Gaussian Process Model-based Short-term Electric Load Forecasting Using Cuckoo Search

TOMOHIRO HACHINO and YOSHIHIRO OKUYA
Kagoshima University
Department of Electrical and Electronics Engineering
1-21-40 Korimoto, Kagoshima 890-0065
JAPAN
hachino@eee.kagoshima-u.ac.jp

Abstract: This paper deals with a Gaussian process model-based short-term electric load forecasting using cuckoo search. 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 cuckoo search is applied to train these Gaussian process models. The results of electric load forecasting for Kyushu district in Japan are shown to demonstrate the effectiveness of the proposed forecasting method.

Key–Words: Short-term electric load forecasting, Gaussian process model, Cuckoo search, Separable least-squares method

1 Introduction

Recently, the independent power producers and the distributed power generators have been increasing. The power systems have been more complicated due to the deregulation and liberalization of the electricity market. It is necessary to forecast electric load demand accurately to operate power systems with high reliability, efficiency and economy. Electric load forecasting is very important for stable starting and halting of generators and reliable load distribution. So far, many methods for electric load forecasting have been developed using multi-layered neural network models [1], fuzzy model [2], Kalman filter [3], H_∞ filter [4], and so forth. These methods are categorized into the parametric forecasting. One needs many weighting parameters to describe the nonlinearity, which makes the training and structure determination of the prediction model complicated. Moreover, any confidence measures of predicted load demands are not given in such forecasting methods.

To overcome these problems, this paper deals with a direct method for short-term electric load forecasting in the Gaussian process (GP) framework. The GP model was originally utilized for the regression problem by O'Hagan [5], and has recently received much attention for use in regression and classification problems [6, 7]. Moreover, this model has been introduced for the modeling of nonlinear dynamic systems [8, 9] and the time series forecasting [10, 11]. Since the GP model has fewer parameters than parametric models, we can easily describe the nonlinearity between the input and output of the prediction model by using a few parameters. The proposed forecasting method gives the predicted electric load demands and the uncertainties of the predicted values as well. The information on the 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. Therefore, the prediction errors are not accumulated as the forecasting horizon increases in the proposed forecasting method.

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 applied the separable least-squares (LS) approach that combines the linear LS method and genetic algorithm (GA) to this problem [12]. However, the GA has many setting parameters and requires a complicated coding technique and genetic operations. In this paper, the separable LS approach that combines the linear LS method and cuckoo search (CS) is applied to train these GP models. The CS is a probabilistic search procedure, which is inspired by the brood parasitic behavior of cuckoo and the Lévy flight behavior of some birds [13]. The CS consists of only the basic arithmetic operations and does not require complicated coding and genetic operations such as crossovers and mutations of the GA. This algorithm has been empirically shown to be very efficient for optimization [13, 14]. These advantages suggest that the use of the CS increases efficiency without deterioration of accuracy for electric load forecasting. In the proposed training algorithm, the hyperparameters of covariance functions are represented by the nests of host birds and the weighting parameters of the prior mean function corresponding to each candidate of hyperparameter, are estimated by the linear LS method.

This paper is organized as follows. In section 2, the problem of short-term electric load forecasting is formulated. In section 3, the multiple GP prior models are derived for every hour ahead predictors. In section 4, the training algorithm of the GP prior models based on CS is proposed. In section 5, short-term electric load forecasting by the GP posterior distribution is described. In section 6, the results of electric load forecasting for Kyushu district in Japan are shown to illustrate the effectiveness of the proposed method. Finally, some conclusions are given in section 7.

2 Statement of the Problem

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

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

where k denotes the time, $y(k)$ is the electric load at the time k, and $y(k+j)$ is the electric load at the j-hours ahead from the time k. $t(k)$ is the temperature at 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.

3 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

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

$\boldsymbol{w}_j$ and $\boldsymbol{f}_j$ are the vector of model outputs and the vector of function values 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 vector of function values $\boldsymbol{f}_j$ can be represented by the GP as

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

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 (5) 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 [7]. 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:

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

where $\tilde{\boldsymbol{X}} = [\boldsymbol{X}, \boldsymbol{e}]$, $\boldsymbol{e} = [1, 1, \cdots, 1]^{\mathrm{T}}$ is the N-dimensional vector consisting of ones, and $\boldsymbol{\theta}_j = [\theta_{j0}, \theta_{j1}, \cdots, \theta_{j48}]^{\mathrm{T}}$ is the unknown weighting parameter vector of the mean function to be trained.

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

$$\boldsymbol{\Sigma}_j(\boldsymbol{X}, \boldsymbol{X}) = \begin{bmatrix} \Sigma_{j(1,1)} & \cdots & \Sigma_{j(1,N)} \\ \vdots & \ddots & \vdots \\ \Sigma_{j(N,1)} & \cdots & \Sigma_{j(N,N)} \end{bmatrix} \tag{7}$$

where the element $\Sigma_{j(p,q)} = cov(f_j(\boldsymbol{x}_p), f_j(\boldsymbol{x}_q)) = s_j(\boldsymbol{x}_p, \boldsymbol{x}_q)$ 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 $\Sigma_{j(p,q)}$:

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

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 $\boldsymbol{h}_j = [\rho_j, \ell_j, \sigma_j]^{\mathrm{T}}$. 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.

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

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

where

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

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.

4 Training of GP Prior Model by CS

To perform electric load forecasting, the proposed direct approach needs 1 to 24 hours ahead prediction models. 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:

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

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

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

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 CS based on the idea of the separable LS approach. The candidates of hyperparameter vector $\boldsymbol{h}_j$ of the covariance are represented by the nests of host birds and searched for by CS, where the candidates of the weighting parameter vector $\boldsymbol{\theta}_j$ are estimated by the linear LS method. The proposed training algorithm is described 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 CS

Generate an initial population of Q host nests $\boldsymbol{\Omega}_{j[i]}$ $(i = 1, 2, \cdots Q)$ for the hyperparameter vector $\boldsymbol{h}_{j[i]}$ randomly. Set the iteration counter l to 0.

step 4: 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 5: 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 6: Evaluation value calculation

Calculate the negative log marginal likelihood of the training data:

$$\begin{aligned} J_{[i]}(\boldsymbol{\Omega}_{j[i]}) &= -\log p(\boldsymbol{w}_j|\boldsymbol{X}, \boldsymbol{\vartheta}_{j[i]}) \\ &= \frac{1}{2}\log|\boldsymbol{K}_{j[i]}| + \frac{1}{2}(\boldsymbol{w}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_{j[i]})^{\mathrm{T}}\boldsymbol{K}_{j[i]}^{-1} \\ &\quad \times(\boldsymbol{w}_j - \tilde{\boldsymbol{X}}\boldsymbol{\theta}_{j[i]}) + \frac{N}{2}\log(2\pi) \\ &\qquad\qquad (i = 1, 2, \cdots, Q) \end{aligned} \tag{14}$$

where $\boldsymbol{\vartheta}_{j[i]} = [\boldsymbol{\theta}_{j[i]}^{\mathrm{T}}, \boldsymbol{h}_{j[i]}^{\mathrm{T}}]^{\mathrm{T}}$.

step 7: Update of host nests

(7-1) Determine the new host nests $\boldsymbol{V}_{j[i]}$ $(i = 1, 2, \cdots Q)$ by Lévy flights:

$$\boldsymbol{V}_{j[i]} = \boldsymbol{\Omega}_{j[i]} + \alpha\boldsymbol{d}_{j[i]} \tag{15}$$

where $\alpha > 0$ is the step size and $\boldsymbol{d}_{j[i]}$ is a random value vector which follows a Lévy distribution:

$$p(\boldsymbol{d}_{j[i]}) = \boldsymbol{d}_{j[i]}^{-\lambda} \quad (1 < \lambda \leq 3) \tag{16}$$

(7-2) Calculate the evaluation values $J_{[i]}(\boldsymbol{V}_{j[i]})$ $(i = 1, 2, \cdots Q)$ in the same way as step 4 $\sim$ step 6.

(7-3) If $J_{[i]}(\boldsymbol{\Omega}_{j[i]}) > J_{[i]}(\boldsymbol{V}_{j[i]})$, update $\boldsymbol{\Omega}_{j[i]}$ with $\boldsymbol{V}_{j[i]}$.

step 8: Reconstruction of host nests

Reconstruct the host nest randomly when the host bird discovers a cuckoo's egg in her nest with probability P_a.

step 9: Repetition for CS

Set the iteration counter to $l = l + 1$ and go to step 4 until the prespecified iteration number l_{max}.

step 10: 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 evaluation value over all the past iterations:

$$m_j(\boldsymbol{x}) = [\boldsymbol{x}^{\mathrm{T}}, 1]\boldsymbol{\theta}_{j[best]} \tag{17}$$

$$\begin{cases} s_j(\boldsymbol{x}_p, \boldsymbol{x}_q) = \rho_{j[best]}^2 \exp\left(-\dfrac{||\boldsymbol{x}_p - \boldsymbol{x}_q||^2}{2\ell_{j[best]}^2}\right) \\ k_j(\boldsymbol{x}_p, \boldsymbol{x}_q) = s_j(\boldsymbol{x}_p, \boldsymbol{x}_q) + \sigma_{j[best]}^2\delta_{pq} \end{cases} \tag{18}$$

where $s_j(\boldsymbol{x}_p, \boldsymbol{x}_q)$ is an element of the covariance matrix $\boldsymbol{\Sigma}_j$, $k_j(\boldsymbol{x}_p, \boldsymbol{x}_q)$ 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 11: Repetition for the GP prior model

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

5 Electric Load Forecasting by GP Model

In section 4, 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.

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

$$\begin{bmatrix} \boldsymbol{w}_j \\ y_*(k+j) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m}_j(\boldsymbol{X}) \\ m_j(\boldsymbol{x}_*) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K}_j & \boldsymbol{\Sigma}_j(\boldsymbol{X}, \boldsymbol{x}_*) \\ \boldsymbol{\Sigma}_j(\boldsymbol{x}_*, \boldsymbol{X}), & s_j(\boldsymbol{x}_*, \boldsymbol{x}_*) + \sigma_{j[best]}^2 \end{bmatrix}\right) \quad (j = 1, 2, \cdots, 24) \tag{19}$$

where $\boldsymbol{\Sigma}_j(\boldsymbol{X}, \boldsymbol{x}_*) = \boldsymbol{\Sigma}_j^{\mathrm{T}}(\boldsymbol{x}_*, \boldsymbol{X})$ 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 (18).

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

$$y_*(k+j)|\boldsymbol{X}, \boldsymbol{w}_j, \boldsymbol{x}_* \sim \mathcal{N}(\hat{y}_*(k+j), \hat{\sigma}_*^2(k+j)) \quad (j = 1, 2, \cdots, 24) \tag{20}$$

where

$$\begin{aligned} \hat{y}_*(k+j) &= m_j(\boldsymbol{x}_*) \\ &\quad + \boldsymbol{\Sigma}_j(\boldsymbol{x}_*, \boldsymbol{X})\boldsymbol{K}_j^{-1}(\boldsymbol{w}_j - \boldsymbol{m}_j(\boldsymbol{X})) \\ \hat{\sigma}_*^2(k+j) &= s_j(\boldsymbol{x}_*, \boldsymbol{x}_*) \\ &\quad - \boldsymbol{\Sigma}_j(\boldsymbol{x}_*, \boldsymbol{X})\boldsymbol{K}_j^{-1}\boldsymbol{\Sigma}_j(\boldsymbol{X}, \boldsymbol{x}_*) + \sigma_{j[best]}^2 \end{aligned} \tag{21}$$

are the predictive mean and the predictive variance at the j-hours ahead, respectively. It is noted that the

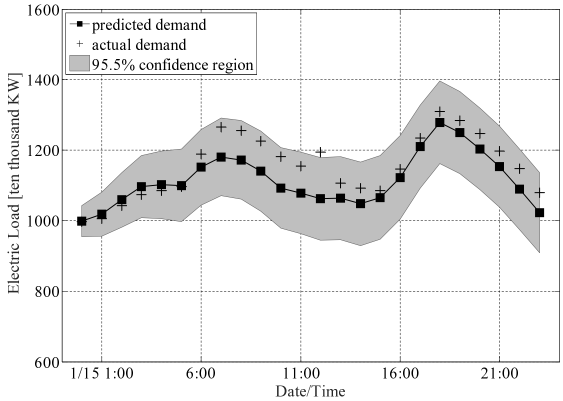

Figure 1: Electric load forecasting result (January, 2016)

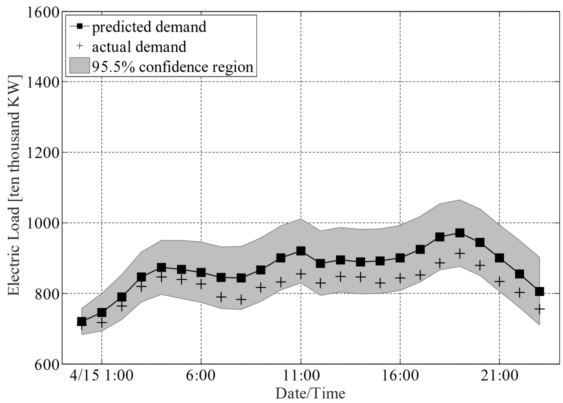

Figure 2: Electric load forecasting result (April, 2016)

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).

6 Electric Load Forecasting for Kyushu District in Japan

Short-term electric load forecasting is performed for Kyushu district in Japan using the proposed forecasting method. The training data is downloaded from the Denki Yohou (Electricity Forecast) released by Kyushu Electric Power Company [15] and Past Weather Data released by Japan Meteorological Agency [16]. The electric load demands in 2015 are utilized for training data. The temperature in Fukuoka, the central city of Kyushu district, is also used for training data. The number of the training input and output data is taken to be $N = 649$ for training

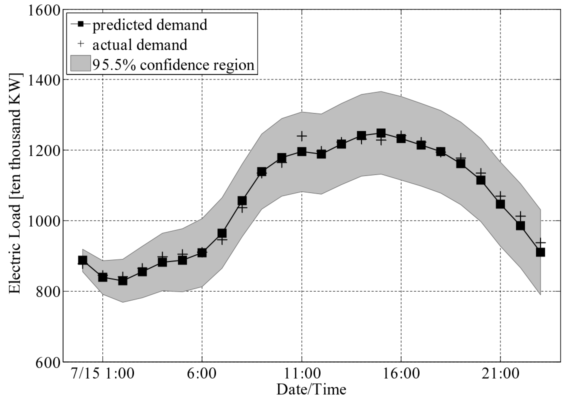

Figure 3: Electric load forecasting result (July, 2016)

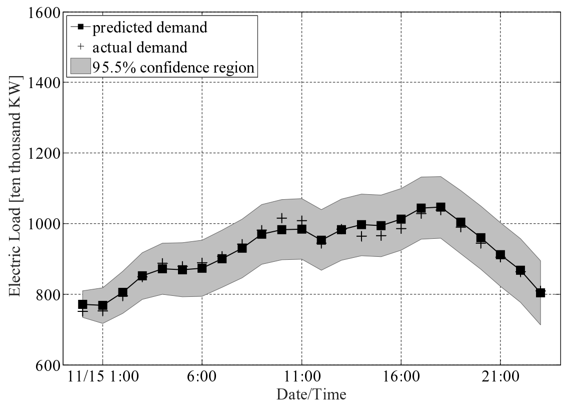

Figure 4: Electric load forecasting result (November, 2016)

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 CS are given as follows:

nest size: $Q = 30$
egg discovery probability: $P_a = 0.125$
termination criteria: $l_{max} = 50$

Figures 1-4 show the results of electric load forecasting on January, April, July, and November, in 2016, respectively. They are typically chosen from 4 seasons. Although the predicted electric demands on January have small errors to the actual demands, the predicted electric demands on April, July, and November are quite 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 practically 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 very useful for management of electric power system.

7 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 CS has been proposed to train the GP prior model. The hyperparameters of covariance functions are represented by the nests of host birds and the weighting parameters of the prior mean function corresponding to each candidate hyperparameter, are estimated by the linear LS method. Forecasting results show that the proposed forecasting method can give accurate predicted electric load demand and the uncertainty of the predicted values as well. This information on the uncertainties of the predicted electric load demands must be very useful for reliable management of electric power system. Development of the forecasting method that is taking another weather data and type of date (weekday or holiday) into consideration is one of the future works.

Acknowledgements: This work was supported by JSPS KAKENHI Grant Number 15K06113.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] T. Hachino, H. Takata, S. Fukushima, Y. Igarashi, Short-term electric load forecasting using multiple Gaussian process models, *International Journal of Electrical and Computer Engineering*, Vol. 8, No. 2, pp. 447–452, 2014.

[13] X.-S. Yang and S. Deb, Cuckoo search via Lévy flights, *Proc. of World Congress on Nature & Biologically Inspired Computing*, pp. 210–214, 2009.

[14] S. Ikeda and R. Ooka, Development of metaheuristic optimization methods for operating energy systems including residual battery and thermal storage tank (in Japanese), *J. Environ. Eng.*, Vol. 79, No. 705, pp. 957–966, 2014.

[15] Denki Yohou (Electricity Forecast), Kyushu Electric Power Co., Inc.
http://www.kyuden.co.jp/power_usages/pc.html

[16] Past Weather Data, Japan Meteorological Agency
http://www.data.jma.go.jp/obd/stats/etrn/