldsymbol{w}^f(t)) = (\boldsymbol{w}^f(t))^{\mathrm{T}}\hat{\boldsymbol{\theta}}_m \tag{25}$$

$$\begin{cases} s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = \hat{\sigma}_y^2 \exp\left(-\dfrac{||\boldsymbol{w}^f(t_p) - \boldsymbol{w}^f(t_q)||^2}{2\hat{\ell}^2}\right) \\ k(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) + \hat{\sigma}_n^2\delta_{pq}, \end{cases} \tag{26}$$

where $s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q))$ is an element of covariance matrix $\boldsymbol{\Sigma}$, $k(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q))$ is an element of covariance matrix $\boldsymbol{K}$, and δ_{pq} is the Kronecker delta, which is 1 if $p = q$ and 0 otherwise.

B. Estimation of the Nonlinear Function

For a new input $\boldsymbol{w}_*^f(t)$ and the corresponding function $f(\boldsymbol{w}_*^f(t))$, we have the following joint Gaussian distribution:

$$\begin{bmatrix} \boldsymbol{y} \\ f(\boldsymbol{w}_*^f(t)) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m}(\boldsymbol{w}) + \boldsymbol{G}\hat{\boldsymbol{\theta}}_l \\ m(\boldsymbol{w}_*^f(t)) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}_*^f(t)) \\ \boldsymbol{\Sigma}(\boldsymbol{w}_*^f(t), \boldsymbol{w}), & s(\boldsymbol{w}_*^f(t), \boldsymbol{w}_*^f(t)) \end{bmatrix} \right) \tag{27}$$

From the formula for conditioning a joint Gaussian distribution [18], the posterior distribution for specific test data is

$$f(\boldsymbol{w}_*^f(t))|\boldsymbol{w}, \boldsymbol{G}, \boldsymbol{y}, \boldsymbol{w}_*^f(t) \sim \mathcal{N}(\hat{f}(\boldsymbol{w}_*^f(t)), \hat{\sigma}_*^2(t)) \tag{28}$$

where the mean function $\hat{f}$ is given as

$$\hat{f}(\boldsymbol{w}_*^f(t)) = m(\boldsymbol{w}_*^f(t)) + \boldsymbol{\Sigma}(\boldsymbol{w}_*^f(t), \boldsymbol{w})\boldsymbol{K}^{-1}(\boldsymbol{y} - \boldsymbol{m}(\boldsymbol{w}) - \boldsymbol{G}\hat{\boldsymbol{\theta}}_l) \tag{29}$$

which is used as the estimated nonlinear function of the objective system. And its covariance function $\hat{\sigma}_*$ is evaluated as

$$\hat{\sigma}_*^2(t) = s(\boldsymbol{w}_*^f(t), \boldsymbol{w}_*^f(t)) - \boldsymbol{\Sigma}(\boldsymbol{w}_*^f(t), \boldsymbol{w})\boldsymbol{K}^{-1}\boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}_*^f(t)) \tag{30}$$

which is used for the confidence measure of the estimated nonlinear function.

V. ILLUSTRATIVE EXAMPLE

Consider an electric power system [19] described by

$$\begin{cases} \ddot{x}(t) + a_1\dot{x}(t) = f(\boldsymbol{z}(t)) \\ f(\boldsymbol{z}(t)) = -\dfrac{P_e}{\tilde{M}} + \dfrac{P_{in}}{\tilde{M}} \\ \qquad = -\dfrac{P_{em}}{\tilde{M}}(1 + u(t))\sin x(t) + \dfrac{P_{in}}{\tilde{M}} \\ y(t) = x(t) + e(t) \end{cases} \tag{31}$$

where $x(t) = \delta(t)$: phase angle, $u(t) = \Delta E_{fd}(t)$: increment of excitation voltage, $\tilde{M}$: inertia coefficient, $\tilde{D}$: damping coefficient, P_e: generator output power, P_{in}: turbine output power. In numerical example, $\tilde{M} = \tilde{D} = 0.06$, $P_{em} = 1.0$, $P_{in} = 0.8$ and $a_1 = \tilde{D}/\tilde{M} = 1.0$ are set. The measurement noise $e(t)$ is white Gaussian noise, where noise-to-signal ratio is about 1.5%. The number of input and output data for identification is taken to be $N = 800$. The third-order Butterworth filter with the cutoff frequency $\omega_c = 10$ [rad/s] is utilized as a delayed state variable filter. The setting parameters of ABC algorithm are chosen as follows:

(i) employed bee size $N_s = 50$

(ii) maximum iteration number $l_{max} = 100$

The hyperparameter vector of the covariance function has been determined by ABC algorithm as $\hat{\boldsymbol{\theta}}_c = [\hat{\sigma}_y, \hat{\ell}, \hat{\sigma}_n]^{\mathrm{T}} = [36.190, 0.342, 0.180]^{\mathrm{T}}$. Estimate of the parameter in the linear term is $\hat{a}_1 = 0.962$, which is very close to the true value $a_1 = 1.0$. The true nonlinear function $f(\boldsymbol{z}(t))$, the estimated nonlinear function $\hat{f}(\boldsymbol{z}(t))$, the absolute error between $f(\boldsymbol{z}(t))$ and $\hat{f}(\boldsymbol{z}(t))$, and the double standard deviation confidence interval (95.5% confidence region) around the

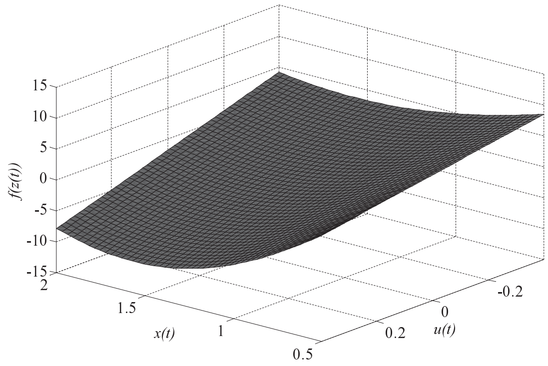

Fig. 1 True nonlinear function

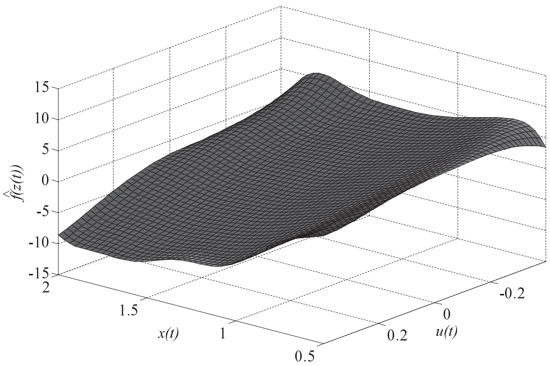

Fig. 2 Estimated nonlinear function

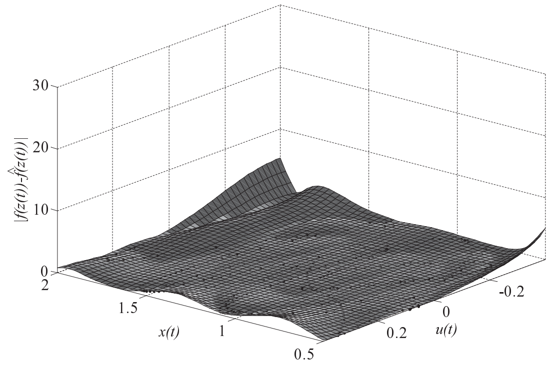

Fig. 3 Absolute error between true and estimated nonlinear functions

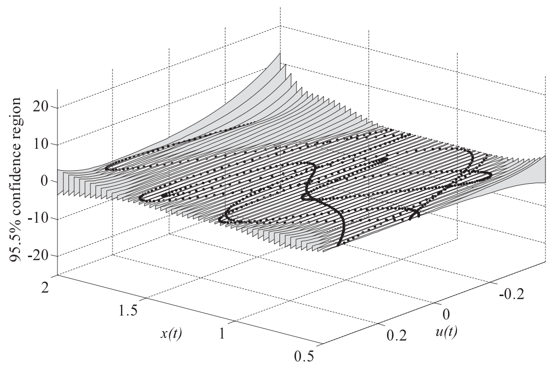

Fig. 4 95.5% confidence region

estimated nonlinear function are shown in Figs. 1 $\sim$ 4, respectively, where the thick curves depict the trajectories of the identification data. Clearly the estimated nonlinear function $\hat{f}(\boldsymbol{z}(t))$ is shown to be very close to the true nonlinear function $f(\boldsymbol{z}(t))$ on the data region. The confidence region of the estimated nonlinear function grows as $\boldsymbol{z}(t)$ goes away from the data region. On the other hand, the confidence region of the estimated nonlinear function is very small on the data region. Fig. 5 shows the true output $x(t)$ and the output $\hat{x}(t)$ by the estimated model, where the outputs were generated by the inputs for validation. This figure indicates that $\hat{x}(t)$ matches $x(t)$ considerably. Consequently, we can confirm that the proposed method gives an accurate model of the objective electric power system.

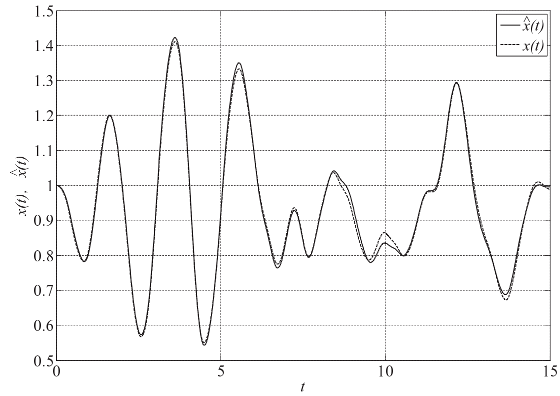

Fig. 5 True output and output by the estimated model

VI. Conclusions

In this paper, we have proposed an identification method of continuous-time nonlinear systems using the GP model. The GP prior model is trained by the aid of ABC algorithm so that the negative log marginal likelihood of the identification data is minimized. The proposed identification method is categorized into the nonparametric identification and does not need the determination of the model structure. Since ABC algorithm has a few setting parameters, the proposed training algorithm is efficient for system identification. Simulation results show that the proposed method can be successfully applied to modeling of the electric power system.

References

[1] C. Z. Jin, K. Wada, K. Hirasawa and J. Murata, Identification of nonlinear continuous systems by using neural network compensator (in Japanese),

IEEJ Trans. C, Vol. 114, No. 5, pp. 595–602, 1994.
[2] K. M. Tsang and S. A. Billings, Identification of continuous time nonlinear systems using delayed state variable filters, *Int. J. Control*, Vol. 60, No. 2, pp. 159–180, 1994.
[3] G. P. Liu and V. Kadirkamanathan, Stable sequential identification of continuous nonlinear dynamical systems by growing radial basis function networks, *Int. J. Control*, Vol. 65, No. 1, pp. 53–69, 1996.
[4] T. Hachino, I. Karube, Y. Minari and H. Takata, Continuous-time identification of nonlinear systems using radial basis function network model and genetic algorithm, *Proc. of the 12th IFAC Symposium on System Identification*, Vol. 2, pp. 787–792, 2001.
[5] T. Hachino and H. Takata, Identification of continuous-time nonlinear systems via local linear equations united by automatic choosing function and genetic algorithm, *Proc. of the 14th World Congress of International Federation of Automatic Control*, pp. 259–264, 1999.
[6] J. Kocijan, A. Girard, B. Banko and R. Murray-Smith, Dynamic systems identification with Gaussian processes, *Mathematical and Computer Modelling of Dynamical Systems*, Vol. 11, No. 4, pp. 411–424, 2005.
[7] T. Hachino and H. Takata, Identification of continuous-time nonlinear systems by using a Gaussian process model, *IEEJ Trans. on Electrical and Electronic Engineering*, Vol. 3, No. 6, pp. 620–628, 2008.
[8] A. Girard, C. E. Rasmussen, J. Q. Candela and R. Murray-Smith, Gaussian process priors with uncertain inputs -application to multiple-step ahead time series forecasting", in *Advances in Neural Information Processing Systems*, Vol. 15, pp. 542–552, MIT Press, 2003.
[9] T. Hachino and V. Kadirkamanathan, Multiple Gaussian process models for direct time series forecasting, *IEEJ Trans. on Electrical and Electronic Engineering*, Vol. 6, No. 3, pp. 245–252, 2011.
[10] A. O'Hagan, Curve fitting and optimal design for prediction (with discussion), *Journal of the Royal Statistical Society B*, Vol. 40, pp. 1–42, 1978.
[11] C. K. I. Williams, Prediction with Gaussian processes: from Linear regression to linear prediction and beyond, in *Learning and Inference in Graphical Models*, Kluwer Academic Press, pp. 599–621, 1998.
[12] M. Seeger, Gaussian processes for machine learning, *International Journal of Neural Systems*, Vol. 14, No. 2, pp. 1–38, 2004.
[13] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
[14] J. M. Wang, D. J. Fleet and A. Hertzmann, Gaussian process dynamical models for human motion, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 2, pp. 283–298, 2008.
[15] B. Likar and J. Kocijan, Predictive control of a gas-liquid separation plant based on a Gaussian process model, *Computers and Chemical Engineering*, Vol. 31, No. 3, pp. 142–152, 2007.
[16] D. Karaboga and B. Basturk, On the performance of artificial bee colony (ABC) algorithm, *Applied Soft Computing*, Vol. 8, No. 1, pp. 687–697, 2008.
[17] I. Iimura and S. Nakayama, Search performance evaluation of artificial bee colony algorithm on high-dimensional function optimization (in Japanese), *Trans. of the ISCIE*, Vol. 24, No. 4, pp. 97–99, 2011.
[18] R. von Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, 1964.
[19] H. Takata, An automatic choosing control for nonlinear systems, *Proc. of the 35th IEEE CDC*, pp. 3453–3458, 1996.

Tomohiro Hachino received the B.S., M.S. and Dr. Eng. degrees in electrical engineering from Kyushu Institute of Technology in 1991, 1993, and 1996, respectively. He is currently an Associate Professor at the Department of Electrical and Electronics Engineering, Kagoshima University. His research interests include nonlinear control and identification, signal processing, and evolutionary computation. Dr. Hachino is a member of IEEJ, SICE, and ISCIE.

Hitoshi Takata received the B.S. degree in electrical engineering from Kyushu Institute of Technology in 1968 and the M.S. and Dr. Eng. degrees in electrical engineering from Kyushu University in 1970 and 1974, respectively. He is currently a Professor emeritus and a part-time lecturer at Kagoshima University. His research interests include the control, linearization, and identification of nonlinear systems. Dr. Takata is a member of IEEJ and RISP.

Shigeru Nakayama received the B.S. degree from Kyoto Institute of Technology in 1972, and the M.S. and Dr. Eng. degrees from Kyoto University in 1974 and 1977, respectively. He is currently a Professor at the Department of Information Science and Biomedical Engineering, Kagoshima University. His research interests include distributed parallel processing, parallel genetic algorithms, parallel image processing, and distributed objects. Dr. Nakayama is a member of IPSJ, IEICE, IEEJ, ISCIE, and JSEE.

Ichiro Iimura received the B.S. and M.S. degrees from Sophia University in 1992 and 1994, respectively, and the Dr. Eng. degree from Kagoshima University in 2004. He is currently a Professor at the Faculty of Administration, the Prefectural University of Kumamoto. His research interests include evolutionary computation, swarm intelligence, distributed parallel processing, and natural user interface. Dr. Iimura is a member of IPSJ, IEICE, IEEJ, ISCIE, JSAI, and JPNSEC.

Seiji Fukushima received the B.S., M.S., and Ph.D. degrees in electrical engineering from Kyushu University in 1984, 1986, and 1993, respectively. He is currently a Professor at the Department of Electrical and Electronics Engineering, Kagoshima University. His research interests include photonics/radio hybrid communication systems and their related devices. Dr. Fukushima is a member of IEICE, IEEE/Photonics Society, Japan Society of Applied Physics, Japanese Liquid Crystal Society, and Optical Society of America.

Yasutaka Igarashi received the B.E., M.E., and Ph.D. degrees in information and computer sciences from Saitama University in 2000, 2002, and 2005, respectively. He is currently an Assistant Professor at the Department of Electrical and Electronics Engineering, Kagoshima University. His research interests include optical CDMA communication systems and the cryptanalysis of symmetric-key cryptography. Dr. Igarashi is a member of IEICE and RISP.