International Journal of Innovative Computing, Information and Control
Volume **17**, Number **6**, December **2021**

 ISSN 1349-4198
pp. 2191–2200

ON-LINE GAUSSIAN PROCESS MODEL IDENTIFICATION OF NONLINEAR SYSTEMS USING PARTICLE SWARM OPTIMIZATION

Tomohiro Hachino

Graduate School of Science and Engineering
Kagoshima University
1-21-40 Korimoto, Kagoshima 890-0065, Japan
hachino@eee.kagoshima-u.ac.jp

Received June 2021; revised September 2021

Abstract. *This paper presents a novel on-line Gaussian process (GP) model identification of nonlinear systems using particle swarm optimization (PSO). The GP is a Gaussian random function whose values follow a multidimensional normal distribution, and is specified by its mean function and covariance function. The identification model of the objective system is derived by using the GP. PSO is applied to the on-line training of the GP prior model by minimizing the negative log marginal likelihood of the time-shifted input-output data set. The nonlinear part of the objective system and its confidence region are evaluated by the predictive mean and predictive covariance of the GP posterior, respectively. The results of on-line identification for a simplified power system are shown to demonstrate the effectiveness of the proposed method.*
Keywords: Identification, On-line, Gaussian process, Nonlinear system, Particle swarm optimization

1. Introduction.

Many practical systems, such as power systems, are continuous-time systems and have nonlinear characteristics. Since an accurate model is required to realize high-performance analysis and control of the target system, the development of identification method for continuous-time nonlinear systems is an urgent issue. Some off-line identification methods of continuous-time nonlinear systems have been reported in [1-4]. However, in many cases, system dynamics may change with time. For such systems, the off-line identification algorithms do not work well. A precise on-line identification method is indispensable to track the time-varying system parameters and nonlinear terms. In on-line framework, several identification methods for discrete-time nonlinear systems have been proposed using radial basis function networks [5,6]. On the other hand, the authors proposed some on-line identification method of continuous-time nonlinear systems [7,8]. In [7], radial basis function model-based identification algorithm was presented combining the recursive least-squares (RLS) method and genetic algorithm. In [8], the authors proposed the moving-window type on-line identification based on the Gaussian process (GP) model. In this method, the hyperparameters of the covariance function were searched for using the firefly algorithm, while the weighting parameters of the mean function and the system parameters of the linear terms were updated by the RLS method. Since each group of parameters was optimized alternately, global optimization of all parameters was not always guaranteed. Recently, the recursive GP has been proposed [9] and it has been applied to nonlinear time-varying system identification [10]. However, these methods do

DOI: 10.24507/ijicic.17.06.2191

not consider the determination of hyperparameters, or require complicated determination algorithms of them.

To realize simple on-line identification that can track time-varying nonlinear systems, this paper proposes an on-line GP model identification method using particle swarm optimization (PSO). The GP model was originally utilized for the regression problem by O'Hagan [11] and has been utilized for regression or classification problem [12]. The GP model is a nonparametric model and fits naturally into Bayesian framework. Since it has fewer parameters than parametric models such as the neural network model, the nonlinearity of the objective system can be described in a few parameters. Moreover, the GP gives us not only the estimated nonlinear function but also its confidence region. In the proposed method, all unknown parameters including the hyperparameters of the covariance function, the weighting parameters of the mean function, and the system parameters of the linear terms, are represented by the particles. Then, PSO is applied to the minimization of the negative log marginal likelihood of the time-shifted input-output data set in on-line form. PSO is a swarm intelligence optimization technique, which was inspired by the social behavior of a flock of birds or a shoal of fish [13], and has been successfully applied to the machine learning applications [14]. The use of PSO might increase the efficiency of on-line identification due to its simple algorithm.

This paper is organized as follows. In Section 2, the system to be estimated is described and the problem is formulated. In Section 3, the identification model is derived using GP prior model. In Section 4, the on-line identification algorithm using PSO is presented. In Section 5, the performance of the proposed identification method is demonstrated through numerical simulations for a simplified power system. Finally, conclusions are given in Section 6.

2. Statement of the Problem.

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

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

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 $(i = 1, 2, \ldots, \alpha)$, m and m_j $(j = 1, 2, \ldots, \beta)$ are assumed to be known. The purpose of this paper is to identify the system parameters $\{a_i\}$ and $\{b_j\}$ of the linear terms and the nonlinear function $f(\cdot)$ with the confidence measure on-line, from the true input and noisy output data.

3. GP Prior Model for Identification.

Equation (1) can be rewritten as

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

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

Multiplying both sides of (2) by the delayed state variable filter $F(p)$ [2] yields

$$p^n y^f(t) = f\left(\boldsymbol{w}^f(t)\right) - \sum_{\substack{i=1 \\ i \neq n_1, n_2, \ldots, n_\alpha}}^{n} a_i p^{n-i} y^f(t) + \sum_{\substack{j=0 \\ j \neq m_1, m_2, \ldots, m_\beta}}^{m} b_j p^{m-j} u^f(t) + \varepsilon^f(t) \tag{3}$$

where $u^f(t) = F(p)u(t)$, $y^f(t) = F(p)y(t)$, and $\boldsymbol{w}^f(t) = F(p)\boldsymbol{w}(t)$ are the filtered signals, and $\varepsilon^f(t)$ is assumed to be zero mean Gaussian noise with variance σ_n^2.

By putting $t = t_1, t_2, \ldots, t_M$ into (3), (3) can be expressed in vector form as follows:

$$\boldsymbol{y} = \boldsymbol{v} + \boldsymbol{G}\boldsymbol{\theta}_l \tag{4}$$

where

$$\begin{aligned}
&\boldsymbol{y} = \left[p^n y^f(t_1), p^n y^f(t_2), \ldots, p^n y^f(t_M)\right]^{\mathrm{T}} \\
&\boldsymbol{v} = \left[f\left(\boldsymbol{w}^f(t_1)\right) + \varepsilon^f(t_1), f\left(\boldsymbol{w}^f(t_2)\right) + \varepsilon^f(t_2), \ldots, f\left(\boldsymbol{w}^f(t_M)\right) + \varepsilon^f(t_M)\right]^{\mathrm{T}} \\
&\boldsymbol{\theta}_l = [a_1, \ldots, a_i, \ldots, a_n, b_0, \ldots, b_j, \ldots, b_m]^{\mathrm{T}} \\
&\boldsymbol{G} = [\boldsymbol{g}(t_1), \boldsymbol{g}(t_2), \ldots, \boldsymbol{g}(t_M)]^{\mathrm{T}} \\
&\boldsymbol{g}(t) = \left[-p^{n-1}y^f(t), \ldots, -p^{n-i}y^f(t), \ldots, -y^f(t), p^m u^f(t), \ldots, p^{m-j}u^f(t), \ldots, u^f(t)\right]^{\mathrm{T}}
\end{aligned} \tag{5}$$

$\boldsymbol{\theta}_l$ is the unknown parameter vector composed of the system parameters of the linear terms.

Let the function value vector $\boldsymbol{f}$ be

$$\boldsymbol{f} = \left[f\left(\boldsymbol{w}^f(t_1)\right), f\left(\boldsymbol{w}^f(t_2)\right), \ldots, f\left(\boldsymbol{w}^f(t_M)\right)\right]^{\mathrm{T}} \tag{6}$$

Then $\boldsymbol{f}$ can be represented by the GP as follows:

$$\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}), \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w})) \tag{7}$$

where

$$\boldsymbol{w} = \left[\boldsymbol{w}^f(t_1), \boldsymbol{w}^f(t_2), \ldots, \boldsymbol{w}^f(t_M)\right] \tag{8}$$

$\boldsymbol{w}$ 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. In this paper, the mean function is given by the first order polynomial, i.e., a linear combination of the input variable:

$$\begin{aligned}
m\left(\boldsymbol{w}^f(t)\right) &= \left(\boldsymbol{w}^f(t)\right)^{\mathrm{T}} \boldsymbol{\theta}_m \\
\boldsymbol{\theta}_m &= \left[\theta_{m1}, \theta_{m2}, \ldots, \theta_{m(\alpha+\beta)}\right]^{\mathrm{T}}
\end{aligned} \tag{9}$$

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

$$\boldsymbol{m}(\boldsymbol{w}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_m \tag{10}$$

The covariance $\Sigma_{pq} = s\left(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)\right)$ 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 as the covariance function:

$$\Sigma_{pq} = s\left(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)\right) = \sigma_y^2 \exp\left(-\frac{\left\|\boldsymbol{w}^f(t_p) - \boldsymbol{w}^f(t_q)\right\|^2}{2\ell^2}\right) \tag{11}$$

where $||\cdot||$ denotes the Euclidean norm.

From (7), the vector $\boldsymbol{v}$ of the noisy function values in (4) can be written as

$$\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}), \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})) \tag{12}$$

where

$$\boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w}) = \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}) + \sigma_n^2 \boldsymbol{I}_M \quad (\boldsymbol{I}_M\text{: } M \times M \text{ identity matrix}) \tag{13}$$

$\boldsymbol{\theta}_c = [\sigma_y, \ell, \sigma_n]^{\mathrm{T}}$ is the unknown hyperparameter vector of the covariance function.

From (4), (10) and (12), the GP model for the identification is derived as

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_m + \boldsymbol{G}\boldsymbol{\theta}_l, \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})\right) \tag{14}$$

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

4. **On-Line Identification.** The GP prior model is trained on-line by optimizing the unknown parameter vector $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_c^{\mathrm{T}}, \boldsymbol{\theta}_m^{\mathrm{T}}, \boldsymbol{\theta}_l^{\mathrm{T}}\right]^{\mathrm{T}}$. In this paper, the following negative log marginal likelihood of the identification data is utilized as the objective function in the training:

$$\begin{aligned} J &= -\log p(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{G}, \boldsymbol{\theta}) \\ &= \frac{1}{2}\log|\boldsymbol{K}| + \frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_m - \boldsymbol{G}\boldsymbol{\theta}_l\right)^{\mathrm{T}}\boldsymbol{K}^{-1}\left(\boldsymbol{y} - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_m - \boldsymbol{G}\boldsymbol{\theta}_l\right) + \frac{M}{2}\log(2\pi) \end{aligned} \tag{15}$$

For this nonlinear optimization problem, we propose an on-line identification algorithm using PSO.

Let k denote the sampling time instants $t = kT$, where T is the sampling period. At the time step k, past input-output data corresponding to the time windows length M is collected. The unknown parameter vector $\boldsymbol{\theta}$ is optimized by PSO in which the objective function is (15). At the next time step $k = k + 1$, the input-output data used for identification is shifted by one time step, where the $\boldsymbol{gbest}$ in the previous time step is retained. Note that this on-line identification algorithm optimizes all unknown parameters simultaneously rather than alternately.

The proposed algorithm is specifically described below in a step-by-step format.

Step 1: Initialization for time step

Let $k = k_s$ be the starting step of identification.

Step 2: Collection of identification data

Collect the input-output data used for identification $\mathcal{D}(k) = \{u(k - M + 1), \ldots, u(k), y(k - M + 1), \ldots, y(k)\}$. Let $t_i = (k - M + i)T$ $(i = 1, 2, \ldots, M)$ in (5).

Step 3: Initialization for PSO

Set the iteration counter l to 1. Generate an initial population of Q particles with random positions $\boldsymbol{\theta}_{[i]}^{(l)} = \left[\boldsymbol{\theta}_{c[i]}^{(l)\mathrm{T}}, \boldsymbol{\theta}_{m[i]}^{(l)\mathrm{T}}, \boldsymbol{\theta}_{l[i]}^{(l)\mathrm{T}}\right]^{\mathrm{T}}$ and velocities $\boldsymbol{V}_{[i]}^{(l)}$ $(i = 1, 2, \ldots, Q)$. If $k \neq k_s$, replace one of the random positions with $\boldsymbol{gbest}$ in the previous time step compulsorily.

Step 4: Construction of covariance matrix

Construct Q candidates of the covariance matrix $\boldsymbol{K}_{[i]}^{(l)}$ using $\boldsymbol{\theta}_{c[i]}^{(l)}$ $(i = 1, 2, \ldots, Q)$.

Step 5: Evaluation value calculation

Calculate the evaluation values which are the values of the negative log marginal likelihood of the identification data $\mathcal{D}(k)$:

$$\begin{aligned} J\left(\boldsymbol{\theta}_{[i]}^{(l)}\right) = \frac{1}{2}\log\left|\boldsymbol{K}_{[i]}^{(l)}\right| + \frac{1}{2}\left(\boldsymbol{y} - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_{m[i]}^{(l)} - \boldsymbol{G}\boldsymbol{\theta}_{l[i]}^{(l)}\right)^{\mathrm{T}}\left\{\boldsymbol{K}_{[i]}^{(l)}\right\}^{-1}\Big(\boldsymbol{y} - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_{m[i]}^{(l)} \\ - \boldsymbol{G}\boldsymbol{\theta}_{l[i]}^{(l)}\Big) + \frac{M}{2}\log(2\pi) \end{aligned} \tag{16}$$

Step 6: Update of the best positions *pbest* and *gbest*

Update $\boldsymbol{pbest}_{[i]}^{(l)}$, which is the personal best position, and $\boldsymbol{gbest}^{(l)}$, which is the global best position among all particles as follows:

If $l = 1$ then

$$\begin{aligned} \boldsymbol{pbest}_{[i]}^{(l)} &= \boldsymbol{\theta}_{[i]}^{(l)} \\ \boldsymbol{gbest}^{(l)} &= \boldsymbol{\theta}_{[i_{best}]}^{(l)} \quad i_{best} = \arg\min_i J\left(\boldsymbol{\theta}_{[i]}^{(l)}\right) \end{aligned} \tag{17}$$

otherwise

$$\boldsymbol{pbest}_{[i]}^{(l)} = \begin{cases} \boldsymbol{\theta}_{[i]}^{(l)} & \left(J\left(\boldsymbol{\theta}_{[i]}^{(l)}\right) < J\left(\boldsymbol{pbest}_{[i]}^{(l-1)}\right)\right) \\ \boldsymbol{pbest}_{[i]}^{(l-1)} & \text{(otherwise)} \end{cases} \tag{18}$$

$$\boldsymbol{gbest}^{(l)} = \boldsymbol{pbest}_{[i_{best}]}^{(l)} \qquad i_{best} = \arg\min_i J\left(\boldsymbol{pbest}_{[i]}^{(l)}\right)$$

Step 7: Update of positions and velocities

Update the particle positions and velocities using (19):

$$\begin{cases} \boldsymbol{V}_{[i]}^{(l+1)} = w^l \cdot \boldsymbol{V}_{[i]}^{(l)} + c_1 \cdot rand_1() \cdot \left(\boldsymbol{pbest}_{[i]}^{(l)} - \boldsymbol{\theta}_{[i]}^{(l)}\right) \\ \qquad\qquad + c_2 \cdot rand_2() \cdot \left(\boldsymbol{gbest}^{(l)} - \boldsymbol{\theta}_{[i]}^{(l)}\right) \\ \boldsymbol{\theta}_{[i]}^{(l+1)} = \boldsymbol{\theta}_{[i]}^{(l)} + \boldsymbol{V}_{[i]}^{(l+1)} \end{cases} \tag{19}$$

where w^l is an inertia factor, c_1 and c_2 are constants representing acceleration coefficients, and $rand_1()$ and $rand_2()$ are uniformly distributed random numbers with amplitude in the range $[0, 1]$.

Step 8: Repetition of PSO

If the iteration counter $l < l_{\max}$, set $l = l + 1$ and go to Step 4.

Step 9: Determination of the GP model

Determine the estimated parameter vector $\hat{\boldsymbol{\theta}}(k) = \left[\hat{\boldsymbol{\theta}}_c^{\mathrm{T}}(k), \hat{\boldsymbol{\theta}}_m^{\mathrm{T}}(k), \hat{\boldsymbol{\theta}}_l^{\mathrm{T}}(k)\right]^{\mathrm{T}}$ at the time step k using the best particle position $\boldsymbol{gbest}^{(l_{\max})}$. Construct the suboptimal prior mean function and the prior covariance function:

$$\hat{m}\left(\boldsymbol{w}^f(t)\right) = \left(\boldsymbol{w}^f(t)\right)^{\mathrm{T}} \hat{\boldsymbol{\theta}}_m(k) \tag{20}$$

$$\begin{cases} \hat{s}\left(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)\right) = \hat{\sigma}_y^2(k) \exp\left(-\dfrac{\left\|\boldsymbol{w}^f(t_p) - \boldsymbol{w}^f(t_q)\right\|^2}{2\hat{\ell}^2(k)}\right) \\ \hat{k}\left(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)\right) = \hat{s}\left(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)\right) + \hat{\sigma}_n^2(k)\delta_{pq} \end{cases} \tag{21}$$

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

Step 10: Estimation of the nonlinear function

Estimate the predictive mean function:

$$\hat{f}\left(\boldsymbol{w}_*^f(t)\right) = \hat{m}\left(\boldsymbol{w}_*^f(t)\right) + \hat{\boldsymbol{\Sigma}}\left(\boldsymbol{w}_*^f(t), \boldsymbol{w}\right)\left\{\hat{\boldsymbol{K}}(k)\right\}^{-1}\left(\boldsymbol{y} - \boldsymbol{w}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_m(k) - \boldsymbol{G}\hat{\boldsymbol{\theta}}_l(k)\right) \tag{22}$$

and the predictive variance:

$$\hat{\sigma}_*^2(t) = \hat{s}\left(\boldsymbol{w}_*^f(t), \boldsymbol{w}_*^f(t)\right) - \hat{\boldsymbol{\Sigma}}\left(\boldsymbol{w}_*^f(t), \boldsymbol{w}\right)\left\{\hat{\boldsymbol{K}}(k)\right\}^{-1}\hat{\boldsymbol{\Sigma}}\left(\boldsymbol{w}, \boldsymbol{w}_*^f(t)\right) \tag{23}$$

where $\boldsymbol{w}_*^f(t)$ is a new input of GP. Equation (22) is the estimated nonlinear function of the system and (23) is utilized as the confidence measure of the estimated nonlinear function, at the time step k.

Step 11: Update of the time step

Set $k = k + 1$ and go to Step 2.

The flowchart of the proposed method is shown in Figure 1.

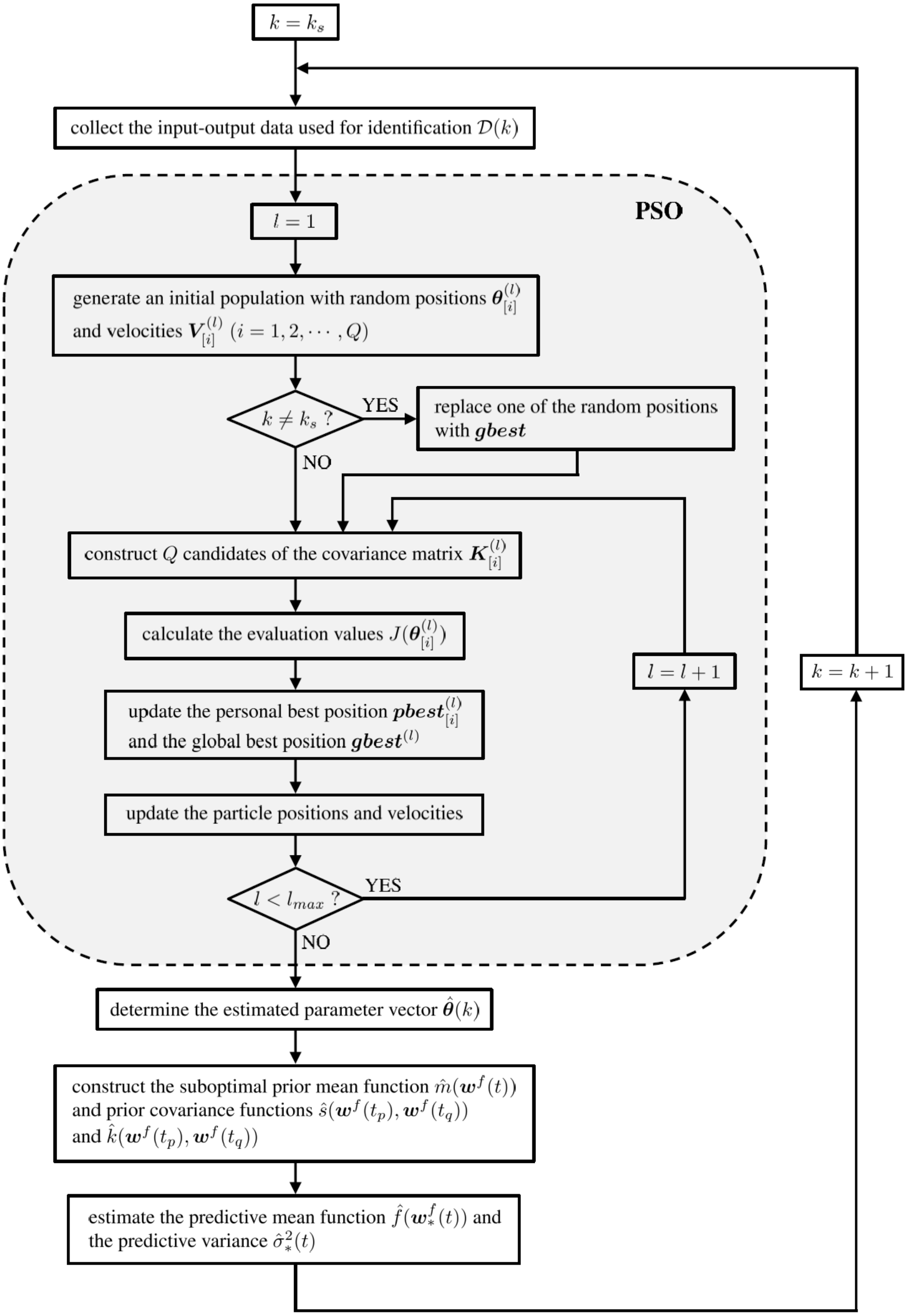


FIGURE 1. Flowchart of the proposed method

5. **Illustrative Example.** The system to be estimated is a simplified electric power system [15] described by

$$\begin{cases} \ddot{x}(t) + a_1\dot{x}(t) = f(\boldsymbol{z}(t)) \qquad \left(a_1 = \tilde{D}/\tilde{M}\right) \\ f(\boldsymbol{z}(t)) = -\dfrac{P_e}{\tilde{M}} + \dfrac{P_{in}}{\tilde{M}} = -\dfrac{P_{em}}{\tilde{M}}(1+u(t))\sin x(t) + \dfrac{P_{in}}{\tilde{M}} \\ \boldsymbol{z}(t) = [x(t), u(t)]^{\mathrm{T}}, \qquad y(t) = x(t) + e(t) \end{cases} \tag{24}$$

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, and P_{em}: maximum output power. a_1 is the system parameter of the linear term. The inertia coefficient and the turbine output power are set to be $\tilde{M} = 0.06$ and $P_{in} = 0.8$, respectively. Due to the time variation of parameters $\tilde{D}$ and P_{em}, the system parameter a_1 of the linear term and the nonlinear function $f(\boldsymbol{z}(t))$ change stepwise as shown in Table 1. The measurement noise $e(t)$ is white Gaussian noise, where the noise-to-signal ratio is about 2.0%. The goal of this simulation is to estimate the system parameter a_1 of the linear term and the nonlinear function $f(\boldsymbol{z}(t))$ on-line from input and noisy output.

Table 1. System parameter and nonlinear function

Interval	t [s]	a_1	$f(\boldsymbol{z}(t))$
1	$[0, 20]$	3.0	$-18.3(1+u(t))\sin x(t) + 13.3$
2	$(20, 50]$	1.0	$-16.7(1+u(t))\sin x(t) + 13.3$
3	$(50, 80]$	2.0	$-20.0(1+u(t))\sin x(t) + 13.3$

The sampling period is taken to be $T = 0.01$ [s] and the time window length is set to be $M = 800$. 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 PSO are chosen as follows:

1) particle size: $Q = 20$
2) inertia factor: $w^l = w_{\max} - (w_{\max} - w_{\min})l/l_{\max}$ ($w_{\max} = 0.9$, $w_{\min} = 0.6$)
3) acceleration coefficients: $c_1 = 1.0$, $c_2 = 1.0$
4) maximum iteration number per time step: $l_{\max} = 5$

In this simulation, the inertia factor and the acceleration coefficients are set within the stable region based on the stability analysis of PSO [16]. Moreover, since it is desirable to reduce the computational burden as much as possible in on-line identification, the particle size and the maximum iteration number are set to be sufficiently small so as not to deteriorate the accuracy of identification.

Figure 2 shows the estimated weighting parameters of the mean function and the estimated hyperparameters of the covariance function. As an example, 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 estimated nonlinear function at the final time $t = 80$ (the end of interval 3) are shown in Figure 3, where the thick curves depict the trajectories of the identification data. 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 becomes very small on the data region and grows as $\boldsymbol{z}(t)$ goes away from the data region. It is one of the features of this method that information on such uncertainty regarding the estimated nonlinear function can be obtained. Note

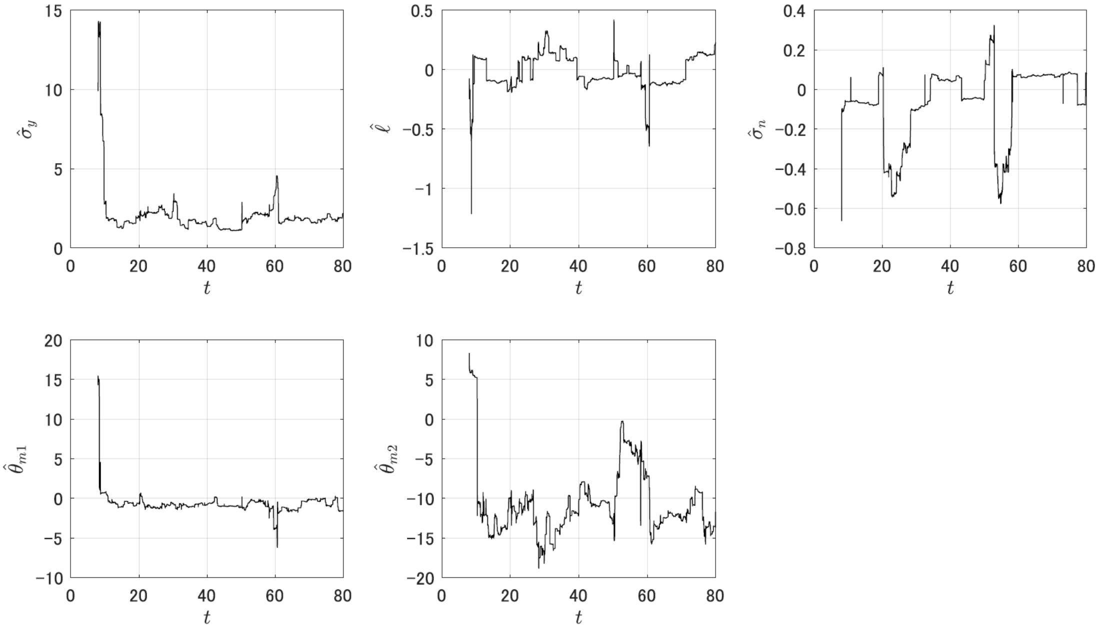
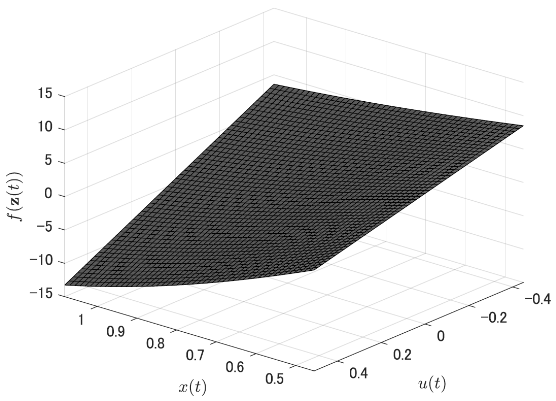
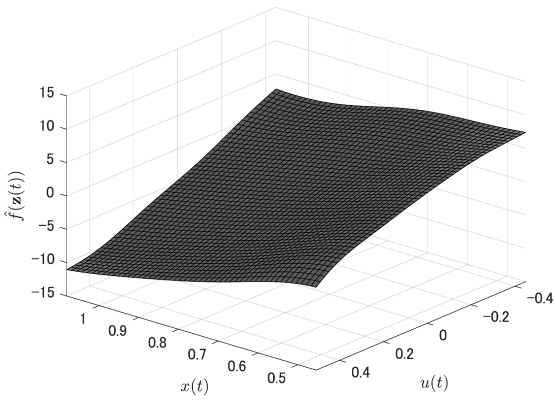


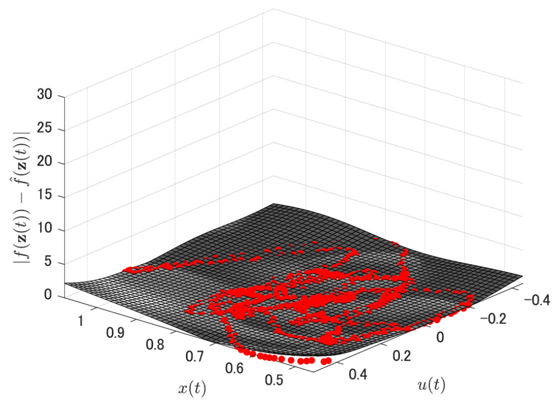
FIGURE 2. Hyperparameters of the covariance function and weighting parameters of the mean function



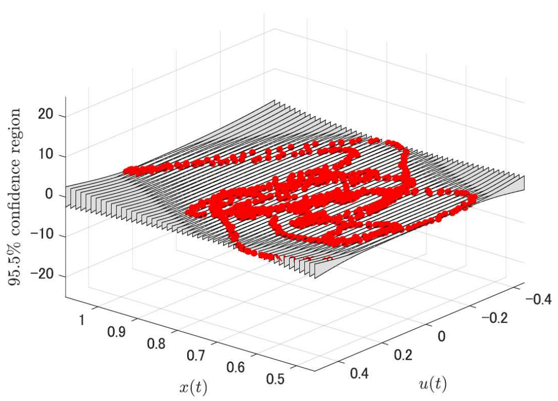
(a) True nonlinear function



(b) Estimated nonlinear function



(c) Absolute error between true and estimated nonlinear functions



(d) 95.5% confidence region around estimated nonlinear function

FIGURE 3. Nonlinear function

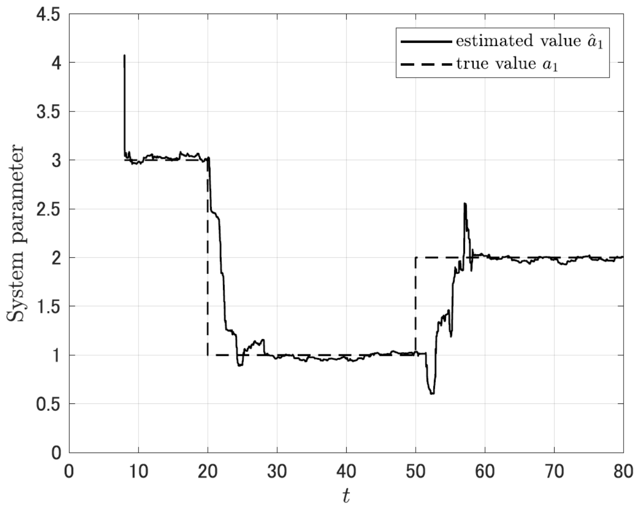


FIGURE 4. System parameter of the linear term

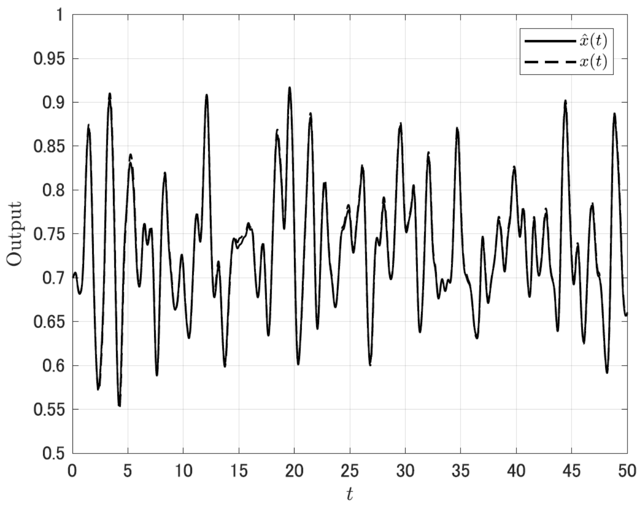


FIGURE 5. True output and output by the estimated model

that this method can estimate the nonlinear function at all time step. The true system parameter a_1 and the estimated system parameter $\hat{a}_1$ of the linear term are shown in Figure 4. From this figure, we can confirm that the proposed identification method can track the time-varying system parameters successfully. Figure 5 shows the true output $x(t)$ and the output $\hat{x}(t)$ by the model estimated at the final time $t = 80$, where $N = 5000$ data of both outputs are generated using the same input for validation. We can confirm that the output $\hat{x}(t)$ by the estimated model almost matches the true output $x(t)$. The mean squares errors of the output $\sum_{k=1}^{N}(x(k) - \hat{x}(k))^2/N$ are 6.95×10^{-5} at $t = 20$ (the end of interval 1), 6.72×10^{-4} at $t = 50$ (the end of interval 2), and 1.75×10^{-5} at $t = 80$ (the end of interval 3), respectively. These results demonstrate the high accuracy of the proposed identification method.

6. **Conclusions.** In this paper, we have proposed an on-line identification method of continuous-time nonlinear systems. The GP prior model for the objective system is derived as the identification model. The unknown parameters consisting of the hyperparameters of the covariance function, the weighting parameters of the mean function, and the system parameters of the linear terms are trained simultaneously on-line using PSO. Simulation

results show that the proposed identification method is effective for the time-varying nonlinear systems. Reducing the computational burden for identification is one of the future works.

REFERENCES

[1] C. Z. Jin, K. Wada, K. Hirasawa and J. Murata, Identification of nonlinear continuous systems by using neural network compensator, *IEEJ Trans. Electronics, Information and Systems*, vol.114, no.5, pp.595-602, 1994 (in Japanese).
[2] K. M. Tsang and S. A. Billings, Identification of continuous time nonlinear systems using delayed state variable filters, *International Journal of Control*, vol.60, no.2, pp.159-180, 1994.
[3] 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.
[4] T. Hachino and H. Takata, Identification of continuous-time nonlinear systems by using a Gaussian process model, *IEEJ Trans. Electrical and Electronic Engineering*, vol.3, no.6, pp.620-628, 2008.
[5] L. Yingwei, N. Sundararajan and P. Saratchandran, Identification of time-varying nonlinear systems using minimal radial basis function neural networks, *IEE Proceedings – Control Theory and Applications*, vol.144, no.2, pp.202-208, 1997.
[6] S. Chen, S. A. Billings and P. M. Grant, Recursive hybrid algorithm for non-linear system identification using radial basis function networks, *International Journal of Control*, vol.55, no.5, pp.1051-1070, 1992.
[7] T. Hachino and H. Takata, On-line identification method of continuous-time nonlinear systems using radial basis function network model adjusted by genetic algorithm, *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol.E87-A, no.9, pp.2372-2378, 2004.
[8] K. Sugiyama and T. Hachino, On-line identification of continuous-time nonlinear systems using Gaussian process models, *Proc. of 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems and 19th International Symposium on Advanced Intelligent Systems*, pp.1219-1224, 2018.
[9] M. F. Huber, Recursive Gaussian process, on-line regression and learning, *Pattern Recognition Letters*, vol.45, pp.85-91, 2014.
[10] K. Seo and M. Yamakita, Nonlinear time-varying system identification with recursive Gaussian process, *Proc. of 2017 American Control Conference*, pp.825-830, 2017.
[11] 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.
[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, 2006.
[13] J. Kennedy and R. C. Eberhart, Particle swarm optimization, *Proc. of IEEE International Conference on Neural Networks*, pp.1942-1945, 1995.
[14] X. Wang, M. Liu, S. Chen and H. Liu, Identification of continuous rotary motor based on improved particle swarm optimizing support vector machine model, *International Journal of Innovative Computing, Information and Control*, vol.14, no.6, pp.2189-2202, 2018.
[15] H. Takata, An automatic choosing control for nonlinear systems, *Proc. of the 35th IEEE Conference on Decision and Control*, pp.3453-3458, 1996.
[16] E. Aiyoshi and K. Yasuda (eds.), *Metaheuristics and Its Applications*, Denki-gakkai, Tokyo, 2007 (in Japanese).

Author Biography

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.

Prof. Hachino is currently a professor at the Graduate School of Science and Engineering, Kagoshima University, Japan. His research interests include nonlinear control and identification, signal processing, and evolutionary computation.