Final Report: Offshore Wind Power Clustering Tool

Walker Hills, Madison Manley, Shane Oh, Matthew Sandry
whills3@gatech.edu,mmanley31@gatech.edu,soh307@gatech.edu,msandry3@gatech.edu

1 Introduction

Energy consumption in the United States (US) has been rapidly increasing each year due to economic and technological developments, therefore further research into optimizing energy production will be needed to decouple the US from its dependency on foreign imports. Additionally, there is a need to revitalize a degrading power grid and to increase the supply of renewable energy to combat climate change. However, barriers such as operations and maintenance costs limit the expansion of wind power. In our work, we will be focusing on Offshore Wind as a specific area of opportunity. One important issue to consider when selecting offshore wind farms is the cost of power transmission. Potential wind farm areas tend to be in close proximity to one another, making power transmission more efficient if power from multiple wind farms is transmitted across long distances on the same lines. For this project, we aim to aid in the design process of wind power grids across the United States by providing an interactive user interface that helps optimize clustering of wind farm locations together for optimal transmission efficiency as well as demonstrate wind power needs across the country and which wind farms would best fill those needs.With careful site selection, the turbine lifetime and energy production can be increased [1].

2 Problem Definition

We will create an interactive visualization tool that aids in the development of wind power farms by demonstrating wind power needs across the United States, clustering wind farms together for reduced transmission costs, and optimizing the wind farms necessary to supply power to individual states. This tool will allow the user to explore different solutions to wind power farm design through an interactive interface, including choosing between the clustering algorithms of Kmeans, Gaussian Mixture Model, and DBSCAN. The primary tools used for this project are Python, JavaScript, and Application Programming Interfaces (APIs).

3 Literature Survey

3.1 Wind Power Integration

In this section, we review current models and studies on wind power to get a deeper understanding of what constraints need to be considered in our model and how our model can be applied to specified regions. In [2], a Renewable Energy Potential or (reV) model is introduced to evaluate onshore wind potential against a set of constraints, such as height ordinances, landscape classifications, roads, and radar exclusion zones. The model assess the resources, costs, and spatial relation they have to power grids. As [2] highlights constraints to a land-based approach, we will consider similar constraints as well in our work. In addition to this, it is important to understand advances in wind power technology. Blaabjerg et. al. include that developments in different technology sectors can improve the capacity, efficiency, and reliability. Improving wind turbine system power electronics has allowed wind turbines to be integrated into the power grid. Furthermore, advances such as back-to-back power converters allowed “[increase in the] energy yield and [reduction of] the mechanical stress" [3].

In [4], a simulation model is designed that estimates the probability for a specified region to experience an energy "drought" due to further integration of wind and solar. The need for long-duration storage due to variability within the supply of wind and solar power generation is also estimated. This will aid in our understanding when measuring the seasonal variability of wind speeds against a user-requested generation threshold. In our work, we are developing an implementation that will involve several US states and their own regulatory controls, therefore it is helpful to consider studies that focus scenarios in electricity demand vs. generation and how the cost is managed. We first introduce the work of Kling et. al. who focuses on how Europe is integrating wind power into their respective power grids and includes the approaches for power balancing, voltage control, and variability that are being used in the EU

[5]. In comparison, [6] focuses on wind power plants in Iowa and Minnesota where several challenges to wind energy integration are addressed. Although wind forecasts are becoming more accurate, there are still challenges in integrating large scale wind power plants due to wind power generators requiring several hours to start up and synchronize with the grid. The study found that a range of 10% to 50% error in day-ahead forecasts can incur costs ranging from $0.39/MWh to $1.44/MWh.

3.2 Wind Power Forecasting

In this section, we review different wind power forecasting methods to help aid in our optimization model. In the work of Hanifi, et. al., the importance of accurate wind power forecasting is highlighted and several approaches of wind power forecasting methods are reviewed. The predictive models include input features such as, wind, speed, and temperature. Overall, an implementation of hybrid methods provide better forecast wind power and define a robust baseline to compared to the performance of newly developed forecast models. In [7], machine learning is used to predict the suitability of a location for long term wind power generation based on prior wind data while [8] analyzes the effect on power grids of wind power's tendency to heavily fluctuate in a non-gaussian manner.

Furthermore, having accurate wind power forecasting can enhance our optimization model by reducing the overall cost of wind power generation. In [9], a novel two-stage wind power forecasting model based on the multi-objective grey wolf optimizer algorithm, the error factor and a nonlinear ensemble method was introduced. Datasets from wind farms in Canada and Spain were used to conduct experiments and it was found that the model achieved better accuracy and stability for forecasting compared to other models. To improve the accuracy and stability wind power forecasting, [10] introduces an ensemble learning model based on [a] stacking framework to make it possible for different models to learn from one another to obtain the optimal wind power forecasting effect.

3.3 Optimization

In this section, we look at various power optimization methods of wind turbines. In [11], an optimization system for utilizing renewable energy systems in supplying power on a regional scale is designed. Multi-objective linear programming is used in order to determine the optimal mix of renewable and nonrenewable resources for power supply. As our work is aimed to find the optimal location of wind farms, it is important to consider how complex terrain will play a role. Song et. al. introduces an efficient lazy greedy algorithm for wind farm micro-siting optimization integrated computational fluid dynamics for wind energy assessment and the virtual particle model for turbine wake flow simulation for complex terrains [12].

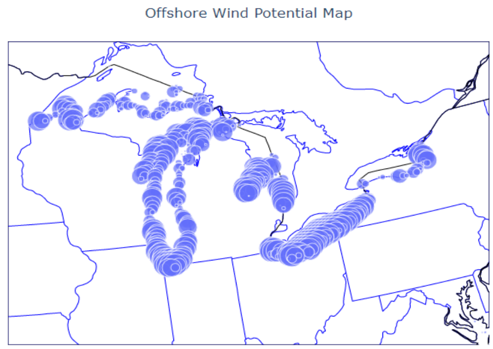
4 Proposed Method

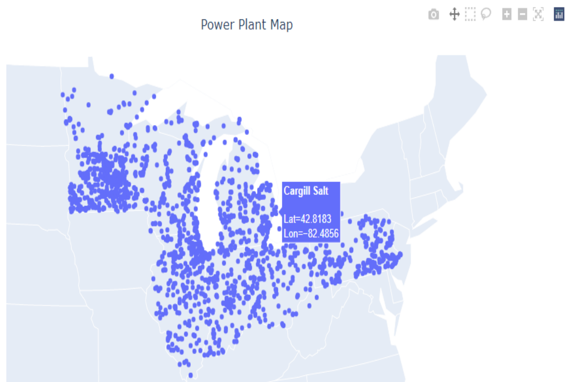
4.1 Intuition

Our approach is unique in a way that we utilize different clustering algorithms mentioned as above, all while visualizing the data in a presentable manner. Planners and elected officials alike can make informed choices with our visualization tool. There are other tools to visualize wind power such as Global Wind Atlas and WindExchange (U.S. Department of Energy), but these tools do not contain temporal data in the same graph, and they do not show potential wind farm locations on the same graph.

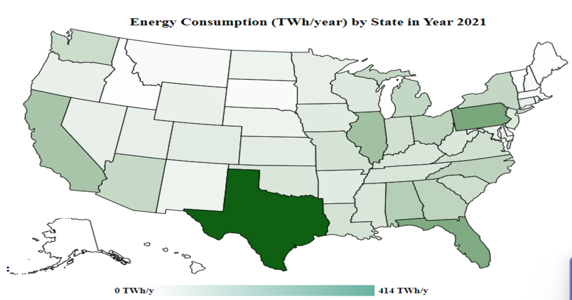
4.2 Data Collection

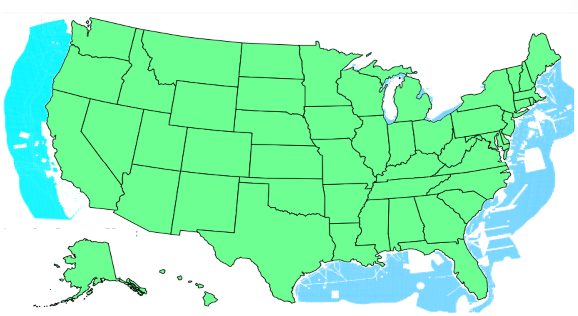
Through the Energy Information Administration's (EIA) API we are able to gather data on all power plants and their net power generation for the years 2001-2021 for our subset of states. We were able to map this data to see where the power plants are and can sum the values to get the yearly demand for each state. We used a data set from the National Renewable Energy Labratory (NREL) that gave us location data for potential offshore wind farms, the square area, the wind speed and their potential for power generation. We then will combine the data from both of these sets to process the data before running it on the optimization model. A bounding box was used to restrict the data to the Great Lakes Region. Maps of the data are shown below. Initially our goal was to focus on just this region for the purposes of implementing an optimization model; however after taking a more clustering focused approach we decided it would be better to perform our analysis on the entire United States and the full dataset of wind sites.





In the following figures below, we can see a heatmap to show the total energy consumption of each state in the year 2021 as well as all the potential offshore wind sites.





4.3 Clustering Algorithms

In this study, three types of clutering algorithms were used: (1) K-means clutering, (2) GMM clustering, and (3) DBSCAN.

K-means clustering is an iterative process with three stages. First, an initial cluster centroids are selected, and the distance from the centroid to all points in the dataset are computed. The points are associated with the nearest mean to create k clusters. The centroids of the k clusters is set as the new mean, and the process is run iteratively until it converges. An example of the algorithm is shown in Figure 3, where k is set to be 6.

GMM clustering is a probabilistic model that assumes that data points are generated with a Gaussian distribution. Unlike k-means algorithm, the covariance structure of the distribution are also considered.

DBSCAN groups points that are closely packed together, and marks points in low-density areas as outliers. Two parameters, epsilon and min points, define density. Epsilon defines the area of a cluster and min points define the minimum number of points to form a cluster.

4.4 Optimization Model - Ideal Case

Objective Function: The following cost function is minimized in order to achieve the optimal solution. The equation is a sum over each cluster, where the first term inside the sum signifies the cost of transmission between clusters and cities, and the second term includes transmission from wind farms to their cluster centers as well as turbine costs.

$$f_{cost} = \sum_{k=1}^{N_k}(\sum_{i=j}^{N_j}(B_{kj}T_{kj}) + \sum_{i=1}^{N_{ki}}(B_{ki}(T_{ki} + W_{ki}(c_t + ym_t))))$$

Constraints

A number of constraints are necessary to properly optimize the model. The following constraint will be added for each state, and ensures that the states receive the necessary amount of power.

For each recipient j:

$$\sum_{k=1}^{N_k}(p_{kj}O_k) \geq D_jP_j$$

The following constraints are added for each cluster. The first defines the energy output of each cluster while the second ensures that the sum of p_{kj}, the proportion of its total output that a cluster gives to each state j, adds up to 1.

For each cluster k:
$O_k = \sum_{i=1}^{N_{ki}} (B_{ki} W_{ki} e_{ki})$
$\sum_{j=1}^{N_j} p_{kj} = 1$

The following limits the number of turbines that can be placed in each wind farm.

For each turbine location i:
$w_{ki} \leq w_{kimax}$

Finally, the following equations define our selection parameters B_x as being Boolean values and dictates that for cluster to state transmission, any transmission percentage requires a value of 1 for that particular Boolean selector. This is unnecessary for farm to cluster transmission since every wind farm transmits all of its power to a single cluster center.

For all B_x:
$B_x \in \{0, 1\}$

If $p_{kj} > 0$ Then $B_{kj} = 1$
Else $B_{kj} = 0$

Variable Descriptions

W_{ki} = the number of turbines placed at location i in cluster k
y = number of years out that the calculation is performed for
B_{kj} = boolean selector for whether or not any power is transmitted from cluster k to state j
B_{ki} = boolean selector for whether or not we place any turbines at location i in cluster k
p_{kj} = the percentage of power output from cluster k that goes towards state j
P_j = the user selected percentage of power demand for state j that will supplied by wind turbines from this model
O_k = total power output of cluster k

Constant descriptions

N_k = number of clusters
N_j = number of recipient states
N_{ki} = number of wind farms i in cluster k
T_{kj} = cost of transmission from cluster k to state j
T_{ki} = cost of transmission from location i to the center of its cluster k
D_j = total power demand of state j
W_{kimax} = maximum number of turbines that can be placed in location i in cluster k
e_{ki} = the amount of power that one turbine produces at location i in cluster k
c_t = cost of building one turbine
m_t = yearly maintenance cost for one turbine

4.5 Optimization Model - Actual Implementation

Due to time constraints and recommendations from the instructors, we have simplified the implementation of the optimization model in order to focus more on the clustering algorithms. In this algorithm, the optimal locations of off-shore wind farms are found by using the Haversine formula (see below) to find the distance between potential wind sites and the coordinates of the capital of each state. It should be noted that this simplifies the model at the expense of accuracy.

$$\begin{aligned} d &= 2r \arcsin\left(\sqrt{\operatorname{hav}(\varphi_2 - \varphi_1) + (1 - \operatorname{hav}(\varphi_1 - \varphi_2) - \operatorname{hav}(\varphi_1 + \varphi_2)) \cdot \operatorname{hav}(\lambda_2 - \lambda_1)}\right) \\ &= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \left(1 - \sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) - \sin^2\left(\frac{\varphi_2 + \varphi_1}{2}\right)\right) \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \\ &= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right). \end{aligned}$$

Below is the pseudo-code of the algorithm we used to find the optimal wind site locations where the user chooses the state, given year, and % of energy consumption needs they want met.

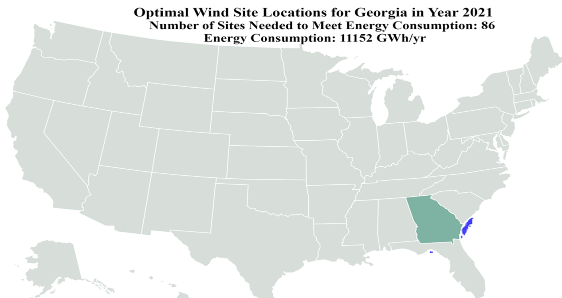
```
Get state, year, % of energy consumption to be met from user
Find energy consumption of given state and year
totalEnergyConsumption = stateEnergyConsumption * % to be met / 100 * 8760
get coordinates of given state
sum = 0

for number of potential offshore wind sites do:
        get coordinates of potential offshore wind site
        Haversine distance(state coordinates, offshore wind site)
        energyGenerated = capacity_mw * capacity_factor * 365 * 24 * 1000
sort by distance, energyGenerated

for number of potential offshore wind sites do:
        if sum > totalEnergyConsumption then:
                break
        else:
                sum += energyGenerated

Plot all potential offshore wind sites before breaking out of loop
```

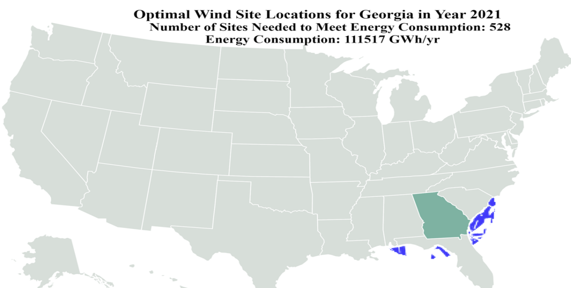
5 Experiments and Evaluations

5.1 Optimization Model Experiment

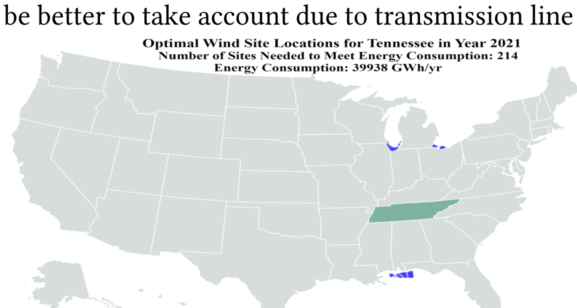
We tested the optimization model with different test cases. In Figure's below, we compared how many wind farms would be needed to meet 10% (first figure) and

100% (second figure) of Georgia's energy consumption in the year 2021. To meet 100% of the energy consumption 214 offshore wind farms would be needed compared to only 86 offshore wind farms needed to meet only 10% of the energy consumption.



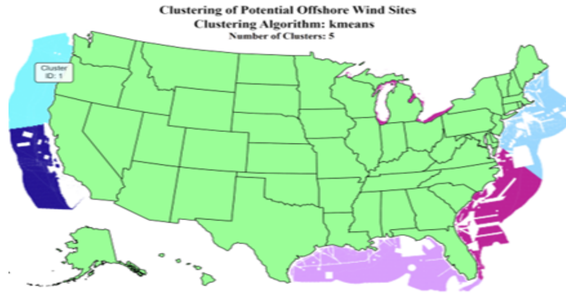


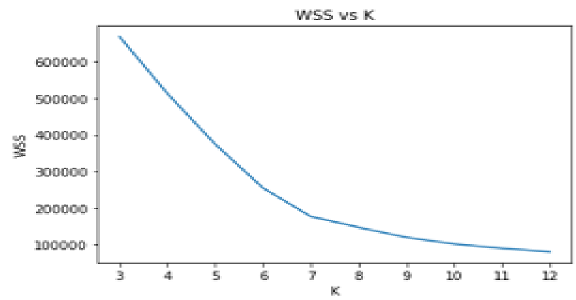
One interesting case we observed is for the state of Tennessee. Our model prioritizes location over the energy generation of a potential wind farm. This is where the clustering algorithm chosen, and its parameters would impact the optimization model. Taking into account the clustering algorithm and centroid locations, the optimal wind farm locations for Tennessee could instead be clustered in one area instead of two, which would be better to take account due to transmission line costs.



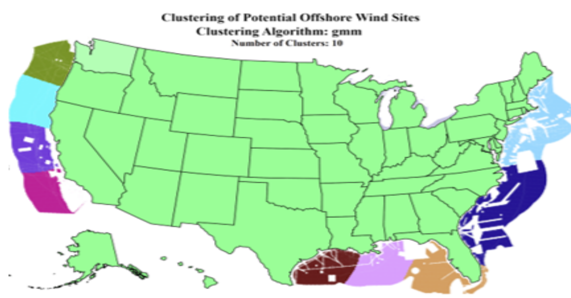
5.2 K-Means Clustering Results

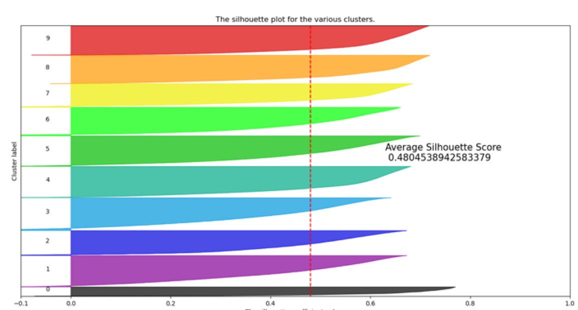
For K-means clustering, the optimal number of points was found by examining an elbow curve. K values of 1-3 were deemed trivial because the data runs the west and east coast (k = 2). At k = 3, another cluster is added in the Southern US around the Gulf of Mexico. The three clusterss are forced by the geographical location of the data. Below is an example of kmeans clustering and the elbow curve.





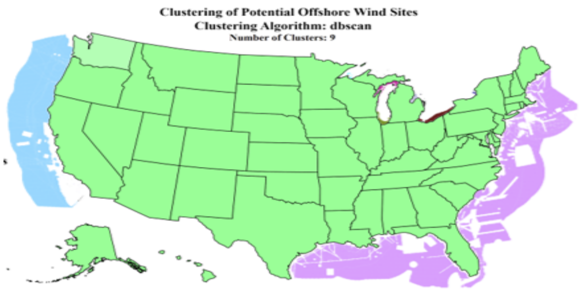
5.3 GMM Clustering Results





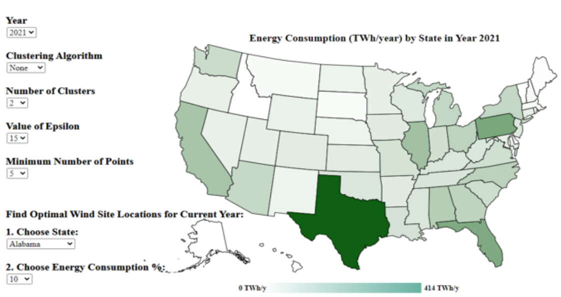
For GMM clustering, we used silhouette analysis to see the separation distance between the clusters. For a cluster size of 10, all the plots are more or less of equal size, therefore the clusters are of similar size. Comparing these plots to other cluster sizes would help find the optimal number of clusters for GMM.

5.4 DBSCAN Clustering Results



For DBSCAN, parameter choice for epsilon and minimum points has a significant impact on the clustering algorithm's performance. Values in the range around 50 for minimum points and 100 miles for epsilon produced the highest performance as measured by silhouette coefficient. Larger numbers of clusters were achievable with smaller values for both parameters, though finding a high-quality solution took much more fine tuning and still resulted in a silhouette coefficient close to 0. DBSCAN was the least successful algorithm in terms of user interactivity due to its parameter choices requiring much more domain knowledge to set properly, and improper setting of the epsilon parameter can easily result in no or only one cluster being created.

5.5 UI Evaluation



In this evaluation, we will assess the efficacy (e.g. Does the UI design reflect the goals of the project?) and the efficiency (e.g. Use of elements, colors, style, etc.) of the UI design. The design should be friendly for first-time users and be both simple and clear. Below are some comments from first time users in the group and first time users outside of the group.

First, the web interface for this project was evaluated by the group members. It seemed that constantly keeping the drop-down for "Value of Epsilon" and "Minimum Number of Points" could be confusing to users who are not familiar with the clustering algorithms, since they could mistakenly think that they need to choose the above value for k-means clustering algorithm. Another thought was that perhaps showing the clustering results and energy consumption map together could be beneficial to the users of the web interface, as that would provide more information. The interface was presented to a person outside the group who has had no experience in the field; the person wished that there was a brief explanation of the different algorithms on the webpage itself.

6 Conclusions and Discussions

6.1 Goals and Future Work

In this project, we have created an interactive visualization tool that aids in the development of wind power farms by demonstrating wind power needs across the United States, clustering wind farms together for reduced transmission costs, and optimizing the wind farms necessary to supply power to individual states. The ideal case optimization model and its integration with the JavaScript visualization tool was the original goal of the project. After speaking with Professor Roozbahani about the issues with the implementation his recommendation was to focus on the clustering within the model. Due to the time constraints of the semester and the complexity of the problem we set out to solve we found that working with clustering the data and a simpler model would be achievable and still accomplish some of our initial goal. The ideal model implementation and its full integration into a user interactive visualization now constitutes our desired future work.

6.2 Statement of Work

Madison designed the UI in Javascript, wrote the algorithm for the implementation of the optimization model, and implemented GMM clustering. Walker worked on the data collection and python notebooks, designed the GUI in python, worked on the theoretical optimization model, and implemented KMeans clustering. Matthew worked on the theoretical optimization model and implemented DBSCAN clustering. Shane worked on the data collection, theoretical optimization model, and python notebooks. All team members have contributed a similar amount of effort for PowerPoint presentation and final paper.

References

[1] A. Kusiak, H. Zheng, and Z. Song, “Power optimization of wind turbines with data mining and evolutionary computation,” *Renewable Energy*, vol. 35, pp. 695–702, 2009.

[2] A. Lopex, T. Mai, E. Lantz, D. Harrison-Atlas, T. Williams, and G. Maclaurin, “Land use and turbine technology influences on wind potential in the united states,” *Energy*, vol. 223, pp. 3183–3198, 2021.

[3] F. Blaabjerg and K. Ma, “Future on power electronics for wind turbine systems,” *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 1, no. 3, pp. 139–152, 2013.

[4] J. D. Amonkar Yash, Farnham and U. Lall, “A k-nearest neighbor space-time simulator with applications to large-scale wind and solar power modeling,” *Patterns 3*, vol. 3, no. 100454, 2022.

[5] W. L. Kling, L. Söder, I. Erlich, and et al., “Wind power grid integration: The european experience,” *17th Power Systems Computation Conference (PSCC)*, 2011.

[6] B. Parsons, M. Milligan, B. Zavadil, D. Brooks, B. Kirby, K. Dragoon, and J. Caldwell, “Grid impacts of wind power: a summary of recent studies in the united states,” *Wind Energy*, vol. 7, no. 2, pp. 87–108, 2022.

[7] H. Demolli, A. Dokuz, S. Ecemis, and M. Alper Gokcek, “Wind power forecasting based on daily wind speed data using machine learning algorithms,” *Energy Conversion and Management*, vol. 198, 2019.

[8] H. Haehne, K. Schmietendorf, S. Tamrakar, J. Peinke, and S. Kettemann, “Propagation of wind-power-induced fluctuations in power grids,” *Physical Review E*, vol. 99, no. 5, 2019.

[9] Y. Hao and C. Tian, “A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting,” *Applied Energy*, vol. 238, pp. 368–383, 2019.

[10] C. Dong, H. Zhang, C. Wang, and X. Zhou, “Wind power forecasting based on stacking ensemble model,decomposition and intelligent optimization algorithm,” *Neurocomputing*, vol. 462, pp. 169–184, 2021.

[11] A. Arnette and C. Zobel, “An optimization model for regional renewable energy development,” *Renewable and Sustainable Energy Reviews*, vol. 16, no. 7, pp. 4606 – 4615, 2012.

[12] M. Song, X. Zhang, and J. Wang, “The lazy greedy algorithm for power optimization of wind turbine positioning on complex terrain,” *Energy*, vol. 80, pp. 567–5745, 2015.