Multistep Ahead Prediction of Electric Power Systems Using Multiple Gaussian Process Models

Tomohiro Hachino, Hitoshi Takata, Seiji Fukushima, Yasutaka Igarashi, and Keiji Naritomi
Kagoshima University, Kagoshima, Japan
Email: {hachino, takata, fukushima, igarashi}@eee.kagoshima-u.ac.jp, k0519282@kadai.jp

***Abstract*—This paper focuses on the problem of multistep ahead prediction of electric power systems using the Gaussian process models. The Gaussian process model is a nonparametric model and the output of the model has Gaussian distribution with mean and variance. The multistep ahead prediction for the phase angle in transient state of the electric power system is accomplished by using multiple Gaussian process models as every step ahead predictors in accordance with the direct approach. The proposed prediction method gives the predictive values of the phase angle and the uncertainty of the predictive values as well. Simulation results for a simplified electric power system are shown to illustrate the effectiveness of the proposed prediction method.**

***Index Terms*—multistep ahead prediction, Gaussian process model, direct method, electric power system**

I. Introduction

In recent years, model predictive control (MPC) has received much attention in both process control and servo control [1]-[5]. The performance of MPC greatly depends on the accuracy of the model used for prediction. Therefore, to improve the performance of MPC, it is urgent to develop an accurate predictor. The Gaussian process (GP) model is one of the attractive models for multistep ahead prediction. The GP model is a nonparametric model and fits naturally into Bayesian framework [6]-[8]. This model has recently attracted much attention for system identification [9], [10], time series forecasting [11]-[13], and predictive control [3], [14], [15]. Since the GP model gives us not only the mean value but also the variance of the conditionally expected value of the output, it is useful for MPC considering the uncertainty of model. Moreover, the GP model has far fewer parameters to describe the nonlinearity than the parametric models such as radial basis function (RBF) model, neural network model, and fuzzy model.

There are two approaches to multistep ahead prediction. One is the direct method that makes multistep ahead prediction directly by using a specific step ahead predictor. The other is the iterated method that repeats one-step ahead prediction up to the desired step. The iterated multistep ahead predictions with propagation of the prediction uncertainty based on the GP model were presented in [11], [12]. Although the computational burden of this approach is not so heavy during the training phase, unacceptable prediction errors are gradually accumulated as the prediction horizon increases especially in the presence of measurement noise.

Therefore, with the aim of MPC, this paper proposes the direct method for multistep ahead prediction of the electric power systems in the GP framework. Multistep ahead prediction for the phase angle in transient state of the electric power system is directly performed by using the multiple trained GP models as every step ahead predictor. The proposed direct method uses not only one-step ahead predictor but also all-step ahead predictors. Therefore, although each step ahead predictor has a systematic error, the prediction errors are not accumulated so much as the prediction horizon increases. The proposed direct method gives the predictive values of the phase angle and uncertainty of the predictive values as well. The uncertainty of the predictive values is usually not obtained by the non GP-based direct methods such as the RBF-based direct method.

This paper is organized as follows. In section II, the problem of multistep ahead prediction is formulated for an electric power system. In section III, the multiple GP prior models are derived for every step ahead predictors and the training method of the GP prior models is briefly described. In section IV, the direct multistep ahead prediction is carried out using the GP posterior distribution. In section V, simulation results are shown to illustrate the effectiveness of the proposed prediction method. Finally, conclusions are given in section VI.

II. Statement of the Problem

Consider a single machine power system described by

$$\begin{cases} \widetilde{M}\ddot{\delta}(t) + \widetilde{D}\dot{\delta}(t) + P_e = P_{in} \\ P_e = \dfrac{e_t E_{fd}}{X_e}\left(1 + \Delta E_{fd}(t)\right)\sin\delta(t) \\ y(t) = \delta(t) + v(t) \end{cases} \tag{1}$$

where $\delta(t)$: phase angle, $y(t)$: phase angle corrupted by the measurement noise $v(t)$, $\Delta E_{fd}(t)$: increment of

Manuscript received June 3, 2014; revised September 18, 2014.

doi: 10.12720/joace.3.4.316-321

excitation voltage, $\widetilde{M}$: inertia coefficient, $\widetilde{D}$: damping coefficient, P_e: generator output power, P_{in}: turbine output power, E_{fd}: excitation voltage, e_t: infinite bus voltage, and X_e: system impedance. The measurement noise $v(t)$ is zero mean white Gaussian noise with variance σ^2. It is assumed that the input $u(k) = \Delta E_{fd}(kT_s)$ and the noisy measurement of the output $y(k) = y(kT_s)$ at $t = kT_s$ are available when the multistep ahead predictors are trained, where T_s is the sampling period. The problem of multistep ahead prediction is usually to estimate the future outputs given the past input and output data. The optimal predictor can be written as

$$\hat{y}(k+j) = \mathrm{E}[y(k+j)|\boldsymbol{x}(k)] \tag{2}$$

where $\mathrm{E}[\cdot]$ is the expectation operator, and

$$\begin{aligned}\boldsymbol{x}(k) = \big[&y(k), y(k-1), \cdots, y\big(k-L_y+1\big),\\ &u(k), u(k-1), \cdots, u(k-L_u+1)\big]^{\mathrm{T}}\end{aligned} \tag{3}$$

Which is the state vector consisting of the past outputs and inputs up to the prespecified lags L_y and L_u. Actually, with the GP framework, not only estimates $\hat{y}(k+j)$ but also its uncertainty, i.e., the variance $\hat{\sigma}^2(k+j)$ are estimated. Therefore, the problem of this paper is to construct the following probability distributions for the multistep ahead prediction

$$\begin{aligned}y(k+j)|\boldsymbol{x}(k) \sim N\big(\hat{y}(k+j), \hat{\sigma}^2(k+j)\big)\\ (j = 1,2,\cdots,M)\end{aligned} \tag{4}$$

And to carry out multistep ahead prediction up to M step based on these distributions, by using the GP framework.

III. GP Prior Model

A. Derivation of GP Prior Models

Consider a j-step ahead predictor as

$$\begin{aligned}y(k+j) = f_j\big(\boldsymbol{x}(k)\big) + \varepsilon_j(k)\\ (j = 1,2,\cdots,M)\end{aligned} \tag{5}$$

where $f_j(\cdot)$ is a function which is assumed to be stationary and smooth. $\varepsilon_j(k)$ is zero mean Gaussian noise with unknown variance σ_j^2. In this paper, this predictor is constructed in the GP framework.

Putting $k = k_s,\ k_s+1, \cdots,\ k_s+N-1$ on (5) yields

$$\boldsymbol{y}_j = \boldsymbol{f}_j + \boldsymbol{\varepsilon}_j \tag{6}$$

where

$$\begin{aligned}\boldsymbol{y}_j &= [y(k_s+j), y(k_s+j+1),\\ &\qquad \cdots, y(k_s+j+N-1)]^{\mathrm{T}}\\ \boldsymbol{f}_j &= \big[f_j(\boldsymbol{x}_1), f_j(\boldsymbol{x}_2), \cdots, f_j(\boldsymbol{x}_N)\big]^{\mathrm{T}}\\ \boldsymbol{\varepsilon}_j &= \big[\varepsilon_j(k_s), \varepsilon_j(k_s+1), \cdots, \varepsilon_j(k_s+N-1)\big]^{\mathrm{T}}\\ \boldsymbol{X} &= [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]^{\mathrm{T}}\\ &= [\boldsymbol{x}(k_s), \boldsymbol{x}(k_s+1), \cdots, \boldsymbol{x}(k_s+N-1)]^{\mathrm{T}}\end{aligned} \tag{7}$$

$\boldsymbol{y}_j$ and $\boldsymbol{f}_j$ are the vector of model outputs and the vector of function values for the j-step ahead predictor, respectively. $\boldsymbol{X}$ is the model input matrix and is common for every step ahead predictors. $\{\boldsymbol{X}, \boldsymbol{y}_j\}$ is the training input and output data for the j-step ahead predictor.

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 which has joint multivariable Gaussian distribution. Therefore, the vector of function values $\boldsymbol{f}_j$ can be represented by the GP as

$$\boldsymbol{f}_j \sim N\big(\boldsymbol{m}_j(\boldsymbol{X}), \boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{X})\big) \tag{8}$$

where $\boldsymbol{m}_j(\boldsymbol{X})$ is the N-dimensional mean function vector and $\boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{X})$ is the N-dimensional covariance matrix evaluated at all pairs of the training input data. Equation (8) means that $\boldsymbol{f}_j$ has a Gaussian distribution with the mean function vector $\boldsymbol{m}_j(\boldsymbol{X})$ and the covariance matrix $\boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{X})$.

The mean function is often represented by a polynomial regression [8]. In this paper, the mean function vector $\boldsymbol{m}_j(\boldsymbol{X})$ is expressed by the first order polynomial, i.e., a linear combination of the model input:

$$\boldsymbol{m}_j(\boldsymbol{X}) = \big[m_j(\boldsymbol{x}_1), m_j(\boldsymbol{x}_2), \cdots, m_j(\boldsymbol{x}_N)\big]^{\mathrm{T}} = \widetilde{\boldsymbol{X}}\boldsymbol{\theta}_j \tag{9}$$

where $\widetilde{\boldsymbol{X}} = [\boldsymbol{X}, \boldsymbol{e}]$, and $\boldsymbol{e} = [1,1,\cdots,1]^{\mathrm{T}}$ is the N-dimensional vector consisting of ones, and $\boldsymbol{\theta}_j = \big[\theta_{jo}, \theta_{j1}, \cdots, \theta_{j(L_y+L_u)}\big]^{\mathrm{T}}$ is the unknown weighting parameter vector of the mean function to be trained. The determination of $\boldsymbol{\theta}_j$ will be discussed in the next subsection.

The covariance matrix $\boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{X})$ is constructed as

$$\boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{X}) = \begin{bmatrix}\Sigma_{j(1,1)} & \cdots & \Sigma_{j(1,N)}\\ \vdots & \ddots & \vdots\\ \Sigma_{j(N,1)} & \cdots & \Sigma_{j(N,N)}\end{bmatrix} \tag{10}$$

where the element $\Sigma_{j(p,q)} = \mathrm{cov}\big(f_j(\boldsymbol{x}_p), f_j(\boldsymbol{x}_q)\big) = s_j\big(\boldsymbol{x}_p, \boldsymbol{x}_q\big)$ is a function of $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$. Under the assumption that the process is stationary and smooth, the following Gaussian kernel is utilized for $\Sigma_{j(p,q)}$:

$$\begin{aligned}\Sigma_{j(p,q)} &= s_j\big(\boldsymbol{x}_p, \boldsymbol{x}_q\big)\\ &= \rho_j^2 \exp\left(-\frac{\big\|\boldsymbol{x}_p - \boldsymbol{x}_q\big\|^2}{2\ell_j^2}\right)\end{aligned} \tag{11}$$

where ρ_j^2 is the signal variance, ℓ_j is the length scale, and $\|\cdot\|$ denotes the Euclidean norm. The free parameters ρ_j and ℓ_j of (11) and the noise standard deviation σ_j are called *hyperparameters* and construct the hyperparameter vector $\boldsymbol{h}_j = \big[\rho_j, \ell_j, \sigma_j\big]^{\mathrm{T}}$. ρ_j can control the overall variance of the random function $f_j(\cdot)$ and determines the magnitude of the function $f_j(\cdot)$. ℓ_j can change the characteristic length scale so that the axis about the model input changes. If ℓ_j is set to be smaller, the function $f_j(\cdot)$ becomes more oscillatory. Therefore, the hyperparameter vector $\boldsymbol{h}_j$ should be appropriately determined according to the training data for precise

prediction. This parameter selection will be also presented in the next subsection.

Since $\boldsymbol{y}_j$ is noisy observation, we have the following GP model for j-step ahead prediction from (6) and (8) as

$$\boldsymbol{y}_j \sim N\big(\boldsymbol{m}_j(\boldsymbol{X}), \boldsymbol{K}_j(\boldsymbol{X},\boldsymbol{X})\big) \tag{12}$$

where

$$\begin{aligned}\boldsymbol{K}_j(\boldsymbol{X},\boldsymbol{X}) &= \boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{X}) + \sigma_j^2 \boldsymbol{I}_N \\ \boldsymbol{I}_N&: N \times N \text{ identity matrix}\end{aligned} \tag{13}$$

In the following, $\boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{X})$ and $\boldsymbol{K}_j(\boldsymbol{X},\boldsymbol{X})$ are written as $\boldsymbol{\Sigma}_j$ and $\boldsymbol{K}_j$, respectively.

B. Training of GP Prior Models

To perform multistep ahead prediction, the proposed direct approach needs 1 to M step ahead prediction models as shown in Fig. 1. The accuracy of prediction greatly depends on the unknown parameter vector $\boldsymbol{\vartheta}_j = \big[\boldsymbol{\theta}_j^{\mathrm{T}}, \boldsymbol{h}_j^{\mathrm{T}}\big]^{\mathrm{T}}$ and therefore $\boldsymbol{\vartheta}_j$ has to be optimized. This training is carried out by minimizing the negative log marginal likelihood of the training data:

$$\begin{aligned}J(\boldsymbol{\vartheta}_j) &= -\log p(\boldsymbol{y}_j|\boldsymbol{X},\boldsymbol{\vartheta}_j) \\ &= \frac{1}{2}\log|\boldsymbol{K}_j| + \frac{1}{2}\big(\boldsymbol{y}_j - \boldsymbol{m}_j(\boldsymbol{X})\big)^{\mathrm{T}} \boldsymbol{K}_j^{-1} \\ &\qquad \times \big(\boldsymbol{y}_j - \boldsymbol{m}_j(\boldsymbol{X})\big) + \frac{N}{2}\log(2\pi) \\ &= \frac{1}{2}\log|\boldsymbol{K}_j| + \frac{1}{2}\big(\boldsymbol{y}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_j\big)^{\mathrm{T}}\boldsymbol{K}_j^{-1}\big(\boldsymbol{y}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_j\big) \\ &\qquad + \frac{N}{2}\log(2\pi)\end{aligned} \tag{14}$$

Since the cost function $J\big(\boldsymbol{\vartheta}_j\big)$ generally has multiple local minima, this training problem becomes a nonlinear optimization one. However, we can separate the linear optimization part and the nonlinear optimization part for this optimization problem. The partial derivative of (14) with respect to the weighting parameter vector $\boldsymbol{\theta}_j$ of the mean function is as follows:

$$\frac{\partial J(\boldsymbol{\vartheta}_j)}{\partial \boldsymbol{\theta}_j} = -\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\boldsymbol{y}_j + \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\tilde{\boldsymbol{X}}\boldsymbol{\theta}_j \tag{15}$$

Note that if the hyperparameter vector $\boldsymbol{h}_j$ of the covariance function is given, then the weighting parameter $\boldsymbol{\theta}_j$ can be estimated by the linear least-squares method putting $\partial J(\boldsymbol{\vartheta}_j)/\partial \boldsymbol{\theta}_j = \boldsymbol{0}$:

$$\boldsymbol{\theta}_j = \big(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\tilde{\boldsymbol{X}}\big)^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\boldsymbol{y}_j \tag{16}$$

However, even if the weighting parameter vector $\boldsymbol{\theta}_j$ is known, the optimization with respect to hyperparameter vector $\boldsymbol{h}_j$ is a complicated nonlinear problem and might suffer from the local minima problem. Therefore, the unknown parameter vector $\boldsymbol{\vartheta}_j$ is determined by the separable least-squares (LS) approach combining the linear LS method and the genetic algorithm (GA) [16], as

$$\begin{aligned}\boldsymbol{\vartheta}_{j[best]} &= \big[\boldsymbol{\theta}_{j[best]}^{\mathrm{T}}, \boldsymbol{h}_{j[best]}^{\mathrm{T}}\big]^{\mathrm{T}} \\ &= \big[\boldsymbol{\theta}_{j[best]}^{\mathrm{T}}, \rho_{j[best]}, \ell_{j[best]}, \sigma_{j[best]}\big]^{\mathrm{T}}\end{aligned}$$

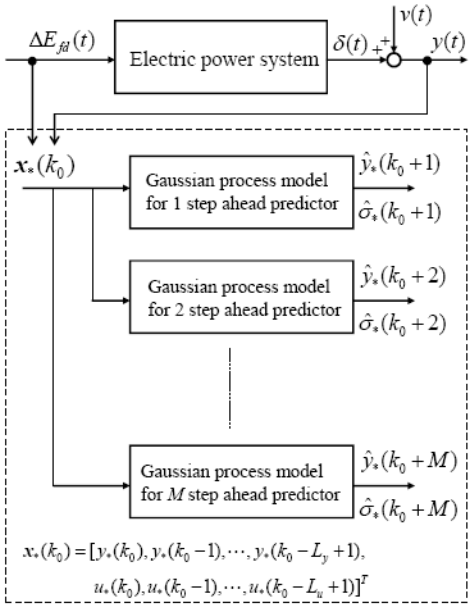

Figure 1. The proposed multistep ahead prediction scheme.

IV. Multistep Ahead Prediction by GP Posterior

In section III, we have already obtained the GP prior models for j $(j = 1,2,\cdots,M)$ step ahead predictors. In the proposed direct approach, multistep ahead prediction up to M step is carried out directly using every GP prior models as shown in Fig. 1.

For a new given test input

$$\boldsymbol{x}_* = \boldsymbol{x}_*(k_0) = \big[y_*(k_0), y_*(k_0-1), \cdots, y_*\big(k_0 - L_y + 1\big), u_*(k_0), u_*(k_0-1), \cdots, u_*(k_0 - L_u + 1)\big]^{\mathrm{T}}$$

And corresponding test output $y_*(k_0+j)$ $(j = 1,2,\cdots,M)$, we have the following the joint Gaussian distribution:

$$\begin{bmatrix}\boldsymbol{y}_j \\ y_*(k_0+j)\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{m}_j(\boldsymbol{X}) \\ m_j(\boldsymbol{x}_*)\end{bmatrix}, \begin{bmatrix}\boldsymbol{K}_j & \boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{x}_*) \\ \boldsymbol{\Sigma}_j(\boldsymbol{x}_*,\boldsymbol{X}) & s_j(\boldsymbol{x}_*,\boldsymbol{x}_*) + \sigma_{j[best]}^2\end{bmatrix}\right) \quad (j = 1,2,\cdots,M) \tag{17}$$

where k_0 is the starting step for prediction, and $\boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{x}_*) = \boldsymbol{\Sigma}_j^{\mathrm{T}}(\boldsymbol{x}_*,\boldsymbol{X})$ is the N-dimensional covariance vector evaluated at all pairs of the training and test data. $s_j(\boldsymbol{x}_*,\boldsymbol{x}_*)$ is the variance of the test data. $\boldsymbol{\Sigma}_j(\boldsymbol{X},\boldsymbol{x}_*)$ and $s_j(\boldsymbol{x}_*,\boldsymbol{x}_*)$ are calculated by the trained covariance function.

From the formula for conditioning a joint Gaussian distribution [17], the posterior distribution for a specific test data is

$$y_*(k_0+j)|\boldsymbol{X},\boldsymbol{y}_j,\boldsymbol{x}_* \sim N(\hat{y}_*(k_0+j), \hat{\sigma}_*^2(k_0+j)) \quad (j = 1,2,\cdots,M) \tag{18}$$

where

$$\begin{aligned}\hat{y}_*(k_0+j) &= m_j(\boldsymbol{x}_*) \\ &\quad + \boldsymbol{\Sigma}_j(\boldsymbol{x}_*,\boldsymbol{X})\boldsymbol{K}_j^{-1}\big(\boldsymbol{y}_j - \boldsymbol{m}_j(\boldsymbol{X})\big)\end{aligned} \tag{19}$$

$$\hat{\sigma}_*^2(k_0+j) = s_j(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{\Sigma}_j(\boldsymbol{x}_*, \boldsymbol{X})\boldsymbol{K}_j^{-1}\boldsymbol{\Sigma}_j(\boldsymbol{X}, \boldsymbol{x}_*) + \sigma_{j[best]}^2$$

Which are the predictive mean and the predictive variance at the j-step ahead, respectively. It is noted that the nonlinearity of the predictive mean can be expressed by the trained hyperparameters even if the prior mean function is set to be a linear combination of the model input as (9).

V. Numerical Simulations

Consider a simplified electric power system [18] described by

$$\begin{cases} \tilde{M}\ddot{\delta}(t) + \tilde{D}\dot{\delta}(t) + P_e = P_{in} \\ P_e = \dfrac{e_t E_{fd}}{X_e}\left(1 + \Delta E_{fd}(t)\right)\sin\delta(t) \\ y(t) = \delta(t) + v(t) \end{cases} \quad (20)$$

where $\tilde{M} = 0.06$, $\tilde{D} = 0.06$, $E_{fd} = 1.0$, $e_t = 1.0$, $X_e = 1.0$, and $P_{in} = 0.8$. These are all per unit values. The training data are sampled with sampling period $T_s = 0.01$ as $u(k) = \Delta E_{fd}(kT_s)$ and $y(k) = y(kT_s)$ at $t = kT_s$. The measurement noise $v(t)$ is zero mean Gaussian noise with standard deviation $\sigma = 0.0036$ (noise to signal ratio (NSR): 1%), $\sigma = 0.0110$ (NSR: 3%), or $\sigma = 0.0185$ (NSR: 5%). The lags for the state vector (3) are chosen as $L_y = 10$ and $L_u = 1$ in the case of $\sigma = 0.0036$, $L_y = 13$ and $L_u = 1$ in the case of $\sigma = 0.0110$, and $L_y = 15$ and $L_u = 1$ in the case of $\sigma = 0.0185$, respectively. The number of the training input and output data is taken to be $N = 300$ for training each j $(j = 1,2,\cdots,M)$ step ahead predictor.

To validate the results of training, the prediction results for 1, 10 and 20 step ahead predictors in the case of $\sigma = 0.0185$ are shown in Figs. 2-4. In these figures, the circles with lines show the predictive mean $\hat{y}_*(k+j)$, the crosses show the measurements (test output) $y_*(k+j)$, and the shaded areas give the double standard deviation confidence interval (95.5% confidence region). From these figures, we can confirm that the error between the test data and the predictive mean is quite small for every step ahead predictors and it does not become so large as the prediction horizon increases.

After training, the multistep ahead prediction up to $M = 20$ step is carried out, where the starting step is set to be $k_0 = 59$ as an example. Figs. 5-7 show the results of the multistep ahead prediction by the proposed method. In these figures, the dotted lines show the true output $\delta_*(k)$. The predictive means $\hat{y}_*(k)$ are quite close to the true output $\delta_*(k)$ for all noise levels. Moreover, the probability that the true measurements $y_*(k)$ are included in the double standard deviation confidence interval is totally 96.7%, which is very close to the expected value 95.5%. This indicates that the proposed prediction method gives the reasonable uncertainty (predictive variance). Therefore, we can say that the proposed multistep ahead prediction can be carried out successfully even in the presence of measurement noise.

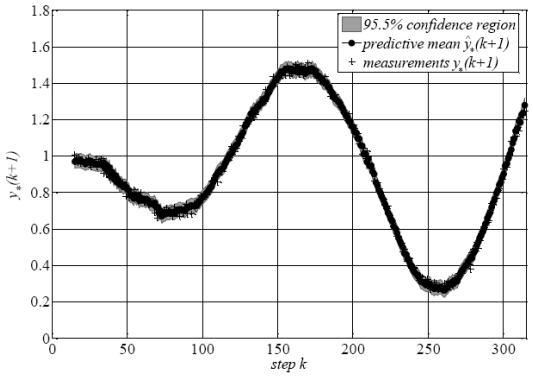

Figure 2. Prediction result for 1 step ahead predictor ($\sigma = 0.0185$).

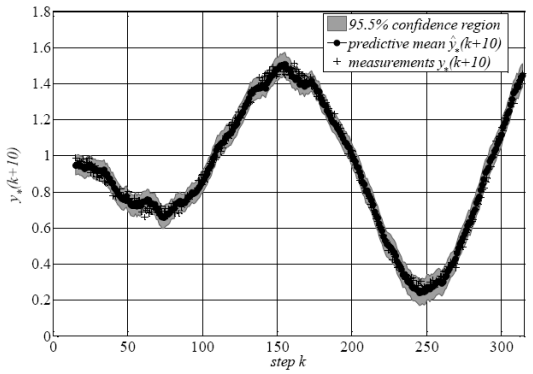

Figure 3. Prediction result for 10 step ahead predictor ($\sigma = 0.0185$).

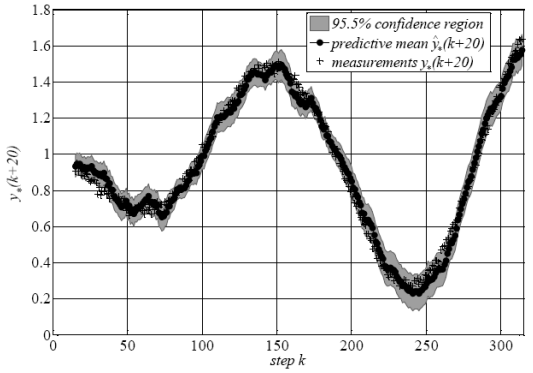

Figure 4. Prediction result for 20 step ahead predictor ($\sigma = 0.0185$).

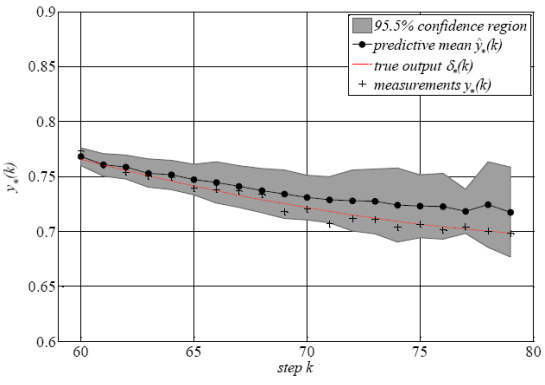

Figure 5. The result of multistep ahead prediction ($\sigma = 0.0036$).

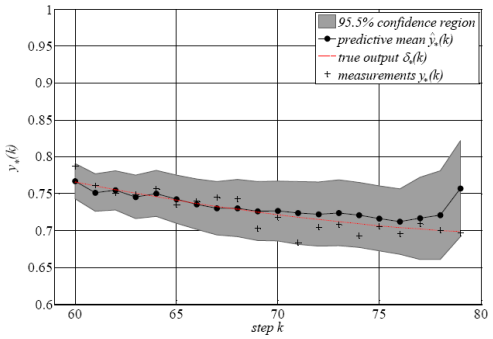

Figure 6. The result of multistep ahead prediction ($\sigma = 0.0110$).

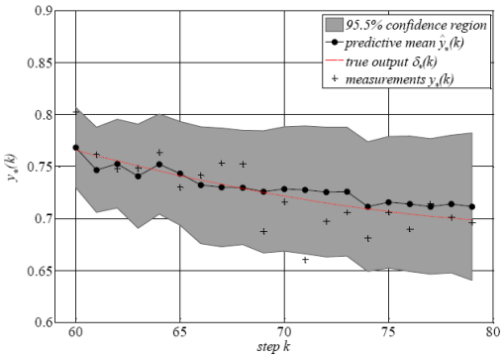

Figure 7. The result of multistep ahead prediction ($\sigma = 0.0185$).

VI. Conclusions

In this paper, we have proposed the multistep ahead prediction of electric power systems by using multiple Gaussian process models. The multistep ahead prediction has been carried out directly by using multiple Gaussian process models as every step ahead predictors. Through the numerical simulations for the simplified electric power system, it has been experimentally demonstrated that the proposed direct method is very accurate even in the presence of measurement noise. Therefore, the proposed prediction method has high potential for MPC. Developing MPC algorithm based on this prediction model is one of the future works.

References

[1] H. Seki, M. Ogawa, S. Ooyama, K. Akamatsu, M. Ohshima, and W. Yang, "Industrial application of a nonlinear model predictive control to polymerization reactors," *Control Engineering Practice*, vol. 9, no. 8, pp. 819-828, 2001.

[2] Z. K. Nagy, B. Mahn, R. Franke, and F. Allgöwer, "Evaluation study of an efficient output feedback nonlinear model predictive control for temperature tracking in an industrial batch reactor," *Control Engineering Practice*, vol. 15, no. 7, pp. 839-850, 2007.

[3] 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.

[4] T. Geyer, "Generalized model predictive direct torque control: Long prediction horizons and minimization of switching losses," in *Proc. of IEEE Conference on Decision and Control*, 2009, pp. 6799-6804.

[5] J. Scoltock, T. Geyer, and U. K. Madawala, "A Comparison of model predictive control schemes for MV induction motor drives," *IEEE Trans. Industrial Informatics*, pp. 909-919, 2013.

[6] A. O'Hagan, "Curve fitting and optimal design for prediction (with discussion)," *J. Royal Statistical Society B*, vol. 40, pp. 1-42, 1978.

[7] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," *Advances in Neural Information Processing Systems*, vol. 8, pp. 514-520, MIT Press, 1996.

[8] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[9] 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.

[10] 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.

[11] 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," *Advances in Neural Information Processing Systems*, vol. 15, pp. 542-552, MIT Press, 2003.

[12] J. Q. Candela, A. Girard, J. Larsen, and C. E. Rasmussen, "Propagation of uncertainty in Bayesian kernel models-application to multiple-step ahead forecasting," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, 2003, pp. 701-704.

[13] T. Hachino and V. Kadirkamanathan, "Multiple Gaussian process models for direct time series forecasting," *IEEJ Trans. Electrical and Electronic Engineering*, vol. 6, no. 3, pp. 245-252, 2011.

[14] J. Kocijan, R. Murray-Smith, C. E. Rasmussen, and B. Likar, "Predictive control with Gaussian process models," in *Proc. IEEE Eurocon 2003: Int. Conference on Computer as a Tool*, 2003, pp. 352-356.

[15] R. Murray-Smith, D. Sbarbaro, C. E. Rasmussen, and A. Girard, "Adaptive, cautions, predictive control with Gaussian process priors," in *Proc. 13th IFAC Symposium on System Identification*, 2003, pp. 1195-1200.

[16] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[17] R. V. Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, 1964.

[18] H. Takata, "An automatic choosing control for nonlinear systems," in *Proc. IEEE Conference on Decision and Control*, 1996, pp. 3453-3458.

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.

Keiji Naritomi received the B.S. degree in electrical and electronics engineering from Kagoshima University in 2014. He is currently a master's student at the Department of Electrical and Electronics Engineering, Kagoshima University. His research interests include nonlinear control and identification.

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.

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.