


An Approach for Erosion and Power Loss Prediction of Wind Turbines Using Big Data Analytics

Dina Fawzy, Sherin Moussa^(✉) , and Nagwa Badr

Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
{dina.fawzy, sherinmoussa, nagwabadr}@cis.asu.edu.eg

Abstract. Due to the huge costs associated with wind energy development, this makes wind farms maintenance and production reliability are of high necessity to ensure sustainability. The continuous evolution of turbines industry has a serious impact on the operation and maintenance costs. Thus, monitoring wind turbines performance and early deterioration prediction are highly required. During the operational life of turbines, some components are persistently exposed to extreme environmental influences that result in their edge erosion. Sensors can be deployed in wind farms to detect such factors, where vast quantities of incomplete, heterogeneous and multi-sourced data are rapidly generated. Hence, wind-related data have been considered as big data that necessitate the intervention of big data analytics for accurate data analysis, which become severely hard to process using traditional approaches. In this paper, we propose the Wind Turbine Erosion Predictor (WTEP) System that uses big data analytics to handle the data volume, variety, and veracity and estimate the turbines erosion rate, in addition to the total power loss. WTEP proposes an optimized flexible multiple regression technique. Experiments show that WTEP achieves high erosion rate prediction accuracy with fast processing time. Thus, it effectively evaluates the accompanied percentage of power loss for wind turbines.

Keywords: Big data analytics · Data mining · Regression analysis · Association rules · Apriori. Principal component analysis · Wind farms reliability · Wind farms maintenance · Erosion. Power prediction

1 Introduction

Wind energy has become a popular source of energy around the world, where its development plants cost huge investments. This requires a keen management of their economic efficiency to ensure higher yields for energy cost reduction [1]. The wind turbine reliability is a critical factor in the success of a wind energy project, which implicates reducing the expensive operation and maintenance (O&M) costs that affect the project's revenue [2]. During the wind turbine's operation, some components, principally the rotor blades, are continuously exposed to certain environmental conditions over time, such as rain, temperature and sand. This results in the deterioration of

the blade's material surface and the increase of its surface roughness if unprotected, leading to its erosion after an average of 2 years of turbines installation and to performance decrease. This requires that in-service maintenance should be performed on the turbine for at least 12 years of operation if it meets its design life, which results in huge maintenance costs [3]. For significant erosion rates (5%–20%), O&M costs are expected to be within \$27–54/MWh. Rain erosion occurs during the processing of turbines in heavy rain. During the high velocity of fallen liquid on a solid target, a high pressure is developed between the solid and liquid, where it varies over many locations [4]. On the other hand, sand erosion exists in the desert environments with the movable dirt and airborne particles affecting turbines' blades, which increase roughness and decrease aerodynamic performance [5, 6]. The high temperature affects wind turbines because the erosion rate increases when the viscosity of liquid reduces [7]. In addition, the increase of wind speed and air density has a positive impact on power production. However, when it exceeds 6 m/s over dry soils, it carries sand and dust towards turbines, leading to erosion. Wind direction is highly effective as well if it is like the sand direction. Whilst a slope that is greater than 20 m affects the angle between the surface and sand/dust, resulting in surface erosion [8]. Sensors can be deployed in the desired location of wind energy plants to monitor such environmental turbines erosion causes, by collecting sensors' data that could be heterogeneous and incomplete massive data [9].

The nature of such wind energy data enforces the desperate need to utilize big data analytics to handle such issues effectively. Big data refers to the collections of so huge and heterogenous datasets that are critically sophisticated to process using customary approaches [10]. This is due to the mainly characterized 4Vs of big data, representing Velocity, Veracity, Variety, and Volume. Big data analytics refers to the usage of advanced analytic techniques against these 4Vs [11]. Wind farm engineers can use big data analytics to manage the risks in order to achieve production goals and recommend activities to address shortfall detected [12]. Thus, the prediction of erosion rate is an efficient way to manage the cost impacts of wind farms through the power usage prediction and the achievement of the supply on demand concept.

In this paper, we introduce the Wind Turbine Erosion Predictor (WTEP) System that uses big data analytics to handle the data volume, variety, and veracity in order to predict turbines' erosion rate. WTEP is built on the top of Trio-V Wind Analyzer system, which is a generic integral system that analyzes the land suitability of a potential location and recommends a distribution layout design, in addition to power prediction using big data analytics prior to wind farms development. WTEP can predict the erosion rate and evaluate its resultant power loss at any spatial region under study based on its environmental factors data rather than other customizable studies. The remaining parts of the paper are organized as follows: Sect. 2 overviews the related works in wind farm reliability, data reduction and power prediction in the wind energy domain. Section 3 presents the proposed system, with a detailed discussion of its architecture. Section 4 explains the experimental approach and the study area. Section 5 discusses the different applied experiments and the associated results. Lastly, Sect. 6 summarizes the conclusion and the future work.

2 Related Work

2.1 Wind Farm Reliability Approaches

Many researches have considered analyzing wind turbines data to maintain wind plants. Most of these studies were done to ensure the reliability of wind farms through extracting the failure history of wind turbines and monitoring their status in order to reduce downtime and increase availability. Authors in [13] monitored the performance of wind farm turbines to detect their downtimes by integrating SCADA system with the turbine's control system and controlling the detected turbines to manage the requirements of power consumption and turbine efficiency. In [14], a platform was developed using the National Reliability Database for turbines' failure detection. Another platform in [15] aimed to discover the hidden patterns in the turbine statuses using the random forest multiclass classification model. SCADA monitoring system was considered in [16] to detect failures by applying an anomaly detection technique. In [17], SCADA data were used to classify the failure events of turbines into severity categories and apply a statistical methodology for each category to decide the wind farm reliability. Since the previous researches have tackled the problem from the engineering perspective, a minimal research effort was dedicated for analyzing the operational and environmental data of wind turbines to raise their performance and reduce the associated maintenance costs. Moreover, most of these studies were poor to process scalable and variable data, since SCADA data are static with a specific format.

2.2 Data Reduction Techniques

Traditional data mining techniques were investigated to fit big data processing. Near Filter Classifier (NFC) upgrades K-Nearest Neighbor (KNN) classification by adding a dimensionality reduction step [18]. It computes the class distribution per every dataset parameter, then sorts the parameters by the calculated value. In [19], Parallel processing was used in the decision tree data mining technique to mine a huge amount of data streams. In addition, "Scalable Advanced Massive Online Analysis (SAMOA)" technique used parallel processing with distributed decision trees for data mining classification over big data [20]. Another upgrade was applied to reduce big data volume using parallel processing by applying K-means on several nodes and combining the results [21]. Although these researches were dedicated to reducing data volume, but they were poor to reach high accuracy that doesn't exceed 60%, with high processing time that reaches 100s with five neighbors [18]. This is in addition to the extra communication time between nodes in the parallel processing approaches, which leads to excessive processing time [19].

2.3 Prediction Techniques for Wind Energy

Several prediction techniques were dedicated for wind energy domain. In [22], the weather prediction used genetic programming. The wind speed and generated power were predicted in [23] using a fuzzy expert system. Artificial neural networks were used in [24] to predict electrical power generated from wind farms. However, such

prediction techniques have just reached 85% accuracy [22] and 20% error rate [24]. The fuzzy system consumes much processing to learn the model that cannot fit big data processing [23].

The contributions in this proposed research can be summarized as follows. (1) We propose the Wind Turbine Erosion Predictor (WTEP) as an integral system for predicting the erosion rate of wind turbines from the data analytical perspective to decrease turbines failure rate. (2) It uses big data analytics to handle wind turbines data volume, variety and veracity, where Double-Reduction Optimum Apriori (DROA) approach is proposed. (3) It presents a new Optimized Flexible Multiple Regression (OFMR) approach to fit big data processing to predict wind turbines erosion rate, taking into consideration the different affecting environmental factors that can be adapted and generalized to wherever the study area is located. Hence, it can fit to evaluate any wind farm irrespective of its location rather than any customized systems to study certain territories, which is one of the main strengths of this proposed system. (4) It predicts the power loss accompanied by the predicted erosion rate.

3 The Proposed Solution

In this section, we present the proposed Wind Turbine Erosion Predictor (WTEP) system. As shown in Fig. 1 representing the system architecture, WTEP is developed in accordance with Trio-V Wind Analyzer system to achieve WTEP functionalities. The study presents the complete work of the proposed system, providing its architecture, the detailed explanation and implementation of all its components, and the associated experimentations. WTEP deals with the data layer managing the factors data of wind farms, in addition to the presentation layer that is connected to the sensors and Google map to manage the user selections and to display the analytical results plotted on the map or generated in reports. WTEP works as shown in Fig. 2. The system user determines the wind farm location and the reduction method to apply on the sensed factors data. WTEP collects the factors data from the sensors in the defined location and then manages their biases and noises using the Variety-Veracity WA Handler [25].

Next, the selected reduction method is applied on the data using the Volume WA Handler. The resultant processed data is then used to analyze the erosion rate and evaluate the associated power loss using Trio-V Power Loss Analyzer in a detailed report, showing each cell with its corresponding erosion and power loss rates. Data velocity, in terms of data processing in the form of a stream, is not handled in this system since it doesn't require real-time processing. Sensors data are accumulated in the data layer, taking into consideration the time representation of data as another data dimension for offline processing and analysis, where the collected data are strongly related to the recording time. Thus, a data stream handler is not required. The main components of WTEP are further explained hereinafter.

3.1 Presentation Layer

This layer provides the User Interface (UI) of the system, which enables the user to determine the wind farm's location and collect the associated environmental factors'

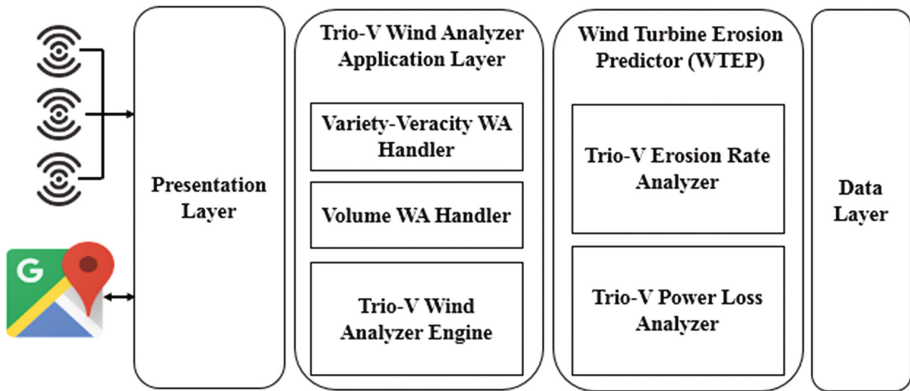


Fig. 1. The proposed system architecture

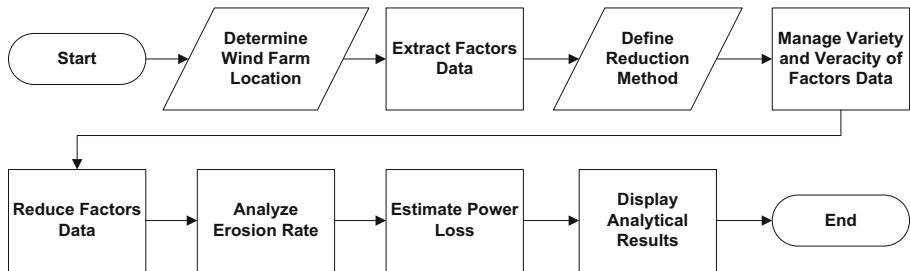


Fig. 2. Wind Turbine Erosion Predictor (WTEP) system flowchart

data from the deployed sensors. It then divides the land into cells of equal size as per a user-defined cell size parameter. In addition, it allows the user to choose a reduction method to manage the huge size of data. Finally, WTEP prediction results are displayed in a detailed report with the suitable graphs per cell, visualizing the expected erosion rate and the corresponding predicted power loss rate.

3.2 Trio-V Wind Analyzer Application Layer

This layer handles the huge Volume, Veracity and Variety (Trio-V) features of the collected environmental factors' data, which are generated from the sensors deployed at the land under study. Then, it evaluates the suitability of this land to establish a wind farm and suggests a distribution layout for the turbines. The main components are explained as follows.

Variety-Veracity WA Handler. This module manages the biases and noise detected in the sensors data while considering its big data nature. It validates the data quality and data inconsistencies before storage into the data layer through several data cleansing processes, including noisy data deletion and filling in missing data with the mean value.

Encoding-decoding processes are considered as well to transform specific factors' data into a certain format to be processed [25].

Volume WA Handler. The deployed sensors generate enormous amounts of data. Thus, this module applies the reduction method that has been selected from the presentation layer. The data layer structure includes different environmental factors to identify each cell, where each factor has excessive amount of data per one cell. WTEP provides several alternative reduction methods merged from different reduction techniques to apply on the cells' factors data. Some of these techniques are responsible for reducing the number of cell factors used for analytics, like Principle Component Analysis (PCA) and Association Rules (ARs) (i.e. column reduction), while others reduce the amount of cell data, like aggregations. PCA is a data reduction technique that uses a mathematical approach to reduce many correlated parameters into a small set of uncorrelated parameters called principal components (PCs). WTEP uses the correlation approach to match the resultant PCs to their corresponding factors in the original dataset by calculating the correlation coefficient between every PC(x) and each factor (y) in the original data using Eqs. (1), (2), (3), and (4) [26]. The factor having the highest correlation coefficient represents the PC.

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad (1)$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} \quad (2)$$

$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \quad (3)$$

$$\text{CorrelationCoff} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (4)$$

Where n represents the number of records for the cell's factors, x represents the resultant PC; and y is the cell's factors data needed to be reduced [26]. The higher the result means that this PC is most correlated to this factor. As for the ARs, we enhanced the original version of the Apriori technique to fit the big data processing by introducing our optimized Apriori algorithm named "Double-Reduction Optimum Apriori" (DROA) to extract the most informative relationships between the factors using the criteria of support and confidence according to Eqs. (5) and (6) [27]. The proposed DROA ARs optimizes the Apriori algorithm to support big data volume by applying two phases before running the basic Apriori; (1) using database scanning time reduction that saves a screenshot of the desired transactions between erosion factors related to a certain area in a supportive map data structure, which decreases the traditional Apriori processing time. (2) Using transactions reduction that reduces transactions by discarding the unsuitable ones that violate erosion values constraints [25]. This allows DROA ARs to work efficiently on a huge number of transactions.

$$\text{Support}_i = \frac{\text{FPi}}{\text{TFP}} \quad (5)$$

$$\text{Confidence } (A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)} \times 100 \quad (6)$$

Where support_i is the support of the i th factor, FPi is the number of times the i th factor is found, and TFP is the total number of factors found. $\text{Confidence } (A \rightarrow B)$ represents the confidence of occurrence; if A occurs, then B will occur too. For more processing efficiency, WTEP allows merging several approaches of reduction methods to additionally reduce data. Thus, the reduction alternatives are: Aggregation functions only, Aggregation followed by PCA, Aggregation followed by DROA ARs, PCA followed by Aggregation, or DROA ARs followed by Aggregation. For instance, aggregation only would be sufficient for small datasets, whereas DROA ARs and PCA are more appropriate for huge datasets.

Trio-V Wind Analyzer Engine. This module is the core of Trio-V Wind Analyzer. It uses big data analytic techniques to perform land suitability analysis for wind farms prior to development. Trio-V determines land suitability through evaluating its environmental factors. Upon the positive evaluation, Trio-V Wind Analyzer recommends the optimum wind farm design that avoids the wake effect problem of turbines and maximizes the generated power by suggesting the suitable turbines' specifications and their distribution layout depending on the analyzed factors of the potential location. Accordingly, it then predicts the expected generated power from this recommended design [25].

3.3 Wind Turbine Erosion Predictor (WTEP)

This module explains the main WTEP functionalities in the following sub-modules:

Trio-V Erosion Rate Analyzer. This module is responsible for determining the erosion rate per one turbine for each land cell by evaluating specific environmental factors; rain, sand, wind speed, slope, wind direction, air density, and temperature [4–7, 33]. These environmental factors have variable values over time per year, where severe edge erosion can be caused if certain thresholds were exceeded as clarified in Table 1. WTEP considers the influence of such variance of values on erosion. For example, the dust storms could be very erosive compared to daily wind. However, the continuous direct exposure of everyday wind can even affect turbines erosion. Thus, all variances of the different factors are considered in the data analytics process to determine the erosion rate with an acceptable accuracy. WTEP aggregates all the previous erosion factors data from the collected sensors data for each cell at the potential land under study. Since WTEP processes one turbine at a time, the number of installed turbines per land cell is not considered as a factor in the erosion rate analysis. The Trio-V Erosion Rate Analyzer module estimates the erosion rate using our proposed Optimized Flexible Multiple Regression (OFMR) technique, which is an enhanced flexible form of the original multiple regression technique to support big data volume by considering a dynamic number of predictors, whereas the original multiple regression analysis is a statistical

Table 1. Erosion factors constraints

Factor	Erosion values constraint
Rain	>100 mm
Sand	1–200 μm
Steel temperature	>160°
Wind Speed	>6 m/s
Slope	>20
Wind direction	–
Air density	–

technique for analyzing relationships between factors using multiple predictors in fixed prediction equation parameters [28]. The following model in Eq. (7) shows the multiple linear regression model with K predictor variables.

$$Y = B_0 + B_1X_1 + B_2X_2 + \dots B_KX_K \quad (7)$$

Where parameter B_0 is the intercept of this plane, while “Y” is the unknown value to be predicted, and parameters $B_1, B_2 \dots B_K$ are referred to as regression coefficients [29]. OFMR supports big data volume by considering a dynamic number of predictors. It can build the model depending on the considered erosion factors based on the land under study, which are additionally reduced in the Volume WA handler, rather than building one fixed model based on all factors. Therefore, OFMR ensures more accurate results than traditional multiple regression technique, where the erosion rate “Y” is correlated only to the existing factors from Volume WA handler. OFMR consumes less processing time due to the flexibility in building the model with any number of predictors. It handles the biases and noise detected in the sensors data by ensuring data quality before building the model using the Variety-Veracity Handler. Thus, OFMR manipulates the overfitting problem in the traditional multiple regression model. Moreover, the factors’ values that are less than the erosion constraint thresholds will be ignored by the OFMR regression model. These features made WTEP adapted and generalized to evaluate any wind farm irrespective of its location, taking into consideration the different affecting environmental factors that would be associated by this location.

Trio-V Power Loss Analyzer. Leading edge erosion poses a major threat to the performance of wind turbines. The modeling of power output per one turbine is a trivial approach that assumes static wind parameters. However, a turbine’s status is inconstant, due to the erosion factors and wind parameters like wind speed, and air density that continuously change and affect the turbine’s status [30]. This module allows WTEP to evaluate the power loss rate according to the predicted erosion rates resulted from Trio-V Erosion Rate Analyzer. Power loss prediction is performed by applying a single linear regression technique using the predicted erosion rate value. Single regression analysis explores relationships statistically, containing one predictor as shown in Eq. (8) [31].

$$Y = B_0 + B_1X \quad (8)$$

Where “Y” is the power loss rate; “X” is the resultant erosion rate; B_0 is the intercept of this plane; and B_1 is the regression coefficient.

4 Case Study Area

Egypt climate is affected by several factors, including its position that lies between Africa and Asia [32]. These factors give Egypt a hot and sunny weather, with a very low humidity. The erosion factors value at the main areas in Egypt are presented in Table 2 [33, 34]. As for the wind direction and air density, their values are continuously changing during the year.

Table 2. Erosion factor values at egyptian areas

Area	Sand particles	Rain	Temperature	Wind speed	Slope
Western desert	<170 microns	20 mm	36 °C	5.3 m/s	32 m
North coast	<120 microns	196 mm	20 °C	6.2 m/s	18 m
Red sea	<80 microns	2.3 mm	30 °C	8 m/s	27 m

5 Experimental Results and Evaluation

WTEP has been developed using JAVA, MS SQL Server and APIs to some scientific libraries and external components. Experiments were held to evaluate WTEP from two points of view: the big data processing efficiency and wind analytics accuracy. Hence, the experimentation is categorized into: the erosion and power loss rates prediction accuracy, and the associated processing time versus the different reduction methods to emphasize that the proposed data reduction and prediction techniques are suitable for big data analysis, supported by a comprehensive comparison with the relevant existing state-of-arts. The evaluation was demonstrated on a machine having core i7, 2.70 GHz, 1T hard disk space, and 8 GB RAM. OFMR prediction accuracy is evaluated using the Root Mean Square Error (RMSE) as per Eq. (9) [29]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2} \quad (9)$$

Where “N” is the number of data points, “ x_i ” the original observed value and “ y_i ” is the predicted value corresponding to the current original data point “ x_i ”. RMSE values vary from 0 to 100 in order to be mapped to percentages, in which the smaller values indicate higher accuracy. RMSE values that are within (0–10) represent an accuracy from 90% and above. A sample of the experimental results of Red Sea area are discussed hereafter, since it is one of the potential areas in Egypt for wind plants. Three dataset sizes are used; small dataset D1 with 100,000 records, medium dataset D2 with

2,2500,000 records, and large dataset D3 with 5,750,000 records. The average temperature is 30 °C with 4 °C variation during winter. The rainfall is low, averaging 2.3 mm per year with average speed 8 m/s, air density equals to 1.2 kg/m³, and slope of 27 m. Red Sea area has occasionally dust storms as well [32]. The values of erosion factors differ depending on the measurement height, representing the height at which the values are detected and recorded. Thus, each dataset is tested for three turbine scale heights; 80, 50 and 30, representing the standard turbines' hub heights in the market. DROA ARs is investigated at confidence and support thresholds: 0.3, 0.5, 0.7 and 0.9, whereas PCA is studied at K-values: 5 and 3. These values have been configured as per many trials of experimental preparation, where their fair representation has been proven to the remaining values.

5.1 Erosion Rate Accuracy vs. Reduction Methods

Previous researches have considered wind farms reliability from the technical fault prediction perspective. In [13], 90% system availability has been achieved using SCADA data monitoring. Random forests data mining was used in [14] to predict turbines' failures with 8.3% error rate. Authors in [15] considered anomaly detection algorithms to detect turbines failures with 90% accuracy. In [24], 88.84% of failures were detected in a detection system of turbine failures using SCADA data. Despite of these previous researches, but they predicted the failures of turbines. To the best of our knowledge, WTEP is the first data analytical system that predicts turbines erosion and power loss rates using big data analytics. Thus, experiments were carried out to evaluate the big data processing efficiency by studying the RMSE results of the erosion rates prediction using OFMR over the different reduction methods. As shown in Fig. 3 for the RMSE results of the three datasets over WTEP reduction methods, the erosion prediction RMSE decreases as the dataset size increases.

Table 3 summarizes RMSE results over D3, representing the largest dataset. The aggregation only has the most accurate results due to the complete number of factors used, then DROA ARs with a reasonable accuracy results, and PCA has the least

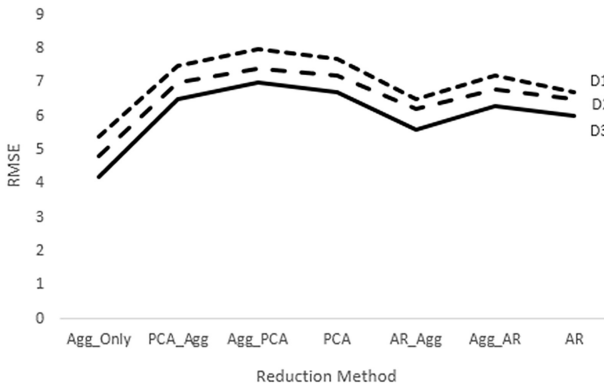


Fig. 3. Erosion prediction RMSE vs. reduction methods

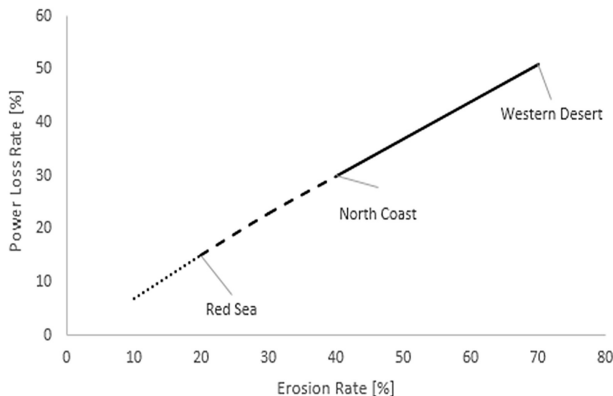
Table 3. Erosion rate prediction at D3

	Agg. Only	ARs+ Agg.	Agg. + ARs	PCA + Agg.	Agg. + PCA
RMSE	4.2	< 5.5	< 6.2	< 6.5	< 7.3

accurate results. Applying DROA ARs or PCA followed by aggregation, the erosion rate prediction error is 10% less than using aggregation first then DROA ARs or PCA, since the erosion rate is calculated from the correlated results generated from DROA ARs or PCA rather than working on all the factors.

5.2 Power Loss Rate vs. Erosion Rate

For the wind analytics evaluation, WTEP have traced the resultant power loss rate over several erosion rate values per three different areas (Western Desert, Red Sea and North Coast) for the largest dataset D3. The higher erosion rate, the more power loss rate as shown in Fig. 4, where each line style represents the erosion rate values interval at a certain area. Erosion rates exceeding 45% represent a major threat to the power production process, as it leads to 30% and more power loss. Figure 3 proves that the erosion rate is high at the Western Desert that reaches 48% and North Coast with 33% due to the increase of sand and fallen rain respectively, whereas a normal erosion rate at Red Sea reaches 17%.

**Fig. 4.** Power loss rate vs. erosion rate

5.3 Processing Time vs. Reduction Methods

Evaluating the big data processing efficiency, WTEP processing time is tested at the different reduction methods for the three datasets as presented in Fig. 5. The processing time increases by enlarging the dataset size. Table 4 shows the results over D3, where

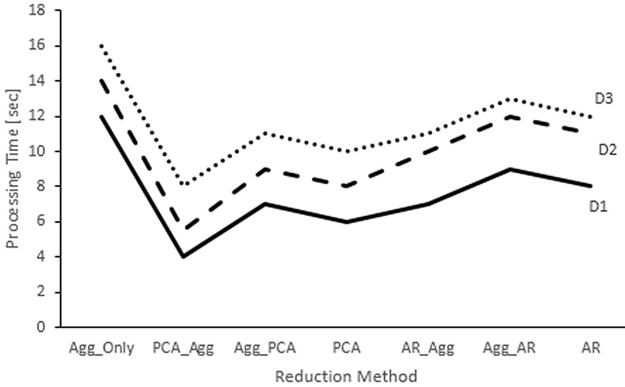


Fig. 5. Processing time vs. reduction methods

Table 4. Processing time over reduction methods

	Agg. Only	ARs + Agg	Agg. + ARs	PCA + Agg.	Agg. + PCA
D3	16	< 11.2	< 12.7	> 8.2	< 11

the largest data can judge the processing time efficiency. The aggregation only consumes the highest processing time due to working on all factors to predict the erosion rate, in contrast to DROA ARs since it works on a less number of factors. The lowest processing time is consumed by PCA. Moreover, using the aggregation first then DROA ARs decreases the number of factors, which reduces the processing time by 20% rather than that of aggregation only that uses all factors in processing. Applying PCA then aggregation, the processing time is 25% less than that of aggregation followed by PCA and 40% less than that of aggregation only. Decreasing K-value by 2, the processing time is reduced by average 2 s. On the other hand, increasing the confidence and support values by 0.2 reduces the processing time by average 3 s, because of reducing the number of factors used for processing.

6 Conclusion

Many researches have considered wind farms reliability evaluation to manage their operation and maintenance costs from the engineering viewpoint. In this paper, we introduce Wind Turbine Erosion Predictor (WTEP) system for predicting the erosion rate of wind turbines from the data analytics perspective to minimize turbines failure rate. WTEP proposes a novel Optimized Flexible Multiple Regression (OFMR) approach for erosion rate prediction that fits big data processing. In addition, it applies a new approach for big data volume handler using Double-Reduction Optimum Apriori (DROA). The Variety-Veracity Handler ensures data quality used for turbines erosion

analysis and power loss prediction. Experiments were performed to evaluate big data processing efficiency and wind analytics at several areas in Egypt, where OFMR reaches >90% in efficient processing time. DROA ARs generates reasonable accurate results in less processing time. The experiments held on the Egyptian locations datasets confirm that the lowest erosion rate is at Red Sea. Our future work is to consider the economic models of wind farm profitability using big data analytics.

References

1. Ackermann, T.: *Wind Power in Power Systems*. Wiley, New York (2005)
2. Walford, C.: *Wind turbine reliability: understanding and minimizing wind turbine operation and maintenance costs*, United States, Department of Energy (2006)
3. Rempel, L.: Rotor blade leading edge erosion-real life experiences. *Wind Syst. Mag.* **11**, 22–24 (2012)
4. Tobin, E., Young, T., Raps, D., Rohr, O.: Comparison of liquid impingement results from whirling arm and water-jet erosion test facilities. *Wear* **271**, 2625–2631 (2011)
5. Poudel, L., Thapa, B., Shrestha, B., Shrestha, N.: Impact of sand on hydraulic turbine material: a case study of Roshi Khola, Nepal. *Hydro Nepal: J. Water Ener. Environ.* **10**, 60–65 (2012)
6. Fiore, G., Selig, M.: Simulation of damage for wind turbine blades due to airborne particles. *Wind Eng.* **39**, 399–418 (2015)
7. Tabakoff, W.: High-temperature erosion resistance coatings for use in turbomachinery. *Wear* **186–187**, 224–229 (1995)
8. Bakirtzis, A.: A probabilistic method for the evaluation of the reliability of stand alone wind energy systems. *IEEE Trans. Ener. Convers.* **7**(1), 99–107 (1992)
9. Fiore, G., Selig, M.: Simulation of damage progression on wind turbine blades subject to particle erosion. In: *54th AIAA Aerospace Sciences Meeting* (2016)
10. Fawzy, D., Moussa, S., Badr, N.: The evolution of data mining techniques to big data analytics: an extensive study with application to renewable energy data analytics. *Asian J. Appl. Sci.* **4**(3), 756–766 (2016). ISSN: 2321–0893
11. Jadhav, D.: Big data: the new challenges in data mining. *Int. J. Innov. Res. Comput. Sci. Technol.* **1**(2), 39–42 (2013)
12. Richardson, D.: Electric vehicles and the electric grid: a review of modeling approaches, Impacts, and renewable energy integration. *Renew. Sustain. Ener. Rev.* **19**, 247–254 (2013)
13. Kim, H., Kim, C., Yue, L.: A study of the growth of single-phase Mg0.5Zn0.5O Films for UV LED. *Int. J. Chem. Nucl. Mater. Metall. Eng.* **8**(7), 674–678 (2014)
14. Hill, R., Jennifer, R., Stinebaugh, A., Briand, D., Benjamin, A.S., Lindsay, J.: *Wind turbine reliability: a database and analysis approach*. Sandia National Laboratories, Albuquerque, New Mexico, 87185 (2008)
15. Verma, A.: *Performance monitoring of wind turbines: a data-mining approach* (2012)
16. Kim, K., Parthasarathy, G., Uluyol, O., Foslien, W., Sheng, S., Fleming, P.: Use of SCADA data for failure detection in wind turbines. In: *ASME 2011, 5th International Conference on Energy Sustainability*, pp. 2071–2079 (2011)
17. Kaidis, C.: *Wind Turbine Reliability Prediction: A Scada Data Processing & Reliability Estimation Tool* (2014)
18. Niu, K., Zhao, F., Zhang, S.: A fast classification algorithm for big data based on KNN. *J. Appl. Sci.* **13**, 2208–2212 (2013)

19. Murdopo, A.: Distributed decision tree learning for mining big data streams. Master of Science Thesis, European Master in Distributed Computing. Barcelona, Spain (2013)
20. De Francisci, M.G.: SAMOA: a platform for mining big data streams. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 777–778 (2013)
21. Cai, X., Nie, F., Huang, H.: Multi-view k-means clustering on big data. In: Twenty-Third International Joint Conference on Artificial Intelligence (2013)
22. Vladislavleva, E., Friedrich, T., Neumann, F., Wagner, M.: Predicting the energy output of wind farms based on weather data: important variables and their correlation. *Renew. Ener.* **50**, 236–243 (2013)
23. Watanabe, T., Fujioka, R.: Fuzzy association rules mining algorithm based on equivalence redundancy of items. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1960–1965. IEEE (2012)
24. Alexiadis, M., Dokopoulos, P., Sahsamanoglou, H.: Wind speed and power forecasting based on spatial correlation models. *IEEE Trans. Ener. Convers.* **14**, 836–842 (1999)
25. Fawzy, D., Moussa, S., Badr, N.: Trio-V wind analyzer: a generic integral system for wind farm suitability design and power prediction using big data analytics. *J. Ener. Res. Technol.* JERT 2017, ASME. <https://doi.org/10.1115/1.4038119>
26. Holland, S.: Principal components analysis (PCA). University of Georgia (2008)
27. Giannotti, F., Lakshmanan, L., Monreale, A., Pedreschi, D., Wang, H.: Privacy-preserving mining of association rules from outsourced transaction databases. *IEEE Syst. J.* **7**(3), 385–395 (2013)
28. Lam, J., Wan, K., Liu, D., Tsang, C.: Multiple regression models for energy use in air-conditioned office buildings in different climates. *Ener. Convers. Manag.* **51**, 2692–2697 (2010)
29. Aranda, A., Ferreira, G., Mainar-Toledo, M., Scarpellini, S., Sastresa, E.: Multiple regression models to predict the annual energy consumption in the Spanish banking sector. *Ener. Build.* **49**, 380–387 (2012)
30. Eminoglu, U., Ayasun, S.: Modeling and design optimization of variable-speed wind turbine systems. *Energies* **7**, 402–419 (2014)
31. Freire, R., Oliveira, G., Mendes, N.: Development of regression equations for predicting energy & hygrothermal performance of buildings. *Ener. Build.* **40**, 810–820 (2008)
32. Eladawy, A., Nadaoka, K., Negm, A., Saavedra, O., Hanafy, M.: Assessment of long term thermal stress on egyptian coral reefs based on remotely sensed sea surface temperature data. *Int. J. Environ. Sci. Dev.* **6**(12), 938–946 (2015)
33. Mortensen, N.G.: Wind atlas for Egypt: measurements, micro-and mesoscale modelling. In: European Wind Energy Conference and Exhibition (2006)
34. El-Barbary, Z., Sallam, D.: Optimizing use of rainfall water in east desert of Egypt. In: Proceedings of 8th International Water Technology Conference, IWTC8, Egypt (2004)