Insights into the Efficiencies of On-Shore Wind Turbines: A Data-Centric Analysis

Gürdal Ertek, Murat Mustafa Tunç, Ece Kurtaraner, Doğancan Kebude

Faculty of Engineering and Natural Sciences

Sabanci University

Istanbul, Turkey

Abstract—**Literature on renewable energy alternative of wind turbines does not include a multidimensional benchmarking studythat can help investment decisions as well as design processes. This paper presents a data-centric analysis of commercial on-shore wind turbines and provides actionable insights through analytical benchmarking through Data Envelopment Analysis (DEA), visual data analysis, and statistical hypothesis testing. The paper also introduces a novel visualization approach for the understanding and the interpretation of reference sets, the set of efficient wind turbines that should be taken as benchmark by inefficient ones.**

Keywords- ***wind turbines; wind energy; data envelopment analysis (DEA); benchmarking; renewable energy***

I.   INTRODUCTION

Wind power refers to the conversion of wind energy into a useful form, such as electricity, using wind turbines. This powertypehas been used for at least 5,500 years; yet the first modern wind turbine is developed at the beginning of 1980's. Wind energy is plentiful, renewable, widely distributed, clean and it reduces greenhouse gas emissions when it displaces fossil fuel derived electricity.

Wind turbines are mechatronic devices that convert wind energy into electrical energy via mechanical energy (Figure 1). Due to the fact that the potential commercially viable wind energy is estimated to be 72 terawatts [1], these environmentally friendly devices started to gain increasing importance by the end of the 1980's. The European Wind Energy Association reports that wind power could generate up to 16% of Europe's electricity by 2020 [2].

The importance of wind power is widely understood and accepted in many countries around the world. At the end of 2010, the worldwide installed wind power capacity has increased to 196,630 megawatts (MW) [3]. In 2011, this number has become 238,351 with a 20.6% increase rate [4]. Top three countries in terms of wind power can be listed as China (62,733 MW), United States (46,919 MW) and Germany (29,060 MW) [4]. Top five wind turbine manufacturers in the world, ranked by annual market share, are the following: Vestas (12.7%), Sinovel (9%), Goldwind (8.7%), Gamesa (8%) and Enercon (7.8%) [5].
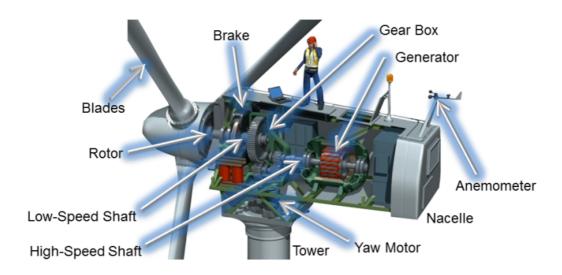

Figure 1.   Components of a wind turbine.

Figure 1 demonstrates the basic components of a wind turbine. A successful wind turbine design depends on five critical issues: load calculations, power control, aerodynamics, wind climate and systems that control the turbine. A successful design includes some certain aspects: Blades are designed with aerodynamic calculations, rotor is a rotational device that turns the shaft and starts the energy producing cycle, the generator is where the energy is finally produced. There are both low-speed shafts and high-speed shafts in a successful wind turbine design, gears in the gear box make the shafts turn at the correct speed and brakes work for emergency stopping or slowing down of the turbines. Anemometer is a device that calculates wind speed. Some turbines include a

controlling device that catches the wind speed from the anemometer and adjust the shaft speed. Also, the yaw control system is another element that increases wind turbine efficiency. This element helps minimize non-symmetrical loads and increasing power output [6].

The goal of this paper is to develop a better understanding of wind turbines through data-centric analytical methods. The methods used are Data Envelopment Analysis (DEA) for multi-dimensional benchmarking; visual data analysis for insight discovery and statistical hypothesis testing for drawing conclusions.

The remainder of the paper is organized as follows: Section 2 provides a brief review of some relevant literature as the background. Section 3 discusses the methodologies used in the study. Section 4 is devoted to the results and analysis of the study, where new insights are obtained with regards to wind turbines benchmarking. Finally, Section 5 presents some conclusive remarks.

II. LITERATURE

In the previous researches, multi-criteria analysis in renewable energy is found to be a popular research topic and an effective tool for policy makers. [7]reviewed over 90 papers of multi-criteria decision making (MCDM) analysis in the renewable energy field, and suggested that multi-criteria analysis can help to justify the choices in the renewable energy sector by providing technical and scientific decision-making support tools. [8]claimed that traditional single criteria analysis can no longer deal with the complexity of current renewable energy systems. [9]provided a selection of criteria and options for the renewable energy technologies assessment based on the analysis and synthesis of parameters. His study also presented a priority rating among the selected options of energy systems. [10]used a multi-criteria method to make an assessment regarding the feasibility of installing mini wind energy turbines.

There exist several algorithms in the literature for selecting the best location for wind turbine installation. [11]developed a genetic algorithm that is capable to handle multi-objective optimization problems, and claimed that the proposed strategy finds out the optimal location for wind farms. [12]also created a decision support system using Geographic Information System (GIS) tool for site selection of wind turbines. [12]claimed that decision makers should consider not only economic feasibility, but also the environmental fitness. [13]also implemented a GIS based multi-criteria site selection to determine suitable locations for wind turbines. [13]analyzed a specific area, namely Eskisehir, Turkey, and studied renewable energy efficiency prospects. [14]presented an integrated approach with traditional Data Envelopment Analysis (DEA) that enables the energy policy makers to select the best possible location in terms of lowest possible cost. [14]used DEA method to rank various locations capabilities with specific input and output indicators. [15]applied a multi-criteria method to select a site to install wind energy turbines in the Italian island of Salina. In [15], not only the energy profile, but also the local environmental conditions were analyzed to evaluate feasible solutions.

In addition to the multi-criteria analysis, there exist other algorithms that can be applied to evaluate the performance of renewable energy systems. [16]evaluated both the energy and environmental performances of a wind farm by implementing a life cycle assessment (LCA). [16]points out that the main environmental impacts of a wind farm caused during the manufacturing. [17]used an integrated DEA - PCA (Principle Component Analysis) approach for performance assessment and optimization of power distribution units in Iran. [18]analyzed the performance of Portuguese wind farms considering ownership and unobserved managerial ability factors. [19]benchmarked the wind farms in the same country by measuring the efficiencies using DEA and Stochastic Frontier Analysis (SFA) methods. [20]compared small-scale and large-scale approaches to find out which one can achieve the most socially, economically and environmentally (SEE) efficient way. [20]indicated that small-scale schemes are the most efficient in terms of SEE.

To the best of our knowledge, ours is the first study in the literature that performs a benchmark of commercial wind turbines. Again, to the best of our knowledge, the visualization of the reference sets in the DEA results is visualized as a directed graph for the first time in literature within this paper. These two applications are the unique contributions of this paper with regards to earlier literature.

III. METHODOLOGY

In this section, the methodologies used, namely, data envelopment analysis (DEA), data visualization, and hypothesis testing,are described.

A. Data Envelopment Analysis (DEA)

Multi-dimensional benchmarking is a fundamental problem in a diverse range of fields, and a multitude of qualitative and quantitative methods have been proposed in the literature for this problem. Among the quantitative methods, Data Envelopment Analysis (DEA) is a very popular choice by both researchers and practitioners [21][22]. DEA uses an optimization based algorithm to determine three types of benchmark results for each of the entities within a group [23]. For each entity in the group, also named as a Decision Making Unit (DMU), the efficiency score value between 0 and 1 represents the relative performance of that entity (DMU) compared to the other entities in the group. The efficiency of a DMU increases with the generation of higher values of its outputs given lower values of its inputs. A DMU is said to be efficient if its efficiency score is equal to 1, and inefficient if it is less 1. The reference set of a DMU is the set of efficient DMUs that the DMU should take as benchmark. The reference set for an efficient DMU consists of itself, whereas the reference set of an inefficient DMU consists of two or more efficient DMUs. The final benchmark result provided by DEA is the set of projections, which tells how much of each input the DMU should decrease and/or each output the DMU should increase so that it becomes efficient.

B. Visual Data Analysis

Data analysis is an indispensable part of all applied research and problem solving in industry. The most fundamental data analysis approaches are data visualization (histograms, scatter plots, surface plots, tree maps, parallel coordinate plots, etc.) [24][25][26], statistics (hypothesis test, regression, PCA, etc.) [27], data mining (association mining, etc.) [28][29], and machine learning methods (clustering, classification, decision trees, etc.) [30]. Among all these approaches data visualization, or in other words, visual data analysis, is the one that relies most on the cognitive skills of human analysts, and allows the discovery of unstructured actionable insights that are limited only by human imagination and creativity. The analyst does not have to learn any sophisticated methods to be able to interpret the visualizations of the data. Data visualization is also a hypothesis generation scheme, which can be, and is typically followed by more analytical or formal analysis, such as statistical hypothesis testing.

Using data visualization to analyze the outputs of DEA was performed in [31][32][33]. Meanwhile, visual benchmarking can be performed using other methods, such as Kohonen maps [34].

C. Hypothesis Testing

Statistical hypothesis testing is the most common form of supporting or refuting hypotheses that are proposed in empirical research. Hypothesis tests are of various forms. In this paper, we will be using two main types of hypothesis tests, goodness of fit test and sample mean comparison test [35]. The Shapiro-Wilk goodness of fit test suggests whether a data sample follows normal distribution or not [36]. This test is crucial for the selection of the appropriate comparison of means tests. The parametric t-test or the nonparametric Mann-Whitney test [37] is used for testing whether two data samples have same mean values or not. The parametric ANOVA or the nonparametric Kruskal-Wallis test is used for testing whether the mean of any sample among a group of samples is different than the others. If the normality of any of the samples involved (in the comparison of means tests) is rejected with a high confidence level, then nonparametric methods are used. Parametric tests are used only if all the samples follow the normal distribution [35].

IV. ANALYSIS AND RESULTS

In this section, we describe how the analysis was performed and present the results obtained.

Our study progressed through the following steps:

1) Data was collected for on-shore wind turbines of leading manufacturers in the world.

2) Two DEA models were constructed and results were obtained.

3) Visual data analysis was performed using scatter plots [38], surface plots, and graph visualization [39][40].

4) Formal statistical hypothesis testing was carried out to test the hypothesis generated through the scatter plots.

Actionable insights, which can help in the better design and improvement of wind turbines, have been identified and documented throughout the process. These insights and their implications are presented in this paper.

A. The Dataset

The data in our study consists of the commercial on-shore wind turbines of the top ten wind turbine manufacturers in the world[5]. On-shore wind turbines are those that are constructed on land, as opposed to the off-shore turbines that are constructed on the sea. While the data for both on-shore and off-shore turbines were collected, the analysis in the study focused only on the on-shore ones, due to the association of the first author with an on-shore wind turbine project. A total of 74 on-shore turbines were listed on the Internet for the following companies: Vestas, GE, Gamesa, Suzlon, Siemens, Goldwind, Sinovel, Ming Yang Power, Mitsubishi Heavy Industries, Enercon.

Missing data was a major issue in the data collection process; none of the companies, except Vestas, had put out their complete technical data openly on the Internet. However, complete data was available for 74 wind turbines for three critical attributes: Diameter, Nominal Wind Speed and Nominal Output. Diameter refers to the diameter of the circle formed by the wind blades. Nominal Wind Speed refers to nominal speed at which the wind turbine operates, and Nominal Output refers to the power (in kW or MW where 1 MW = 1000 kW) generated by the wind turbine. These three attributes are critical because they can be sufficient to build a basic DEA benchmarking model. While price is a very important benchmarking attribute, it was not included, since it was not readily available and since it would be very difficult to obtain it. Once the data collection was completed, the data was cleaned based on the taxonomy given in [41].

B. DEA Results

Two DEA models were constructed, and the complete data for the models and their results are made available online in the supplement file package [42]. Both models take each wind turbine as a DMU, and consider Diameter and Nominal Wind Speed as the inputs and Nominal Output as the output. The first model, hereafter referred to as Model A, includes all the 74 on-shore turbines. The second model, hereafter referred to as Model B, includes only the 32 turbines that can operate under low wind (regardless of the whether they operate in medium or high wind). The BCC-O (BCC Output Oriented) model was selected, since a major design goal for wind turbines is output oriented, trying to maximize the power generated under given physical and natural constrains. The constructed DEA models were analyzed and their results were generated using the SmartDEA Solver software [43].

Model A resulted in 4 efficient DMUs among the 74 included. Model B resulted in 5 efficient DMUs among the 32 included.

C. Visual Analysis: Scatter Plots

The first step of visual data analysis involved scatter plots using Orange data mining software [44]. Figures 2-8 visualize the DEA results for Model A (all on-shore turbines).

Figure 2 illustrates the efficiency scores for the wind turbines of different manufacturers. It can be visually seen that the turbines of Enercon and GE tend to have higher efficiency scores. Thus one can hypothesize that there exist significant differences in the mean efficiency scores for turbines of different manufacturers. This hypothesis, together with others, will be formally investigated through statistical hypothesis testing.

Figure 3 plots efficiency score vs. nominal output. Two visual insights can be observed in this figure: (a) One can find efficient turbines, regardless of nominal output values; (b) While efficiency seems to correlate with nominal output, there is a cluster of turbines with 1500 kW nominal output, which have higher efficiencies than the regular linear pattern.

Figure 4 plots efficiency score vs. cut-in wind speed. Cut-in wind speed is the minimum wind speed at which aturbine can start to operate. While there are very few observations for cut-in speeds of 2 and 2.5, there are many observations for cut-in speeds of 3, 3.5, and 4. Thus one can hypothesize difference in means of the efficiency scores for different cut-in wind speeds.
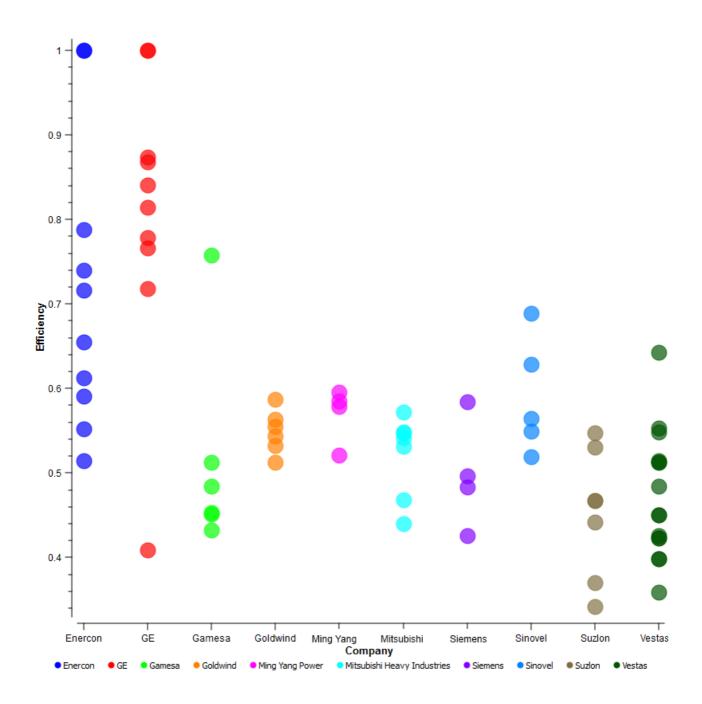
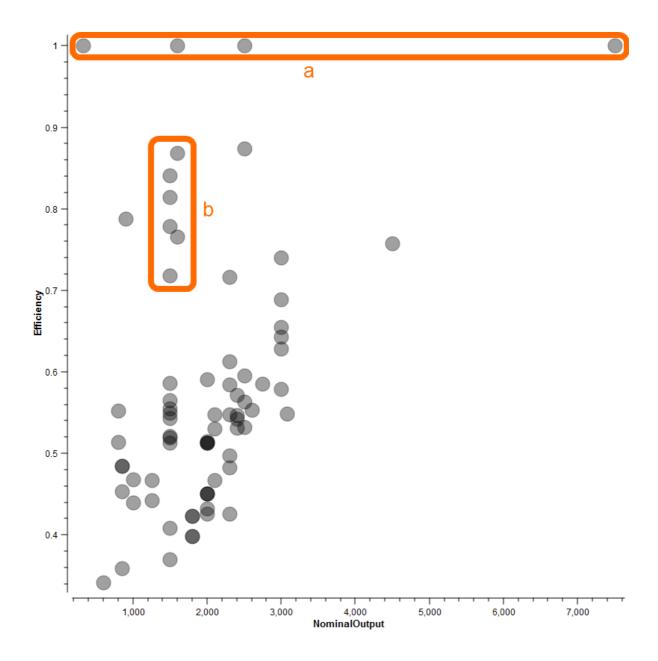Figure 2.   Efficiencies of the wind turbines of different manufacturers.

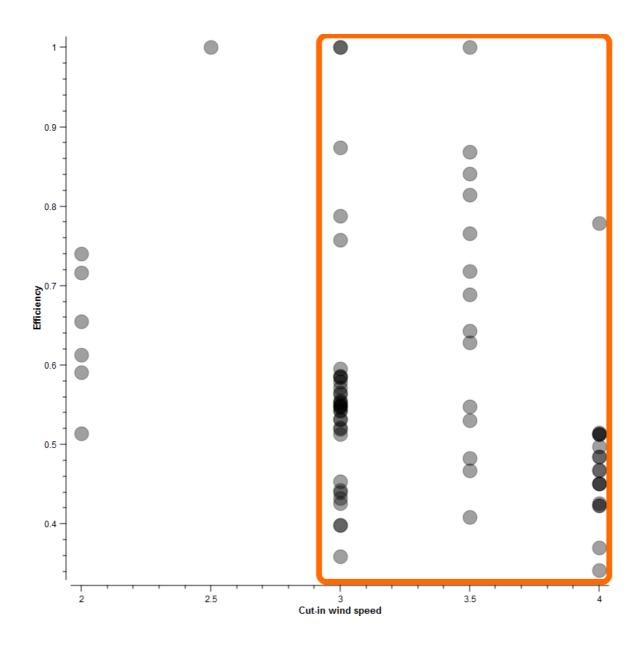Figure 3.   Efficiency vs. Nominal Output.

Figure 4.   Efficiency vs. Cut-in Wind Speed
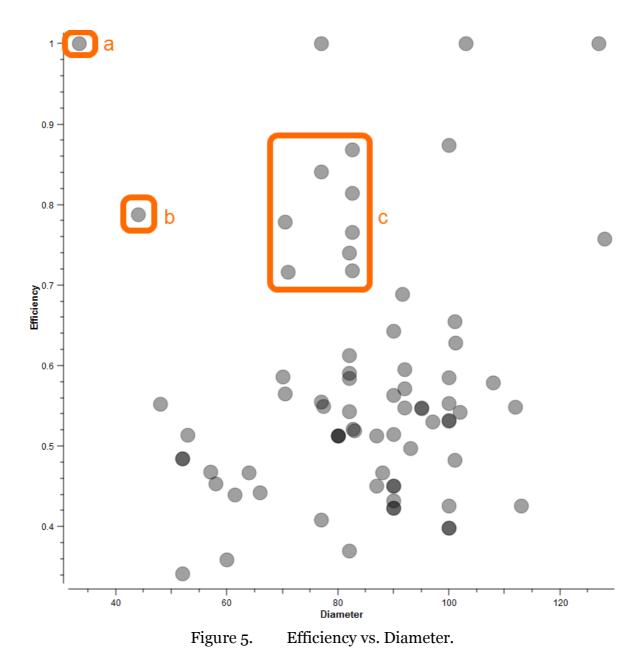
Figure 5.   Efficiency vs. Diameter.

Figure 5 plots efficiency score vs. diameter. There are three outlier patterns in this figure: (a) The turbine with the minimum diameter is efficient; (b) The turbine with the second minimum diameter has a high efficiency score; (c) While efficiency seems to correlate with diameter, there is a cluster of turbines with diameter 70-85 meters, which have higher efficiencies than the regular linear pattern.
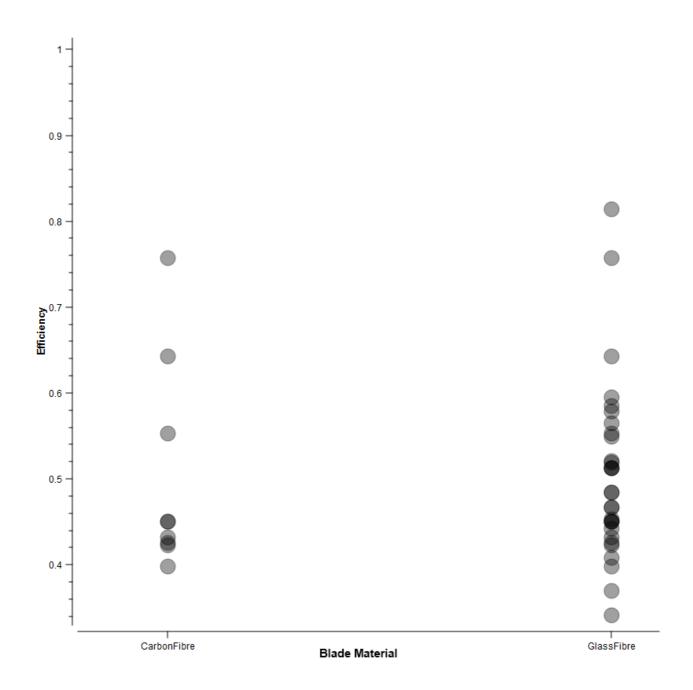
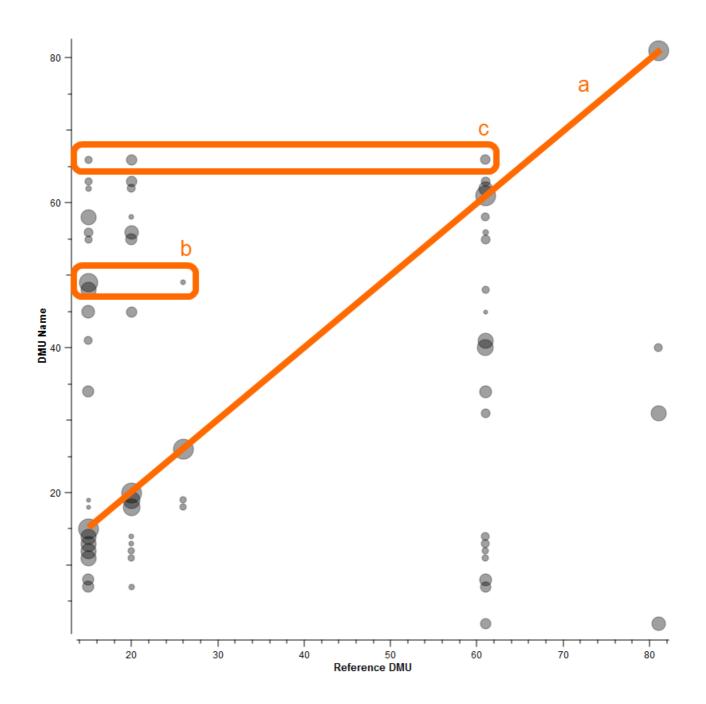Figure 6. Efficiency vs. Blade Material.

Figure 7.   Visualization of reference sets.

Figure 6 plots efficiency score vs. blade material. Wind turbines with blades that include glass fiber material seem to have higher efficiency scores, calling for a formal hypothesis test.

Figure 7 visualizes the reference sets of each of the DMUs. The y-axis shows each DMU, the x-axis shows the DMUs in the reference sets, and the sizes (areas) of the points reflect the weights for that reference. There are three patterns that can be observed in here: (a) The DMUs (turbines) on this line are the efficient ones, with efficiency scores of 1; (b) Some DMUs, such as DMU_49, should primarily take as example only one of the efficient DMUs; (c) Some other DMUs, such as DMU_66, have a balanced weight distribution for the efficient DMUs in their reference sets.

Figure 8 visualizes the projections for each of the DMUs. The y-axis shows the efficiency of the DMUs. Each DMU is represented by three points, which refer to the projections for the total of

three input plus output attributes. The x-axis shows the percentage change that is needed for each of the attributes, so that the DMU becomes efficient. There are two basic insights from this figure: (a) Some of the DMUs require both a decrease in the inputs, and an increase in the outputs, and this can be detected very quickly through visualization; (b) Most of the DMUs in this particular case study need only to increase their outputs.
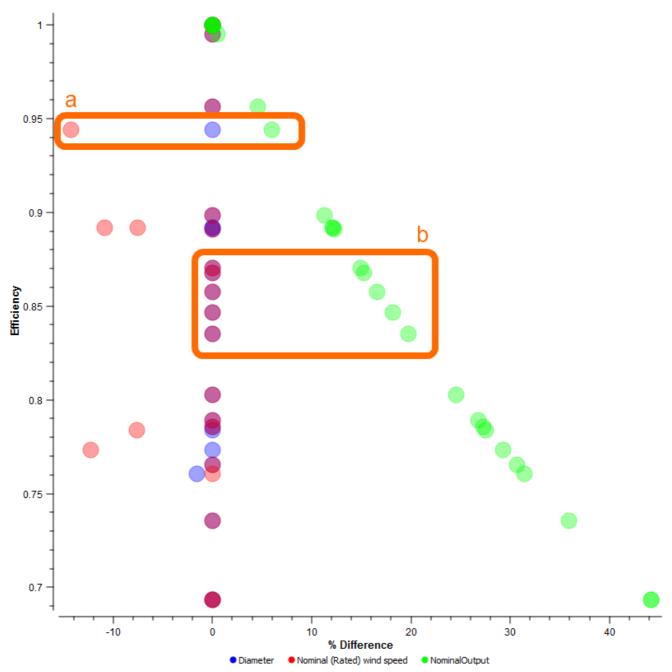

Figure 8.   Visualization of the projections.

A. Visual Analysis: Surface Plots

The next type of visual analysis is surface plots, where a surface is constructed to reflect the z values, based on the x and y values. In our two surface plots, we have also mapped efficiency scores to the color of the surface, which enabled new insights. Figures 9 and 10 visualize the DEA results for Model A (all on-shore turbines) using Miner3D software [45].

Figure 9 presents the surface plot where one can observe the interaction between diameter, nominal wind speed, and nominal output, and how this interaction affects efficiency scores. The highlighted region and the design parameter intervals are observed to result in low efficiency scores, and thus should be avoided.

Figure 10 explores the impact of diameter, cut-in wind speed, and nominal output on the efficiency scores. Regions (a) and (c) should be avoided as much as possible. While the efficiency scores are better in region (b), there are other regions where the efficiency scores are much better. So the most fundamental insight from this surface plot is that higher values of cut-in wind speed bring inefficiencies and one should try to design or invest in wind turbines with cut-in wind speeds of 3 or 3.5.

B. *Visual Analysis: Graph Visualization*

The final visualization is novel in the field of DEA, and suggests the representation of the reference sets as a graph. In this newly proposed scheme, each DMU is represented as a node in a graph, and each directed arc represents a reference relationship from an inefficient DMU node to an efficient DMU node.
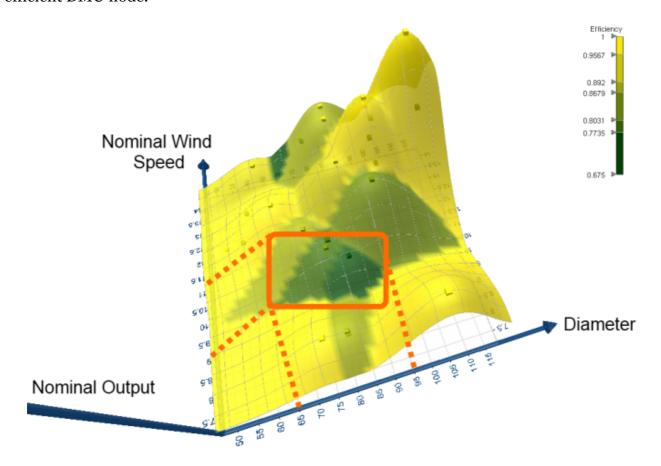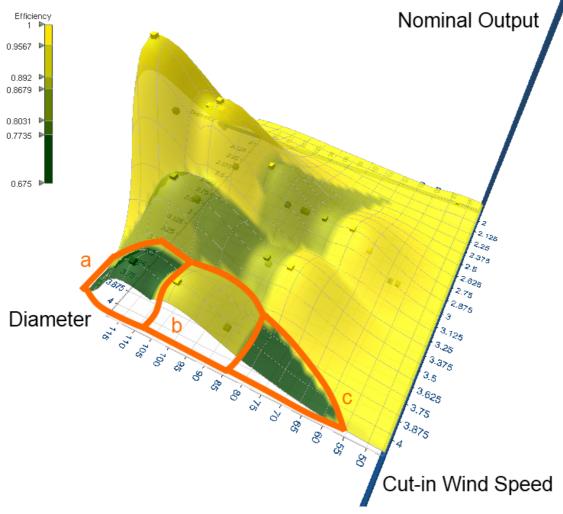

Figure 9.   Surface Plot.

Figure 10.   Surface Plot.

Figure 11 visualizes the reference sets for Model B (on-shore turbines that can operate under low-wind), using yEd software [46]. One can immediately observe a central cluster of DMUs, where DMU_15, _18, and _49 are "best" designs (the reader is referred to [23] for the appropriate interpretation of efficiency scores). Also, one can observe two small independent outlying clusters where DMU_24 or _68 are the best designs.

C. Hypothesis Testing

While the hypothesis tests accept or reject the proposed hypothesis, these results cannot be interpreted as the ultimate proof or disproof of causal relationships. Rather, they only support or oppose causal relationships. Statistical support for differences between group means only suggest that the differences are due to the groups [47].

Statistical tests allow us to formalize the insights that were obtained through the visualizations. Our tests confirmed some of the visual insights: There are statistically significant differences in mean efficiency scores with respect to manufacturers (Figure 2) and cut-in wind speed (Figure 4), as well as cut-out speed (not visualized in this paper). The difference with respect to blade material was not statistically significant. The detailed analysis and significance values are available in the supplement file package [42].
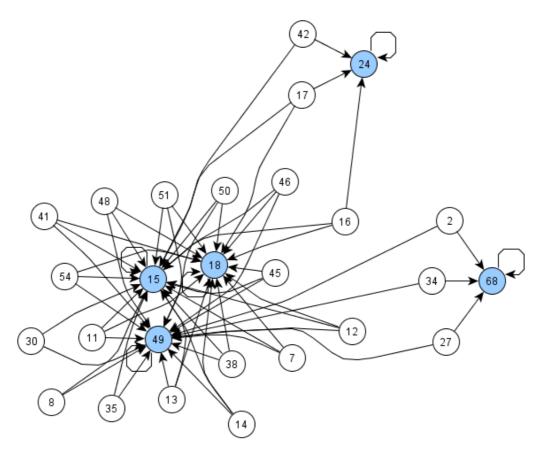
Figure 11.   Graph visualization of the reference sets.

V.   CONCLUSIONS

This paper presented a data-centric benchmarking and analysis of commercial wind turbines in the global energy market for the first time, using data envelopment analysis and data visualization. A new graph-based visualization scheme was proposed for displaying reference sets in DEA, and its usefulness was demonstrated, in addition to the application of traditional visual methods and statistical tests. The discovered insights can help managers in their investment decision making and design engineers in improving their turbine designs.

VI.   REFERENCES

[1]   CEOE."Mapping the global wind power resource," Accessed on March 22, 2012. Available under http://bit.ly/GGvqmJ

[2]   Energy Portal,"Wind energy could be 16 percent of Europe's electricity by 2020," Accessed on March 22, 2012. Available under http://bit.ly/GFxqtC

[3]   World Wind Energy Association,"World Wind Energy Report 2010," Accessed on March 22, 2012. Available under http://bit.ly/fV4RQS

[4]   Global Wind Energy Council, "Wind Energy Report 2011,"Accessed on March 22, 2012. Available under http://bit.ly/x8GWgF

[5]   IHS Emerging Energy Research, "Global Wind Turbine Supply Market Share Evolution,"Accessed on March 22, 2012. Available underhttp://bit.ly/GJW7px

[6]   Windpower. Wind Turbine Design.Accessed on March 22, 2012. Available under http://bit.ly/GFXJ1Y

[7]   R. Abu-Taha, "Multi-criteria applications in renewable energy analysis: A literature review," In Proceedings of Technology Management in the Energy Smart World (PICMET) '11, IEEE, pp. 1-8, 2011.

[8]   J.R. San Cristóbal, "Multi-criteria decision-making in the selection of a renewable energy project in spain: The Vikor method," Renewable Energy, 36, pp. 498-502, 2011.

[9]   N.H. Afgan and M.G. Carvalho, "Multi-criteria assessment of new and renewable energy power plants," Energy, 27 (8), pp. 739-755, 2002.

[10] G. Aghilone, F. De Felice and A. Petrillo, "A Comperative Analysis Based on Analytic Network Process for Selection of a Mini Wind Station Plant," 2011.

[11] M. Alonso, H. Amaris and C. Alvarez-Ortega, "A multiobjective approach for reactive power planning in networks with wind power generation," Renewable Energy, 37, pp. 180-191, 2011.

[12] N.Y. Aydin, E. Kentel and S. Duzgun, "GIS-based environmental assessment of wind energy systems for spatial planning: A case study from Western Turkey," Renewable and Sustainable Energy Reviews, 14(1), pp. 364-373, 2010.

[13] A.E. Cengiz, Y. Guney and A. Cabuk, (2006). "Determining Renewable Energy Efficiency in Eskisehir, Turkey: A GIS Based Solution," Energy Sources, 2010, pp. 2020-2030.

[14] A. Azadeh, S.F. Ghaderi and M.R. Nasrollahi, "Location optimization of wind plants in Iran by an integrated hierarchical Data Envelopment Analysis," Renewable Energy, 36, pp. 1621-1631, 2010.

[15] F. Cavallaro and L. Ciraolo, "A multicriteria approach to evaluate wind energy plants on an Italian island," Energy Policy, 33 (2), pp. 235-244, 2005.

[16] F. Ardente, M. Beccali, M. Cellura, and V. Lo Brano, "Energy performances and life cycle assessment of an Italian wind farm," Renewable and Sustainable Energy Reviews, 12(1),pp. 200-217, 2008.

[17] A. Azadeh, S.F. Ghaderi and H. Omrani, "A deterministic approach for performance assessment and optimization of power distribution units in Iran," Energy Policy, 37, pp. 274-280, 2008.

[18] C.P. Barros and O.S. Antunes, "Performance assessment of Portuguese wind farms: Ownership and managerial efficiency," Energy Policy, 39, pp. 3055-3063, 2011.

[19] G. Iglesias, P. Castellanos and A. Seijas, "Measurement of productive efficiency with frontier methods: A case study for wind farms," Energy Economics, 32 (5), pp. 1199-1208, 2010.

[20] J. Burton and K. Hubacek, "Is small beautiful? A multicriteria assessment of small-scale energy technology applications in local governments," Energy Policy, 35 (12), pp. 6402-6412, 2007.

[21] S. Gattoufi, M. Oral, A. Kumar and A. Reisman, "Content analysis of data envelopment analysis literature and its comparison with that of other OR/MS fields," Journal of the Operational Research Society, 2004a, 55, pp. 911-935.

[22] A. Emrouznejad, B. Parker and G. Tavares, "Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA," Journal of Socio-Economic Planning Sciences, 42(3), pp. 151–157, 2008.

[23] W. W. Cooper, L. M. Seiford and K. Tone, "Introduction to data envelopment analysis and its uses: With DEA solver software and references," Springer, 2006.

[24] R. Spence, "Information visualization," ACM Press, 2001.

[25] D. A. Keim, "Information visualization and data mining," IEEE Transactions on Visualization, 2002.

[26] R. Lengler and M. Eppler, "Towards A Periodic Table of Visualization Methods for Management," In IASTED Proceedings of the Conference on Graphics and Visualization in Engineering (GVE), Florida, USA, 2007.

[27] D. Wackerly, M. Mendenhall, and R. L. Scheaffer, "Mathematical Statistics with Applications," Duxbury Press, 2007.

[28] J. Han and M. Kamber, "Data mining: concepts and techniques," Morgan Kaufmann Publishers, 2006.

[29] O. Z. Maimon and L. Rokach. "Data mining and knowledge discovery handbook," Springer-Verlag New York Inc, 2005.

[30] E. Alpaydin, "Introduction to Machine Learning," The MIT Press, 2009.

[31] F. Ulus, O. Kose, G. Ertek and S. Sen, "Financial benchmarking of transportation companies in the New York Stock Exchange (NYSE) through data envelopment analysis (DEA) and visualization," In 4th International Logistics and Supply Chain Congress, Izmir, Turkey, 2006.

[32] G. Ertek, M. A. Can, and F. Ulus, "Benchmarking the Turkish apparel retail industry through data envelopment analysis (DEA) and data visualization," In EUROMA 2007, Ankara, Turkey, 2007.

[33] S. El-Mahgary, and R. Lahdelma, "Data envelopment analysis: Visualizing the results," European Journal of Operational Research, 83(3), pp. 700–710. and Computer Graphics, 8(1), pp. 1–8, 1995.

[34] G. Deboeck and T. Kohonen "Visual Explorations in Finance: with Self-Organizing Maps,"Springer, 1998.

[35] W. J. Conover, "Practical Nonparametric Statistics," John Wiley and Sons, 1998.

[36] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, 1965, 52 (3/4), pp. 591-611.

[37] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other", The Annals of Mathematical Statistics, 1947, 18 (1), pp. 50-60.

[38] J. Chambers, W. S. Cleveland, B. Kleiner, P. A. Tukey, "Graphical Methods for Data Analysis," Springer , 1983.

[39] I. Herman, G. Melançon and M.S. Marshall. "Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics," 6(1), pp. 24–43, 2000.

[40] I. Tollis, P. Eades, G. Di Battista and L. Tollis. "Graph drawing: algorithms for the visualization of graphs," Prentice Hall, 1998.

[41] W. Kim, B.J. Choi, E.K. Hong, S.K. Kim and D. Lee, "A taxonomy of dirty data. Data Mining and Knowledge Discovery," 7(1). pp.81–99, 2003.

[42] G, Ertek, M.M. Tunç, E. Kurtaraner, D. Kebude, "Supplement document for 'Insights into the Efficiencies of On-Shore Wind Turbines: A Data-Centric Analysis'," 2012. Available under http://people.sabanciuniv.edu/ertekg/papers/supp/08.zip

[43] A. E. Akcay, G. Ertek and G. Buyukozkan, "Analyzing the solutions of DEA through information visualization and data mining techniques: SmartDEA framework," Expert Systems with Applications, 39, pp. 7763–7775, 2012.

[44] T. Curk, J. Demšar, Q. Xu, G. Leban, U. Petrovič, I. Bratko, G. Shaulsky, and B. Zupan. "Microarray data mining with visual programming," Bioinformatics. 21(3):396-398. 2005. Orange software available under http://orange.biolab.si

[45] Miner3D. http://www.miner3d.com

[46] yEd. http://www.yworks.com

[47] S. Nowak, "Some problems of causal interpretation of statistical relationships," Philosophy of science, 27(1), pp. 23–38, 1960.