International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:10, No:1, 2016

Using Gaussian Process in Wind Power Forecasting

Hacene Benkhoula, Mohamed Badreddine Benabdella, Hamid Bouzeboudja, Abderrahmane Asraoui

***Abstract*—**The wind is a random variable difficult to master, for this, we developed a mathematical and statistical methods enable to modeling and forecast wind power. Gaussian Processes (GP) is one of the most widely used families of stochastic processes for modeling dependent data observed over time, or space or time and space. GP is an underlying process formed by unrecognized operator's uses to solve a problem. The purpose of this paper is to present how to forecast wind power by using the GP. The Gaussian process method for forecasting are presented. To validate the presented approach, a simulation under the MATLAB environment has been given.

***Keywords*—**Forecasting, Gaussian process, modeling, wind power.

I. Introduction

NOWADAYS, wind power is set to become one of the major energy resources because it is clean. Also, it is one of the available nontraditional energy-giving.

In recent years, environmental considerations have prompted the use of wind power as a renewable energy resource. However, the biggest challenge in integrating wind power into the electric grid is its intermittency. One approach to deal with wind intermittency is forecasting future values of wind power production.

Wind power forecast is crucial for the design and operation of the power system incorporating large scale of wind power. There have been reports regarding the design of power system considering the wind power forecast [2].

Wind power forecasting is an important part in the structure of the Smart Grid. The term Smart Grid elicits a mental image of a fully automated power distribution system, capable of monitoring usage and voltage levels, constantly making adjustments to keep everything running at optimal levels.

Wind power forecasting improve the economic and technical integration of large amounts of wind energy into the existing electricity grid. Trading, balancing, grid operation, controllability and safety issues increase the importance of predicting power output from wind power operators. Therefore, wind power forecasting systems have to be integrated into the monitoring and control systems of the transmission system operator (TSO) and wind farm operators/ traders [6]

Statistical algorithms based on computational learning are typically used to convert weather forecasts and observational data to wind power point forecasts [6]

Wind power forecasting is a regression problem [2]; regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables.

In this paper, Gaussian process regression (GPR) model is presented. A simulation under MATLAB was carried to demonstrate the performances of the proposed methodology.

II. The Methods of Regression

A. Linear Regression

Linear regression is a statistical modeling technique used to describe a continuous response variable as a function of one or more predictor variables. It can help understand and predict the behavior of complex systems or analyze experimental, financial, and biological data [5].

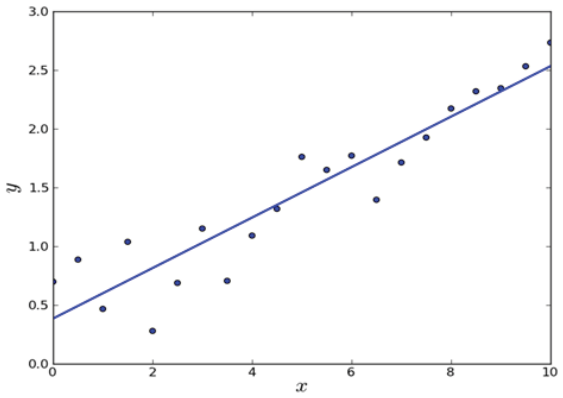

Fig. 1 Linear regression

Linear regression techniques are used to create a linear model. The model describes the relationship between a dependent variable y (also called the response) as a function of one or more independent variables $\mathbb{x}_n$(called the predictors). The general equation for a linear regression model is:

$$y = \beta 0 + \sum \beta n \mathbb{x}_n + \mathcal{E}n \quad (1)$$

where β represents linear parameter estimates to be computed and $\mathcal{E}$ represents the error terms.

The least-squares regression (linear) line is regression method, for a given set of co-ordinate values,

H. Benkhoula is with the Faculty of Electrical Engineering, Electrotechnical Department, USTO, Algeria (phone: 0213-558-415657; e-mail: belharche@gmail.com).
M. B. Benabdella is with the Faculty of Electrical Engineering, Electrotechnical Department, USTO, Algeria (phone: 213 552535506; e-mail: benabdallah22@yahoo.fr)
H. Bouzeboudja is with the Faculty of Electrical Engineering, Electrotechnical Department, USTO, Algeria (phone: 213 673372018; e-mail: hbouzeboudja@yahoo.fr)
A. Asraoui is with the Faculty of Electrical Engineering, Electrotechnical Department, USTO, Algeria (phone: 0213-777-597709; e-mail: asraoui.usto@gmail.com).

$(X1, Y1), (X2, Y2), \ldots, (Xn, Yn)$ let the X values be the independent variables and the Y-values are the dependent values. Also let $D1, \ldots, Dn$ be the vertical distances between the line shown as PQ in Fig. 2 and the points representing the co-ordinate values. The least squares regression line, i.e. the line of best fit, is the line which makes the value of $D_1^2 + D_2^2 + \cdots + D_n^2$ a minimum value.

B. Nonlinear Regression

Nonlinear regression is a statistical technique that helps describe nonlinear relationships in experimental data. Nonlinear regression models are generally assumed to be parametric, where the model is described as a nonlinear equation. Typically, machine learning methods are used for non-parametric nonlinear regression [5].

Parametric nonlinear regression models the dependent variable (also called the response) as a function of a combination of nonlinear parameters and one or more independent variables (called predictors). The model can be univariate (single response variable) or multivariate (multiple response variables).

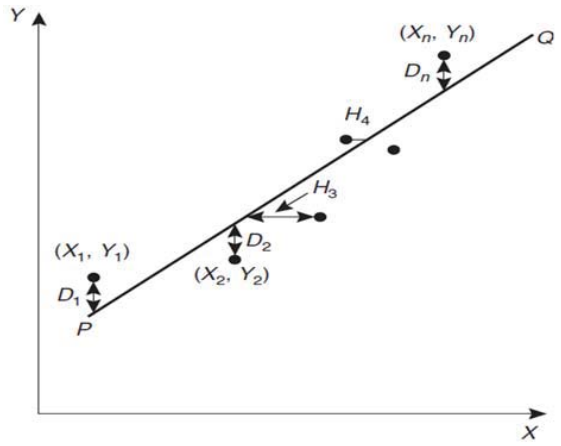

Fig. 2 The least-squares regression lines

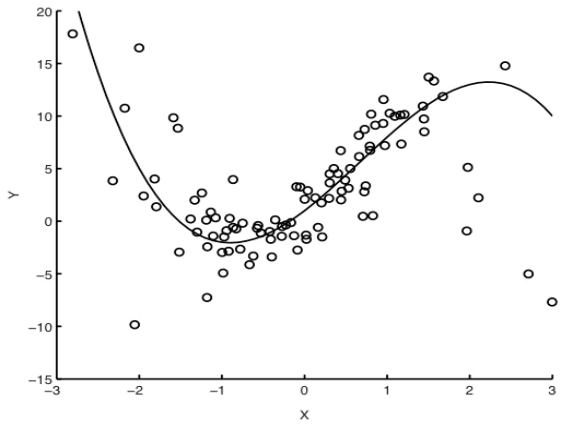

Fig. 3 Nonlinear regression

The parameters can take the form of an exponential, trigonometric, power, or any other nonlinear function. To determine the nonlinear parameter estimates, an iterative algorithm is typically used.

$$y = f(x, \beta) + \mathcal{E} \tag{2}$$

where, β represents nonlinear parameter estimates to be computed and $\mathcal{E}$ represents the error terms.

III. Support Vector Regression

Support vector regression (SVR) is an extension of support vector machines (SVM), a family of popular supervised machine learning tools based on statistical learning theory originally proposed by Vapnik [7]

SVM is based on the principle of statistical machine learning process and structural risk minimization, which minimizes the upper bound generalization error rather than local training error, which is usual approach in other traditional machine learning methodologies. This is one of the advantages of SVM over other soft computing learning algorithms. Other advantages include; unique solution due to the convex nature of the optimal problem, the use of high dimensional spaced set of kernel functions which discreetly comprise non-linear transformation, hence no assumption in functional transformation which makes data linearly separable indispensable.

Given a set of data points represented by = $\{x_i, d_i\}_i^n$, where x_i is the input space vector of the data sample, d_i is the target value and n is the number of data points. SVM equations based on Vapnik's theory [8] approximates the function as:

$$y = \omega\varphi(x) + b \tag{3}$$

$$R_{SVMs}(C) = \frac{1}{2}\|w\|^2 + C\frac{1}{n}\sum_{i=1}^{n} L(x_i, d_i) \tag{4}$$

where φ(x) represents high dimensional-space features that map the input space vector x, w is a normal vector, b is a scalar, and $C\frac{1}{n}\sum_{i=1}^{n} L(x_i, d_i)$ represents the empirical risk. The parameters w and b can be estimated by minimization of regularized risk function after introduction of positive slack variables ξ_i and ξ_i^* that represent upper and lower excess deviation [8]

$$\text{Minimize } R_{SVMs}(C, \xi_i, \xi_i^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} L(\xi_i, \xi_i^*) \tag{5}$$

$$\text{Subject to } \begin{cases} d_i - w\varphi(x_i) + b_i \ll \varepsilon + \xi_i \\ w\varphi(x_i) + b_i - d_i \ll \varepsilon + \xi_i \\ \xi_i, \xi_i^* \gg 0, 1, \ldots, L \end{cases}$$

where $\frac{1}{2}\|w\|^2$ represent the regularization term, C is the error penalty factor used to control the trade-off between the regularization term and empirical risk, ε is the loss function, which equates to approximation accuracy of the training data point, and l is the number of elements in the training data set.

Equation (3) can be solved with the introduction of Lagrange multiplier and optimality constraints, hence obtaining a generic function given by

$$f(x) = \sum_{i=1}^{l} (\beta_i - \beta_i^*) K(x_i + x_j) + b \quad (6)$$

where $K(x_i, x_j) = \varphi(x_i)\varphi(x_j)$ and the term $K(x_i, x_j)$ is called the kernel function.

IV. Gaussian Process for Regression (GPR)

Wind forecast is a regression problem [2]. GP is fully specified by the mean function and the covariance function (Kernel function) representing the correlation between two variables given a data set $\mathcal{D} = (\mathbb{X}, y)$ consisting of N Input vectors $\mathbb{X} = \{\mathbb{x}_n\}_{n=1}^{N}$ and corresponding targets= $\{y_n\}_{n=1}^{N}$the actual wind power generation is the corresponding real valued targety. In standard GP, we assume that the relationship between the input vector and the target is given by:

$$y_n = f(\mathbb{x}_n) + \mathcal{E}_n \quad (3)$$

$\mathcal{E}_n$ is Gaussian noise that follows: $\mathcal{N}(0, \sigma^2)$; f is an underlying function, we assume that, $F = [f(\mathbb{x}_1), f(\mathbb{x}_2), \dots f(\mathbb{x}_N)]^T$ behave according to a Gaussian process, that is, $p(f|\mathbb{x}_1, \mathbb{x}_2, \dots, \mathbb{x}_N) = \mathcal{N}(0, \mathbb{K})$. Here $\mathbb{K}$ is the kernel matrix with elements

$$\mathbb{K}(y|f, \mathbb{x}) = \mathcal{N}(f, \sigma^2 I)_{nm} = k(\mathbb{x}_n, \mathbb{x}_m)$$

k being the kernel function for GP is the automatic relevance determination (ARD)function [2], that is:

$$k(\mathbb{x}, \mathbb{x}^*) = \theta_0 \exp\{-\tfrac{1}{2}\textstyle\sum_{d=1}^{\mathcal{D}} \theta_d (x_d - x'_d)^2\} \quad (4)$$

where $\theta_0 = \{\theta_0, \theta_1, \dots, \theta_{\mathcal{D}}\}$ are the hyper parameters. Then, the distribution of y conditioned onthe value of f is given by an isotropic Gaussian:

$$p(y|f, \mathbb{x}) = \mathcal{N}(f, \sigma^2 I) \quad (5)$$

From the property of the Gaussian distribution, we can see that the marginal distribution of y is:

$$p(y|\mathbb{x}) = \int p(y|f, \mathbb{x}) p(f|\mathbb{x}) df = \mathcal{N}(0, k + \sigma^2 I) \quad (6)$$

Our objective is to predict $y_*$and f_* at a testing point$\mathbb{x}_*$. To prepare for GPR, we calculate the kernel function among all combinations of these points, summarizing our findings in three matrices [1]–[3].

$$\mathbb{K} = \begin{bmatrix} k(\mathbb{x}_1, \mathbb{x}_1) & \cdots & k(\mathbb{x}_1, \mathbb{x}_n) \\ \vdots & \ddots & \vdots \\ k(\mathbb{x}_n, \mathbb{x}_1) & \cdots & k(\mathbb{x}_n, \mathbb{x}_n) \end{bmatrix} \quad (7)$$

$$\mathbb{K}_* = [k(\mathbb{x}_*, \mathbb{x}_1)\, k(\mathbb{x}_*, \mathbb{x}_2)\, k(\mathbb{x}_*, \mathbb{x}_n)$$

$$\mathbb{K}_{**} = k(\mathbb{x}_*, \mathbb{x}_*) \quad (8)$$

The distribution of y conditioned on the value of y is given by

$$p(y_*|y) = \mathcal{N}(\mathbb{K}_* \mathbb{K}^{-1} y, \mathbb{K}_{**} - \mathbb{K}_* \mathbb{K}^{-1} \mathbb{K}_*^T) \quad (9)$$

Our best estimate for $y_*$is the mean of this distribution:

$$\bar{y}_* = \mathbb{K}_* \mathbb{K}^{-1} y, \quad (10)$$

Fig 4 illustrates a typical example of a prediction problem.

Given six noisy data points (error bars are indicated with vertical lines), we are interested in estimating a seventh at $x_* = 0.2$ [1]-[4].

V. Simulation

A data sample is used to train the model, specified as a vector (n=201).

Different kernel functions are used in this simulation test.

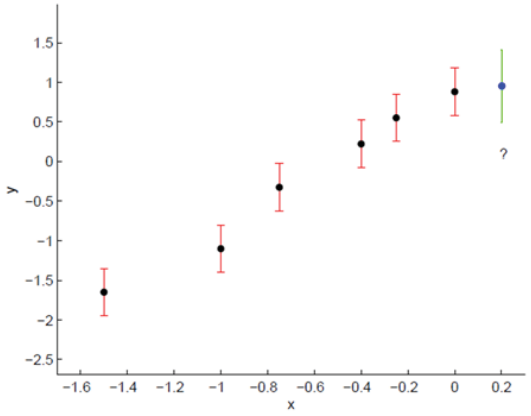

Fig. 4 Estimation of y

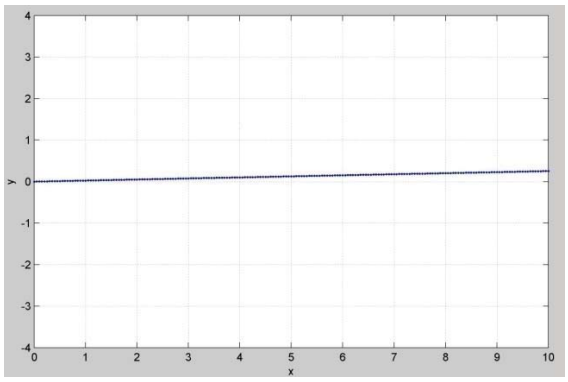

Fig. 5 Estimation of y_*, using linear kernel function

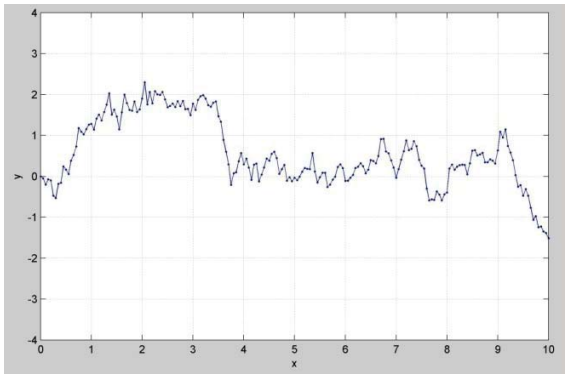

Fig. 6 Estimation of y_* using squared exponential kernel function

The choice of the kernel function is very important for an e goodness estimate.

Choosing the most appropriate kernel highly depends on the problem at hand – and fine tuning its parameters can easily become a tedious and cumbersome task.

The squared exponential kernel function is widely used in wind power forecasting [2].

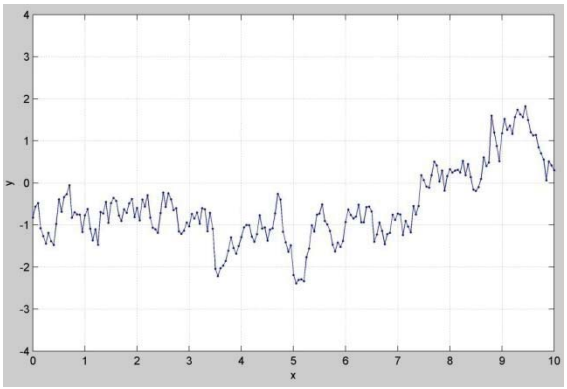

Fig. 7 Estimation of y_* using squared exponential kernel function

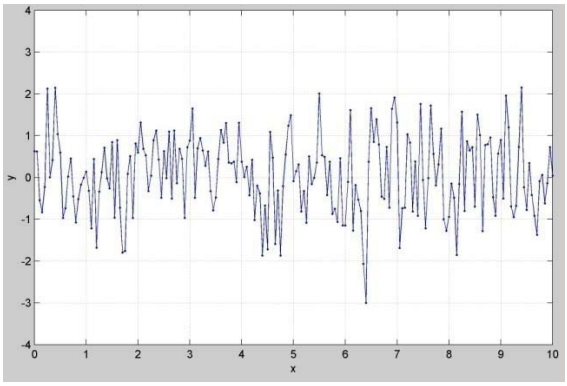

Fig. 8 Estimation of y_*, using Brownian kernel function

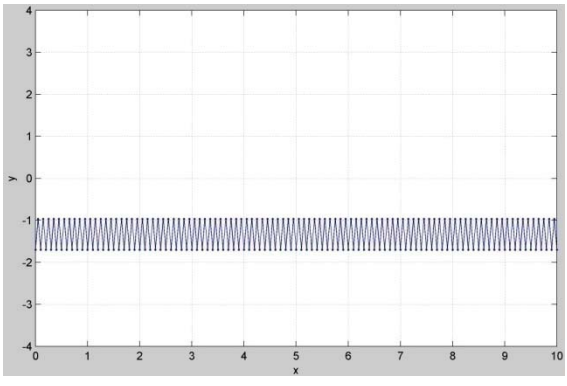

Fig. 9 Estimation of y_*, using Periodickernel function

VI. Conclusion

Wind power forecast is very useful tool for designing a communicative system achievable, realistic and timed between wind power and power grid.

Our future objective is the development a real-time forecasting system and makes it more realistic and intractable.

References

[1] M. Ebden, "Gaussian process for regression; a Quick Introduction" *Robotics Research Group, a tutorial*.

[2] P. Kou, F. Gao, X.Guan "Prediction Intervals for Wind Power Forecasting: Using Sparse Warped Gaussian Process "*Power and Energy Society General Meeting, 2012 IEEE*, pp. 1 - 8, July. 2012.

[3] E. Snelson, Z. Ghahramani, "Sparse Gaussianprocesses using pseudo-inputs," *in 2006 Advances in Neural InformationProcessing Systems 18*, pp. 1257-1264.

[4] Rasmussen, Carl Edward, "Gaussian processes for machine learning." (2006).

[5] http://www.mathworks.com