International Mathematical Forum, Vol. 6, 2011, no. 53, 2615 - 2624

A Nonparametric Circular-Linear Multivariate Regression Model with its Application to Wind Energy

Xu Qin and Jiang-She Zhang

Faculty of Science and State Key Laboratory for Manufacturing Systems Engineering
Xi'an Jiaotong University, Xi'an 710049, P.R. China

Xiao-Dong Yan

Key Laboratory of Regional Climate-Environment Research for Temperature East Asia,
Institute of Atmospheric Physics, Chinese Academy of Science, Beijing 100029, P.R. China

Abstract

A nonparametric circular-linear multivariate regression model using kernel weighted local linear method is proposed. The bandwidths are selected using the cross-validation method. The suitability of models is judged from the coefficient of determination R^2. The model is applied in this paper to wind data extracted from National Centers for Environmental Prediction (NCEP) reanalysis data. The conclusion reached is that compared with the existing model, the nonparametric model used in this paper provides a better fit and prediction for wind energy generated given wind speed, time and wind direction.

Mathematics Subject Classification: 62G08, 62H10, 62H11

Keywords: nonparametric regression, circular variable, kernel, cross-validation, wind

1 Introduction

The combined effect of the widespread depletion of fossil fuels and the gradually emerging consciousness about environmental degradation have given priority to the use of conventional and renewable alternative energy sources such as solar, wind and solar-hydrogen energies [1,4]. Among the renewable sources of energy, wind energy is not only freely and widely available but also inexhaustible. The rapid development in wind energy technology has made it an alternative to conventional energy systems in recent years [6].

In this paper, we mainly consider the regression and prediction of wind energy given time, wind speed and wind direction. In other words, we are interested in the multivariate regression models with linear and circular predictors.

In the available literature on multivariate regression, many researchers have proposed many parametric and nonparametric regression models based on some linear predictors ([3,10] for example). In 2009, Marzio M. D. et al have extended least squares local polynomial to the case of d-dimensional circular predictors [9].

In the above mentioned literature, the predictors are either completely linear or completely circular. However, study on the case of linear and circular predictors is scare. In earlier literature, researchers proposed several symmetric regression models involving some circular and linear predictors ([5], for example). In 2006, SenGupta and Ugwuowo proposed several asymmetric circular-linear multivariate regression models which were motivated by the need to predict some environmental characteristics based on some circular and linear predictors [11]. In that paper, authors chose the following model

$$Y_i = M + \beta_1 v_i + \beta_2 t_i + A\ cos(\theta_i + \nu\ cos\theta_i) + \epsilon_i \tag{1}$$

to fit wind energy generated (Y_i) given the wind speed (v_i), time (t_i), and wind direction (θ_i).

However, the above mentioned models have been only limited to fit parametric models. And we have to choose a proper model among many parametric models for fitting in practice. With improper model, the fitted results may be unsatisfactory. As stated by Ruppert and Wand(1994) [10], nonparametric regression has become a rapidly developing field as researchers have realized that parametric regression is not suitable for adequately fitting curves to many data sets that arise in practice.

In this paper, our primary interest is in proposing a nonparametric multivariate model with circular and linear predictors. We consider the data set $(L_i, Y_i), i = 1, 2, ..., n$, where $L_i = (\mathbf{X}_i^T, \Theta_i^T)^T$, the $\mathbf{X}_i$ and the Θ_i are d_1-dimensional linear and d_2-dimensional circular predictor variables respectively, the Y_i are scalar response variables. We will assume the model

$$Y_i = m(L_i) + \epsilon_i,\ i = 1, ..., n \tag{2}$$

where ϵ_i's are mutually independent and identically distributed real-valued random variables with zero mean and unit variance and are independent of the L_i's.

Much of our attention will be devoted to the local linear least squares kernel estimator of m, i.e. $\widehat{m}(l; H, \mathbf{C})$. Thereinto, H denote the smoothing parameters of linear kernel and $\mathbf{C}$ denote the concentration parameters of circular kernel. In this paper, these parameters are all regarded as bandwidth parameters. We use cross-validation (CV) method to select bandwidths. The CV method is one of the most widely used bandwidth selectors for kernel smoothing. And the CV is data-driven and adaptive [7]. The suitability of the models is judged from the coefficient of determination R^2.

The remainder of this article is organized as follows. Section 2 we consider the kernel weighted local linear regression estimator. The CV method is described in Section 3. In Section 4, a comparison is made between the existing model proposed by SenGupta and the new proposed model in the paper. These two models are applied to fit and predict a series of wind energy data whose initial data is extracted from the NCEP reanalysis data at 850Pha in 2005. Finally, Section 5 provides some concluding remarks.

2 Nonparametric circular-linear multivariate regression model

Let $\mathbf{y} = (Y_1, ..., Y_n)^T$ be the response vector,

$$L_l = \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T & \sin(\Theta_1 - \Theta)^T \\ \vdots & \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T & \sin(\Theta_n - \Theta)^T \end{bmatrix}$$

the design matrix, and

$$W_l = \mathrm{diag}\{K_{H\mathbf{C}}(L_1 - l), ..., K_{H\mathbf{C}}(L_n - l)\}$$

the weight matrix, where $K_{H\mathbf{C}}$ is the linear-circular function,

$$\begin{aligned} K_{H\mathbf{C}}(L_i - l) &= K_H(X_i - x) \cdot K_{\mathbf{C}}(\Theta_i - \theta) \\ &= \frac{1}{\sqrt{h_1 \cdots h_{d_1}}} \prod_{j=1}^{d_1} K^{(h_j)}({h_j}^{-1/2}(X_{ij} - x_j)) \cdot \prod_{p=1}^{d_2} K_{\kappa_p}(\Theta_{ip} - \theta_p) \end{aligned}$$

where $K^{(h_j)}$ is a standard linear kernel [8] and K_{κ_p} is a second-order circular kernel [9]. The local linear least squares kernel estimator of $m(l)$ is given by the first entry of the vector

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\boldsymbol{\beta}} \end{pmatrix} = \arg\min_{\alpha, \boldsymbol{\beta}} \sum_{i=1}^{n} \{Y_i - \alpha - \boldsymbol{\beta}^T (L_i - l)\}^2 K_{H\mathbf{C}}(L_i - l) = ({L_l}^T W_l L_l)^{-1} {L_l}^T W_l \mathbf{y}$$

Thus the local linear least squares kernel estimator of $m(l)$ is

$$\widehat{m}(l; H, \mathbf{C}) = \mathbf{e}_1^T ({L_l}^T W_l L_l)^{-1} {L_l}^T W_l \mathbf{y} \tag{3}$$

where $\mathbf{e_1}$ is the $(d_1 + d_2 + 1) \times 1$ vector having 1 in the first entry and all other entries 0.

In (3), there are $d_1 + d_2$ bandwidth parameters. In this paper, we use the classical CV method to select these parameters.

3 Bandwidth selection

In this paper, the CV method is used to select bandwidths. We consider

$$CV(h_1, ..., h_{d_1}, \kappa_1, ..., \kappa_{d_2}) = \sum_{i=1}^{n} (Y_i - \hat{m}_{-i}(l_i))^2 \tag{4}$$

where $\hat{m}_{-i}(l_i)$ is the leave-one-out estimator with l_i deleted. We choose h_1,...,h_{d_1}, κ_1,...,κ_{d_2} to minimize (4). The method of non-linear least squares is applied to the above computation.

In this paper, the circular predictor we consider is 1-dimensional. Thus our main aim is to select proper $h_1, ..., h_{d_1}, \kappa$ for minimizing (4). The bandwidth selection steps in this paper are the following: In this paper, the stopping

---

S1. For j=1, fix a κ value $\kappa^{(j)}$and minimize the equation (4). Then we get a smoothing parameters vector $h^{(j)}$ and a CV value, where $h^{(j)} = (h_1^{(j)}, ..., h_{d_1}^{(j)})^T$.

S2. Let j=j+1 and repeat the step 1 until the stopping criterion is satisfied. Then a series of CV values are yielded.

S3. Find the minimum value of the above CV value series and the corresponding bandwidths $h_1^{(*)}, ..., h_{d_1}^{(*)}, \kappa^{(*)}$ are to be the finally chosen bandwidths.

---

criterion used is the maximum iteration number which is based on a pre-given range of κ.

4 Experiments

In this part, we will carry numerical experiments for the parametric model (1) proposed by SenGupta and the nonparametric model (2) proposed in this paper. All programs are written in Matlab. The extraction of data used in this paper resorts to GrADS software. In order to show the effectiveness of the new proposed model to fit the regression of wind energy given the wind data and time, a study is carried out of sample taken from NCEP reanalysis data at 850 Pha in 2005. We choose 65 daily wind speed, wind direction and air temperature from January 1 to March 7 in 2005 at the lattice point (110^{o}E, 37.5^{o}N). The pre-60 data are used to fit models and the post-5 are used to predict. The initial data available are daily wind speed, wind direction and air temperature. In this paper, the wind direction corresponding to the northerly direction is taken as direction 0. These initial data are plotted in Fig. 1. In this paper, the linear predictors are 2-dimensional and the circular predictor is 1-dimensional. The wind energy per unit area is:

$$W = \frac{1}{2}\rho v^3$$

where v is the wind speed (m/s) and ρ is the air density (kg/m^3) which is computed by Clapeyron equation [2]. The Clapeyron equation is based on atmospheric pressure, air temperature and other two constants.

Then we apply these data to fit multivariate regression models (1) and (2). For parametric model (1), the nonlinear least square method is used to estimate the parameters. The fitted model is

$$\hat{Y}_i = -286.6874 + 98.5473 v_i + 1.4780 t_i - 22.8765 \ \cos(\theta_i - 1.1962 \ \cos\theta_i) \quad (5)$$

In nonparametric model, the linear kernel we take is Gauss kernel and the circular kernel is von Mises kernel. The per-given range of κ is from 0.2 to 10 with the interval 0.2. Then we use CV method to select bandwidths. Fig. 2 displays the different CV values according to different concentration parameters. When $\kappa = 4.8$, the corresponding CV value is the minimum. The finally chosen bandwidth results are listed in Table 1. With these bandwidth values, we can get the local linear least square kernel estimator of wind energy through equation (3).

Table 1: Bandwidth values chosen in this paper

	time(days)	wind speed(m/s)	wind direction(rad)
bandwidth values	6.8402	0.8213	4.8000

For illustration of the effectiveness of the new model to fit wind energy, we compare the fitted and predictive results of model (1) and model (2). It can be observed from Fig. 3 that the fitted effect of the nonparametric model is significantly better than the one of the existing parametric model. Furthermore, we compute the coefficient of determination R^2. For model (1), R^2 is 0.7704.

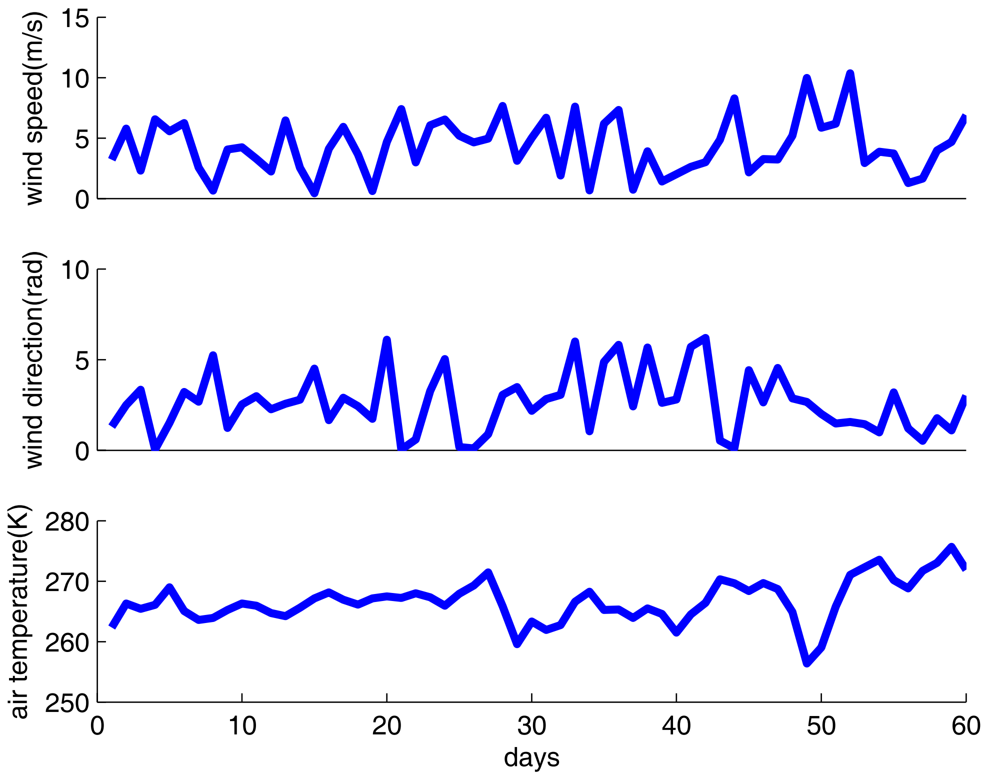

Figure 1: The initial wind speed, wind direction and air temperature series

For model (2), the value is close to 1. Fig. 4 and Table 2 show the predictive results of the two models. The results also show that the nonparametric model provides a better prediction than the existing model for the wind energy data under study.

Table 2: The comparison of predictive results

True value (KW)	Predictive value (KW)	
	SenGupta Model (relative error)	New Model (relative error)
0.3731	0.4896 (0.31)	0.3624 (0.03)
2.6935	1.1021 (0.60)	1.9016 (0.29)
1.1618	0.7965 (0.31)	1.0329 (0.11)
0.0889	0.2396 (1.70)	0.0741 (0.17)
0.2261	0.4027 (0.78)	0.2998 (0.33)

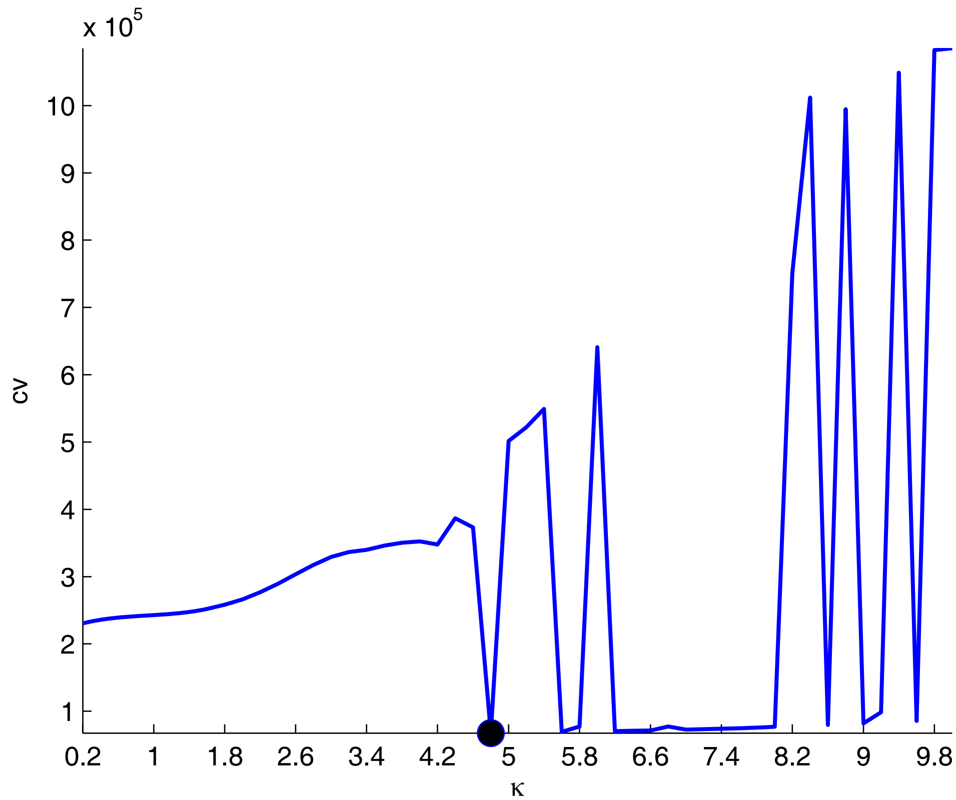

Figure 2: The CV values according to different κ

5 Conclusions and discussion

In this paper a nonparametric model for wind energy analysis is presented. The CV method is used to select bandwidths. A series of lattice point data is carried to compare the effectiveness of the proposed model with an existing parametric model. The results show that both in the fitting and prediction, the proposed model in this paper provides a better choice when we consider wind energy generated given time, wind speed and wind direction.

With the bandwidth selection process, the nonparametric model proposed in this paper is relatively time-consuming in the calculation compared with the parametric model. We will further seek more time-saving algorithm. In addition, the CV method is one of classical methods to select bandwidth. There are other methods for bandwidth selection. We will discuss the effectiveness of these methods in our another paper.

ACKNOWLEDGEMENTS. The first author's research was supported by the National Natural Science Foundation of China (Grant Nos. 60675013,

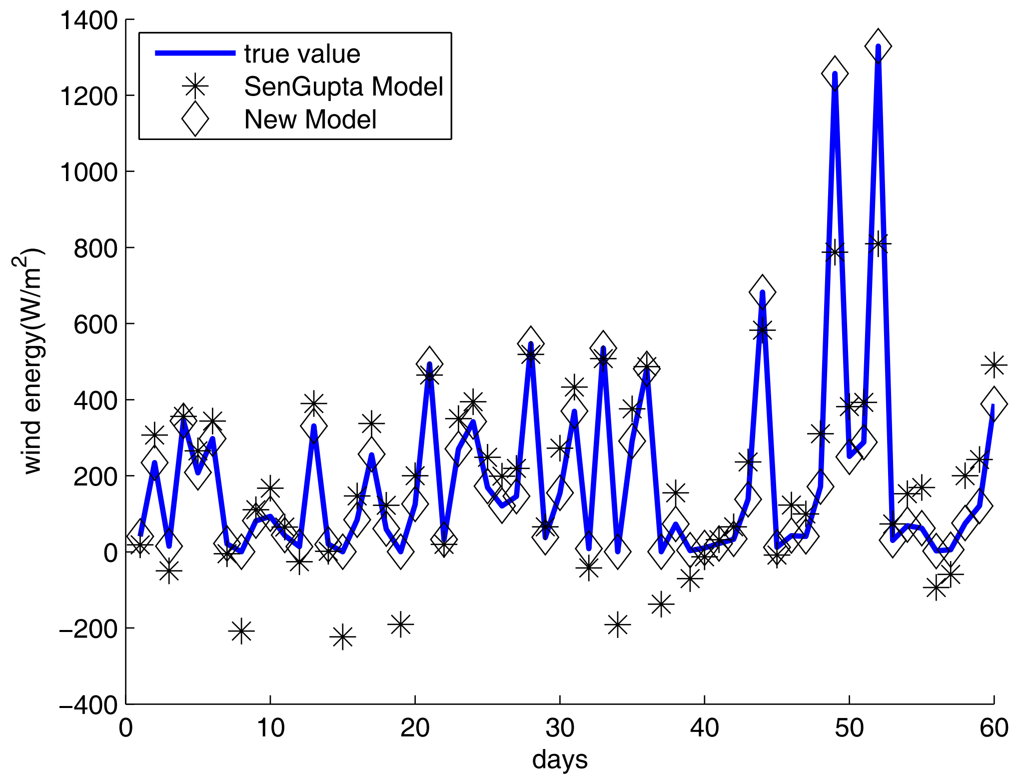

Figure 3: The fitted results of two models

10531030) and the National Basic Research Program of China (973Program)(Grant No. 2007CB311002).

References

[1] S. Akpinar, E.K. Akpinar, Wind energy analysis based on maximum entropy principle (MEP)-type distribution function, *Energy Conversion and Management*, **48** (2007), 1140-1149.

[2] Y.A. Cengel, M.A. Boles, *Thermodynamics C An Engineering Approach, McGraw-Hill Series in Mechanical Engineering (3rd ed.)*, Boston, 1998.

[3] W. Härdle, M. Müller and S. Sperlich, et al, *Nonparametric and Semiparametric Models*, Springer, Berlin, 2004.

[4] A. Hepbasli, O. Ozgener, A review on the development of wind energy in Turkey, *Renew Sust Energy Rev*, **8**(2004), 257-276.

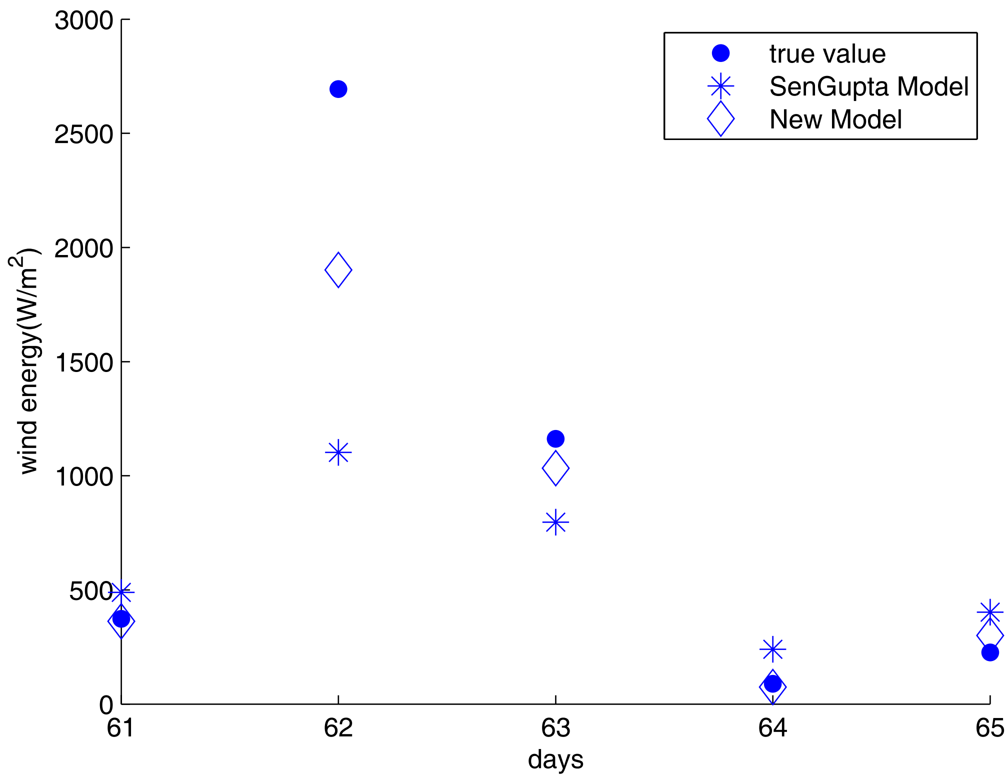

Figure 4: The predictive results of the two models

[5] R.A. Johnson, T.E. Wehrly, Some angular-linear distributions and related regression models, *Journal of the American Statistical Association*, **73**(1978), 602-606.

[6] C.G. Justus, *Winds and wind system performance*, The Franklin Institute Press, Philadelphia, Pennsylvania, 1978.

[7] Q. Li, J. Racine, Cross-validated local linear nonparametric regression, *Statistica Sinica*, **14**(2004), 485-512.

[8] Q. Li, J.S. Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, 2007.

[9] M.D. Marzio, A. Panzera and C.C. Taylor, Local polynomial regression for circular predictors, *Statistics and Probability Letters*, **79**(2009), 2066-2075.

[10] D. Ruppert, M.P. Wand, Multivariate locally weighted least squares regression, *The Annals of Statistics*, **3**(1994), 1346-1370.

[11] A. SenGupta, F.I. Ugwuowo, Asymmetric circular-linear multivariate regression models with applications to environmental data, *Environmental and Ecological Statistics*, **13**(2006), 299-309.

Received: April, 2011