

# Short-term wind power prediction based on GPR-BSO model

Tao Chen<sup>1\*</sup>, Xinjian Li<sup>2</sup>, Zhemeng Zhang<sup>1</sup>, Tongguang Yang<sup>3</sup>, Shengtao He<sup>1</sup>, Xiwen Shao<sup>1</sup>, Jing Liao<sup>1</sup>

<sup>1</sup> College of Electrical and Information Engineering, Hunan University, Changsha, Hunan Province, 410082, China

<sup>2</sup> Yiyang Power Supply Company, State Grid Hunan Electric Power Company, Yiyang, Hunan Province, 413000, China

<sup>3</sup> College of Machinery and Electrical Engineering, Hunan City University, Yiyang, Hunan Province, 413049, China

**Abstract.** Wind power forecasting is a crucial part for the safe and stable operation of wind power integration, which is under the influence of different factors such as wind speed, wind direction, atmospheric pressure. These factors bring randomness and volatility to wind power which makes it less predictable. While, there are very limited studies on describing the uncertainty of wind power. Therefore, to providing additional information on the uncertainty and volatility, a kernel-based on Gaussian Process Regression (GPR) incorporating the hyper-parameters intelligent optimization method is proposed in this paper. Firstly, the hyper-parameters solution of GPR is formulated as a nonlinear optimization with constraints. Then, an intelligent algorithm named Brain-storming optimization (BSO) is adopted to obtain the optimal hyper-parameters of GPR. Furthermore, the performance is examined on short-term wind power data. Most importantly, the GPR incorporating BSO can avoid the hyper-parameters at local optimum.

## 1 Introduction

The uncertainty and stochastic nature of wind power has brought significant challenges to power system operation and planning [1]. The wind power forecasting can mitigate the negative impact on the grid, which is critical to the security and economics of the power system operation [2]. Currently, wind power forecasting methods fall into two main categories, statistical and physical methods. Among them, the physical method is based on surface information and meteorological data as the initial boundary conditions, using mathematical models for inference to solve [3]-[5]. Statistical methods are used to build statistical models from the characteristics of historical data to further predict wind power at future times and mainly include time series methods and artificial intelligence methods [6]-[8]. As all methods have different aspects of error, weighting various forecasting methods or fusing the advantages of different methods to construct combined forecasting models is also a research direction aimed at reducing the upper limit of error of a single method.

The focus of this paper is on the study of statistical method for wind power forecasting, such as multivariate linear regression (MLR), Autoregressive Integrated Moving Average (ARIMA) [9], artificial neural network (ANN), support vector regression (SVR) [10], neuron-fuzzy system (NFS). The main process of this network is clustering the data in the first step, then chose the best data as the training data to the SVM to fit the different situation. Comparing with the New York Independent System Operator and single SVM result, the hybrid network outperforms the ISO and single SVM in MAE and MAPE.

The above methods can be summarized as linear and nonlinear methods. Simple linear regression and multivariable linear regression were used to forecast the load. Reddy [11] predicted the short-term load with the linear regression technique and found its coefficients. Amral [12] used a Multiple Linear Regression method for the short-term load forecasting, but this model was very sensitive to the temperature and also affected by other weather factors, therefore, it requires a very accurate temperature forecast and error coefficients to increase the accuracy of forecasting [12].

Besides, another nonparametric machine learning named Gaussian Process Regression (GPR) has been successfully applied in power load forecasting [13]. The superiority of GPR comes from the use of the Bayesian Inference theory, which evaluates the confidence interval and indicates where the model is unreliable. However, the determination of the hyper-parameters needs to solve a non-convex optimization, which easily traps in local optimum [14]. Xiao et al. optimized the hyper-parameters of GPR using Particle Swarm Optimization (PSO) [15]. However, the PSO has characteristics of slow convergence rate and early-maturing problem, which affect the training speed and accuracy of the PSO-GPR algorithm. The contribution of this paper is to propose a new methodology for short-term load forecasting using GPR with Brainstorming Optimization (BSO), and its demonstration in a real-load case study.

This paper is organized as follows: In section 2, the theories of GPR and BSO are introduced respectively. In section 3, the nonlinear optimization with constraints is formulated for hyper-parameters of GPR, and the BSO approach used to design the optimal solution is described. In section 4, a real case study are used to test the proposed

\*Corresponding author's e-mail: itachitao@outlook.com

methodology. Finally, the main conclusions are summarized in section 4.

## 2Methodology

### 2.1Gaussian Process Regression

The gaussian process is also called normal stochastic process from the perspective of probability theory. The main idea is to convert point predictions into predictions of posterior means and variance. The Gaussian process regression models a function to generate a nonparametric model directly, which is regarded as a supervised learning algorithm widely used in many kinds of fields in the past decades.

Given a input vector  $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ , where  $x_i \in \mathfrak{R}^d, X \in \mathfrak{R}^{n \times d}$ , and the corresponding output matrix  $Y = \{y_1, y_2, \dots, y_n\}$ , the unknown input-output mapping  $f: \mathfrak{R}^d \rightarrow \mathfrak{R}$  is as follows,

$$Y = f(X) + \varepsilon \quad (1)$$

Where:  $\varepsilon \sim N(0, \sigma_n^2)$ . The general GPR model is defined as:

$$f(X) \sim GP(m(X), K(X, X)) \quad (2)$$

$$m(X) = E[f(X)] \quad (3)$$

$$K(X, X) = E[(f(X) - m(X))(f(X) - m(X))] \quad (4)$$

Generally,  $m(X)$  sets to zero mean function. The output vector  $Y$  is determined due to  $\varepsilon \sim N(0, \sigma_n^2)$  as follows,

$$Y \sim GP(0, K(X, X) + \sigma_n^2 I) \quad (5)$$

Assuming that there is a new query point  $X^*$ . The joint distribution of the training output  $Y$  and the query output  $Y^*$  is formulated as:

$$\begin{bmatrix} Y \\ Y^* \end{bmatrix} \sim N \left( 0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix} \right) \quad (6)$$

According to the Bayesian rules,  $p(Y^* | X^*, Y, X)$  can be calculated as follows,

$$p(Y^* | X^*, Y, X) \sim N(\bar{m}(X), \Sigma + \sigma_n^2 I)$$

$$s.t. \begin{cases} \bar{m}(X) = K(X^*, X)(K(X, X) + \sigma_n^2 I)^{-1} Y + m(X^*) \\ \Sigma = K(X^*, X^*) - K(X^*, X)(K(X, X) + \sigma_n^2 I)^{-1} K(X, X^*) \end{cases} \quad (7)$$

$$p(Y^* | X^*, Y, X) \sim N(\bar{m}(X), \Sigma + \sigma_n^2 I) \quad (8)$$

$$\bar{m}(X) = K(X^*, X)(K(X, X) + \sigma_n^2 I)^{-1} Y + m(X^*) \quad (9)$$

$$\Sigma = K(X^*, X^*) - K(X^*, X)(K(X, X) + \sigma_n^2 I)^{-1} K(X, X^*) \quad (10)$$

The covariance function  $K(X, X)$  which commonly called the kernel function describing the correlation of  $X$ . The kernel function maps the nonlinear relationship between sampling to the high-dimensional feature space

to convert it into a linear relationship, which simplifies the process. The squared exponential covariance function is universally used which is formulated as follows,

$$K(x_i, x_j) = \sigma_f^2 \exp \left( -\frac{1}{2l^2} (x_i - x_j)^T \Lambda^{-1} (x_i - x_j) \right) \quad (11)$$

where  $x_i$  and  $x_j$  are two input vectors,  $\sigma_f^2$  is signal variance, which controls the smoothness of function,  $l$  is the characteristic length scale which controls vertical fluctuation of function. So the hyper-parameter  $\theta = [\sigma_f^2, l]$  determines the performance of GPR. To achieve suitable values, a log marginal likelihood is defined as,

$$L(\theta) = \frac{1}{2} Y^T (K(X, X) + \sigma_f^2 I)^{-1} Y + \frac{1}{2} \log |K(X, X) + \sigma_f^2 I| + \frac{n}{2} \log(2\pi) \quad (12)$$

where the first term is data-fit, the second term is a complexity penalty, and the last term is a normalizing constant, where  $n$  is the number of training samples.

The common method for solving non-convex optimization problems is taken the gradient descent method (i.e. Newton method). However, it is easy to trap in local optimization and it is sensitive to the initial value. To solve this problem, Brainstorming Optimization algorithm is proposed to obtain the optimal values of hyper-parameters.

### 2.2Brainstorming Optimization Algorithm

Different from traditional optimization algorithms inspired by animals' behaviors or natural phenomena (such as ant colony algorithm, simulated annealing algorithm, genetic algorithm.), BSO simulates the thinking process of human beings [16]. In the brainstorming process, a group of people are divided into several teams to solve a problem. To get solutions as many as possible, the individuals in each team are demanded that not to give comments on others' solutions before this term is over. In order to bring more inspiration collision, the individuals are encouraged to use divergent thinking on others' solutions. And then choosing the best in these solutions. The BSO can be summarized as three main processes:

(1) Clustering process

Cluster the  $N$  individuals into  $M$  clusters and choose a center from each cluster according to the value of fitness function.

(2) Replacing process

Randomly choosing an individual to replace the center and change its value using eq. (13):

$$x_{new} = x_{org} + \xi \times N(0, 1) \quad (13)$$

$$\xi = rand() \times \log sig((0.5 Iter_{max} - Iter_{cur}) \times s^{-1}) \quad (14)$$

where  $x_{org}$  is the selected individual,  $x_{new}$  is the updated individual,  $\xi$  is the dynamic update weight,  $rand()$  is a random value between  $[0, 1]$ .  $\log sig()$  is a logarithmic sigmoid transfer function.  $Iter_{max}$  is the max value of

iteration.  $Iter_{cur}$  is the current value of iteration.  $s$  is slope-changing value of  $\log sig()$  function.

(3) Generating process

The new individual has three characteristics which are dominant and flexibility. The dominant means that the current individual with good performance will lead the next generation, which means that the clustering center participates in generating new individual with a high probability. The flexibility means that the new individual could also be a combination of different individuals from two classes eq.(17).

$$x^{new} = x_j^a \tag{15}$$

$$x^{new} = x_j^b \tag{16}$$

$$x^{new} = rand() \times x_j^a + (1 - rand()) \times x_j^b \tag{17}$$

where,  $x_j^a$  and  $x_j^b$  are two individuals from different classes.

(4) Updating process

During each iteration, the current individual generates a new individual. After comparison the *fitness* value, the better individuals will be preserved and entered into the next iteration process. In this way, before reaching the maximum number of iterations, the algorithm is always searching for the optimal solution.

### 2.3The BSO-GPR Framework

To achieve the best hyperparameters of GPR kernel function, the BSO is embedded into the nonlinear optimization and the fitness function is defined as:

$$fitness(k) = \frac{1}{n} \sum_{i=1}^n (y_i^* - y_i)^2 \tag{18}$$

where,  $y_i^*$  is the prediction, and the  $y_i$  is the actual value.

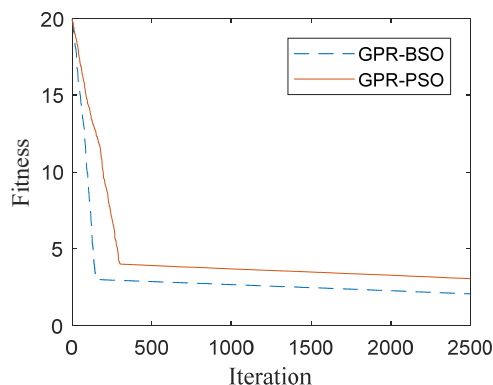


Figure 1. Convergence Information of BSO and PSO

### 2.4The Evaluation index

The model prediction ability is rated by the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of determination ( $R^2$ )

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2} \tag{19}$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - y_i^*}{y_i} \right| \tag{20}$$

$$R^2 = \left( 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \right) \times 100 \tag{21}$$

### 3Case Study

In order to evaluate the performance of the proposed methodology, a wind power data obtained from National Weather Service (NWS) is used in this section. It contains 15 minutes of wind speed, wind direction, wind power and other meteorological data. After the normalizing data, the data from July 1,2017 to August 30, 2017 are used for training and the data on August 31, 2017 are used for forecasting. The input factors include 15 variables such as wind speed, wind direction, atmospheric pressure. The model uses the squared exponential as the covariance function, BSO to optimize the hyper-parameter, and the prediction error as the optimization objective function.

The convergence value comparing between BSO and PSO is shown in Figure.1. The load prediction curve is shown in Figure.3 and Figure.4. Additionally, Support vector machine (SVR) and Random Forest (RF) are taken as comparisons which are shown in Figure.2.

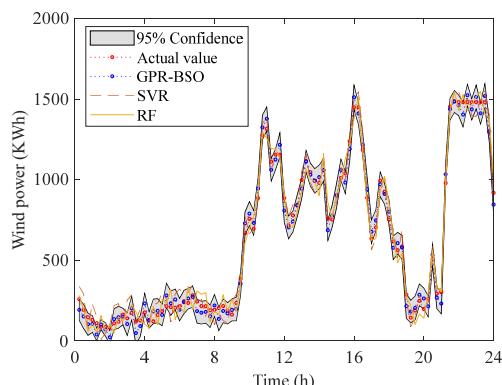
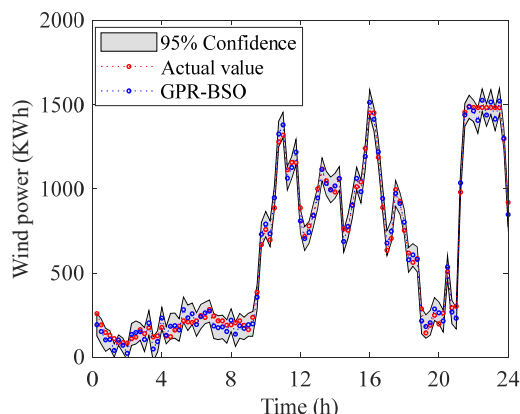
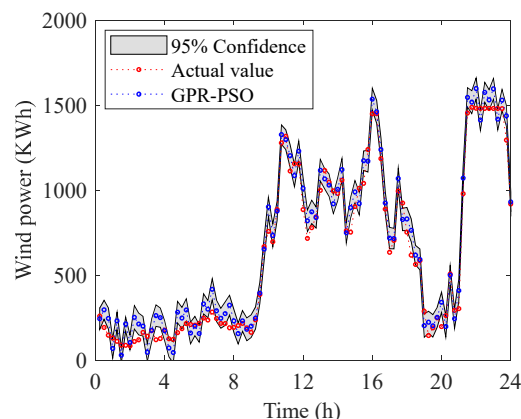


Figure 2. Error comparisons with several methods



**Figure 3.** Prediction of GPR-BSO



**Figure 4.** Prediction of GPR-PSO

**Table 1.** Error information of several methods

Algorithm	RMSE/KW	MAPE/%	R-squared/%
BSO-GPR	43.08	12.52	99.18
PSO-GPR	79.10	23.45	97.25
SVR	67.54	21.27	98.00
RF	58.71	17.20	98.49

According to Figure.1, BSO converges to 3.13 from iteration 143 and PSO converges to 4.24 from iteration 297. BSO has a fast convergence speed and a better ability of global searching than PSO. From the results of Figure.3 and Figure.4, the GPR-BSO probability density distribution including more actual value than the PSO which means that BSO has more probability to optimize the hyper-parameter than PSO in the same situation. From the result of Figure.2 and Table 1, average error results of multiple experiments indicate that GPR-BSO outperforms other methods in all evaluation indexes. Therefore, BSO has a better ability than PSO in optimization and GPR can provide reliable distribution for wind power prediction.

#### 4Conclusion.

A GPR model for short-term wind power forecasting is presented in this paper. Wind power presents uncertainty, nonlinearity, and complexity, making it difficult to forecast with a linear model. While, the parameters of the model such as neural network are too complex, and GPR as a non-parameter model can describe this mapping relationship easily. In addition, the GPR model can also predict the posterior probability density distribution to provide more information. The hyper-parameters of GPR are obtained by using the brainstorming optimization and the accuracy of GPR model is improved.

#### Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant, Key dispatching and controlling technologies for large-scale offshore wind power centrally connecting to local power grids (2019YFE0114700).

#### References

1. Fang S, Chiang H D (2016) A high-accuracy wind power forecasting model IEEE Transactions on Power Systems, 32(2): 1589-1590.
2. Yan J, Zhang H, Liu Y, et al. (2017) Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping. IEEE Transactions on Power Systems, 33(3): 3276-3284.
3. Huibin L, Wenxiao Q, Qianqian Y, et al. (2016) Research on the Short-term prediction model of wind power generation based on GIS. In:2016 China International Conference on Electricity Distribution (CICED). IEEE, 1-4.
4. Wang Y, Liu Y, Song P, et al. (2019) Impact of the accuracy of NWP wind speed forecasts on wind power forecasting. In: 8th Renewable Power Generation Conference Shanghai China
5. Terciyanl E, Demirci T, Küçük D, et al. (2013) Enhanced nationwide wind-electric power monitoring and forecast system. IEEE Transactions on Industrial Informatics, 10(2): 1171-1184.
6. Zhang H, Liu Y, Yan J, et al. (2020) Improved deep mixture density network for regional wind power probabilistic forecasting. IEEE Transactions on Power Systems, 35(4): 2549-2560.
7. Wu J L, Ji T Y, Li M S, et al. (2015) Multistep wind power forecast using mean trend detector and mathematical morphology-based local predictor. IEEE Transactions on Sustainable Energy, 6(4): 1216-1223.
8. Hong W C. (2011) Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. Energy, 36(9): 5568-5578.

9. Hor C L, Watson S J, Majithia S. (2016) Daily load forecasting and maximum demand estimation using ARIMA and GARCH 2006 International Conference on Probabilistic Methods Applied to Power Systems. IEEE, 2006: 1-6.
10. Ko C N, Lee C M. (2013) Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter Energy, 49: 413-422.
11. Dinesh Reddy M. (2017) Load forecasting using linear regression analysis in time series model for RGUKT, RK valley campus HT feeder. Int. J. Eng. Res. Technol, 6(5): 2278-0181.
12. Amral N, Ozveren C S, King D. (2007) Short term load forecasting using multiple linear regression 2007 42nd International universities power engineering conference. IEEE, 2007: 1192-1198.
13. M. Alamaniotis, S. Chatzidakis and L. H. Tsoukalas (2014) Monthly load forecasting using kernel based Gaussian Process Regression, MedPower 2014, Athens, pp. 1-8.
14. L. Chong, J. Rong, D. Wenqiang, S. Weicheng and M. Xiping (2016) Short-term PV generation forecasting based on weather type clustering and improved GPR model, 2016 China International Conference on Electricity Distribution (CICED) Xi'an China, pp. 1-5.
15. Xiao Z, Zhan S, Xiang Z, et al. (2016) A GPR-PSO incremental regression framework on GPS/INS integration for vehicle localization under urban environment 2016 IEEE 27th annual international symposium on personal, indoor, and mobile radio communications (PIMRC). IEEE, pp. 1-6.
16. Cheng S, Qin Q, Chen J, et al. (2016) Brain storm optimization algorithm: a review. Artificial Intelligence Review, 46(4): 445-458.