Comparing Decision Rules for Siting Interconnected Wind Farms

Paras Choudhary
Pennsylvania State University
puc134@psu.edu

Seth Blumsack
Pennsylvania State University
sethb@psu.edu

George Young
Pennsylvania State University
young@meteo.psu.edu

Abstract
The variability and non-dispatchability of wind power creates many challenges for the operators of electric transmission systems. Current U.S. wind energy policies are focused on encouraging quantities of wind power without much attention paid to quality of the power produced. Using detailed meteorological data from 113 different weather stations in Oklahoma, we simulate power production from a large number of interconnected wind farms and devise a variance-minimizing rule for successively adding farms over a wide geographic area. Our variance-minimizing rule reduces the standard deviation of five-minute averaged wind power output decreases by 27% after grouping of 8 stations. We compare our variance-minimizing decision rule with two other decision rules for incremental wind investment: a nearest-neighbor rule that has been suggested in previous literature and a profit-maximization rule that reflects decentralized decision-makers. All interconnection decision rules reduce the aggregate variance of wind power output, particularly after several stations are interconnected. We find that the nearest-neighbor rule reduces variance by less than half that of the variance-minimization rule. The profit-maximization rule achieves 75% of the variance reduction attained through the variance-minimization rule. We also evaluate and compare wind power variability over hourly and daily time scales and analyze the sensitivity of variance-minimizing wind energy investment patterns to wind-speed measurement frequency. This work is a first step in a larger project, in which we plan to compare the intermittency costs of wind power that arise from different siting policies or decision rules.

1. Introduction

Wind energy is considered to be an advantageous renewable electricity generation source because it is non-polluting, economic and allows the generation site to be used for agriculture or forestry. According to the US Energy Information Administration (EIA), wind penetration in the US electricity production has increased from 0.15% to 1.8% over the last 10 years. Federal tax credits like the Production Tax Credit (PTC), Investment Tax Credit (ITC) and other incentives have encouraged growth in the U.S. wind energy sector [1], as have state-level Renewable Portfolio Standards (RPS). RPS programs provide market mechanisms to ensure that a growing percentage of electricity is produced from renewable sources [2]. It provides cost-effective and competitive markets for renewable generators. RPS policies currently exist in 28 U.S. states plus the District of Columbia, but not at the national level.

The variable nature of wind power, its non-dispatchability and errors in wind generation predictions create many challenges for the system operators. In particular, if wind penetration keeps increasing at the current rate, the cost of wind generation will increase by uncertain amounts due to requirements for fill in generation capacity and ancillary services. Current U.S. renewable energy policies seem designed to increase the penetration of wind but little policy attention has been paid to the quality of wind power produced and the system cost of obtaining it. Most system operators distribute the cost incurred due to wind integration evenly across their entire system. Distributing these costs over wind generators based on variability of power produced may have implications on determining profitability of wind power projects. The Bonneville Power Authority (BPA) has already introduced a wind integration charge of 0.6¢/kWh to cover for the cost of maintaining reliability [3].

Previous research has examined the variability of wind power output over temporal and spatial scales [3-5]. Studies have been done to examine the effect of interconnecting wind farms [6-9] and geographic smoothing of wind power variability is a widely accepted idea. Much of this previous literature has
suggested that interconnecting nearby wind farms can reduce the variability of output significantly. Our analysis compares this “nearest-neighbor” algorithm with two additional policy options for selecting wind farms to be interconnected to a common grid. The first option sequentially adds wind farms in a manner that minimizes the aggregate variance of wind power output. The second option adds wind farms in order of profitability.

We assume that lossless transmission lines connect the wind farms, and that new wind farms connect to the grid at locations close to where the wind energy is produced, thus allowing us to neglect the cost of transmission lines. This assumption may be fair for a single state such as Oklahoma, but perhaps not for a wider geographic area. Finally, we assume that the system cost of wind variability is at least weakly monotonically increasing in the magnitude of variability. This assumption is consistent with the existing literature on wind interconnection (see, for example, [6], [10], [11]).

Our variance minimizing algorithm, described further in Section 3, sequentially adds wind farms that are least correlated with the group of existing stations. Our analysis for the standard deviation of latitudes and longitudes of variance-minimizing wind additions for Oklahoma shows that the stations are spread out so as to make the power least correlated, which is consistent with earlier findings [9, 12].

There is no significant effect on the average power output of the group, but after grouping of 8 stations, the variability in five-minute wind power output decreases by 27%, in hourly averaged power by 25% and daily averaged power by 22%. This suggests that the costs of fast-response fill-in power (such as regulation power) are likely to stabilize after a small number of wind farms are built, as long as those wind farms are geographically dispersed as indicated by our algorithm.

We then compare the results of our variance minimization site selection process with those of revenue-based connections and those of distance based connections of wind farms. Our results show that standard deviation of 5-minute power data decreases by 27% after addition of 8 stations whereas this value is only 12% for nearest stations connections and 22% for revenue based connection.

We also evaluate and compare wind power variability over different time scales and analyze the impact of wind measurement frequency on sensitivity of variance minimizing wind energy investment patterns. Comparison of standard deviation of 5 minute, hourly and daily averaged wind power shows that the 5 minute and hourly characteristics are almost identical. This suggests that we may not need high-resolution data for our algorithm to reach valid conclusions. Further analysis shows that sequences of addition of wind farms calculated using 5 minute and hourly wind data are interchangeable for calculation of standard deviation and coefficient of variation.

2. Data

We are working with wind speed data provided by Mesonet in Oklahoma. The Mesonet data accounts for over 110 automated stations across Oklahoma. There is at least one Mesonet station in each of Oklahoma's 77 counties as shown in figure 1 [11]. We use 5-minute averaged wind speed data (10 m height) from 113 weather stations in the state.

![Figure 1. Location of 113 Mesonet weather stations in Oklahoma](image)

We use the data from each station for 366 days (Jan 01, 2002 to Jan 01, 2003, chosen to represent a normal wind year). There are approximately 105,408 data points for each weather station; the dataset has no missing data. We use a logarithmic velocity profile to estimate wind speeds at a height of 80 m (equation 1, 2) [13]. It assumes that the surface layer is adiabatic and the wind velocity profiles depend on surface roughness length \( Z_0 \). We use \( Z_0 \) values provided for each station by Mesonet. This physically-based equation has the advantage over empirical power law formulations of being valid at all levels within the atmospheric surface layer and being easily extended to include diabatic effects if the necessary surface flux data are available [14].

\[
(1) \quad u(h_r) = \frac{u_*}{k} \ln \frac{h_r}{Z_0}
\]

For each weather station,

\[
(2) \quad \bar{u}(80m) = \bar{u}(10m) \cdot \frac{80}{\ln \frac{10}{Z_0}}
\]
Where
\[ u(h_r) = \text{Wind speed at reference height} \]
\[ h_r = \text{Reference height} \]
\[ z_0 = \text{Surface roughness length} \]
\[ \kappa = \text{von Karman constant (~0.4)} \]

We estimate the power output for a wind energy installation at each Mesonet station location by using the power curve for a GE 1500 kW wind turbine (model GE1.5 S) with a cut-in speed of 8.9 mph and cut-off speed of 55.9 mph. The power curve is given in figure 2 [17].

Of these 113 weather stations, some stations turned out to be very low power output wind sites, with less than a 10% capacity factor. The low wind power stations have very low coefficients of correlation with other stations as they have zero power output most of the time (zero value entries in their power matrix). Thus, the low correlation of these stations with other stations arises from the fact that their power output is zero most of the time, not because of inherent differences in wind characteristics. Including these low power sites in our analysis will lead to two problems: first, our algorithm selects these sites first and therefore average power output of our connected wind farm cluster falls rapidly. Second, we doubt that it would be economically feasible to place a wind farm at these sites, since the turbines would not be producing electricity very often. Thus, we selected the top 50 wind power stations that have moderate to high capacity factors. As seen in figure 3, there is no significant decrease in average power output of our wind farm cluster when we work with top 50 wind sites. But if we work with all 113 available sites, the average power output decreases by almost 25% after grouping of 8 wind farms.

The capacity factors of these top 50 wind sites vary from 21.9 to 39.3 percent. Figure 4 shows the locations of the top 50 stations. We note that excluding the lowest-power stations in Oklahoma removes large portions of the eastern part of the state from our analysis. The portion of Oklahoma that we do not consider has a highly forested and somewhat hilly topography [18]. The low wind-speed and power readings at these locations may arise from the placement of the Mesonet stations, rather than an inherent deficiency in the wind resource. These sites are located in mountain valleys, thus being sheltered from the prevailing winds, particularly at night when cold air pools in the valley floors. This phenomenon is one of several reasons that many wind turbines are situated on ridge crests rather than in the intervening valley floors.

---

1 In reality, most large wind farms will be spread over a sufficiently large area that this geographic dispersion alone may contribute somewhat to reduction in the variation of output. A method for incorporating these micro-climate effects is suggested in [16], but the method cannot be implemented without data on turbulence intensity.
3. Variance minimization site selection algorithm

Our goal is to develop an algorithm to determine the order in which a group of wind farms located away from each other should be interconnected so that the overall variability of their joint power output is minimized. This algorithm will help us determine the best sites among available sites for setting up new wind farms and will provide a benchmark for comparison with the nearest-neighbor and profit-maximization decision rules.

We assume that we have wind farms of equal installed capacity at each of these weather stations. We can find the 5-minute averaged wind power by using the 80 meter wind speed data and the power curve as mentioned earlier. Also assume that we start interconnecting these wind farms starting at some station $k$ ($k$ being between 1 and 50). We want to find the wind farm, which when connected to the one at station $k$, will minimize the percent standard deviation of their combined power output. By interconnecting we mean that they are connected by lossless uncongested transmission lines. In effect, it means that rather than having two separate wind farms, we can analyze their combination as a single unit.

To find the variance minimizing station, we calculate the correlation coefficient of the 5-minute power data of station $k$ with all the other 49 stations. We quantify this correlation using Pearson’s linear correlation coefficient given by equation (3)

$$\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y \sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} ; (-1 \leq \rho \leq 1).$$

For the purposes of illustration, suppose that station $j$ has the lowest correlation coefficient with station $k$. By connecting stations $j$ and $k$, the variability of their combined power output will be minimized.

$$\text{var}(k + j) = \text{var}(k) + \text{var}(j) + 2 \rho_{kj} \sigma_k \sigma_j$$

Our variance-minimizing site selection algorithm proceeds as follows.

**Step 1:** Select a single station $j$, $j \in 1, \ldots, 50$.

**Step 2:** Find the station $k \neq j$ that solves the following minimization problem:

$$\min_k \text{var}(P_k + P_j),$$

Where $P_k$ and $P_j$ are the power outputs of stations $j$ and $k$.

**Step 3:** Group stations $j$ and $k$ into a single station $j\ast$, defined by:

$$P_{j\ast} = P_j + P_k.$$

Steps 2 and 3 are repeated until all 50 stations are interconnected.

Since we have assumed equal installed capacity at each wind farm, we can also divide $P_{j\ast}$ by the number of stations (n) aggregated to form $P_{j\ast}$ at each iteration. This normalizes the power output, so that it is easier to compare with other stations.

3.1. Pseudo code for finding station addition sequence algorithm

$P_n = \text{Wind power output vector (105408 X 1)}$

containing power output values from a 1500 kW wind turbine at station n; data taken every 5 minutes for 366 days

$TP = \text{Wind power output vector (105408 X 50)}$

formed by concatenating the column vectors $P_1 P_2 \ldots P_n$

$N = \text{number of stations}$

$\text{Sequence} = \text{Sequence of addition of stations for variability minimization}$

If we want to start addition of stations starting at $j^{th}$ station,

$TP \leftarrow [P_1 P_2 P_3 \ldots P_n]$
Sequence[1] ← j
Standard_Deviation[1] ← stdev(TP[column j])
Power_Average[1] ← mean(TP[column j])
  do exchange TP[column 1] ↔ TP[column j]
    for i ← 1 to N-1
      Corr ← Correlation Coefficient of column 1 of TP with all other columns
      I ← index of minimum (Corr)
      Sequence[i+1] ← I
      Power1 ← TP [Column 1]
      Power2 ← TP [Column I]
      NP = ((P1*i)+P2)/(i+1) // average the power
      // Remove The Column I as station at I has already been added and replace Column 1 with NP (New Power)
    end for
  end do

4. Results

4.1. Standard deviation of longitudes and latitudes

We examine the geographical spread of the sites that were selected and how the sites are positioned so as to minimize aggregate variance of simulated wind farm power output in Oklahoma. We calculate the differences in latitudes and longitudes of the stations from that of Oklahoma City (which is more or less situated in central Oklahoma). We then calculate the standard deviation of these differences as the wind farms are progressively connected. Figure 5 shows the standard deviation in latitudes and longitudes with increasing number of stations.

![Figure 5. Standard deviation of longitudes and latitudes with incremental addition of wind farms (averaged over 50 iterations)](image)

We note from the longitudes plot that the stations tend to have high geographical dispersion initially and the dispersion decreases as the number of stations added increases. Thus, the variance-minimizing wind energy site selection algorithm will select the site that is geographically furthest from the group of existing stations. For example, choosing an initial station in eastern Oklahoma will lead to that station initially being aggregated with a station in far western Oklahoma. The third station added will be in eastern Oklahoma, but closer to Oklahoma City. The latitudes show similar behavior, but the effect is not apparent as Oklahoma itself has less latitudinal spread. The results provide another piece of evidence in support of findings from earlier studies (that did not include Oklahoma) about geographical smoothing of wind variability.

4.2. Results with 5-minute data

First, we look at the effect of interconnecting wind farms on the standard deviation of power output. Figure 6 shows the average standard deviation of combined power (normalized as discussed in Section 3) with increasing number of wind farms connected together.
A point to note is that the overall variability (or standard deviation) increases as we add more stations. This is due to the fact that our range of output power values increases as we add wind farms together. For 1 wind farm, the power output values will be between 0 to 1500 kW, and for 10 connected wind farms, the combined power can vary between 0 to 15,000 kW.

Since we have 50 stations and the sequence in which we should add wind farms depends on the starting station, we ran 50 simulations of our algorithm, each time beginning with a different station. To see the general trend, we plotted the mean value of standard deviation averaged over 50 simulations; the y-value on figure 6 at station 10 on the x-axis shows the mean value of standard deviation seen over all 50 simulations after interconnection of 10 wind farms.

We see that there is a decrease of more than 27 percent in the standard deviation after interconnection of 8 stations. The average standard deviation of the wind farm group decreases rapidly, then reaches a minimum and then increases very slowly.

To investigate the reduction in variability without averaging the output power after adding stations, we also considered the coefficient of variation given by equation (6)

\[ Cv = \frac{\sigma}{\mu} \]

Where 
- \( \sigma \) = standard deviation 
- \( \mu \) = mean

This helps us in comparing the data sets with widely differing means. So even if we do not average the output power, the coefficient of variation captures the effect of connecting wind farms together. Figure 7 shows the mean of \( Cv \) versus increasing number of stations. The values shown are means over 50 different simulations. We note that Coefficient of variation decreases from 1.22 to 0.82 after 8 stations are added, a 32% decrease.

4.3. Comparison with other methods to connect wind farms

We compared the results from our variance-minimizing site selection algorithm with two other algorithms:

Nearest-neighbor Decision Rule: Previous literature has suggested that interconnecting nearby wind farms can reduce the aggregate variation in wind power output [3-9]. We implement the nearest-neighbor rule by choosing new wind farms to minimize the distance from the centroid of existing wind farms. This method suppresses some low frequency variations as weather in close by areas shows less variation on a longer time scale.

Profit-maximization Decision Rule: An alternative decision rule would be to choose the highest-revenue locations first (we note that since we neglect transmission costs in our analysis and assume that wind farms connect to the transmission grid at nearby locations, revenue maximization and profit maximization will be similar goals). We first estimate the energy prices and thus revenues at all the potential sites by generating Locational Marginal Prices (LMP) within Oklahoma. Then we connect stations in descending order of their revenue, so the
highest-revenue stations are connected first, with lower-revenue stations connected subsequently.

Figure 8 and 9 show the comparison of standard deviation and coefficient of variation when the wind farms are connected using different algorithms. We note that standard deviation decreases by 27 percent with variance minimization algorithm after addition of 8 stations, whereas this value is 12 percent for nearest stations based and 22 percent for revenue-based algorithm.

Figure 8. Comparison of standard deviation of power when wind farms are connected using different algorithms

A point to note here is that the revenue based connection algorithm generates just one sequence (based on net revenues in a year), whereas the other two algorithms generate 50 different sequences, depending on which wind farm is the first to be constructed. We calculate the standard deviation and coefficient of variation for all 50 sequences and then average them so as to get general trends. In figure 9, for example, the coefficient of variation for the revenue-based rule is sometimes lower than the coefficient of variation of the average of the 50 sequences generated by the variance-minimization rule.

Figure 9. Comparison of coefficient of variation of power when wind farms are connected using different algorithms

4.4. Evaluation of variability over multiple time scales

In many U.S. states, electricity is traded on centralized day ahead and spot markets. Ancillary services (for providing fill-in power for wind energy and for other purposes) are also procured and dispatched over multiple time scales. In this section, we compare the variability in five-minute power to hourly and daily power produced by our simulated Oklahoma wind farms.

To get the hourly and daily wind power, we take the hourly and daily average of 5-minute power data. Then we follow the same steps as we did for 5-minute data to calculate the standard deviation and coefficient of variation. Figure 10 shows the mean standard deviation for 5-minute, hourly and daily averaged power data.

Our results show that the decrease in standard deviations for hourly and daily power show similar characteristics as 5-minute data. This would be expected since the hourly and daily data sets are simple averages of the original 5-minute data stream. But we see that the 5 minute and hourly curves almost coincide, which is consistent with the results in [3] based on a frequency spectrum analysis of very high resolution wind data. We note here that our results in figure 10 are not meant to downplay the importance of high-frequency data in evaluating the costs of providing fill-in power for intermittent wind farms. Our results do, however, suggest that data at resolutions of five-minute or more frequent intervals may have limited value in differentiating low-quality wind sites from high-quality wind sites, and that wind farm siting decisions based on five-minute average

Proceedings of the 44th Hawaii International Conference on System Sciences - 2011
data and hourly average data may be similar if not identical.

As evident from figure 10, the 5-minute and daily characteristics are also similar, but daily data standard deviation curve is displaced vertically by approximately 85 kW. This difference is approximately constant, regardless of the number of stations considered (the position of the x-axis in figure 10). So the daily standard deviation starts from a lower base and follows the same trend as 5-minute standard deviation.

**Figure 10. Standard deviation of output power with incremental addition of wind farms for 5-minute, hourly and daily averaged power.**

Computing the coefficient of variation, we see similar trends, as shown in Figure 11.

**Figure 11. Coefficient of variation of output power with incremental addition of wind farms for 5-minute, hourly and daily averaged power.**

4.5. The sensitivity of variance-minimizing wind energy investment patterns to wind-speed measurement frequency

Figure 10 suggests that the magnitude of five-minute wind energy variations is nearly identical to the magnitude of hourly wind energy variations. Thus, it is not clear that five-minute data (which may be expensive to obtain relative to hourly measurements) has significant informational value when evaluating the costs of intermittent wind energy. We have found, however, that the variance-minimizing order (sequence) of the addition of wind farms will differ somewhat depending on whether the variance is calculated based on 5-minute, hourly or daily averaged powers. To quantify the difference, we calculate the average displacement of every station in different sequences.

To explain displacement, suppose that we have two sequences for incremental addition of stations, one for 5-minute data and one for hourly data, both starting at station 3. This “wind investment stack” is illustrated in Figure 12. We note the position of each wind farm in both the sequences and calculate the absolute value of the difference between the indices of their positions.

**Figure 12. Illustrating differences in the variance-minimizing incremental wind farm investment stack for analysis based on five-minute wind data and hourly wind data.**

As an example, if wind station 23 appears 5th in 5-minute sequence and 18th in hourly sequence, its displacement is

\[ D_{23} = |5-18| = 13 \]

We calculate the mean displacement by averaging displacements of all 50 stations in a sequence, as shown in equation (7) below.

\[
D_{av} = \frac{\sum_{i=1}^{50} |index_{5\text{min}} - index_{\text{hour}}|}{50}
\]

\[
(7)
\]
Table 1 shows the range of average displacement values for different sequences in different time scales. We note that the average displacement varies from a minimum of 0.32 to 3.52. Uncorrelated sequences should see an average displacement equal to 25 (since in a random site selection, a given station should be equally likely to move up or down in the incremental investment stack). Thus we see that the order of addition of wind stations do not vary too much for variability minimization of 5-minute, hourly or daily wind power.

Table 1. Pair-wise average displacements of stations in different time sequences.

<table>
<thead>
<tr>
<th></th>
<th>5-min</th>
<th>Hourly</th>
<th>Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-min</td>
<td>-</td>
<td>0.32 – 2.08</td>
<td>2.48 – 3.76</td>
</tr>
<tr>
<td>Hourly</td>
<td>0.32 – 2.08</td>
<td>-</td>
<td>2.32 – 3.52</td>
</tr>
</tbody>
</table>

To observe the impact of wind data used for our calculations, we created a matrix of standard deviation characteristics using 5 minute, hourly averaged and daily averaged wind data and respective sequences of addition. We take the station addition sequence produced by 5 minute data and use it to calculate the standard deviation of wind power output based on five-minute interval data, hourly interval data and daily interval data. Similarly, we repeat the process for the investment sequences produced by analysis of hourly and daily data. This exercise generates a 3 x 3 “matrix” of wind power variance profiles based on five-minute, hourly and daily standard deviations of wind power output arising from wind farm investment decisions based on analysis of five-minute, hourly and daily average wind data. These matrices are presented in Figures 13 (standard deviation) and 14 (coefficient of variation). In the figures, the “rows” of each matrix represent the granularity of data used to determine wind-farm investment sequences, and the “columns” of each matrix indicate the time scales over which variances are calculated.

We observe that the first and second rows in figures 13 and 14 are almost identical. The first row shows the 5 minute, hourly and daily data characteristics when we connect wind farms using sequences generated by 5 minute data. Similarly, second row shows these characteristics calculated using sequences generated by hourly averaged wind data. This reinforces our findings in the displacement section that 5 minute and hourly sequences of connecting wind farms are very similar and will provide same results even if interchanged. Thus, if we have hourly sampled data for a region, that should suffice to determine variation minimizing sites and connection sequences.

5. Conclusions and Policy Implications

U.S. policy at the state and federal level has been encouraging the development of wind energy. Incremental additions to wind energy fleets are based on analyses of site feasibility and profitability. Using data from Oklahoma that is fine-grained over spatial and temporal dimension, we confirm results of the
previous studies [6-9] about geographical smoothing of wind power variation. The standard deviation of latitudes and longitudes show that least correlated stations tend to spread out geographically. The intermittence and variation in wind increases the system costs of using wind power. We have initially examined the variance minimization as a decision criterion for location of wind farms. We found that if wind farms are connected according to our algorithm, there is a decrease of more than 25 percent in the standard deviation and 30 percent in coefficient of variation of output power after grouping of 8 wind farms. The comparison of variance minimization, revenue based and distance based site selection methods shows that the decrease in standard deviation is only 12% for distance based and 22% for revenue based after 8 wind farms are connected.

Our work in this area is ongoing, but our results thus far have important policy implications. System operators currently do not consider the quality of wind resources when establishing queues for interconnection; these queues are established based on decisions made by individual wind developers (similar to our profit-maximization decision rule for sequentially adding wind farms). Allowing system operators to use simple quality metrics, such as expected reduction in aggregate wind variance, to prioritize wind investments in the interconnection queue could considerably reduce the costs of meeting social goals related to the utilization of wind energy. A cost-minimizing system operator would not want to permit the sequential interconnection of geographically proximate wind farms; dispersing them over a geographical area according to our variance minimization algorithm will lead to minimization of the cost of ancillary services. The magnitude of the cost savings is an important area of future work.

6. Acknowledgements
Support for this project was provided by the Penn State Electricity Markets Initiative and the College of Earth and Mineral Sciences at Penn State. The authors would like to thank Andrew Kleit for providing helpful comments on earlier drafts; Paulo Raposo for assistance in generating maps; and Gary McManus of the Oklahoma Mesonet for assistance with data gathering and interpretation.

7. References