

An Effective Wind Power Prediction using Latent Regression Models

^{1,2}Benamar Bouyeddou, ³Fouzi Harrou, ⁴Ahmed Saidi, ³Ying Sun

¹*Abou Bekr Belkaid University, STIC Lab., Department of Telecommunications, Tlemcen, Algeria*

²*University of Saida-Dr Moulay Tahar, Department of Electronics, Faculty of Technology, Saida, Algeria*

³*King Abdullah University of Science and Technology, CEMSE Division*

Thuwal, 23955-6900, Saudi Arabia, fouzi.harrou@kaust.edu.sa

⁴*Ahmed Draia University, Department of Science and Technology, Laboratory of Sustainable Development and Computer Science, Adrar 01000, Algeria*

Abstract—Wind power is considered one of the most promising renewable energies. Efficient prediction of wind power will support in efficiently integrating wind power in the power grid. However, the major challenge in wind power is its high fluctuation and intermittent nature, making it challenging to predict. This paper investigated and compared the performance of two commonly latent variable regression methods, namely principal component regression (PCR) and partial least squares regression (PLSR), for predicting wind power. Actual measurements recorded every 10 minutes from an actual wind turbine are used to demonstrate the prediction precision of the investigated techniques. The result showed that the prediction performances of PCR and PLSR are relatively comparable. The investigated models in this study can represent a helpful tool for model-based anomaly detection in wind turbines.

Index Terms—Power prediction, Wind turbine, regression methods, Latent variable models.

I. INTRODUCTION

Wind power production's main critical and challenging problem is its intermittent volatility, mainly to weather conditions [1], [2]; making the integration of wind turbines into the power grid not an easy task [3], [4]. Hence, accurately predicting wind power is of great interest to cope with the impacts of wind power fluctuation on power system operation. Furthermore, it is helpful for safety and managing the power grid, delivery, storage. Accurate wind power prediction plays a core role in efficiently and safely integrating wind power in a power grid. Essentially, the necessity for a precise wind power prediction has been raised with the increase of installed capacity. Thus, several prediction methodologies have been designed in the literature and can generally be grouped into two classes that are physical and statistical-based methods [5]. Generally speaking, the essence of physical models consists in the application of atmospheric motion equations to estimate future trends of meteorological measurements, then forecasting wind power by using some forecasted meteorological variables (e.g., wind speed) [6], [7]. Importantly, a physical model is generally performed using numerical weather estimation, which is implemented into two steps: prediction of wind speed, then transformation of this wind speed to wind power [8]. This mechanism is usually accomplished by using wind power curve which could be modeled using parametric or non-

parametric methods [9]. Although physical models provide acceptable features in predicting the long-term trend of wind variance, they are costly to develop and time-consuming and provide low prediction accuracy for a local area. For more details refer to [9]. On the other hand, statistical models use data mining methods to construct a model that expresses the interaction between wind power and other input variables [10]–[12].

Over last decades, several researches focused on applying artificial intelligence techniques for wind power prediction [13]–[15]. In [13], the k-nearest neighbor classifier (kNN) is used for predicting wind power in short-term horizon using multi-tupled meteorological input measurements. Based on genetic programming and ensemble of neural networks, authors in [15] proposed a robust approach to predict wind power. In li2020wind, a wind power prediction method is proposed by combining the advantages of wavelet decomposition, SVM, and atomic search algorithm. In this hybrid approach, wavelet decomposition has been used to preprocess the input wind power measurement, and the Atomic search algorithm is applied to optimize the SVM for improved wind power prediction. The study in li2021improving designed an optimized SVR model to achieve more reliable wind power prediction. Importantly, in this approach, cuckoo search arithmetic-based optimization is used to determine the SVR model's optimal hyper-parameters. Data collected from a French wind farm has been used to show the prediction efficiency of this approach. In [16], a method integrating a data mining technique and enhanced SVM algorithm is introduced for a short-term of wind power. This method used data mining to explore the correlation linking wind powers and wind speed and correct the invalid original data. However, this method is not verified for long-term prediction [16]. The authors in [17] propose an approach for forecasting wind power based on machine learning and kernel density estimation. This approach use information from nearby measurements and meteorological forecast values to enhance prediction precision. In [18], ensemble learning-driven methods, including Random Forest (RF) and Boosted Trees, have been implemented to predict wind power. Measurements from four wind turbines placed in

Turkey and France have been adopted to verify the prediction accuracy of these techniques. Results reveal the suitable performance of the ensemble methods in predicting wind power and their superior performance compared to Support Vector Regression and Gaussian process regression. In addition, the findings show that incorporating lagged variables improves the prediction accuracy of the ensemble models. In [19], a hybrid generative adversarial network (GAN)-driven model is proposed for short-term wind power forecasting. Specifically, the prediction problem has been addressed as a min-max game using GAN. The authors in shahid2020novel introduced an approach to predict wind power by embedding wavelet kernels with long short-term memory (LSTM). This approach, called WN-LSTM, aims to catch the dynamic behavior of wind power measurements. Prediction results using actual data from wind farms in Europe recommends using WN-LSTM compared to other existing machine learning techniques. Recently, a deep learning-driven approach called stacked independently recurrent autoencoder (IRAE) has been presented in [20] to effectively predict wind power. At first, the variational mode decomposition approach is applied to decompose data into subsequences to reduce noise in data. Then, IRAE is utilized for wind power prediction. This approach exhibited better prediction compared to some commonly used prediction models. The authors in [21] proposed an autoregressive dynamic adaptive (ARDA) model to predict wind power in real-time. This ARDA approach is dynamically adapted to catch data variations. This approach exhibited enhanced performance compared to ARIMA and LSTM models.

Precise prediction of wind power is crucial to sustainably integrate the wind power in a power grid. The overarching aim of this work is to design an approach enabling an efficient forecast of wind power production using times series SCADA data from a wind turbine. This paper investigated the performance of two latent variables regression methods, i.e., partial least squares (PLS) and principal components regression (PCR) [22]–[24]. To the best of our knowledge, these latent variables regression (LVR) models have not been widely exploited for wind power prediction. Multivariate data-derived models, e.g., PCR and PLS, are frequently implemented to correlated multivariable by performing regression on a reduced number of uncorrelated variables (i.e., latent variables), which consists of linear combinations of the original variables [25]–[27]. When implementing LVR models, only fewer latent variables are utilized instead of using raw measurements. Generally, LVR models result in well-conditioned parameter estimations and reliable model predictions. One of the main contributions of this paper is bringing this LVR method to the attention of the renewable energy community and show how it can be applied for predicting wind power. Real measurements recorded every 10 minutes from an actual wind turbine are adopted to show the prediction quality of the investigated techniques.

The wind power data used in this study is described in Section II. Then, the regression models are briefly reviewed in

Section III. In Section IV, the prediction results are discussed. Lastly, conclusions are presented in Section V.

II. WIND TURBINE DATASETS

The wind power data studied in this article is gathered from a high-wind-speed wind turbine Senvion MM82 in France. The principal characteristics of this wind turbine are given in Table I. The hub heights of this wind turbine of 80 meters make it a choice for sites with height restrictions (Table I). The dataset is recorded every ten minutes. The recorded measurements contain twelve input variables and the active power as the response variable.

TABLE I
WIND TURBINE MAIN CHARACTERISTICS.

Senvion MM82	
Rated power	2,050.0 kW
Rotor diameter	82.0 m
Hub height	80 m
Number of blades	3
Rotor speed, max	7.1 U/min
Generator Speed, max	1,800.0 U/min
Cut-in wind speed	3.5 m/s
Rated wind speed	14.5 m/s
Cut-out wind speed	25.0 m/s

The training measurements gathered from September 1st, 2013 to May 14, 2014, is employed for constructing the prediction models. Table II presents summary statistics of the training dataset. Table II indicates that several variables were negatively skewed with relatively large kurtosis.

III. METHODS

This section describes the prediction methodologies PLSR and PCR used for predicting wind power.

A. Principal Component Regression (PCR)

The PCR approach is implemented into two stages: first, we apply a principal component analysis on the matrix of the explanatory variables \mathbf{X} . Then, Ordinary Least Squares regression is applied to link the retained principal component and the response variable [28], [29] (Figure 1).

By using PCA, the data matrix \mathbf{X} can be expressed as a sum of the approximated matrix, $\hat{\mathbf{X}}$, and residual data, \mathbf{E} .

$$\mathbf{X} = \mathbf{T}\mathbf{W}^T = \sum_{i=1}^k t_i w_i^T + \sum_{i=k+1}^m t_i w_i^T = \hat{\mathbf{X}} + \mathbf{E} \quad (1)$$

where $\mathbf{T} \in R^{n \times m}$ denotes a matrix of the principal components (PCs) and $\mathbf{W} \in R^{m \times m}$ refers to the loading matrix. In the case of cross-correlated data, \mathbf{X} , the first ' k ' PCs (where $k < m$) are enough to preserve pertinent information in the input data. Here cumulative percentage variance(CPV) procedure is utilized to select the number of PCs to retain in the model [30]. Let $\hat{\mathbf{T}} \in R^{n \times k}$ is the matrix of the k retained

TABLE II
SUMMARY STATISTICS OF THE TRAINING MEASUREMENTS.

	mean	std	min	Q.25	Q.5	Q.75	max	skewness	kurtosis
Wind speed (m/s)	5.697	2.512	0	4.29	5.70	7.12	19.44	0.27	3.88
Torque (Nm)	2206.75	2310.09	-734.41	350.24	1625.31	3223.94	10875.70	1.37	4.72
Generator converter speed (rpm)	1137.29	596.29	1.86	971.80	1257.59	1640.26	1804.78	-0.81	2.43
Generator speed (rpm)	1135.81	596.57	-0.07	970.07	1256.11	1639.08	1803.15	-0.81	2.43
Converter torque (Nm)	2342.14	2265.35	-9.32	641.63	1838.41	3288.11	10915.70	1.36	4.82
Rotor speed (rpm)	10.83	5.70	0	9.23	11.96	15.64	17.21	-0.81	2.42
Pitch angle (deg)	11.01	25.23	-37.06	-0.98	-0.95	0.11	92.68	2.10	6.38
Gearbox oil sump temperature (°C)	52.95	8.84	11.47	52.08	56.14	57.89	63.16	-2.53	9.69
Absolute wind direction corrected (deg)	172.43	100.22	0.01	66.39	190.51	249.15	359.99	-0.11	1.84
Rotor bearing temperature (°C)	27.13	6.18	5.07	23.53	27.94	31.30	41.51	-0.57	3.25
Generator bearing 1 temperature (°C)	39.70	7.24	6.09	37.61	41.40	44.28	61.56	-1.78	6.87
Generator bearing 2 temperature (°C)	39.17	7.36	4.28	37.23	40.73	43.66	60.52	-1.90	7.30
Active power (kW)	564.01	626.06	-12.38	79.15	320.11	873.02	2266.90	1.20	3.34

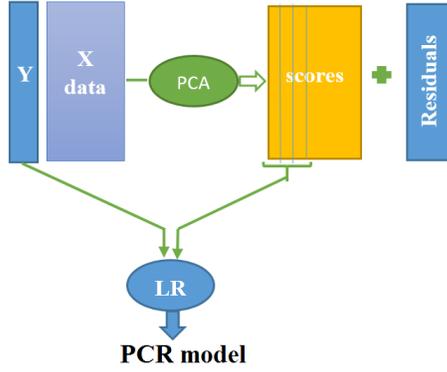


Fig. 1. A conceptual representation of PCR model.

principal components. In the PCR model, the linear regression between $\hat{\mathbf{T}}$ and the response variable \mathbf{y} is obtained via solving the optimization problem (2).

$$\hat{\beta} = \arg \min_{\beta} \left(\|\hat{\mathbf{T}}\beta - \mathbf{y}\|_2^2 \right) \quad (2)$$

The least squares solution is expressed as:

$$\hat{\beta} = \left(\hat{\mathbf{T}}^T \hat{\mathbf{T}} \right)^{-1} \hat{\mathbf{T}}^T \mathbf{y}. \quad (3)$$

B. Partial Least Square (PLS)

Consider the matrix of the explanatory variables $\mathbf{X} \in \mathbb{R}^{n \times m}$ and the response matrix $\mathbf{Y} \in \mathbb{R}^{n \times p}$. The core idea of PLS is extracting the latent variables iteratively via the maximization of the covariance of the extracted latent variables. PLS comprises two models: the inner model and outer model [31]–[33] (Figure 2). The outer model is given as:

$$\begin{cases} \mathbf{X} = \sum_{i=1}^l \mathbf{t}\mathbf{p}_i^T = \mathbf{T}\mathbf{P}^T + \mathbf{G} \\ \mathbf{Y} = \sum_{i=1}^l \mathbf{u}\mathbf{q}_i^T = \mathbf{U}\mathbf{Q}^T + \mathbf{F}, \end{cases} \quad (4)$$

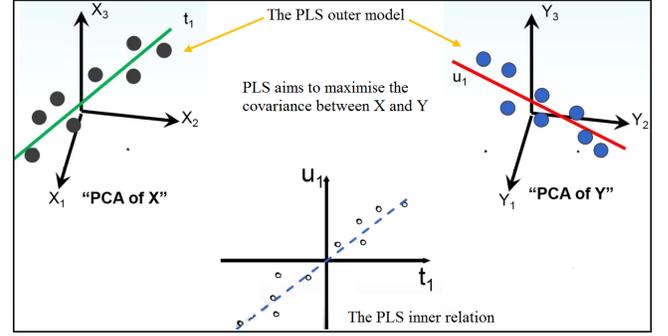


Fig. 2. A conceptual representation of PLSR.

where, $\mathbf{T} \in \mathbb{R}^{n \times l}$ and $\mathbf{U} \in \mathbb{R}^{n \times q}$ denotes a matrix of the transformed uncorrelated variables. The loading matrices of input and output space are $\mathbf{P} \in \mathbb{R}^{m \times l}$ and $\mathbf{Q} \in \mathbb{R}^{p \times q}$, respectively. Here, \mathbf{G} and \mathbf{F} refer to the model residuals. The cross-validation procedure is applied to determine the number of PCs, l . The maintained PCs of the input and output space, i.e., \mathbf{T} and \mathbf{U} , are linked through the following linear inner model:

$$\mathbf{U} = \mathbf{T}\beta + \mathbf{E}, \quad (5)$$

where β is a regression matrix, and \mathbf{E} is a residual part. The input \mathbf{Y} can now be expressed as:

$$\mathbf{Y} = \mathbf{T}\beta\mathbf{Q}^T + \mathbf{F}^*, \quad (6)$$

The construction of the PLS model is performed in a sequential manner [34], [35].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The wind power prediction based on regression models (e.g., PCR and PLSR) is performed into three stages as summarized in Figure 3. Using training data, prediction models

have been constructed. Here, we used 5-fold cross-validation to train the models and avoid overfitting. Then, these models are employed for predicting the response based on unseen input data. The quality of the predicted is judged by computing R-squared (R^2), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad (8)$$

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \text{mean}(Y))^2}, \quad (9)$$

A. Preliminary data analysis

Generally speaking, wind turbines are constructed to work within an interval of wind speeds and with maximum power. This power value is the nominal value of the generator, but it is rarely achieved. The operation of the wind turbine relies on wind speed. This wind speed will constantly vary, as shown in Figure 4(a), which represents the evolution of wind speed and power production from the 2,05 MW Senvion MM82 wind turbine. The wind power curve (Figure 4(a)) is usually used to detect abnormalities in wind turbines by comparing the empirical wind power curve with the theoretical one provided by the constructor. Figure 4(b) illustrates the plot of rotor speed against wind speed. One can see that the rotor speed is increasing in accordance with wind speed until reaching the cut-in wind speed. Another important curve to characterize the operating condition of a wind turbine is the pitch curve which plot winds speed against pitch angle (Figure 4(c)). Similar to the power curve analysis, deviations in characteristic operation in the pitch curve can be identified, and further analysis for anomaly detection can be performed. Figure 4(c) displays a wind speed variation as a function of blade pitch angle degree under normal conditions.

B. Models design

The training measurements are utilized to design prediction methods. Here, we used the 5-fold cross-validation procedure to find the model parameters. One crucial step in PCR and PLSR models consists in the selection of the number of PCs. To this end, the cumulative percentage variance(CPV) method is adopted because of its simplicity and efficiency. Figure 5(a-b) shows respectively the CPV explained in the input data \mathbf{X} for both PLSR and PCR, as well as the CPV explained in the output for the PLSR model. From Figure 5(a-b), It can be seen that two PCs are sufficient to describe around 99% of the variability in \mathbf{X} in the two models. We used two PCs to build the PCR model, while three PCs are used for the PLSR model.

Table III summarizes the prediction accuracy of the PLSR and PCR models when using the testing dataset. As shown in

Table III and Figure 6, the measured wind power prediction match well with the predicted power from the PLSR and PCR models.

TABLE III
SUMMARY OF MODELS QUALITY.

	R^2	RMSE	MAE
PCR	0.949	146.102	101.601
PLSR	0.949	146.007	101.577

The prediction results of the two models based on testing data are illustrated in Figure IV and Table V. The results in Table III indicate that the prediction performances of PCR and PLSR are relatively comparable. Also, this result confirms that the LVR models are relatively appropriate in predicting wind power.

TABLE IV
MODELS COMPARISON USING THE TRAINING DATASETS.

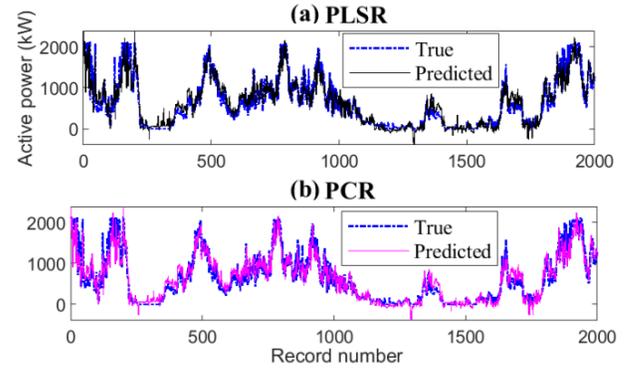


TABLE V
SUMMARY OF MODELS QUALITY USING TESTING DATA.

	R^2	RMSE	MAE
PCR	0.931	156.451	113.348
PLSR	0.930	157.702	114.026

V. CONCLUSION

A reliable prediction of wind power production may be a tool for facilitating wind turbine health monitoring. This paper presented a comparative study between two latent variable regression approaches (i.e., PCR and PLSR) for predicting wind power using SCADA dataset. The data used to verify the quality of these two methods has been collected from a 2,05 MW Senvion MM82 wind turbine. The conducted experiments demonstrated the PCR and PLSR achieve acceptable prediction performance of wind power. These two models achieved an R^2 of 93% for predicting wind power. This study demonstrated the feasibility of using LVR methods to predict wind power.

As future work, we plan to use these models to develop statistical monitoring schemes able to detect faults in wind tur-

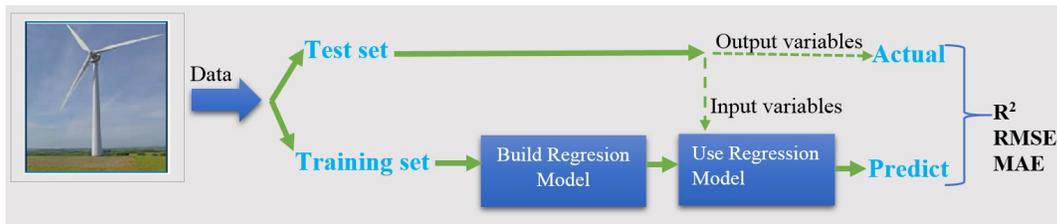


Fig. 3. General data flow diagram of regression models.

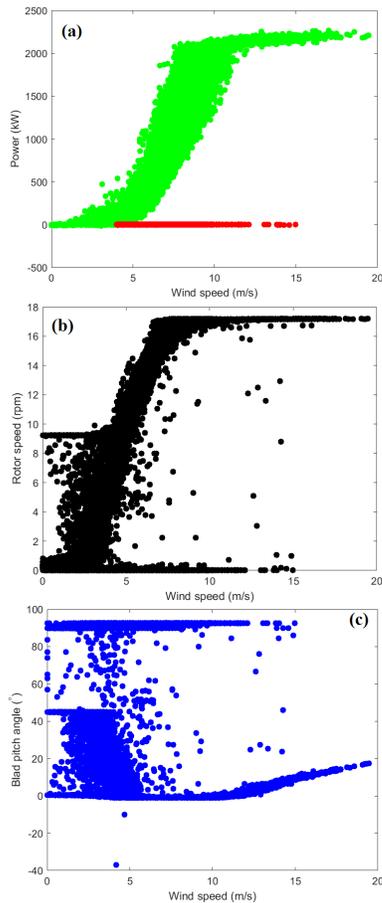


Fig. 4. (a) Active power, (b) rotor speed, and (c) blade pitch angle as a function of the wind speed for the 2,05 MW Senvion MM82 wind turbine obtained from SCADA data.

bines. Specifically, LVR models will be combined with monitoring charts, such as CUSUM and generalized likelihood tests to detect sensors and process anomalies in wind turbines [22]. LVR models will generate residuals that monitoring charts will check to uncover potential anomalies in the inspected wind turbine [36].

ACKNOWLEDGEMENT

This publication is based upon work supported by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2019-CRG7-3800.

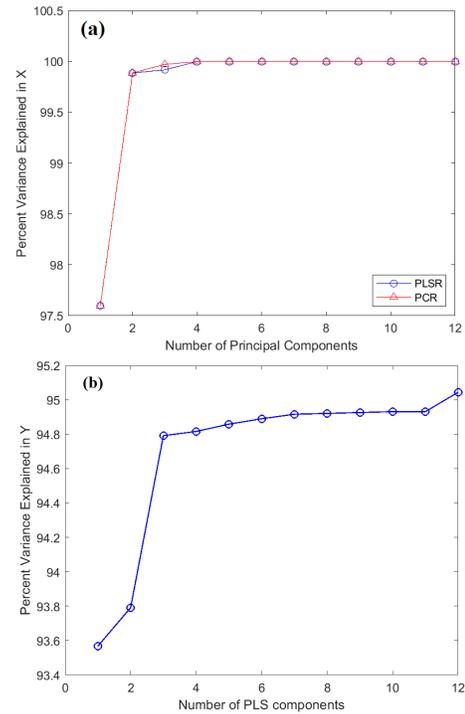


Fig. 5. Selection of the number of principal components in PLSR and PCR.

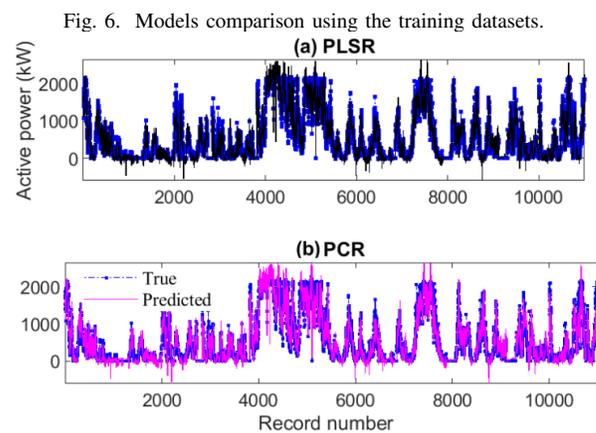


Fig. 6. Models comparison using the training datasets.

REFERENCES

- [1] T. Ouyang, X. Zha, and L. Qin, "A combined multivariate model for wind power prediction," *Energy Conversion and Management*, vol. 144, pp. 361–373, 2017.

- [2] A. Kisvari, Z. Lin, and X. Liu, "Wind power forecasting—a data-driven method along with gated recurrent neural network," *Renewable Energy*, vol. 163, pp. 1895–1909, 2021.
- [3] E. T. Renani, M. F. M. Elias, and N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," *Energy Conversion and Management*, vol. 118, pp. 193–203, 2016.
- [4] Q. Wu, F. Guan, C. Lv, and Y. Huang, "Ultra-short-term multi-step wind power forecasting based on CNN-LSTM," *IET Renewable Power Generation*, vol. 15, no. 5, pp. 1019–1029, 2021.
- [5] T. Ouyang, X. Zha, L. Qin, Y. He, and Z. Tang, "Prediction of wind power ramp events based on residual correction," *Renewable Energy*, vol. 136, pp. 781–792, 2019.
- [6] F. Ding, Z. Tian, F. Zhao, and H. Xu, "An integrated approach for wind turbine gearbox fatigue life prediction considering instantaneously varying load conditions," *Renewable energy*, vol. 129, pp. 260–270, 2018.
- [7] S.-l. Feng, W.-s. Wang, C. Liu, and H.-z. Dai, "Study on the physical approach to wind power prediction," *Proceedings of the CSEE*, vol. 30, no. 2, pp. 1–6, 2010.
- [8] S. Han, Y.-h. Qiao, J. Yan, Y.-q. Liu, L. Li, and Z. Wang, "Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network," *Applied Energy*, vol. 239, pp. 181–191, 2019.
- [9] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 243–254, 2014.
- [10] X. Yuan, C. Chen, Y. Yuan, Y. Huang, and Q. Tan, "Short-term wind power prediction based on lssvm-gsa model," *Energy Conversion and Management*, vol. 101, pp. 393–401, 2015.
- [11] Z. Su, J. Wang, H. Lu, and G. Zhao, "A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting," *Energy conversion and management*, vol. 85, pp. 443–452, 2014.
- [12] J. Yan and T. Ouyang, "Advanced wind power prediction based on data-driven error correction," *Energy Conversion and Management*, vol. 180, pp. 302–311, 2019.
- [13] M. Yesilbudak, S. Sagioglu, and I. Colak, "A novel implementation of knn classifier based on multi-tupled meteorological input data for wind power prediction," *Energy conversion and management*, vol. 135, pp. 434–444, 2017.
- [14] X. Jiang, S. Day, D. Clelland, and X. Liang, "Analysis and real-time prediction of the full-scale thrust for floating wind turbine based on artificial intelligence," *Ocean Engineering*, vol. 175, pp. 207–216, 2019.
- [15] A. Zameer, J. Arshad, A. Khan, and M. A. Z. Raja, "Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks," *Energy conversion and management*, vol. 134, pp. 361–372, 2017.
- [16] C. Li, S. Lin, F. Xu, D. Liu, and J. Liu, "Short-term wind power prediction based on data mining technology and improved support vector machine method: A case study in northwest china," *Journal of Cleaner Production*, vol. 205, pp. 909–922, 2018.
- [17] J. Lv, X. Zheng, M. Pawlak, W. Mo, and M. Miśkiewicz, "Very short-term probabilistic wind power prediction using sparse machine learning and nonparametric density estimation algorithms," *Renewable Energy*, vol. 177, no. November, pp. 181–192, 2021.
- [18] J. Lee, W. Wang, F. Harrou, and Y. Sun, "Wind power prediction using ensemble learning-based models," *IEEE Access*, vol. 8, pp. 61 517–61 527, 2020.
- [19] B. Zhou, H. Duan, Q. Wu, H. Wang, S. W. Or, K. W. Chan, and Y. Meng, "Short-term prediction of wind power and its ramp events based on semi-supervised generative adversarial network," *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106411, 2021.
- [20] L. Wang, R. Tao, H. Hu, and Y.-R. Zeng, "Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder," *Renewable Energy*, vol. 164, pp. 642–655, 2021.
- [21] F. Zhang, P.-C. Li, L. Gao, Y.-Q. Liu, and X.-Y. Ren, "Application of autoregressive dynamic adaptive (arda) model in real-time wind power forecasting," *Renewable Energy*, vol. 169, pp. 129–143, 2021.
- [22] F. Harrou, Y. Sun, M. Madakyaru, and B. Bouyedou, "An improved multivariate chart using partial least squares with continuous ranked probability score," *IEEE Sensors Journal*, vol. 18, no. 16, pp. 6715–6726, 2018.
- [23] M. Madakyaru, F. Harrou, and Y. Sun, "Improved anomaly detection using multi-scale pls and generalized likelihood ratio test," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2016, pp. 1–6.
- [24] F. Harrou and Y. Sun, "Enhanced anomaly detection via pls regression models and information entropy theory," in *2015 IEEE Symposium Series on Computational Intelligence*. IEEE, 2015, pp. 383–388.
- [25] G. Wang and J. Jiao, "Quality-related fault detection and diagnosis based on total principal component regression model," *IEEE Access*, vol. 6, pp. 10 341–10 347, 2018.
- [26] F. Harrou, Y. Sun, A. S. Hering, M. Madakyaru *et al.*, *Statistical process monitoring using advanced data-driven and deep learning approaches: theory and practical applications*. Elsevier, 2020.
- [27] F. Harrou, J. F. Ramahaleomiarantsoa, M. N. Nounou, and H. N. Nounou, "A data-based technique for monitoring of wound rotor induction machines: A simulation study," *Engineering science and technology, an international journal*, vol. 19, no. 3, pp. 1424–1435, 2016.
- [28] M. Stone and R. J. Brooks, "Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression," *J. Roy. Stat. Soc. B*, vol. 52, no. 2, p. 237, 1990.
- [29] I. Frank and J. Friedman, "A statistical view of some chemometric regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–148, 1993.
- [30] F. Harrou, M. N. Nounou, and H. N. Nounou, "Detecting abnormal ozone levels using pca-based glr hypothesis testing," in *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. IEEE, 2013, pp. 95–102.
- [31] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual reviews in control*, vol. 36, no. 2, pp. 220–234, 2012.
- [32] Y. Wang, Y. Wei, T. Liu, T. Sun, and K. T. Grattan, "Tdlas detection of propane/butane gas mixture by using reference gas absorption cells and partial least square approach," *IEEE Sensors Journal*, vol. 18, no. 20, pp. 8587–8596, 2018.
- [33] F. Harrou and Y. Sun, "Pls-based memory control scheme for enhanced process monitoring," in *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*. IEEE, 2016, pp. 477–480.
- [34] P. Geladi and B. R. Kowalski, "Partial least square regression: A tutorial," *Analytica chimica Acta*, vol. 185, pp. 1–17, 1986.
- [35] Y. Hiroyuki, Y. B. Hideki, F. C. E. O. Hiromu, and F. Hideki, "Canonical correlation analysis for multivariate regression and its application to metabolic fingerprinting," *Biochemical Engineering Journal*, vol. 40, pp. 199–204, 2008.
- [36] F. Harrou, Y. Sun, A. S. Hering, M. Madakyaru, and A. Dairi, "Linear latent variable regression (lvr)-based process monitoring," in *Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches*. Elsevier, 2021, pp. 19–70.