World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:8, No:2, 2014

Short-Term Electric Load Forecasting Using Multiple Gaussian Process Models

Tomohiro Hachino, Hitoshi Takata, Seiji Fukushima, Yasutaka Igarashi

Abstract—This paper presents a Gaussian process model-based short-term electric load forecasting. The Gaussian process model is a nonparametric model and the output of the model has Gaussian distribution with mean and variance. The multiple Gaussian process models as every hour ahead predictors are used to forecast future electric load demands up to 24 hours ahead in accordance with the direct forecasting approach. The separable least-squares approach that combines the linear least-squares method and genetic algorithm is applied to train these Gaussian process models. Simulation results are shown to demonstrate the effectiveness of the proposed electric load forecasting.

Keywords—Direct method, electric load forecasting, Gaussian process model, genetic algorithm, separable least-squares method.

I. Introduction

RECENTLY, power systems have been more complicated and their uncertainties have been increasing due to the deregulation and liberalization of the electricity market. It is indispensable to forecast electric load demand accurately to operate power systems with high reliability and economy. Short-term electric load forecasting is very important for starting and halting problem of generator, and economical load distribution. So far, many methods for electric load forecasting have been developed using multi-layered neural network models [1]–[3], fuzzy model [4], Kalman filter [5], $\mathcal{H}_\infty$ filter [6], and so forth. However, since these methods are categorized into the parametric forecasting, one needs many weighting parameters to describe the nonlinearity, which makes the training of the prediction model and the determination of the model structure complicated. Moreover, any confidence measures for predicted load demands are not given in such forecasting methods.

To overcome these problem, in this paper, we propose a direct method for short-term electric load forecasting in the Gaussian process (GP) framework. This electric load forecasting is carried out by the GP-based time series forecasting [7]. The GP model was originally utilized for the regression problem by O'Hagan [8], and has recently received much attention for use in regression and classification problems [9]–[11]. Moreover, this model has been introduced for the modeling of nonlinear dynamic systems [12]–[14] and the time series forecasting [7], [15], [16]. Some applications such as human motion modeling [17] and predictive control in gas-liquid separation plants [18] have also been reported by using the GP model. Since the GP model has fewer parameters than parametric models such as the neural network model and fuzzy model, we can describe the nonlinearity between the past electric load and the future electric load by using a few parameters. The uncertainties of the predicted electric load demands are usually not obtained by the non-GP-based method such as the neural network model-based method. The proposed forecasting method gives the predicted electric load demands and the uncertainties of the predicted values as well. This information about uncertainties of the predicted electric load demands must be very useful for reliable management of electric power system. Moreover, in the proposed method, the forecasting is directly performed by using the multiple trained GP models as every hour ahead predictors. As the proposed direct method uses not only one-hour ahead predictor but also all-hours ahead predictors, the prediction errors are not accumulated as the forecasting horizon increases.

To perform electric load forecasting in the GP framework, the GP prior models have to be trained by minimizing the negative log marginal likelihood of the training data. Unfortunately the cost function generally has multiple local minima, therefore, one has to handle a nonlinear optimization method which is very complicated. The gradient based optimization algorithm still suffers from the local minima problem unless the initial guess is suitable. We adopt the genetic algorithm (GA) [19] to train the GP models in this paper. The hyperparameters of covariance functions are coded into binary bit strings in the GA, and the weighting parameters of the prior mean function corresponding to each candidate hyperparameter, are estimated by the linear least-squares (LS) method. This training is based on the separable LS approach [20] which has been utilized for linear and nonlinear system identification [21], [22].

This paper is organized as follows. In section II, the problem of short-term electric load forecasting is formulated. In section III, the multiple GP prior models are derived for every hour ahead predictors. In section IV, the training method of the GP prior models based on the separable LS approach is proposed including the use of the GA. In section V, short-term electric load forecasting by the GP posterior distribution is presented. In section VI, simulation results are shown to illustrate the effectiveness of the proposed forecasting method. Finally,

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).

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).

some conclusions are given in section VII.

II. Statement of the Problem

Assume that a j-hours ahead electric load predictor is described as

$$\begin{aligned} y(k+j) &= f_j(\boldsymbol{x}(k)) + \epsilon_j(k) \qquad (j = 1, 2, \cdots, 24) \\ \boldsymbol{x}(k) &= [y(k), y(k-1), \cdots, y(k-23)]^{\mathrm{T}} \end{aligned} \tag{1}$$

where k denotes the time, $y(k)$ is the electric load, and $y(k+j)$ is the electric load at the j-hours ahead from the time k. $f_j(\cdot)$ is a function which is assumed to be stationary and smooth. $\epsilon_j(k)$ is zero mean Gaussian noise with variance σ_j^2.

The problem of this paper is to construct the following probability distributions for the multiple ahead prediction

$$\begin{aligned} y(k+j)|\boldsymbol{x}(k) \sim \mathcal{N}(\hat{y}(k+j), \hat{\sigma}^2(k+j)) \\ (j = 1, 2, \cdots, 24) \end{aligned} \tag{2}$$

and to carry out electric load forecasting up to 24 hours ahead based on these distributions, by using the GP framework.

III. GP Prior Model

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

$$\boldsymbol{w}_j = \boldsymbol{f}_j + \boldsymbol{\epsilon}_j \tag{3}$$

where

$$\begin{aligned} \boldsymbol{w}_j &= [y(k_s+j), y(k_s+j+1), \\ &\qquad \cdots, y(k_s+j+N-1)]^{\mathrm{T}} \\ \boldsymbol{f}_j &= [f_j(\boldsymbol{x}_1), f_j(\boldsymbol{x}_2), \cdots, f_j(\boldsymbol{x}_N)]^{\mathrm{T}} \\ \boldsymbol{\epsilon}_j &= [\epsilon_j(k_s), \epsilon_j(k_s+1), \\ &\qquad \cdots, \epsilon_j(k_s+N-1)]^{\mathrm{T}} \\ \boldsymbol{X} &= [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]^{\mathrm{T}} \\ &= [\boldsymbol{x}(k_s), \boldsymbol{x}(k_s+1), \cdots, \boldsymbol{x}(k_s+N-1)]^{\mathrm{T}} \\ &= \begin{bmatrix} y(k_s) & y(k_s+1) & \cdots & y(k_s+N-1) \\ y(k_s-1) & y(k_s) & \cdots & y(k_s+N-2) \\ \vdots & \vdots & \cdots & \vdots \\ y(k_s-23) & y(k_s-22) & \cdots & y(k_s+N-24) \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{4}$$

$\boldsymbol{w}_j$ and $\boldsymbol{f}_j$ are the model output vector and the function value vector for the j-hours ahead predictor, respectively. $\boldsymbol{X}$ is the model input matrix and is common for every hour ahead predictors. $\{\boldsymbol{X}, \boldsymbol{w}_j\}$ is the training input and output data for the j-hours 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 function value vector $\boldsymbol{f}_j$ can be represented by the GP as

$$\boldsymbol{f}_j \sim \mathcal{N}(\boldsymbol{m}_j(\boldsymbol{X}), \boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{X})) \tag{5}$$

where $\boldsymbol{m}_j(\boldsymbol{X})$ is the N-dimensional mean function vector and $\boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{X})$ is the N-dimensional covariance matrix evaluated at all pairs of the training data. Equation (5) means that $\boldsymbol{f}_j$ has a Gaussian distribution with mean vector $\boldsymbol{m}_j(\boldsymbol{X})$ and covariance matrix $\boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{X})$.

The mean function is often represented by a polynomial regression [11]. 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 input:

$$\begin{aligned} \boldsymbol{m}_j(\boldsymbol{X}) &= [m_j(\boldsymbol{x}_1), m_j(\boldsymbol{x}_2), \cdots, m_j(\boldsymbol{x}_N)]^{\mathrm{T}} \\ &= \tilde{\boldsymbol{X}} \boldsymbol{\theta}_j \end{aligned} \tag{6}$$

where $\tilde{\boldsymbol{X}} = [\boldsymbol{X}, \boldsymbol{e}]$, $\boldsymbol{e} = [1, 1, \cdots, 1]^{\mathrm{T}}$ is the N-dimensional vector of ones, and $\boldsymbol{\theta}_j = [\theta_{j0}, \theta_{j1}, \cdots, \theta_{jL}]^{\mathrm{T}}$ is the unknown weighting parameter vector of the mean function to be trained. The estimation of $\boldsymbol{\theta}_j$ will be discussed in section IV. It might be natural to use the first order polynomial of the input variable as the mean function in the case that no prior information about the mean function of the process generating time series data is available. This setting does not limit the electric load forecasting because the posterior mean function is modified to appropriate nonlinear function according to electric load data by the trained covariance function (see (20)).

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

$$\boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{X}) = \begin{bmatrix} S_{j(1,1)} & S_{j(1,2)} & \cdots & S_{j(1,N)} \\ S_{j(2,1)} & S_{j(2,2)} & \cdots & S_{j(2,N)} \\ \vdots & \vdots & & \vdots \\ S_{j(N,1)} & S_{j(N,2)} & \cdots & S_{j(N,N)} \end{bmatrix} \tag{7}$$

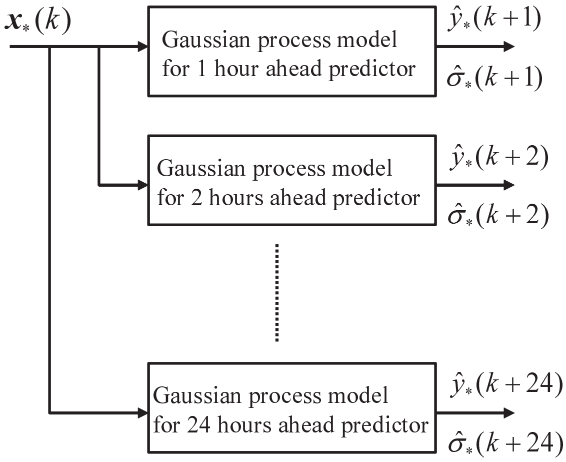
where the element $S_{j(p,q)} = cov(f_j(\boldsymbol{x}_p), f_j(\boldsymbol{x}_q)) = s_j(\boldsymbol{x}_p, \boldsymbol{x}_q)$ is a function of $\boldsymbol{x}_p$ and $\boldsymbol{x}_q$. Under the assumption that the process is stationary and smooth, the following Gaussian kernel is utilized for $S_{j(p,q)}$:

$$\begin{aligned} S_{j(p,q)} &= s_j(\boldsymbol{x}_p, \boldsymbol{x}_q) \\ &= \rho_j^2 \exp\left(-\frac{||\boldsymbol{x}_p - \boldsymbol{x}_q||^2}{2\ell_j^2}\right) \end{aligned} \tag{8}$$

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 (8) and the noise standard deviation σ_j are called *hyperparameters* and construct the hyperparameter vector $\boldsymbol{h}_j = [\rho_j, \ell_j, \sigma_j]^{\mathrm{T}}$. The role of the covariance function of the GP is similar to that of the kernels of the support vector machines. 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 $\boldsymbol{h}_j$ should be appropriately determined according to the training data for precise electric load forecasting. This parameter selection will be also presented in section IV.

Since $\boldsymbol{w}_j$ is noisy observation, we have the following GP model for j-hours ahead prediction from (3) and (5) as

$$\boldsymbol{w}_j \sim \mathcal{N}(\boldsymbol{m}_j(\boldsymbol{X}), \boldsymbol{K}_j(\boldsymbol{X}, \boldsymbol{X})) \tag{9}$$



$$\boldsymbol{x}_*(k) = [y_*(k), y_*(k-1), \cdots, y_*(k-23)]^T$$

Fig. 1 The proposed electric load forecasting scheme

where

$$\begin{aligned} \boldsymbol{K}_j(\boldsymbol{X},\boldsymbol{X}) &= \boldsymbol{S}_j(\boldsymbol{X},\boldsymbol{X}) + \sigma_j^2 \boldsymbol{I}_N \\ \boldsymbol{I}_N &: N \times N \ \ \text{identity} \ \ \text{matrix} \end{aligned} \tag{10}$$

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

IV. Training of GP Prior Model

To perform electric load forecasting, the proposed direct approach needs 1 to 24 hours ahead prediction models as shown in Fig. 1. The accuracy of forecasting greatly depends on the unknown parameter vector $\boldsymbol{\vartheta}_j = [\boldsymbol{\theta}_j^{\mathrm{T}}, \boldsymbol{h}_j^{\mathrm{T}}]^{\mathrm{T}}$ 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:

$$\begin{aligned} J(\boldsymbol{\vartheta}_j) &= -\log P(\boldsymbol{w}_j|\boldsymbol{X},\boldsymbol{\vartheta}_j) \\ &= \frac{1}{2}\log|\boldsymbol{K}_j| + \frac{1}{2}(\boldsymbol{w}_j - \boldsymbol{m}_j(\boldsymbol{X}))^{\mathrm{T}}\boldsymbol{K}_j^{-1} \\ &\quad \times(\boldsymbol{w}_j - \boldsymbol{m}_j(\boldsymbol{X})) + \frac{N}{2}\log(2\pi) \\ &= \frac{1}{2}\log|\boldsymbol{K}_j| + \frac{1}{2}(\boldsymbol{w}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_j)^{\mathrm{T}}\boldsymbol{K}_j^{-1}(\boldsymbol{w}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_j) \\ &\quad + \frac{N}{2}\log(2\pi) \end{aligned} \tag{11}$$

Since the cost function $J(\boldsymbol{\vartheta}_j)$ 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 (11) with respect to the weighting parameter vector $\boldsymbol{\theta}_j$ of the mean function is as follows:

$$\frac{\partial J(\boldsymbol{\vartheta}_j)}{\partial \boldsymbol{\theta}_j} = -\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\boldsymbol{w}_j + \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\tilde{\boldsymbol{X}}\boldsymbol{\theta}_j \tag{12}$$

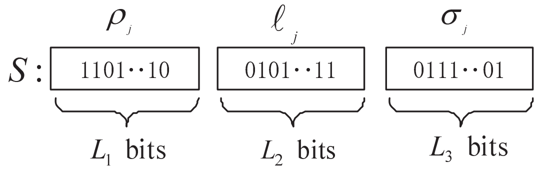


Fig. 2 Coding

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 LS method putting $\partial J(\boldsymbol{\vartheta}_j)/\partial\boldsymbol{\theta}_j = 0$:

$$\boldsymbol{\theta}_j = (\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{K}_j^{-1}\boldsymbol{w}_j \tag{13}$$

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, in this paper, we propose a method that combines the linear LS method with GA based on the idea of the separable LS approach [20]. Only the hyperparameter vector $\boldsymbol{h}_j$ of the covariance is coded into binary bit strings as shown in Fig. 2 and searched by the GA which has a high potential for global optimizations [19].

ρ_j is decoded logarithmically as follows:

$$\begin{aligned} \rho_j &= 10^r \\ r &= \frac{\log_{10}\rho_{max} - \log_{10}\rho_{min}}{2^{L_1}-1}\mathcal{R} + \log_{10}\rho_{min} \end{aligned} \tag{14}$$

where $\mathcal{R}$ is the decimal value of the binary representation of the first block of the string S and $[\rho_{min}, \rho_{max}]$ is the search range of ρ_j. ℓ_j and σ_j are also decoded logarithmically in the same manner.

The proposed training algorithm is as follows:

step 1: Initialization for training

Set $j = 1$ and let the training input data be $\boldsymbol{X}$.

step 2: Preparation of training output data

Let the training output data be $\boldsymbol{w}_j$.

step 3: Initialization for GA

Generate an initial population of Q binary bit strings for the hyperparameter vector $\boldsymbol{h}_j$ randomly.

step 4: Decoding

Decode Q strings into real values $\boldsymbol{h}_{j[i]}$ $(i = 1, 2, \cdots, Q)$ by the decoding method as in (14).

step 5: Construction of covariance matrix

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

step 6: Estimation of $\boldsymbol{\theta}_j$

Estimate Q candidates of the weighting parameter vector $\boldsymbol{\theta}_{j[i]}$ of the mean function corresponding to $\boldsymbol{h}_{j[i]}$ $(i = 1, 2, \cdots, Q)$ from (13).

step 7: Fitness value calculation

Calculate the negative log marginal likelihood of the training data:

$$\begin{aligned} J_{[i]} &= -\log P(\boldsymbol{w}_j|\boldsymbol{X},\boldsymbol{\vartheta}_{j[i]}) \\ &= \frac{1}{2}\log|\boldsymbol{K}_{j[i]}| + \frac{1}{2}(\boldsymbol{w}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_{j[i]})^{\mathrm{T}}\boldsymbol{K}_{j[i]}^{-1} \\ &\quad \times(\boldsymbol{w}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_{j[i]}) + \frac{N}{2}\log(2\pi) \\ &\qquad (i = 1, 2, \cdots, Q) \end{aligned} \tag{15}$$

and the fitness values $F_{[i]} = D - J_{[i]}$ [19], where $\boldsymbol{\vartheta}_{j[i]} = [\boldsymbol{\theta}_{j[i]}^{\mathrm{T}}, \boldsymbol{h}_{j[i]}^{\mathrm{T}}]^{\mathrm{T}}$ and D is a positive constant value.

step 8: Reproduction

Reproduce each of individual strings with the probability of $F_{[i]}/\sum_{\zeta=1}^{Q} F_{[\zeta]}$. Practically, the linear fitness scaling [19] is utilized to avoid undesirable premature convergence.

step 9: Crossover

Select two strings randomly and decide whether or not to cross them over according to the crossover probability P_c. Exchange strings at a crossing position if the crossover is required. The crossing position is chosen randomly.

step 10: Mutation

Alter a bit (0 or 1) of string according to the mutation probability P_m.

step 11: Repetition for GA

Repeat *step 4* $\sim$ *step 10* from generation to generation so that the fitness value of the population increases. In simulations, the genetic operations will be repeated until prespecified G-th generation.

step 12: Determination of the GP prior model

Construct the suboptimal prior mean and prior covariance for the j-hours ahead predictor by using $\boldsymbol{\vartheta}_{j[best]} = [\boldsymbol{\theta}_{j[best]}^{\mathrm{T}}, \boldsymbol{h}_{j[best]}^{\mathrm{T}}]^{\mathrm{T}} = [\boldsymbol{\theta}_{j[best]}^{\mathrm{T}},\ \rho_{j[best]}, \ell_{j[best]}, \sigma_{j[best]}]^{\mathrm{T}}$ with the best fitness value over all the past generations:

$$m_j(\boldsymbol{x}) = [\boldsymbol{x}^{\mathrm{T}}, 1]\boldsymbol{\theta}_{j[best]} \tag{16}$$

$$\begin{cases} s_j(\boldsymbol{x}_p, \boldsymbol{x}_q) = \rho_{j[best]}^2 \exp\left(-\dfrac{||\boldsymbol{x}_p - \boldsymbol{x}_q||^2}{2\ell_{j[best]}^2}\right) \\ k_j(\boldsymbol{x}_p, \boldsymbol{x}_q) = s_j(\boldsymbol{x}_p, \boldsymbol{x}_q) + \sigma_{j[best]}^2\delta_{pq} \end{cases} \tag{17}$$

where $s_j(\boldsymbol{x}_p, \boldsymbol{x}_q)$ is an element of the covariance matrix $\boldsymbol{S}_j$, $k_j(\boldsymbol{x}_p, \boldsymbol{x}_q)$ is an element of the covariance matrix $\boldsymbol{K}_j$, and δ_{pq} is a Kronecker delta which is 1 if $p = q$ and 0 otherwise.

step 13: Repetition for the GP prior model

If $j < 24$ then $j = j + 1$ and go to *step 2*.

V. Electric Load Forecasting by GP Model

In section IV, we have already obtained the GP prior models for j $(j = 1, 2, \cdots, 24)$ hours ahead predictors. In the proposed direct approach, short-term electric load forecasting up to 24 hours ahead is carried out directly using every GP prior models as shown in Fig. 1.

For a new given test input $\boldsymbol{x}_* = \boldsymbol{x}_*(k) = [y_*(k), y_*(k-1), \cdots, y_*(k-23)]^{\mathrm{T}}$ and corresponding test output $y_*(k+j)$ $(j = 1, 2, \cdots, 24)$, we have the following the joint Gaussian distribution:

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{w}_j \\ y_*(k+j) \end{bmatrix} \\ &\sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m}_j(\boldsymbol{X}) \\ m_j(\boldsymbol{x}_*) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}_j & \boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{x}_*) \\ \boldsymbol{S}_j(\boldsymbol{x}_*, \boldsymbol{X}), & s_j(\boldsymbol{x}_*, \boldsymbol{x}_*) + \sigma_{j[best]}^2 \end{bmatrix}\right) \\ &\qquad (j = 1, 2, \cdots, 24) \end{aligned} \tag{18}$$

where $\boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{x}_*) = \boldsymbol{S}_j^{\mathrm{T}}(\boldsymbol{x}_*, \boldsymbol{X})$ 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. $\boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{x}_*)$ and $s_j(\boldsymbol{x}_*, \boldsymbol{x}_*)$ are calculated by the trained covariance function (17).

From the formula for conditioning a joint Gaussian distribution [23], the posterior distribution for a specific test data is

$$\begin{aligned} &y_*(k+j)|\boldsymbol{X}, \boldsymbol{w}_j, \boldsymbol{x}_* \sim \mathcal{N}(\hat{y}_*(k+j), \hat{\sigma}_*^2(k+j)) \\ &\qquad (j = 1, 2, \cdots, 24) \end{aligned} \tag{19}$$

where

$$\begin{aligned} \hat{y}_*(k+j) &= m_j(\boldsymbol{x}_*) \\ &\quad + \boldsymbol{S}_j(\boldsymbol{x}_*, \boldsymbol{X})\boldsymbol{K}_j^{-1}(\boldsymbol{w}_j - \boldsymbol{m}_j(\boldsymbol{X})) \\ \hat{\sigma}_*^2(k+j) &= s_j(\boldsymbol{x}_*, \boldsymbol{x}_*) \\ &\quad - \boldsymbol{S}_j(\boldsymbol{x}_*, \boldsymbol{X})\boldsymbol{K}_j^{-1}\boldsymbol{S}_j(\boldsymbol{X}, \boldsymbol{x}_*) + \sigma_{j[best]}^2 \end{aligned} \tag{20}$$

are the predictive mean and the predictive variance at the j-hours 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 input as (6).

VI. Simulations

Short-term electric load forecasting is performed for the Kanto area in Japan using the proposed forecasting method. The training data is downloaded from the TEPCO ELECTRICITY FORECAST [24] released by Tokyo Electric Power Company. The electric load demands in 2011 are utilized for training data. The number of the training input and output data is taken to be $N = 200$ for training each j $(j = 1, 2, \cdots, 24)$ hours ahead predictor. Electric load forecasting up to 24 hours ahead is carried out for each season. The design parameters of the GA are given as follows:

- population size: $Q = 30$
- string length: $L_1 = L_2 = L_3 = 10$
- crossover probability: $P_c = 0.8$
- mutation probability: $P_m = 0.03$
- search range of ρ_j: $[\rho_{min}, \rho_{max}] = [10^{-3}, 10^2]$
- search range of ℓ_j: $[\ell_{min}, \ell_{max}] = [10^{-3}, 10^2]$
- search range of σ_j: $[\sigma_{min}, \sigma_{max}] = [10^{-3}, 10^2]$
- termination criteria $G = 100$-th generation

Figs. 3 – 6 show the results of electric load forecasting on February, May, August, and November, in 2012, respectively. They are typically chosen in each 3 months from 4 seasons. Although the predicted electric demands on February have small errors to the actual demands, the predicted electric

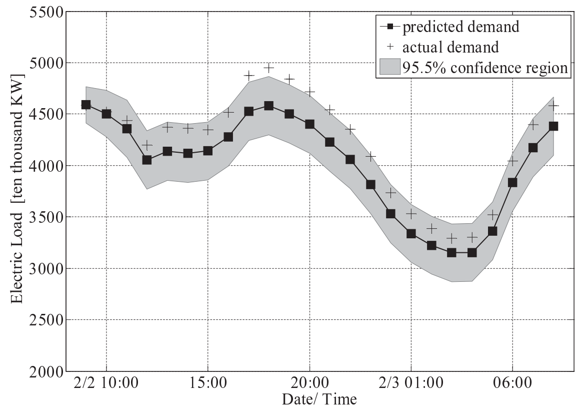


Fig. 3 Electric load forecasting result (February, 2012)

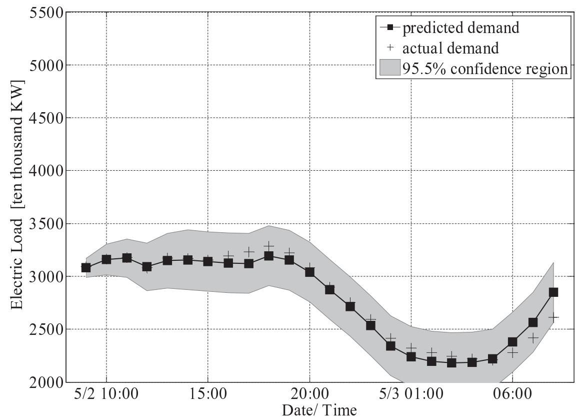


Fig. 4 Electric load forecasting result (May, 2012)

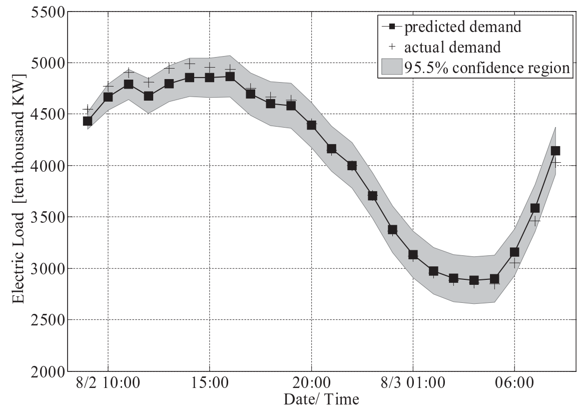


Fig. 5 Electric load forecasting result (August, 2012)

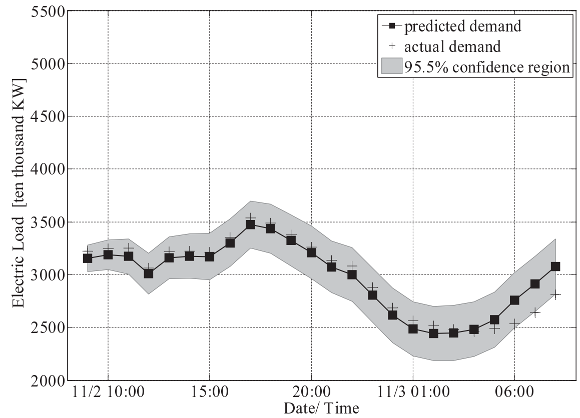


Fig. 6 Electric load forecasting result (November, 2012)

demands on May, August, and November are very close to the actual demands. Moreover, the 95.5% confidence regions are quite reasonable for all seasons. Note that these uncertainties for the predicted electric demands are usually not obtained by the non-GP-based method such as the neural network model-based method. Since the proposed forecasting method gives not only the predicted electric demands but also the uncertainties, we can utilize the upper value of the confidence region $\hat{y}_{max}(k+j) = \hat{y}_*(k+j) + 2\hat{\sigma}_*(k+j)$ as the maximum value of the predicted electric demand. This information must be useful for management of electric power.

VII. Conclusions

In this paper, a GP model-based short-term electric load forecasting has been proposed. The short-term electric load forecasting has been carried out directly by using multiple GP model as every hour ahead predictors. The separable LS approach combining the linear LS method with GA has been proposed to train the GP prior model. Simulation results show that the proposed forecasting method can give accurate predicted electric load demand and the uncertainty of the predicted values as well. Development of the forecasting method that is taking the weather data and the type of date into consideration is one of the future works.

References

[1] T. Ishida and S. Tamaru, Daily electric load forecasting using artificial neural network (in Japanese), *IEEJ Trans. B*, Vol. 114, No. 11, pp. 1109–1115, 1994.

[2] H. Takata, K. Sonoda and T. Hachino, Daily electric load forecasting using ANN and GA in Tanegashima (in Japanese), *The Research Reports of the Faculty of Engineering Kagoshima University*, No. 40, pp. 55–58, 1998.

[3] O. Ishioka, Y. Sato, T. Ishihara, Y. Ueki, T. Matsui and T. Iizaka, Development of electric load forecasting system using neural networks (in Japanese), *IEEJ Trans. B*, Vol. 120, No. 12, pp. 1550–1556, 2000.

[4] H. Mori and H. Kobayashi, Optimal fuzzy inference for short-term load forecasting, *IEEE Trans. Power Syst.*, Vol. 11, No. 1, pp. 350–356, 1996.

[5] H. M. A. Hamadi and S. A. Soliman, Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model, *Trans. on Electric Power Systems Research*, No. 68, pp. 47–59, 2004.

[6] T. Namerikawa and Y. Hosoda, $\mathcal{H}_\infty$ filter-based short-term electric load prediction considering characteristics of load curve (in Japanese), *IEEJ Trans. C*, Vol. 132, No. 9, pp. 1446–1453, 2012.

[7] 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.

[8] 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.

[9] 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.

[10] M. Seeger, Gaussian processes for machine learning, *International Journal of Neural Systems*, Vol. 14, No. 2, pp. 1–38, 2004.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.
[12] 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.
[13] G. Gregorčič and G. Lightbody, Gaussian process approach for modelling of nonlinear systems, *Engineering Applications of Artificial Intelligence*, Vol. 22, No. 4-5, pp. 522–533, 2009.
[14] 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.
[15] A. Girard, C. E. Rasmussen, J. Q. Candela and R. Murray-Smith, Gaussian process priors with uncertain inputs -application to mutiple-step ahead time series forecasting", in *Advances in Neural Information Processing Systems*, Vol. 15, pp. 542–552, MIT Press, 2003.
[16] J. Q. Candela, A. Girard, J. Larsen and C. E. Rasmussen, Propagation of uncertainty in Bayesian kernel models -application to multiple-step ahead forecasting, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. II-701–704, 2003.
[17] 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.
[18] 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.
[19] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.
[20] G. H. Golub and V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, *SIAM Journal of Numerical Analysis*, Vol. 10, No. 2, pp. 413–432, 1973.
[21] J. Bruls, C. T. Chou, B. R. J. Haverkamp and M. Verhaegen, Linear and non-linear system identification using separable least-squares, *European Journal of Control*, Vol. 5, pp. 116–128, 1999.
[22] F. Previdi and M. Lovera, Identification of non-linear parametrically varying models using separable least squares, *International Journal of Control*, Vol. 77, No. 16, pp. 1382–1392, 2004.
[23] R. von Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, 1964.
[24] TEPCO ELECTRICITY FORECAST, Tokyo Electric Power Company, http://www.tepco.co.jp/forecast/index-j.html, accessed November 22, 2012.

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 control and identification for nonlinear systems. 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.

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.