Reducing Variability through Geographic Diversity: A Tri-Metric Evaluation of Wind and Solar Site Pairings

Dulf Vincent Genis
Ameren Innovation Center
Ameren Services
Champaign, Illinois, United States
dvjgenis@gmail.com

Shil Prasenjit
Ameren Innovation Center
Ameren Services
Champaign, Illinois, United States
pshil@ameren.com

Kevin Ficek
Ameren Innovation Center
Ameren Services
Champaign, Illinois, United States
kficek@ameren.com

ABSTRACT

This paper introduces a trimetric methodology combining production diversity, correlation analysis, and absolute difference evaluations to assess the geographic diversity of wind and solar energy sites. Using data from Ameren Missouri's prospective wind and solar expansions and coverage statistics that validate data completeness, the study evaluates how pairing sites across different locations can reduce simultaneous low-output hours and smooth out power generation.

The production diversity component measures how often one site generates power while the other is idle, highlighting complementary behavior. Correlation coefficients reveal whether two sites tend to follow the same output patterns, while absolute differences quantify the magnitude of that variability over time. These metrics present a reproducible framework for identifying site pairs that minimize collective volatility and maximize overall efficiency.

Key findings demonstrate that highly mismatched, low-correlation pairs (e.g., wind farms separated by hundreds of miles, or solar sites in distinct climates) consistently exhibit more stable combined output. This outcome supports regulatory and utility discussions about locating renewable infrastructure beyond local state lines, particularly where policy resistance exists. By rigorously quantifying site-level diversity, this approach equips utilities and regulators with actionable insights for optimizing geographically distributed renewables.

INTRODUCTION

Ameren Missouri faces a critical challenge in its renewable energy expansion: balancing regulatory preferences for in-state development with the demonstrated benefits of geographic diversity. While MISO has approved the utility's renewable growth, skepticism persists about locating assets across state lines, despite evidence that distributed wind and solar sites can reduce volatility and enhance grid stability by leveraging complementary weather patterns. The crux of the issue lies in proving that specific out-of-state pairings offer measurable advantages over localized clusters.

Despite widespread agreement on the value of geographic diversity, a systematic analytical framework for assessing how effectively two sites complement each other is still lacking. Much of the existing literature focuses on broad system-level models or provides generic guidelines for spacing wind and solar farms. Missing, however, is a replicable methodology that can zero in on specific site pairs and quantitatively measure (1) production diversity–how often one site produces when the other does not; (2) correlation–how synchronized or unsynchronized two sites' energy outputs are; and (3) absolute differences–the magnitude of power output fluctuations. Each metric provides a distinct perspective on whether two locations are "geographically diverse" and how that diversity may translate into greater efficiency and operational reliability.

Hence, this project introduces a tri-metric approach intended to identify and validate geographically diverse pairs of renewable energy sites in a practical, reproducible manner. Combining production diversity metrics, Pearson correlation coefficients, and absolute difference evaluations, the framework highlights which site pairings are most likely to smooth out each other's energy dips, maintain consistent production, and reduce the risk of simultaneous low-output events. This is especially relevant for Ameren Missouri, which needs robust, data-driven arguments to justify the placement of renewables outside state lines. The theoretical framework driving this analysis lies in two central ideas:

1. Diversity-Driven Stability: Sites in climatically distinct regions exhibit anti-correlated generation patterns, smoothing aggregate output.

2. Quantifiable Complementarity: Statistical modeling (correlation coefficients, threshold analyses) can rank pairings by their diversity potential.

Research Hypothesis

- The proposed tri-metric framework reliably identifies highly diverse site pairings (e.g., pairs with low correlation and high mismatch).

- Site pairs with greater geographic separation will exhibit stronger diversity metrics, compared to closely spaced pairs, due to reduced climatological overlap.

Under the umbrella question, *"How can we measure site-level geographic diversity and demonstrate that it enhances efficiency?"* this project's objectives include:

- Defining a simple, data-driven approach for identifying whether two wind or solar farms exhibit sufficient diversity to warrant combined planning.

- Quantifying how strongly or weakly correlated their power outputs are, how often they offset each other's downturns, and whether large power differences consistently help the grid.

- Generate evidence to inform Ameren's cross-state siting negotiations with MISO and state regulators.

Paper Structure

Looking forward to the structure of this writing, Section 2 (Literature Review) provides an overview of prior work on geographic diversity for renewable energy. It highlights the theoretical basis for multi-site synergy, the economic justifications for distributed renewable installations, and the research gaps this study aims to fill. Section 3 (Methodology) details the tri-metric approach, explaining how data is collected, cleaned, and analyzed to measure production diversity, correlation, and absolute differences. Section 4 (Results) discusses the coverage of each site, followed by metrics for wind and solar pairs, culminating in insights on how distance and local climate shape output patterns. Section 5 (Discussion and Future Work) interprets the findings, addresses methodological limitations, and identifies potential areas for aggregate pair analysis and advanced correlation. Finally, Section 6 (Conclusion) summarizes the project's key takeaways, emphasizing the policy and practical ramifications for utilities like Ameren Missouri considering broader geographic deployment of renewables.

By establishing a reproducible tri-metric methodology, this study aims to inform utility planners, support regulatory discussions, and illustrate how differently located renewable energy sites can collectively improve power reliability and efficiency. Suppose the results convincingly show that out-of-state projects complement in-state resources. In that case, Ameren Missouri may successfully navigate regulatory barriers and optimize its renewable expansion across a broader geographical spectrum.

LITERATURE REVIEW

1. Theoretical Background: Geographic Diversity in Renewable Energy

The concept of geographic diversity has gained prominence in renewable-energy research as a means to mitigate the inherent variability of wind and solar power. By distributing generation sources over large areas, fluctuations in power output can be smoothed, helping to meet grid reliability and efficiency standards. Early systematic approaches to this idea date back to stochastic-geometry models and multi-dimensional analyses [1,10]. For example, Diakov employs a vector-based angle metric to assess how wind and solar generation align with load in high-dimensional space [1]. Although this perspective is valuable for system-wide balancing, it does not explicitly focus on diversity at the individual site-pair level.

Meanwhile, applications of the Central Limit Theorem (CLT) underpin much of the rationale: aggregating multiple (partially) independent renewable generators reduces variability. Handschy et al. apply Monte Carlo simulations to hypothetical wind-farm arrays and show that the probability of simultaneous low-output events decreases exponentially as independent sites are added [5]. This highlights the benefit of dispersing wind farms over hundreds of miles to lower inter-site correlations.

Together, these foundations demonstrate that geographic diversity can reduce short-term variability [10], diminish extremely low-generation events [5], and enhance overall system reliability [1]. However, most existing studies either focus on the entire power system or broad geographic expansions, without offering a simple, reproducible “tri-metric” framework tailored to practical, site-pair decision-making.

---

2. Key Concepts

2.1 Production Diversity

Production diversity measures how often one site produces energy while another does not—a direct indicator of complementary behavior. Mills and Wiser analyze short-term insolation data across multiple solar sites and find that aggregated PV output smooths sub-hourly ramps [10]. A related “zero-output mismatch” metric is often applied to wind [4]. While these metrics highlight the value of non-synchronous generation, none unify mismatch hours with correlation and absolute differences in a single standardized framework.

2.2 Correlation Analysis

Correlation quantifies the linear relationship between outputs. Tutorials from Kent State University Libraries define Pearson’s r for this purpose [7], while the pandas documentation details its implementation [12]. In renewable-energy studies, low Pearson correlation between site pairs indicates higher diversity—if wind drops at Site A, Site B may remain stable [10,9]. Although rank-based measures (e.g., Spearman) can capture nonlinear dependencies [3], Pearson’s r remains the standard due to its interpretability and widespread tooling support.

2.3 Absolute Power Differences

Absolute power difference captures the magnitude of output discrepancies between sites. Mills and Wiser introduce “step changes” to gauge sub-hourly solar ramps [10], and Diakov’s angle metric similarly explores differences in multi-dimensional generation space [1]. Focusing on pairwise deltas—e.g., a consistent 0.05 MW output gap—yields an intuitive measure for site-level planning. Handschy et al. note that real-world correlations may temper assumed independence [5], but large absolute differences often signal valuable complementarity.

---

3. Related Work

3.1 Renewable Site Selection Strategies

Policy and economics have long driven site selection. Gómez-Quiles applies modern portfolio theory to wind farms, showing a 5 % increase in annual returns when capacity is spread across multiple sites versus a single zone [4]. Ghorbani et al. argue that single-issue decarbonization overlooks broader sustainability dimensions (biodiversity, equity, local acceptance), advocating for globally inclusive yet locally adaptive siting [3]. While they don't propose a tri-metric evaluation, their emphasis on regional context aligns with the need for metrics that go beyond energy output alone.

3.2 Limitations in Existing Policies

Utility-scale renewables often face regulatory barriers. Mills and Wiser observe that state-level policies can restrict the cost-saving potential of wide-area solar deployment [10]. Single-utility "monopsony" rules similarly limit aggregated wind-farm expansions [4]. Although the literature supports broad geographic dispersion [1,5,9], policy constraints still hamper implementation at the scale needed for maximum diversity benefits.

---

4. Research Gaps and Contributions

Despite ample evidence that geographic diversity enhances system reliability and can boost financial returns, most existing studies adopt either broad system-level models [1] or purely economic frameworks [4], without providing a simple, reproducible approach for individual site-pair evaluation. In particular, there is no unified method that simultaneously quantifies:

1. Production Diversity: the frequency with which one site generates power while its partner is idle.
2. Correlation Coefficients: the linear alignment or independence of their outputs.
3. Absolute Differences: the magnitude and frequency of their output discrepancies.

By merging these three metrics into a single framework, we can capture complementary behavior (mismatch), linear synergy or divergence (correlation), and the scale of variability (absolute delta) in one analysis. Applying this "tri-metric" approach to Ameren Missouri's prospective wind and solar expansions enables straightforward ranking and optimization of site pairs, directly addressing stakeholders' need for practical, site-level decision support.

Moreover, while some authors touch on aggregate variability, such as combined standard deviations or coefficients of variation [5,10], they stop short of standardizing these into a replicable procedure for utilities. Nor do they explore how incorporating alternative correlation measures (e.g., Spearman's ρ) might refine evaluations in the presence of nonlinear dependencies [3]. Our study fills these gaps by:

- Providing a step-by-step, Python-friendly workflow for computing all three metrics on any pair of time-series generation datasets.
- Demonstrating the approach on Ameren-sourced wind and solar data to show how out-of-state pairings yield fewer correlated dips.
- Extending the methodology to include optional rank-based correlation calculations and combined-variability aggregations for robust, year-to-year planning.

Novelty and Significance This tri-metric framework goes beyond traditional single-metric studies by delivering a reproducible, site-focused tool that regulated utilities can immediately adopt. It quantitatively illustrates how geographic dispersion reduces joint low-output events and smooths overall power delivery, directly confronting policy barriers that favor in-state siting [10]. By offering clear, data-driven evidence that out-of-state sites can outperform local ones on diversity metrics, our approach equips MISO, regulators, and local communities with actionable insights to support more geographically diverse renewable deployments.

METHODOLOGY

Data Collection and Preprocessing

This section outlines the data sources, preprocessing procedures, and power output estimation techniques used to analyze geographic diversity in solar and wind energy production across the United States, specifically, the Midwest.

We collected historical solar and wind data from multiple locations spanning 2018 to 2023 [11, 8]. For solar energy, we sourced high-resolution irradiance and meteorological data from the National Renewable Energy Laboratory's National Solar

Radiation Database (NSRDB), which offers hourly records at 2 km spatial and 10-, 30-, or 60-minute temporal resolution [11], providing a robust and standardized foundation for solar energy analysis.

For wind energy, we utilized the Meteostat Python library to access historical weather records—wind speed and air temperature—from thousands of stations worldwide [8], making it a trusted source for wind modeling.

Each dataset includes key environmental variables necessary for power estimation—Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI), and wind speed—along with timestamps and site metadata.

Data Preprocessing Steps

The data preprocessing stage consisted of several critical steps to ensure consistency, accuracy, and relevance across all sites:

- Fetching Data:
  Solar and wind datasets were programmatically retrieved for each site using API credentials stored in a secure Excel file. This streamlined process ensured reproducibility and traceability of the data collection workflow.

- Time Standardization:
  All datetime values were converted to a uniform format and adjusted for local time zones when appropriate. This allowed for accurate temporal alignment across geographically diverse sites.

- Filtering Operational Hours:
  To focus on periods of meaningful energy production:

  - For solar sites, we retained only experimental daylight hours (07:00–19:00).

  - For wind sites, we retained all operational hours, reflecting the 24/7 nature of wind energy generation.

- Cleaning and Calculating Power Output:
  Raw environmental variables were used to estimate power output for both solar and wind using standardized, academically validated approaches:

  - Solar Power Calculation:
    Power output was estimated using the sum of Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI) rather than Global Horizontal Irradiance, to avoid needing the exact tilt for every station (we assume panels remain perpendicular to the sun or use sun-tracking). An assumed panel efficiency of 21% and a 1 m² area yields:

    Power = (DNI + DHI) × Efficiency × Area

    - This conversion method aligns with established solar PV modeling practices [2, 13].

    - Note: moving to true global irradiance by incorporating measured panel tilt is flagged as Future Work.

  - Wind Power Calculation:
    Wind speeds (measured in meters per second) were mapped to power output using a standardized turbine power curve based on industry models and academic methods [6].

  - Scaling and Normalization:
    Estimated outputs were converted to megawatts (MW) to ensure consistency across all sites and energy types. Unit normalization allowed for valid comparisons.

- Ensuring Data Completeness:
  To assess data quality and availability, we generated coverage statistics for each site.

These statistics quantify the proportion of valid records over the analysis period and provide transparency regarding data gaps and reliability. (See *Coverage Data* in the subsection below.)

Coverage Data

In addition to the basic data cleaning steps, we compiled coverage statistics for each site to quantify the proportion of hours that contained valid data. This coverage data was collected in separate CSV files for both solar and wind packages (e.g., yearly_coverage_stats.csv). Each file indicates, for every site and year:

- ActualHours: Number of hours where data was successfully retrieved and validated.
- ExpectedHours: Number of hours expected for that year (e.g., 8,760 for a non–leap year).
- CoveragePercent: (ActualHours / ExpectedHours) × 100%.

Locations with CoveragePercent below a predetermined threshold (80% in this study, based on visual inspection and data stability needs) were flagged for caution. This threshold balances data reliability with the desire to retain enough sites for meaningful comparison.

Note: The Results section highlights how coverage data is used to adjust visualization opacity or otherwise annotate sites with lower coverage, enabling readers to interpret the data's reliability and completeness.

Determining Geographic Diversity

Geographic diversity analysis identifies location pairs with complementary power output patterns through three analyses: Production, Correlation, and Absolute Difference Evaluation.

1. Production Analysis: Wind vs. Solar

This analysis evaluates how geographic diversity impacts renewable energy production by identifying instances where one site produces power while another does not. Wind and solar are analyzed separately due to fundamental generation differences.

A. Wind Production Analysis (Zero Output Comparison)

Wind production can drop to zero when speeds are insufficient. This analysis highlights complementary pairings where one location produces power while another fails to meet the threshold.

Method:

- Compare hourly power outputs for location pairs (e.g., Howard, IA vs. Atchison, MO).
- Define zero-output hours as time intervals where power output strictly equals 0 MW (no generation). This absolute threshold identifies complete production stoppages.
- Visualize mismatches (see Figure X) and log instances where one site produces while the other is at zero.

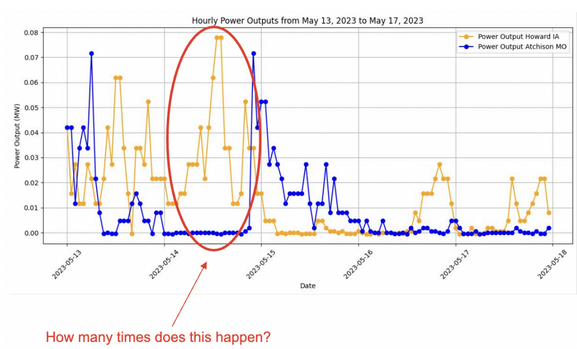

Figure X: Hourly power outputs (May 13–17, 2023) for Howard, IA (blue) and Atchison, MO (orange). Red highlights indicate zero-output mismatches—critical for geographic diversity assessment.

Metrics Generated:

- Total Hours with Zero-Output Mismatch
- Percentage of Mismatch Hours
- Total Available Hours from data sources

Interpretation:
Pairs with higher mismatch percentages exhibit stronger geographic diversity, ensuring more consistent generation when one site underperforms.

B. Solar Production Analysis (Baseline Performance Comparison)

Unlike wind, solar plants rarely have zero output during daylight hours. Instead, variability arises due to cloud cover, atmospheric turbidity, and seasonal changes. This analysis assesses solar diversity by comparing actual solar output to a clear-sky baseline model supplied by the NSRDB, which represents expected performance under ideal conditions.

Method:

- Compare hourly power output for each location pair against a modeled clear-sky irradiance baseline (see Figures Y and Z).

- Log a mismatch if one site matches its clear-sky baseline (ideal production) while the other deviates.

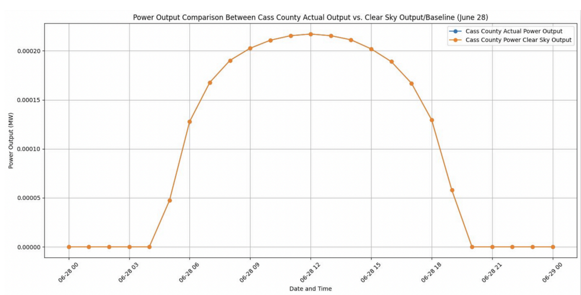

Figure Y: Cass County actual vs. clear-sky solar output (June 28), to which no gaps are seen, indicating perfect alignment with Cass County's actual output to expected (ideal).

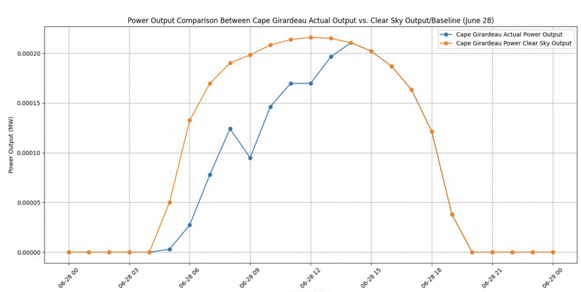

Figure Z: Cape Girardeau actual vs. clear-sky solar output (June 28). Highlighted regions show mismatches where one site meets baseline while the other underperforms.

Metrics Generated:

- Total Hours of Mismatch

- Percentage of Mismatch Hours

- Total Available Hours

Interpretation:
A higher percentage of mismatch hours indicates stronger geographic diversity in solar production. Figures Y and Z illustrate this phenomenon: when one site maintains near-ideal output (e.g., Cass County's close alignment with clear-sky baseline), while another shows significant deviations (e.g., Cape Girardeau's early dips), the pair demonstrates complementary generation patterns. This metric avoids false diversity signals from nighttime non-production.

2. Correlation Analysis

This analysis quantifies how similarly power outputs behave at different sites using the Pearson correlation coefficient (*r*), which measures the strength and direction of linear relationships between two datasets [7].

Why Pearson?

- Python Default: Pandas/NumPy uses Pearson as the default correlation method (df.corr()), making it computationally efficient and widely adopted [12].

- Interpretability: Pearson's *r* clearly indicates:

    - r ≈ 1: Near-perfect positive correlation (sites produce power similarly).

    - r ≈ 0: No linear relationship (sites vary independently, suggesting diversity).

- ○ r < 0: Inverse correlation (rare for renewables; one site's high output aligns with another's low output).

- Alignment: Solar/wind outputs tend to have positive correlations (since weather systems affect nearby regions similarly), but geographic spread can reduce this effect.

Method:

- Compute Pearson's *r* for hourly power outputs of each site pair.
- Analyze multiple years of data to capture long-term trends (avoiding short-term noise).

Metrics Generated:

- Pearson Correlation Coefficient (*r*) value and experimental value interpretation.
  - *r* > 0.7: High similarity (low diversity).
  - 0.3 < *r* < 0.7: Moderate similarity and potential geographic diversity.
  - *r* < 0.3: Good geographic diversity.

Interpretation:

- Low/negative *r* indicates strong diversity (e.g., clouds over Site A but clear skies at Site B).
- Why negative *r* is rare: Solar/wind outputs rarely oppose each other completely—regional weather patterns typically create, at best, weak positive correlations.

3. Absolute Difference Evaluation

This analysis measures how much and how often power outputs diverge between sites, complementing correlation by quantifying magnitude gaps rather than just patterns.

Method:

- Hourly Absolute Differences:
  - Calculate |Site_A - Site_B| each hour (see Figure A for wind sites Atchison, MO vs. Howard, IA).
  - Figure B shows these differences over time, with peaks indicating significant mismatches.

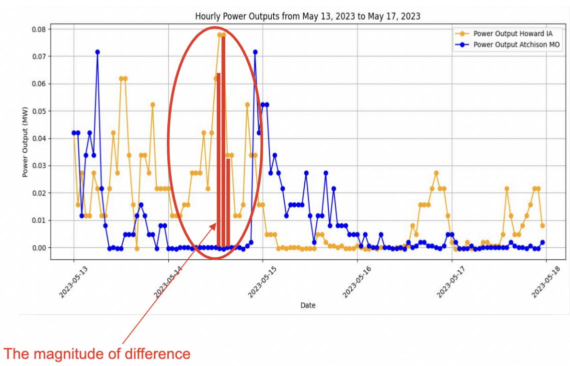

Figure A: Hourly power outputs for Atchison, MO (blue) vs. Howard, IA (orange), May 13–17, 2023. Gaps between lines represent absolute differences.

- Experimental Threshold Filtering:
  - Solar: Flag differences > 0.00005 MW (sensitive to small PV variations).
  - Wind: Flag differences > 0.05 MW (accounts for turbine scalability).

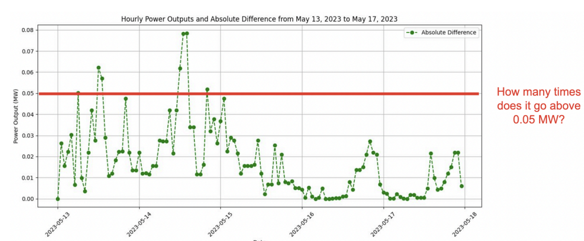

Figure B: Atchison, MO vs. Howard, IA Absolute differences (MW) over time. The red dashed line marks the wind experimental threshold (0.05 MW); peaks above it indicate significant mismatches.

Metrics Generated:

- Hours Above Threshold: Total count of meaningful deviations.
- Percentage of Hours Above Threshold

Interpretation:

- High values suggest strong geographic diversity:
    - Example: Figure B's peaks (>0.05 MW) show hours where one site significantly outperformed the other.
- Low values indicate consistent output parity and suggest balanced generation (even if correlation is weak).

Visualization and Output

The outputs from these three analyses are CSV files containing all key metrics and results, which can be used to recreate any scatter plots, bar charts, or other visualizations presented here (or to build new ones). Note that the packaged graphs themselves aren't included in the reproducible repository, but the CSVs are supplied for full rebuild capability. Additionally, interactive Plotly maps were developed to show how solar irradiance and wind speed change over daily, weekly, monthly, and yearly intervals, helping to intuitively explore site relationships and identify the most geographically diverse location pairs visually.

Key Visual Outputs:

- Production Analysis: CSVs with wind power mismatch metrics and solar output vs. clear-sky baseline; visualizations can include threshold-based or time series plots.
- Correlation Analysis: CSVs with solar and wind correlation metrics; potential visualizations include heatmaps and scatter plots showing correlation vs. distance.
- Absolute Difference Evaluation: CSVs detailing absolute difference distributions; visualizations can include boxplots, histograms, or time series for selected site pairs.

Data & Code Ownership

The entire reproducible analysis package (CSV data, scripts, notebooks) was developed for this study but remains the intellectual property of Ameren Innovation Center; Ameren retains all ownership rights.

Summary of Methodology

This tripartite approach—production mismatches, correlation strength, and difference magnitude—systematically identifies site pairs that maximize geographic diversity. The visual outputs transform complex metrics into actionable insights for:

- Grid reliability: Prioritizing pairs that minimize simultaneous generation drops.
- Efficiency: Allocating resources to sites with complementary output profiles.

RESULTS

1. Data Coverage and Reliability

Validating data completeness is essential before interpreting production or correlation metrics. We assessed coverage gaps using two heatmaps—one for wind sites from Meteostat (Figure 1a) and one for solar from the NSDRB (Figure 1b)—which quantify the percentage of available hourly data per site per year.

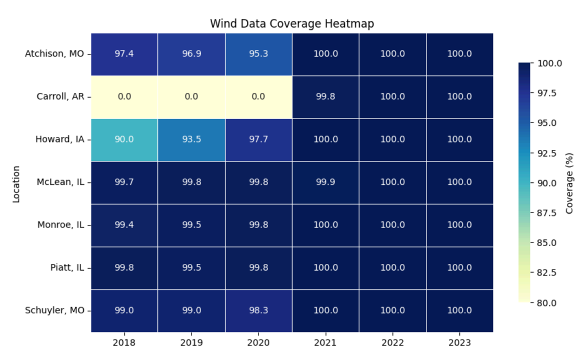

Figure 1a: Wind data coverage heatmap (2018–2023). Carroll, AR (2018–2020) shows critical gaps (0% coverage), while Howard, IA, and Illinois sites maintain ≥90% coverage. Most locations achieve ≥95% by 2021.

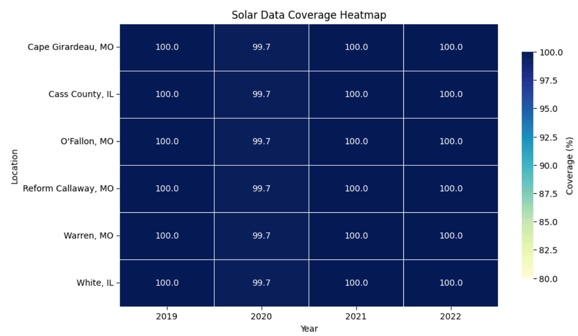

Figure 1b: Solar data coverage heatmap (2019–2022). All Missouri/Illinois sites maintain near-complete coverage (≥99.7%), ensuring high reliability for solar analyses.

For low-coverage wind sites (e.g., Carroll, AR), we:

- Reduced their weight in the visualizations (alpha transparency).
- Subject them to scrutiny in our tripartite approach and metrics generated for the affected years.

Conclusion: With ≥95% coverage for most of the site-years, the dataset supports robust analysis of production diversity, correlation, and absolute differences.

---

2. Production Diversity (Zero Output and Mismatch Analyses)

2.1 Wind Production Diversity

Wind farms frequently experience zero-output hours when wind speeds fall below operational thresholds. Our analysis quantifies how effectively site pairs compensate for these gaps by identifying hours where one site generates power while the other is idle.

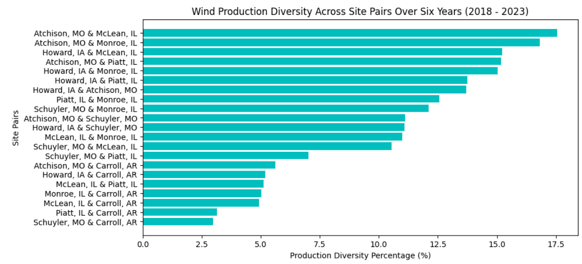

Figure 2a: Wind production diversity across site pairs (2018–2023). Atchison, MO & McLean, IL, lead with 17.55% mismatch hours (9,230 hours), demonstrating strong complementarity.

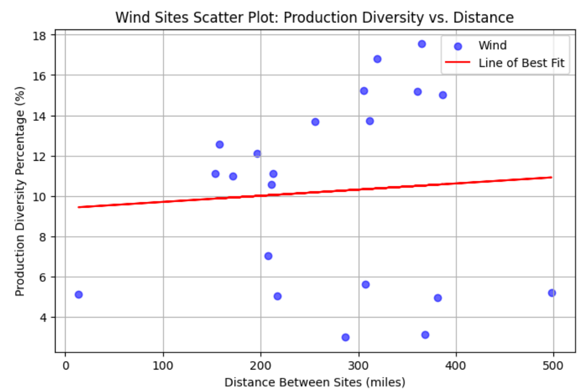

Figure 2b: Production diversity vs. distance between sites, with a line of best fit added to clarify the overall trend. Most lower outliers involve Carroll, AR (coverage < 5% in 2018–2020), which pulls down the diversity metrics.

Key Findings:

- Top performers:
  - Atchison, MO & McLean, IL (17.55% mismatch) and Atchison, MO & Monroe, IL (16.81%).
  - These pairs show consistent complementarity during low-wind events.
- Distance Relationship:
  - Diversity generally increases with distance (Figure 2b trendline).
  - Outlier Note: Pairs with Carroll, AR (e.g., Atchison, MO & Carroll,

AR) underperform due to poor data coverage, not a true lack of diversity.

Interpretation:

High mismatch percentages indicate sites that stabilize aggregate output—when one falters, the other often produces. This is particularly valuable for:

- Grid Reliability: Reducing coincident downtime.
- Resource Planning: Prioritizing pairs like Atchison-McLean for joint deployment.

2.2 Solar Production Diversity

Unlike wind farms, solar sites rarely drop to zero output during daylight hours. Instead, diversity emerges from mismatches in cloud cover and atmospheric conditions, measured by deviations from clear-sky baseline generation.

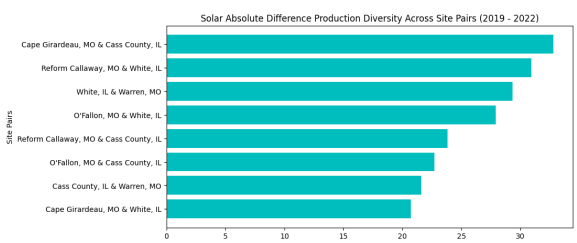

Figure 3a: Solar absolute difference diversity (2019–2022). Cape Girardeau, MO & Cass County, IL, lead with 28.49% mismatch hours (4,991 hours), where one site met clear-sky targets while the other underperformed.

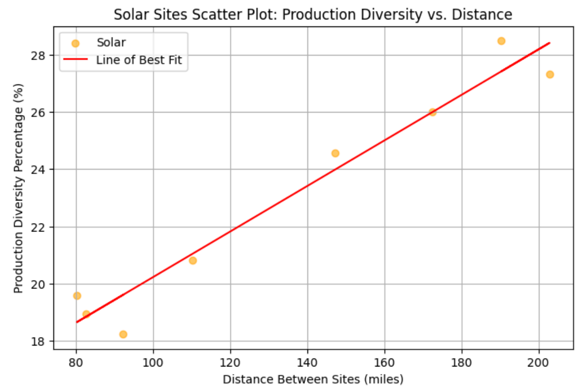

Figure 3b: Diversity vs. distance for solar pairs. The positive trend confirms that spacing sites extensively enhances mismatch frequency, with optimal gains at 150–200 miles.

Key Findings:

- Top performers:
  - Cape Girardeau, MO & Cass County, IL (28.49% mismatch).
  - Reform Callaway, MO & White, IL (27.31%).
  - These pairs show resilience to localized cloud cover.
- Distance Relationship:
  - Sites spaced closely show limited diversity, as they frequently experience similar cloud cover and irradiance conditions.
  - Diversity increases steadily with distance, peaking for pairs 150–200 miles apart (22–28% mismatch). Beyond 200 miles, gains a small plateau, suggesting an optimal range for Midwest solar deployments.

Interpretation:

High mismatch percentages (>20%) indicate:

- Microclimate Leverage: Cloud systems often miss one site while hitting another.

- Grid Benefits: Closer pairs see fewer mismatch hours, possibly limiting resilience benefits.

---

3. Correlation Analysis

Correlation quantifies how similarly two sites' power outputs fluctuate. Lower Pearson coefficients (r) indicate stronger geographic diversity, as sites are less likely to experience simultaneous production dips.

3.1 Correlation Results

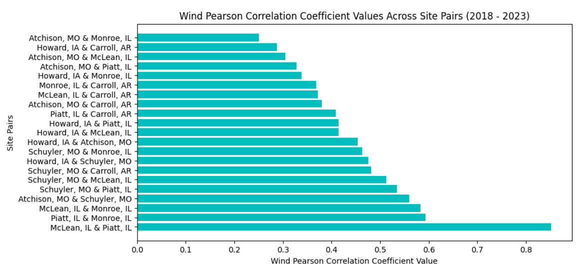

Figure 4a: Wind correlation coefficients (2018–2023). Atchison, MO & Monroe, IL show notably low correlation (r=0.25), making them ideal complements.

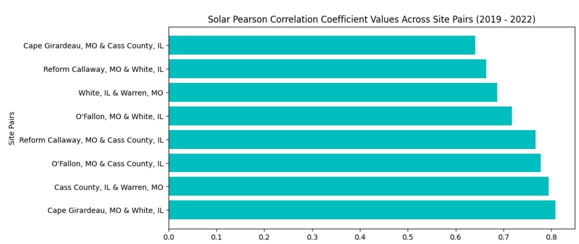

Figure 4b: Solar correlation coefficients (2019–2022). Cape Girardeau, MO & Cass County, IL (r=0.64) offer moderate diversity, while Cape Girardeau & White, IL (r=0.81) are nearly synchronized.

Key Findings:

- Wind:
  - High-diversity pairs: Atchison, MO & Monroe, IL (r=0.25) and Howard, IA & Carroll, AR (r=0.28) (Up to be discarded due to Carroll's data coverage gaps).
  - Low-diversity pairs: McLean, IL & Piatt, IL (r=0.85) and Piatt, IL & Monroe, IL (r=0.59).
- Solar:
  - Relatively high-diversity pairs: Cape Girardeau, MO & Cass County, IL (r=0.64) and Reform Callaway, MO & White, IL (r=0.66).
  - Relatively low-diversity pairs: Cass County, IL & Warren, MO (r=0.79) and Cape Girardeau, MO & White, IL (r=0.81).

3.2 Correlation vs. Distance

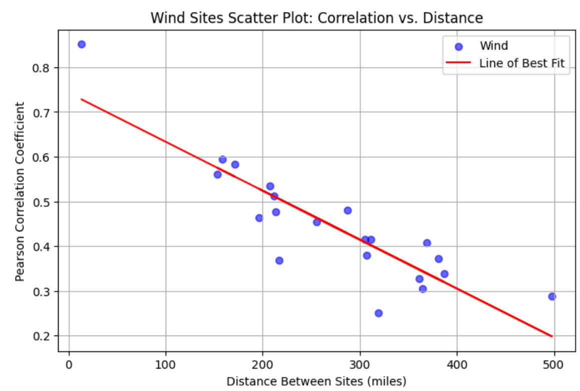

Figure 5a: Wind correlation declines sharply as distance increases.

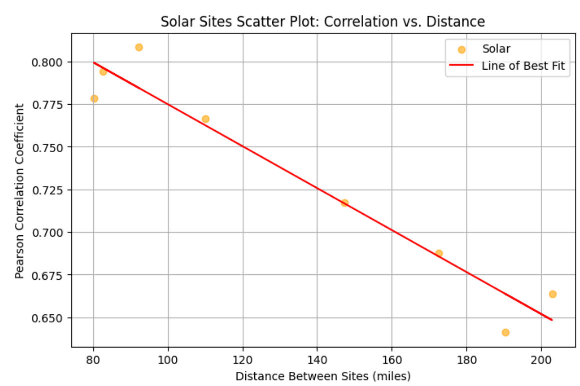

Figure 5b: Solar correlation declines moderately as distance increases.

Interpretation:

- Wind:

- ○ Strong distance effect: Wind patterns vary significantly between spaced-out sites.
- ○ Exception: Local terrain (e.g., river valleys or plains) can override trends.

- Solar:
  - ○ Cloud cover variability drives moderate correlations even at shorter distances.

---

4. Absolute Difference Evaluation

This analysis identifies how frequently and dramatically two sites' power outputs diverge—a critical measure of their ability to compensate for each other's fluctuations.

4.1 Hours Above Threshold

We quantify how often absolute power output differences exceed experimental thresholds:

- Wind: >0.05 MW
- Solar: >0.00005 MW

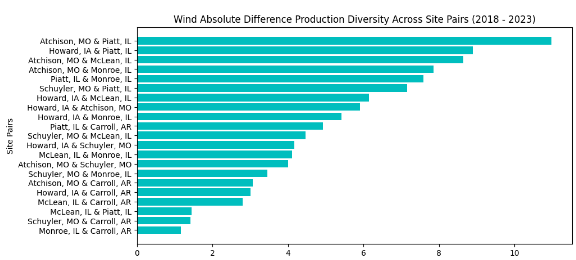

Figure 6a: Wind absolute difference diversity (2018–2023). Atchison, MO & Piatt, IL lead with 10.97% of hours (5,769) exceeding the threshold—ideal for offsetting low-wind events.

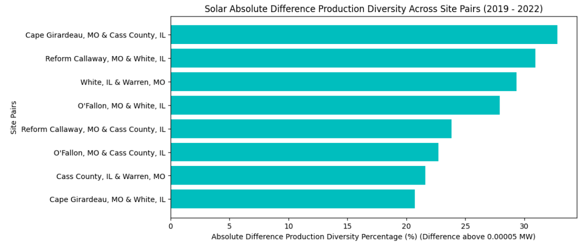

Figure 6b: Solar absolute difference diversity (2019–2022). Cape Girardeau, MO & Cass County, IL, show the highest variability (17.61% of hours), demonstrating cloud-cover resilience.

Key Findings:

- Wind:
  - ○ High-diversity pairs: Atchison, MO & Piatt, IL, and Howard, IA & Piatt, IL.
  - ○ Low-diversity pairs: McLean, IL & Piatt, IL, and Schuyler, MO & Carroll, AR (Low coverage pair).
- Solar:
  - ○ Relatively high-diversity pairs: Cape Girardeau, MO & Cass County, IL, and Reform Callaway, MO & White, IL.
  - ○ Relatively low-diversity pairs: Cass County, IL & Warren, MO, and Cape Girardeau, MO & White, IL.

Distance Relationship:

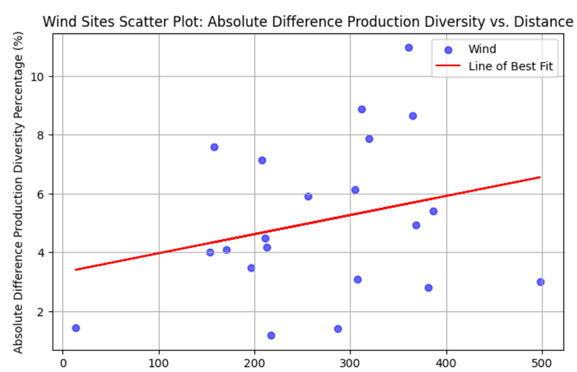

Figure 7a (Wind): Diversity has a general upward trend, but is scattered along with outlier values of low coverage pairs (Carroll, AK).

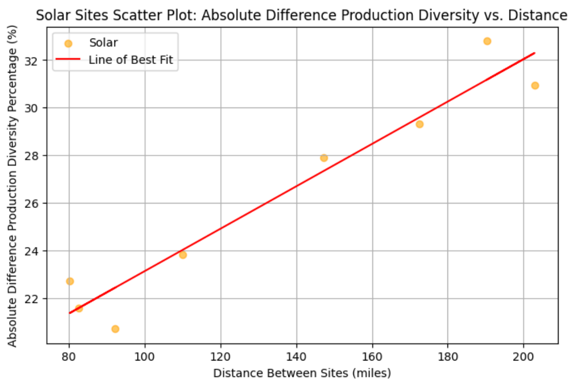

Figure 7b (Solar): Strong positive trend, with optimal gains at 150–200 miles (25–35% threshold hours).

Interpretation:

- Wind:
  - Atchison, MO & Piatt, IL (10.97% threshold hours) show strong complementary generation – when one underperforms, the other compensates by >0.05 MW ~11% of the time
  - Optimal spacing: >200 miles, especially across state lines.
- Solar:
  - Cape Girardeau, MO & Cass County, IL (17.61% threshold hours) demonstrate cloud-cover resilience, with one site offsetting the other by >0.00005 MW ~18% of daylight hours.
  - Optimal spacing: 150-200 mile spacing, across state lines.

Summary of Results

Our analysis confirms that geographic diversity is a powerful tool for stabilizing renewable energy generation. Wind farm pairs like Atchison, MO, and Monroe, IL, with their low correlation (r=0.25) and high production mismatch (15.2% of hours), demonstrate how strategic siting can create complementary generation profiles. Similarly, solar pairs such as Cape Girardeau, MO, and Cass County, IL, achieve significant production offsets (17.61% threshold hours), effectively mitigating localized cloud cover impacts. These high-performing pairs contrast sharply with closely clustered sites like McLean, IL, and Piatt, IL (r=0.85), which exhibit nearly synchronized output patterns and minimal diversification benefits. Energy planners can now identify optimal site pairings that maximize grid reliability by systematically evaluating production mismatches, correlation coefficients, and absolute differences. This tri-metric approach provides utilities like Ameren Missouri with data-driven evidence to advocate for strategically distributed renewable assets, particularly across state lines where weather patterns diverge most significantly. The results underscore a critical opportunity: Renewable portfolios across diverse geographies can reduce reliance on fossil fuel backups while maintaining more consistent power delivery.

DISCUSSION AND FUTURE WORK

Discussion

Interpretation of Findings

1. Variability Reduction in Select Pairs
   Several site pairings exhibit low correlation and high mismatch hours, implying they do not slump simultaneously. Notably, Atchison, MO & Monroe, IL (wind) and Cape Girardeau, MO & Cass County, IL (solar) are prime examples of geographic diversity. This mitigates sudden drops in total output, potentially reducing reliance on costly reserves.

2. Geographic Distance: Helpful but Not Absolute
   While distance correlates negatively with output similarity, local climate zones can override pure distance. Some solar pairs

100–150 miles apart remain moderately correlated; conversely, pairs 300–400 miles apart can remain surprisingly in sync if they share similar weather patterns.

3. Diversity vs. Integration Challenges
   Large absolute differences can minimize joint lulls, but they might produce sharper combined ramps. Operators must balance the benefit of offsetting dips against the cost of volatile or "peaky" aggregates.

Theoretical Contributions

1. Empirical Methodology for Site Complementarity
   This study provides a structured, tri-metric framework by merging zero-output mismatch analysis, correlation, and absolute difference. Previously, much analysis occurred at a broad system scale (Diakov, 2012) or hinged on single metrics like correlation or step changes.

2. Foundation for Aggregate Pair Analysis
   Although the current approach focuses on mismatch frequencies and correlation, it sets the stage for advanced aggregate metrics, such as standard deviation of combined outputs, coefficient of variation, and aggregated low-output hours, indicating how pairs behave as a unit.

Practical Implications

1. Data-Driven Site Selection
   Utility planners can systematically rank potential site pairs by mismatch percentages, correlation coefficients, and absolute differences. Instead of broad climate assumptions, they obtain empirical synergy indicators for each pair.

2. Informed Renewable Energy Policy
   Out-of-state renewable projects may overcome local regulatory resistance if data shows that such sites effectively offset in-state shortfalls. This supports cross-border expansions and the rethinking of state-centric restrictions that hamper grid stability.

3. Infrastructure and Grid Planning
   Where synergy is high, additional transmission lines or storage might be cost-effective. Conversely, pairs with minimal synergy do not merit robust interconnection unless additional factors (e.g., local load) justify it.

Limitations

1. Grid Integration
   The study does not incorporate transmission constraints, real-time load matching, or economic dispatch modeling. Thus, feasibility remains partially theoretical until grid operators account for the costs of connecting these distant sites.

2. Historical Weather Dependence
   The analysis relies on 2018–2023 data. Future climate patterns (e.g., increasingly frequent extreme weather) could shift synergy levels. Regular reassessment with updated data is needed.

3. Scope: Wind & Solar
   The tri-metric approach is valid for major renewables but does not yet include hydro, geothermal, or energy storage. Expanding the method to multi-resource synergy would deepen its applicability.

Future Work

1. Efficiency Analysis Using Aggregate Pair Analysis

Evaluating Site Pair Stability Using Aggregate Standard Deviation, Coefficient of Variation, and Low-Output Hours:

- If Std_Aggregate < each site's std dev, pairing reduces volatility.
- If CV_Aggregate < site CVs, the combined output stabilizes more effectively.

- If Low_Output_Hours_Aggregate < that of each site individually, the pair significantly cuts blackout risks.

2. Advanced Correlation Metrics

Spearman Rank or Mutual Information could detect non-linear or more complex dependencies. Where wind speed or solar irradiance forms nonlinear relationships, these measures might expose complementary patterns overlooked by Pearson's r.

3. Refined "Three-Pronged" Approach

Alongside absolute difference, consider net difference (ensuring consistent coverage) and efficiency (aggregated total production). Even high diversity can be unhelpful if the pair lacks adequate overall generation.

4. Real-Time Data Integration

Moving from historic data to live monitoring or frequent updates would allow dynamic site pairing to reflect seasonal or even weekly weather shifts, optimizing scheduling and grid dispatch.

5. Global Solar Irradiance Calculation

Springer defines Total Radiation as the "sum of beam and diffuse radiation," and Global Radiation as "the sum of direct, diffuse, and reflected radiation" [13]. Incorporating Global Radiation into our solar power estimates would account for ground-reflected irradiance and better capture real-world panel performance. In this study, however, we did not use Global Radiation because we lacked data on panel tilt angles and surface albedo needed to model reflected components, so we approximated solar output using only Direct Normal Irradiance (DNI) + Diffuse Horizontal Irradiance (DHI) or Total Radiation.

6. Ongoing Package Configurability

The existing analysis packages for wind and solar can gain:

- Aggregate Pair Analysis Modules—to display aggregated standard deviations, CVs, and low-output hours.
- Flexible Correlation Options—(Pearson, Spearman, etc.).
- Enhanced Visualization, interactive dashboards for stakeholders, and policy briefings.

CONCLUSION

The analyses presented in this study confirm that geographic diversity plays a pivotal role in improving renewable energy reliability and overall generation stability. By employing a tri-metric approach, this research offers a structured framework for identifying which site pairings are most likely to offset each other's downturns and deliver a more continuous energy supply. Notably, certain wind pairs (e.g., Atchison, MO & Monroe, IL) demonstrate strong mismatch hours and low correlation, indicating true complementarity. In the solar domain, Cape Girardeau, MO & Cass County, IL, exemplify how sites 100–200 miles apart can produce markedly different generation patterns that collectively smooth power output.

The implications for utilities and policymakers are extensive. First, local or state-bound restrictions may reduce the full potential benefits of renewables if high-potential out-of-state sites are barred or discouraged. Second, a data-driven tri-metric methodology allows for transparent ranking and selection of site pairs, ensuring decisions are rooted in empirical evidence rather than broad assumptions about climate zones. Third, the methodology can be expanded beyond wind and solar to include hydro, geothermal, or even hybrid resources, as long as reliable hourly (or sub-hourly) data are available.

Looking forward, integrating real-time data—or at least frequently updated data streams—would allow continuous optimization of site pairing and power scheduling. Moreover, advanced correlation measures and aggregate pair analysis can refine the framework further, potentially leading to a robust tool for grid operators. Ultimately, these approaches can strengthen the argument for widely distributed renewable projects, reduce the risk of simultaneous low-output periods, and help utilities cost-effectively meet sustainability goals.

This research provides a concrete, reproducible methodology to evaluate the reliability and complementarity of wind and solar site pairings. As utilities and regulators confront the twin challenges of climate variability and policy constraints, this tri-metric approach offers a practical tool to guide siting decisions that balance resilience, efficiency, and regulatory acceptance.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions of several individuals whose efforts and guidance were invaluable to this project. In particular, we thank Sulay Ahmed and Samuel Gerstein, former Ameren Innovation Center interns, for their early work on the initial wind and solar analyses. We also extend our sincere appreciation to Amanda C. Barbieri (Senior Manager, Renewable Origination & Business Development) and Mahendra Mishra (Director, Renewable Business Development & Acquisitions) for their sponsorship and expert guidance on solar and wind renewables, respectively, throughout this study. This research was led by the current Ameren Innovation Center Data Science intern and lead author, Dulf Vincent Genis, under the management of Shil Prasenjit and Kevin Ficek of the Ameren Innovation Center. Their collective support made this tri-metric evaluation possible.

REFERENCES

[1] V. Diakov, "The value of geographic diversity of wind and solar: Stochastic geometry approach," presented at the 2012 World Renewable Energy Forum. [Online]. Available: https://www.nrel.gov/docs/fy12osti/54707.pdf

[2] M. Fedkin and J. A. Dutton, "1.2 Efficiency of Conversion," in EME 812: Utility Solar Power and Concentration, Dutton Institute. [Online]. Available: https://www.e-education.psu.edu/eme812/node/4

[3] Y. Ghorbani, "A diversified approach to the green energy transition: Beyond decarb-centrism," J. Cleaner Prod., vol. 434, p. 140414, 2024, DOI:10.1016/j.jclepro.2023.140414.

[4] C. Gómez-Quiles, "The value of wind resource geographic diversity for wind farm profitability," IEEE Trans. Power Syst., vol. 27, no. 4, pp. 2074–2083, 2012. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6203391

[5] M. Handschy, "Reduction of wind power variability through geographic diversity," in Variable Renewable Energy and the Electricity Grid, Chapter 12, 2014. [Online]. Available: https://arxiv.org/pdf/1608.06257

[6] D. Kasper and J. A. Dutton, "Wind Energy and Power Calculations," in EM SC 470: Applied Sustainability in Contemporary Culture, Dutton Institute. [Online]. Available: https://www.e-education.psu.edu/emsc297/node/649

[7] Kent State University Libraries, "SPSS tutorials: Pearson correlation." [Online]. Available: https://libguides.library.kent.edu/spss/pearsoncorr

[8] C. Lamprecht, "Meteostat: The Weather's Record Keeper." [Online]. Available: https://meteostat.net/en/

[9] B. Marion, "Clear sky irradiance year-to-year variations and trends: Preprint," Nat. Renew. Energy Lab., 2020. [Online]. Available: https://www.nrel.gov/docs/fy21osti/75923.pdf

[10] A. D. Mills, "Implications of geographic diversity for short-term variability and predictability of solar power," Lawrence Berkeley Nat. Lab., 2010. [Online]. Available: https://www.osti.gov/servlets/purl/1196768

[11] National Renewable Energy Laboratory, "National Solar Radiation Database (NSRDB)." [Online]. Available: https://nsrdb.nrel.gov/

[12] pandas development team, "pandas.DataFrame.corr," pandas. [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html

[13] G. N. Tiwari and A. T. Shyam, Handbook of Solar Energy. Springer, 2016. [Online]. Available: https://link.springer.com/book/10.1007/978-981-10-0807-8