ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
23 October 2025

REVISED
26 March 2026

ACCEPTED FOR PUBLICATION
14 April 2026

PUBLISHED
28 April 2026

LETTER

Factors shaping the siting of utility-scale solar and wind projects in the United States

Grace C Wu[1,*], Yohan Min[1], Ranjit Deshmukh[1,2], Paloma Cartwright[2], Joseph DeCesaro[2], Daniel Kerstan[2], Desik Somasundaram[2] and Henry Strecker[2]

[1] Environmental Studies, University of California Santa Barbara, Santa Barbara, CA 93106, United States of America
[2] Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA 93106, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: gracecwu@ucsb.edu

Keywords: onshore wind, solar PV, siting, land use

Supplementary material for this article is available online

Abstract

The location of a wind or solar project is one of the most consequential decisions a renewable energy developer can make when planning a new project, as it shapes numerous aspects of project economics and local impacts on ecosystems and host communities. This study examines key factors influencing the siting of utility-scale solar and wind projects in the contiguous United States using multiple statistical and machine learning models and a robust sampling approach. We find that while proximity to transmission lines or substations and existing projects is critical for both technologies, solar locations are more flexible and are shaped by a greater diversity of factors including higher population density, greater road accessibility, and lower ecological impact. In contrast, wind locations are primarily driven by wind resource quality and agricultural land use. There are notable regional nuances in these national trends, with some variables like population density having greater effects in the Midwest. While individual models emphasize different predictors and regional patterns, the ensemble reveals consistent tendencies that increase confidence in interpretation. Lastly, the probabilities of wind and solar being sited in disadvantaged (DAC) and non-DAC areas vary across regions, suggesting that the Midwest and Northeast in particular may see disproportionately less development pressure from solar and wind projects in disadvantaged community census tracts.

1. Introduction

The rate of development of utility-scale solar and wind projects must increase several-fold to achieve decarbonization targets [1], such that large areas will be required for renewable generation [2, 3]. Equally consequential, however, is where these projects are built: siting decisions influence everything from plant profitability and interconnection costs to social acceptance and ecological impact [4–7]. Anticipating which areas are likely to attract development—by identifying the factors that drive siting—can support renewable energy, transmission, and land-use planning, guide mitigation of environmental and social impacts, and help address barriers to deployment.

Most site suitability studies adopt a multi-criteria decision analysis approach within a geographic information system, relying largely on expert opinions or industry standards of technical, economic, and environmental suitability [8, 9]. Only a handful of studies have attempted to empirically and systematically examine siting factors for solar or wind development using historical data and statistical methods with the aim of identifying the most important drivers of successful siting outcomes—whether in the United States (U.S.) or in other countries [10–17]. Among these, machine learning approaches such as random forest, gradient boosting and those developed and largely applied in ecological contexts (e.g. Maximum Entropy) have been used [11–13, 16, 17]. Traditional statistical approaches such as multivariate regression (e.g. logistic) have also been used in other studies to predict future locations of wind farms in the U.S. [10, 15] and to understand the role that environmental justice indicators have played in solar and wind project siting in the U.S. [14].

In this study, we advance the existing US literature [2, 14, 15] by systematically comparing results from multiple machine learning and statistical-based approaches and seek to examine not just global (national-scale) relationships [2, 14] between siting factors and siting outcomes, but also how those relationships may vary regionally within the U.S. We use a more comprehensive set of siting factors compared to other studies that examined both wind and solar—including substation, roads, renewable portfolio standards which were omitted in [14] and land use land cover and socio-demographic factors which were omitted in [2]. Lastly, instead of using presence/-absence of development at the grid-cell level [11, 15] or the census tract level [14] as in previous studies, we generated a balanced sample of presence and pseudo-absence data using each project as a data point and randomly sampled pseudo-absence locations from areas that meet minimum techno-economic suitability criteria for development. This robust approach ensures that data points are independent (vs treating abutting grid-cells within the same power plant as separate data points) and allows quantification of the most important factors driving development among only techno-economically feasible locations [18] (vs including all land area).

In this study, our objective was to (1) understand technical, socio-economic, environmental, and political drivers of utility-scale solar and wind siting in the U.S. and how the influence of key factors varies regionally, (2) predict areas of likely future development so as to be able to assess their potential environmental impacts and social implications, (3) inform the design of more equitable and lower impact solar and wind build-out by understanding regional differences in siting probabilities in designated disadvantaged communities.

A comprehensive review of the literature of studies on the suitability of solar and wind sites guided the selection of the technical, environmental, and socio-economic factors we examined (table 1, supplementary information (SI) figures S1–S3). For technical criteria, we include electricity generation potential (i.e. capacity factors); slope; distance to nearest substation, transmission line, and road. Environmental factors capture both land cover and ecological sensitivity, with the latter measured through an aggregate score that incorporates features such as wetlands and habitats for endangered or threatened species. Socio-economic and political factors span a wide range, from state-level renewable portfolio standard targets to land acquisition costs, population density, and demographic indicators such as poverty, minority share, and unemployment. Lastly, we include a spatial lag variable, which measures the extent to which project siting is influenced by the presence of nearby projects. Including this variable accounts for spatial autocorrelation and allows us to quantify the extent to which solar or wind projects cluster.

We employed four machine learning and statistical methods—random forest, XGBoost (gradient boosting), logistic regression, and lasso regression—and compared variable importance and predictive performance. Additionally, we conducted a random slope regression to examine how the relationships between key factors and siting outcomes vary across geographic regions. To explore one possible social implication of continued siting practices, we used the trained predictive models to compare the probabilities of siting utility-scale solar and wind project development in disadvantaged and non-disadvantaged communities.

2. Results

2.1. Key siting factors for utility-scale solar PV and onshore wind projects

To identify the most important siting factors for solar PV and wind, we generated a comprehensive dataset of locations of utility-scale solar PV and wind projects developed in the U.S. from 2017 to 2023 and their site characteristics. We applied a paired presence and pseudo-absence sampling approach, following the guidance of [19], where 10 replicates of pseudo-absence locations were randomly selected from areas without existing solar or wind development but that meet minimum site suitability criteria—ensuring these locations met minimal requirements to be technically, economically, or environmentally feasible for utility-scale development using the maps developed in [18] and drawing on methods developed in [2]. To enhance robustness, we generated 10 replicates of randomly sampled pseudo-absence locations (figure 1), which is an optimal number of replicates for our sample size (200–500 presence locations) to achieve reliable estimates using both regression and machine learning models [19]. This presence and pseudo-absence dataset was used in random forest, XGBoost, logistic regression, and lasso regression models. K-fold cross validation was used to evaluate model performance. Overall, results reveal multiple and distinct significant siting factors for wind and solar (figure 2). Variable importance values differ between models, but the ensemble or average across models, however, reveals consistent tendencies in overall variable rankings (figure 2).

Significant siting factors for solar PV include proximity to infrastructure such as roads, transmission lines, and substations, higher population density, non-forested areas, flatter terrain, spatial lag, and areas with lower environmental impacts (figure 2(A), SI figure S4). These factors indicate that solar projects are more likely to be developed in areas with high infrastructure accessibility, closer to load centers, and lower environmental impacts

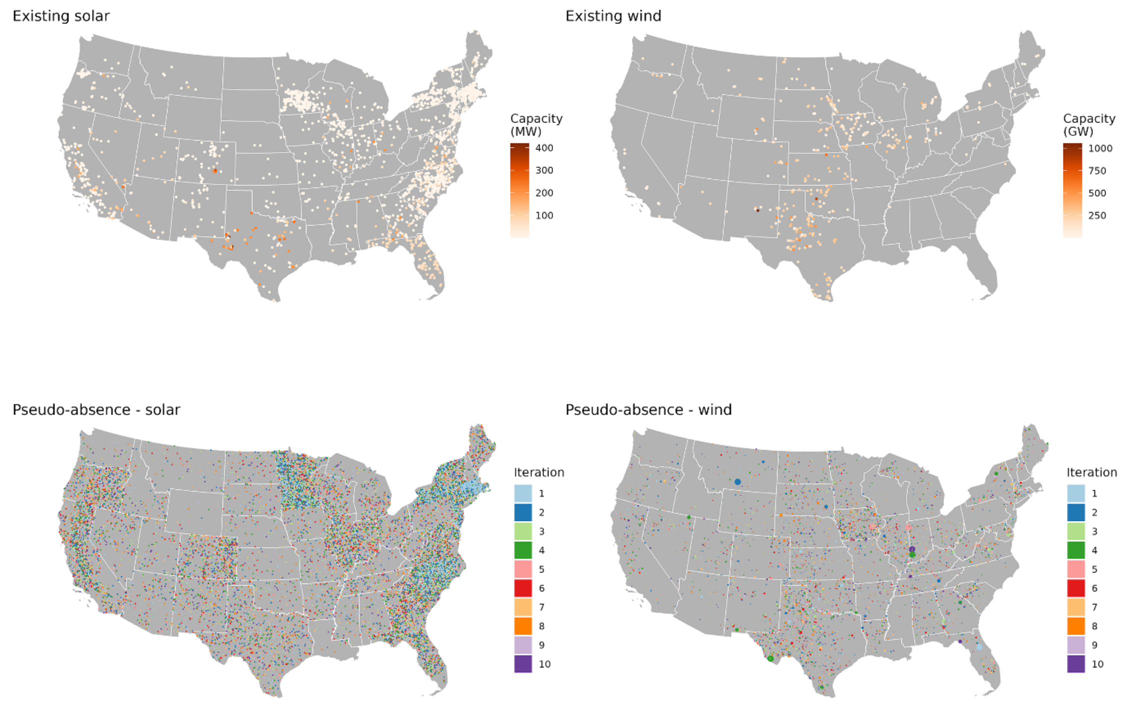

Figure 1. Locations of utility-scale solar and onshore wind projects in operation by the end of 2023, with colors indicating each project's nameplate capacity, and 10 randomly generated replicates of pseudo-absence locations for each technology with equivalent sample size and project footprint area within each state (indicated by the size of the polygons).

(e.g. ecological sensitivities and forested land). Note that highly densely populated areas (i.e. metropolitan areas) were excluded when randomly sampling for pseudo-absence locations since it would be infeasible to site utility-scale solar or wind power in such areas. For wind projects, critical siting variables include agricultural land, capacity factor (a measure of wind power output based on wind speed profiles), spatial lag, proximity to transmission lines, forest land, and slope (figure 2(B), SI figure S5). Wind projects are more commonly located in areas that are windier, have favorable land cover (agricultural and non-forested), closer to transmission lines, and on more sloped terrain, reflecting the importance of site-specific wind resource quality and infrastructure accessibility. Land acquisition cost is also a statistically significant factor, but the effect size is small.

While the above variables reflect the average result across the four models, there is notable disagreement across models for certain variables. For example, there is disagreement between lasso and XGboost reporting differing levels of importance for key variables like spatial lag and regional dummy variables (West and Mountain-West) for wind. GLM ranked agricultural land use eighth for wind projects, while it is the most important variable in the other three models (figure 2(B)). For solar, lasso ranked environmental score and substation distance much lower than the other three models (figure 2(A)). These differences illustrate why an ensemble of models provides a more reliable basis for interpretation than any single approach.

The spatial lag variable is significant for both solar and wind in the GLM (logistic) regression and has high variable importance scores in the two machine learning models—a strong indication that project locations have a tendency to be spatially clustered (figure 2). The odds ratios of the regional dummy variables, which are all significant in the logistic regression model, indicate regions in which there is greater or less than proportionate siting of wind or solar relative to the potential based on the number of randomly chosen pseudo-absence locations. For example, there are disproportionately more wind projects in the West (compared to the reference, Mountain-West), and less in the South, Midwest, and Texas. For solar, the trend is similar—there are disproportionately more solar projects in the Mountain-West (and West, though not significant) compared to other regions. Land cover type is more strongly predictive of wind sites than solar sites, with disproportionately more wind projects located on agricultural land and disproportionately fewer projects on shrubland, grassland, forest land, and riparian land compared to the reference, other land cover types (i.e. snow-ice, exotic vegetation, and open water). Solar only shows significantly lower odds of being located in forested and riparian areas compared to other land cover types. None of the social variables, e.g. percent minority or unemployment, were significantly

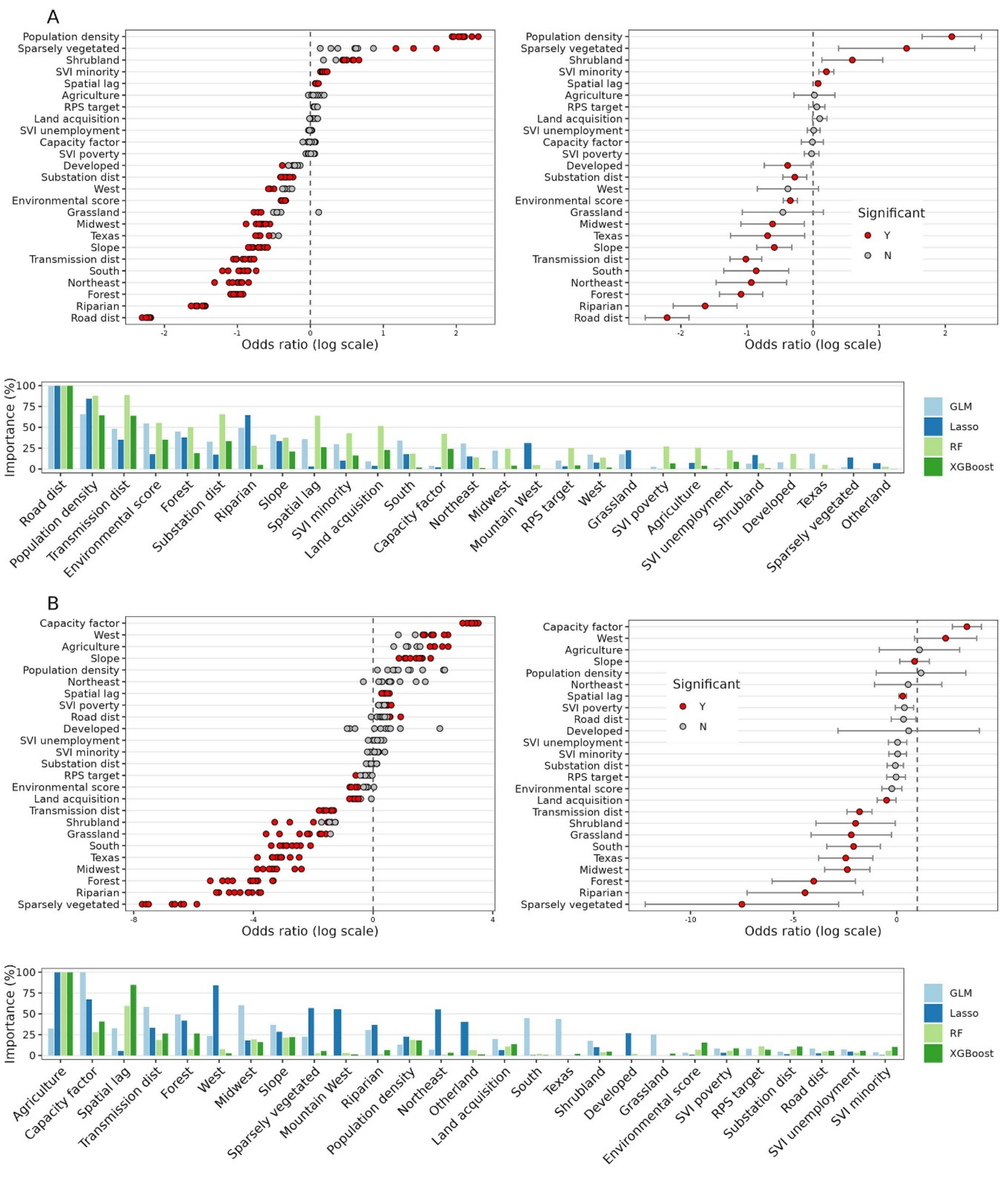

Figure 2. Variable importance for utility-scale solar (A) and onshore wind (B) projects. Estimates of predictor coefficients expressed as log-odds ratios are derived from logistic regression models (generalized linear model), with each point indicating the result from one of 10 randomly generated sets of pseudo-absence locations. Red points indicate significant predictor variables, grey points indicate non-significant predictor variables, and grey error bars show 95% confidence intervals. Dotted black horizontal lines represent an odds ratio baseline of no effect. The single set of log-odds ratio values in the right-side figures is from the pseudo-absence set with the median model performance (chosen using Akaike's Information Criterion). Bar plots show the predictor importance (%), with variables ordered along the x-axis by average percentage importance across four models: logistic generalized linear model (GLM), lasso regression, random forest (RF), and XGBoost).

correlated with wind or solar project locations, nor was the state renewable portfolio standard target a significant predictor (figure 2).

2.2. Probability of new project development

Using models trained on existing projects' site characteristics, we estimated the probability of siting a wind and solar project on each 250 meter grid cell in the contiguous U.S. by replacing the historical transmission dataset (transmission dataset published in 2016 before all projects in the sample were operational) with a dataset of current and planned transmission infrastructure (year 2023). Suitable locations for utility-scale solar and wind projects are identified using site suitability maps from the Power of Place—National study (see SI figure S7 for spatial distribution of suitable areas). This process generated siting probability surfaces for each technology and model (figure 3(A), see SI figure S8 for predictions, including those non-suitable areas). GLM and lasso,

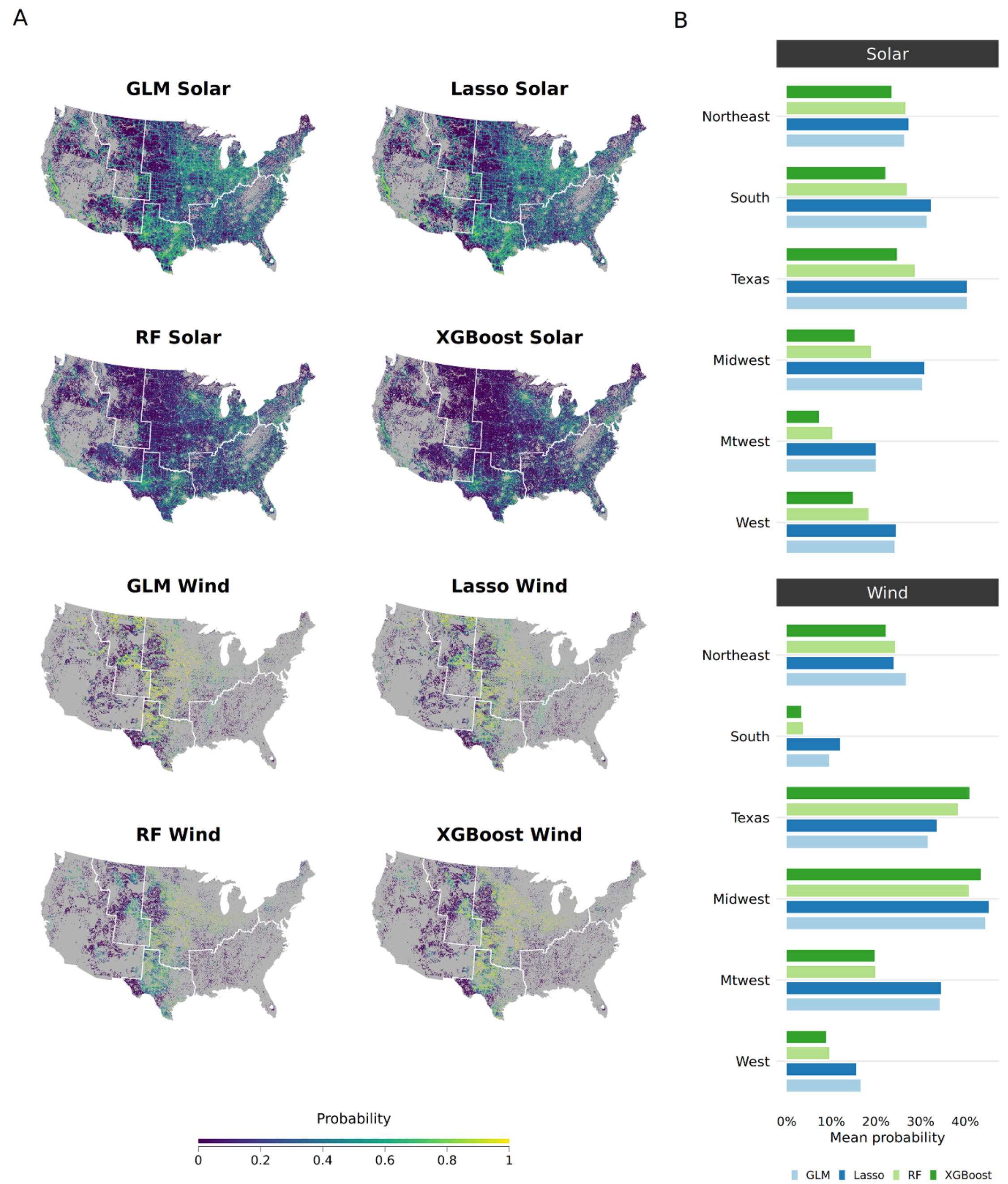

Figure 3. Model predictions for utility-scale solar and onshore wind project locations in the contiguous U.S. from four models: (logistic) generalized linear model (GLM), lasso regression, random forest (RF), and XGBoost) (A). Bar charts show the mean siting probability score in each region for each technology (B).

both statistical approaches, produced similar prediction patterns, showing more spatial heterogeneity, with higher probabilities for project locations across the US. In contrast, random forest and XGBoost, both machine learning approaches, yielded similar spatial patterns but are more exclusive, limiting high-probability areas to fewer regions that closely follow the general locations of existing projects (figure 1).

The differences between machine learning and statistical methods are more apparent for solar than they are for wind. Receiver operating characteristic (ROC) scores, which is a measure of a model's ability to distinguish project from non-project locations over the range of possible probability cutoffs, were higher for the machine learning models when applied to solar data (random forest and XGBoost: 0.88–0.91) than the statistical models (GLM and lasso: 0.81-0.82; SI figure S9). A similar pattern is seen for sensitivity (the share of true project locations correctly identified) and specificity (the share of true non-project locations correctly identified) metrics—for solar, machine learning 0.79–0.83 sensitivity and 0.83–0.86 specificity, compared with about 0.7-0.74 sensitivity and 0.76-0.80 specificity for GLM and lasso. In contrast, differences among methods were smaller for wind and in general, the models performed better (SI figure S9). ROC values were 0.92–0.95 across all models and sensitivity values (0.89–0.94) and specificity values (0.89–0.94) were also very similar. These results suggest that model choice is more consequential for solar predictive performance. For both technologies, random forest and XGBoost models provide more precise predictions for existing project locations, but may reflect a greater tendency to overfit compared to statistical approaches.

Regional trends show that the average solar siting probability scores are highest in Texas, Northeast, and South, with the Mountain-West and West having the lowest average scores (figure 3(B)). There is some disagreement across methods with respect to the relative ranking of the top three regions, but consensus that Mountain-West has the lowest average siting probability score. For onshore wind, there is generally more agreement across the methods in having consensus on the top two high siting probability regions (Midwest, followed by Texas) and the lowest two probability regions (South, followed by West). Average siting probability scores for solar show less variation across regions, particularly for the statistical approaches, compared to wind, which shows mean probabilities in the most favorable regions for siting wind being several times greater than the lowest.

2.3. Regional differences in the effects of key factors

Having identified the key siting factors using the four national-scale models, we examined whether the effect sizes of these siting factors varied regionally. To do so, we ran a random slope regression with regions as the groups, allowing the effect of key factors to differ across regions. We chose six major U.S. regions rather than individual states, as the sample size in many states was too limited or entirely absent to reliably estimate state-level variation. We chose to examine three of the most important variables for each technology: population density, distance to transmission lines, and environmental score for solar; and capacity factor, spatial lag, and transmission distance for wind. Although distance to road was an important factor for solar (figure 5), it poses less of a constraint on development than transmission access or environmental considerations. Results show that for solar, the relationship between population density and project location is strongest in the Midwest and weakest in Texas (figure 4(A)), with more projects located in areas with higher population density outside of metropolitan areas. Proximity to transmission lines shows a similar regional pattern: the projects in the South tend to favor locations closer to transmission lines, while the Northeast and Western regions had a weaker effect of transmission proximity. The environmental score is an aggregate of dozens of datasets that represent a location's ecological characteristics, with higher scores indicating higher ecological sensitivity and habitat importance. While solar projects across the U.S. tend to be sited in locations with lower ecological value, this effect is greatest in the Northeast and lowest in the Mountain-West and the South (figure 4(A)).

For wind projects, those located in the South tend to be sited in places with the highest capacity factor, whereas projects in the West and Mountain-West show the lowest sensitivity to capacity factors (figure 4(B)). Spatial clustering of wind projects, represented by the spatial lag variable, is the most pronounced in the Northeast, indicating a stronger tendency for wind projects to be located near each other. Proximity to transmission lines is influential for wind project siting in the Northeast and the South, while transmission proximity's effect is less pronounced in the West and Mountain-West.

Overall, projects in the West and Mountain-West appear to follow similar siting trends for wind, but diverge for solar. Otherwise, individual regions display distinct patterns of relative factor importance, suggesting that siting outcomes are a complex interplay of several local and regionally specific criteria that conventional siting criteria may not adequately capture (e.g. cultural, political, or institutional norms). For explanations of the most salient regional trends that draw on contextual information, please see the discussion section.

2.4. Siting probability differences between disadvantaged and non-disadvantaged communities

Using the siting probability surfaces and tract-level identification of disadvantaged communities (DAC; defined by socioeconomic, environmental, and health vulnerabilities identified using the climate and economic justice screening tool (CEJST)) we determine whether siting likelihood or success differs between designated disadvantaged (DAC) and non-disadvantaged (non-DAC) communities and whether differences are consistent across regions (see SI figure S10 for DAC designation).

On the whole, we find that siting probabilities for wind and solar are higher in non-DAC communities in regions where there are significant differences (figure 5, see SI figure S11 for the average probabilities). Specifically, there are significantly higher solar PV siting probabilities in non-DAC census tracts than in DAC census tracts in the Midwest, the South, and the Northeast, with the most dramatic differences being in the Northeast (figure 5(B)). The differences are less consistent for wind siting probabilities, with only the Midwest showing significantly higher siting probabilities in non-DAC communities across all modeling approaches. The higher wind siting probabilities are significant for Texas, the South, the Northeast, and the Mountain West only for predictions made using the machine learning approaches. The West is the only region where

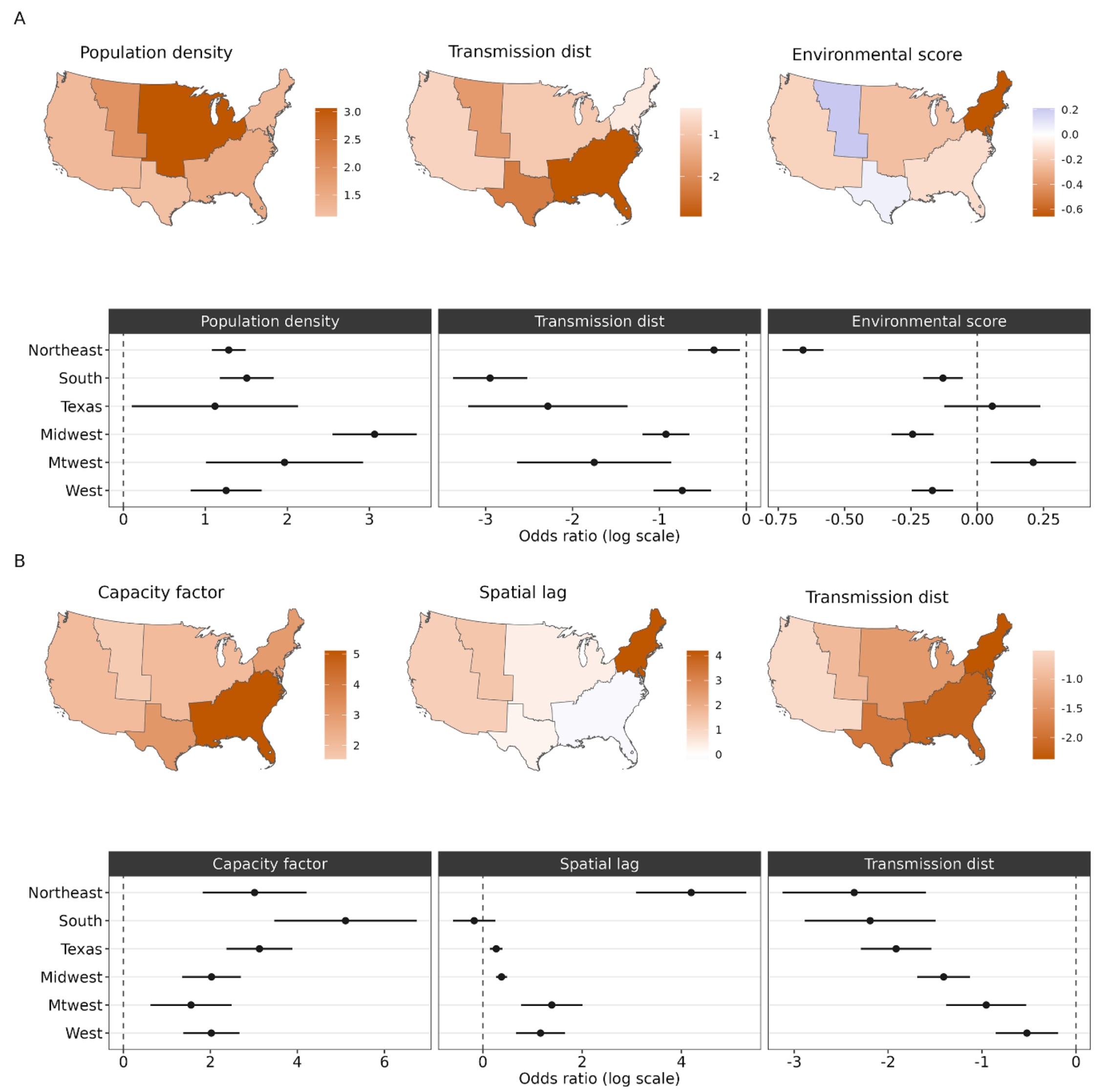

Figure 4. Regional effects, expressed as log odds ratios (logistic regression combined fixed and random coefficients), of significant predictors for utility-scale solar (A) and land-based wind (B) projects. The estimates represent the second-order relative impacts of predictors across regions after accounting for their global effects. Warmer (red) colors indicate higher predictor impacts, while cooler (blue) colors represent lower impacts. For predictors with negative coefficients (e.g. transmission lines and environmental impacts), the color scale was adjusted to reflect the direction of their influence. Sample sizes in each region is as follows (S = solar, W = wind): Northeast - 710 (S), 30 (W); Midwest - 622 (S), 198 (W), West - 376 (S), 53 (W); South - 679 (S), 31 (W); Texas - 93 (S), 106 (W); Mtwest - 92 (S), 28 (W)..

the siting probabilities for both wind and solar are not statistically different between non-DAC and DAC communities. These regional trends are explained by both the spatial distribution of DAC vs non-DAC areas within regions and how it coincides with high or low siting probability areas, as well as by the spatial distribution across regions and how certain regions like the Northeast simply have fewer DAC compared to non-DAC census tracts, whereas in the West the proportion of DAC to non-DAC is more balanced (figure 5(A)).

3. Discussion

Using a diversity of machine learning and statistical models, we identified the most important factors that determine the locations of utility-scale wind and solar PV projects in the contiguous U.S. and used these models to estimate siting probability for other suitable land areas in the U.S.

3.1. Resource quality drives wind siting, while solar siting favors proximity to load and infrastructure within suitable areas

Our results generally agree with previous studies that renewable resource quality (wind speeds) matters the most for wind siting [11, 14, 17, 20]. However we find that solar siting is less sensitive to marginal variations in solar radiation and instead shows a significantly positive association with population density within technically suitable areas for utility-scale solar. While this appears to contrast with O'Shaughnessy *et al* [14], who found a negative correlation with population

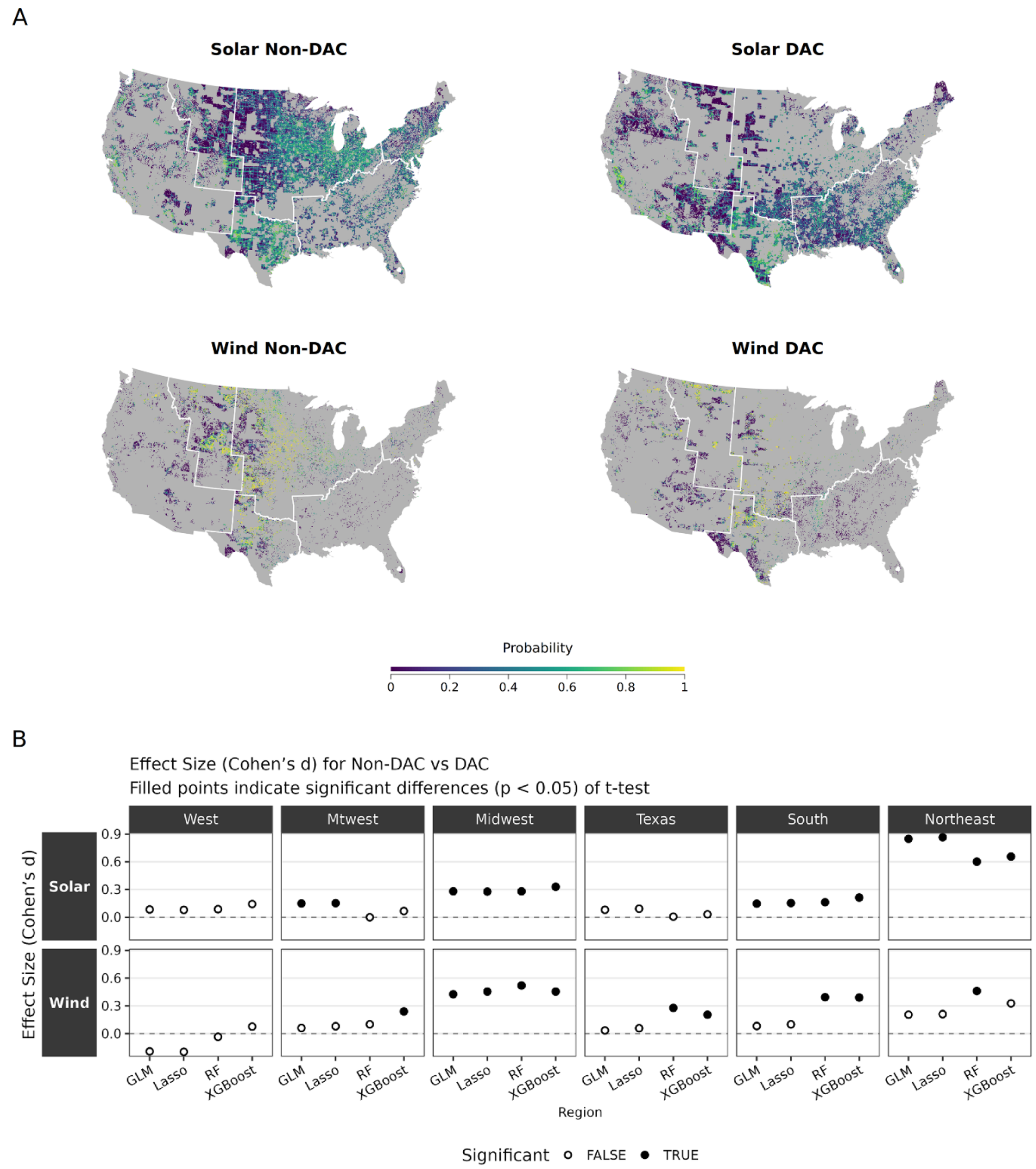

Figure 5. Predicted probabilities of utility-scale solar and land-based wind project development locations for disadvantaged and non-disadvantaged communities. (A) Siting probability surfaces generated using a Generalized Linear Model (GLM). Higher values indicate higher siting probabilities of onshore wind or solar project development. (B) Standardized effect sizes (Cohen's (d) comparing adoption rates between Non-DAC and DAC communities across six U.S. regions (West, Mountain-West, Midwest, Texas, South, and Northeast). Positive values indicate higher siting probability in non-DAC areas, whereas negative values indicate higher siting probability in DAC areas. Results are presented for four distinct modeling frameworks: generalized linear model (GLM), lasso regression, random forest (RF), and XGBoost. Filled points denote statistically significant disparities between DAC and non-DAC areas ($p < 0.05$, t-test)..

density, this discrepancy is likely a function of study design. By treating all U.S. census tracts at potential sites, previous regressions inevitably find that solar is sited away from the highest density metropolitan areas where utility-scale wind or solar development is physically infeasible. In contrast, our use of a suitability mask for pseudo-absence sampling specifically compares existing projects against available and developable land. The positive association we find between population density and solar siting suggests that developers preferentially site solar projects in the 'peri-urban' fringe or more settled rural areas rather than in remote, sparsely populated regions. This likely reflects a strategic preference for siting close to load centers, existing road and transmission networks, which reduces project costs. These factors carry more weight for solar due to its higher siting flexibility compared to wind. This interpretation is supported by the regional results, which show the weakest correlation with population density in Texas. A large portion of Texas's solar capacity is located in Competitive Renewable Energy Zones (CREZ), which are concentrated in the least populated west and panhandle counties. CREZs are connected to major load centers by 3500 miles of high voltage transmission lines, which were built specifically to move large amounts of remotely sited renewable electricity in the west to cities in the east [21]. Regional variation in the importance of wind capacity factors may be due to the fact that the Midwest and Texas have higher quality wind resources compared to the West, and thus, developers are able to choose sites with higher capacity factors relative to the minimum economically viable threshold (30%).

3.2. Contrasting land use and ecological impacts in solar vs wind siting

We find that land use and land cover type are important considerations for both wind and solar development, with each technology preferring or avoiding specific land types. Other studies generally find, and our results agree, that solar projects in the U.S. are sited disproportionately more on sparsely vegetated land [13, 14] and disproportionately less on forestland and riparian areas. Wind is disproportionately sited on agricultural land, similarly to results from O'Shaughnessy *et al* 14 and Wimhurst *et al* 2023, and tends to avoid shrubland, forestland, and riparian areas. These results suggests that wind project siting in agricultural land is a more compatible land use since farming can continue largely as usual between turbines, whereas development on other land cover types results in land conversion. Solar projects tend to be sited in areas with lower ecological or conservation value, with very little regional variation in the relative importance of this siting factor. However, environmental considerations were not a significant factor for wind farm locations. These results are also reasonable. These findings align with the differing variables of importance for solar and wind. Solar development is less constrained by site-specific resource quality compared to wind and thus, solar can preferentially avoid areas of high conservation value. Wind projects, in contrast, are more dependent on limited high-quality wind resource areas, reducing the influence of environmental considerations.

3.3. Transmission accessibility drives siting, with solar favoring substations and wind favoring power lines

We find that proximity to the nearest transmission line is a critical siting criteria for both technologies, which is consistent with other studies for solar [13] and wind [15, 20], but inconsistent with O'Shaughnessy *et al* [14]. Proximity to the nearest substation was significant only for solar and was slightly less important compared to transmission. This is surprising given that utilities and ISOs may prefer that utility-scale projects interconnect to an existing substation via an upgrade (i.e. adding a new bay) over interconnecting to an existing transmission line via a line tap since the latter requires a new switching station, outages, and land use permits. However, these factors differ depending on the size of the project or the geography as line taps may be the faster and less expensive solution when there is excess line capacity and when the nearest substation is too far away. Our results are consistent with this general rule of thumb for interconnection—wind farms tend to be located in remote areas where the transmission network is relatively sparse, making line taps with switchyards more practical, whereas solar is more flexible and can be more optimally sited closer to load centers and substations. In terms of the regional effects, solar projects in the South and wind projects in the Midwest tend to be located closer to transmission lines. This trend is intuitive for wind since the Midwest is the 'wind-belt', offering many more windy areas to choose from compared to other regions, thus providing more opportunities to develop windy sites closer to existing transmission lines. This trend for solar is also consistent with the fact that solar development in the south is relatively recent and that the best sites close to transmission are the ones that were the first to be developed.

3.4. Solar projects display more siting flexibility compared to wind projects

In general, a comparison of variable importance results for wind and solar reveals the relative flexibility of solar siting. Unlike wind projects, which need to be sited in windy locations, solar PV can generate electricity cost competitively even in areas with lower solar radiation and thus can be sited in ways that reduce infrastructure and mitigation costs—specifically, reducing substation and road proximity and ecological impacts. There is lower spatial variability in solar irradiance than in wind speeds, thus generation output differences across solar sites are not as significant over the same area compared to wind sites. Although each technology has its own unique set of siting criteria, we find significant spatial autocorrelation, or project clustering for both technologies. The tendency for wind projects to be sited close to one another suggests that inter-project global blockage effects, in which the presence of turbines reduces the average wind speeds on a larger scale, beyond local wind obstruction for downwind turbines (wake effects) [22], may continue to be or grow in importance as an issue for the wind industry. Regional variable relationships with siting outcomes suggest that states that are politically more favorable to clean energy development (West and Mountain-West) have disproportionately more solar and wind development compared to other regions. Although this may seem counterintuitive given that Texas has the highest installed onshore wind capacity in the U.S., followed by several Midwestern states, western states have developed a greater fraction of their overall wind potential than these other wind-abundant states.

3.5. Model ensembles improve confidence in siting drivers and probability maps reveal technology-specific spatial patterns

Individual machine learning and statistical models emphasize different variables and regional patterns, underscoring the uncertainty of relying on a single model. The ensemble, however, revealed consistent tendencies that strengthen confidence in interpretation. For instance, while GLM assigned low importance to agricultural land use, other three models ranked it highly, supporting the conclusion that agriculture is a strong predictor of wind project locations. For siting probability estimates, machine learning methods, random forest and XGBoost, outperform classical statistical methods, GLM (logistic) and lasso regression, in predicting existing project locations. While machine learning models exhibit high accuracy, their tendency to overfit current patterns may limit their utility for forecasting future development in unexplored areas. If the objective is to identify promising locations for future projects, GLM and lasso regressions provide valuable insights, as their greater inclusivity better accommodates planning scenarios beyond current practices and trends in existing projects. Siting probability surfaces for wind mirror actual wind project siting trends to date—with high probability appearing in states with the most wind installed. Solar siting probability maps point to Texas, the Midwest, and the South as areas with high probability of development. This contrasts with the observed concentration of solar projects in a number of western states such as California and Nevada. However, the solar probability surfaces are aligned with variable importance results that reveal higher odds of siting in the West and Mountain-West than would otherwise be the case given all other non-regional siting factors.

3.6. Siting probability surfaces reveal possible differing development pressure between disadvantaged and non-disadvantaged communities

The observed disparities in siting probabilities between DAC and non-DAC census tracts—most notably in the Northeast, Midwest, and South—raise questions about the equitable distribution of renewable energy development impacts. Whether communities perceive or receive net benefits or net burdens from a solar or wind project in their vicinity is context-dependent and project-specific. Our results suggest that in regions like the Northeast and Midwest, DACs may face lower 'development pressure' than their non-DAC neighbors. While this may protect DACs from land use pressures (e.g. converted farmland), it could potentially exclude them from the economic benefits (e.g. tax revenue, employment). Whatever the outcome—net-benefit or net-burden—this suggests that the impacts may be concentrated within non-DAC communities in these regions. In the West, we observe no significant differences in development pressure between DAC and non-DAC. These probability surfaces provide a means to generate an 'equity baseline' by which to measure whether certain communities are being disproportionately or equitably prioritized as host communities for utility-scale wind or solar projects.

3.7. Evolving institutional constraints and siting applications

While our study offers valuable insights into the siting patterns of utility-scale solar and wind projects in the U.S., it reflects a snapshot of project development over a six-year period, during and after which industry dynamics and regulatory conditions have continued and will continue to evolve. This 2017-2023 time frame likely captures a transitional phase of market maturation. While many prime sites (e.g. high resource quality and close to existing grid infrastructure) were developed in earlier years for wind, developers are now navigating more constrained opportunities. However, this has also been tempered by the technological advancements in wind—larger rotor diameters and higher hub heights—that have allowed development in areas previously considered marginal [23]. In this same time period, the solar industry experienced a major inflection point in which solar expanded outside of the desert southwest and became a nationally competitive technology. Decision criteria for site suitability will continue to be influenced by diverse individual land management decisions and regulatory frameworks, which complicate land acquisition for project development. Local ordinances often impose restrictions, such as stipulations or outright bans, on utility-scale solar and wind projects, significantly reducing the availability of land [6]. The number of counties with solar and wind zoning ordinances have been surging, with 25% of counties (as of 2026) with some restriction against wind or solar siting, up from 15% in 2023 [24]. Our results suggest that many high probability areas are located in counties with siting ordinances, e.g. with Oregon and Ohio standing out as having many counties with both high siting probability scores and zoning ordinances.

Other institutional constraints such as interconnection availability, cost, and time to interconnect also play important roles that were not considered in this study. The interconnection queue data through 2023 suggest that the independent system operator (ISO) regions with the most solar capacity in the queue is the West, followed by PJM (Eastern U.S.), California (CAISO), MISO (Midwest), and then Southeast, and ERCOT (Texas) [25]. This geographic trend contrasts with our finding that the

Mountain-West has the lowest average siting probability score, but agrees with our random-effects results showing that there are disproportionately more wind and solar projects in the West and disproportionately fewer in the South, Midwest, and Texas. This suggests that while the queues are very long and backlogged in some ISOs like the West, these regions also have disproportionately more existing projects and suggests that they have the political will, institutional support, or community acceptance to clear projects in the queue.

As land scarcity intensifies, understanding the interplay of institutional constraints, social opposition, and land market dynamics becomes crucial, as these factors may drive up land acquisition costs and hinder renewable energy expansion. Future research that incorporates these variables can identify more reliable suitable areas and enhance the accuracy of siting projections, ultimately supporting realistic decarbonization pathways. Results from this study can enable decision-makers to anticipate where transmission investments, streamlined permitting, and environmental or social safeguards may be needed to ensure sustainable and equitable outcomes. Moreover, the approaches and maps developed here can facilitate spatial planning of renewable energy infrastructure in ways that are aligned with state and national energy growth projects and ultimately enable more proactive, coordinated, and conflict-sensitive renewable energy development.

4. Methods

4.1. Siting variable selection and preprocessing

Using U.S.-specific solar PV and onshore wind site suitability studies [26–28] and existing literature that examined empirical relationships between siting decisions and siting criteria [10–15], we selected 13 siting factors as predictor variables, categorized into the following: renewable resource (capacity factors), technical (slope, distance to existing substations, transmission lines, and roads), socioeconomic (land acquisition cost, environmental sensitivity, seven land use and land cover classes, percentage living below poverty, percentage minority, percentage unemployed), political (renewable portfolio standard) (table 1). Land use and land cover variables were treated as dummy variables. Additionally, we included six regional dummy variables (Northeast, Midwest, West, South, Texas, Mountain-west) to examine coarse geographic variation in relationships. Although we only include one dataset representing environmental sensitivity, this single dataset is an aggregation and scoring of hundreds of individual datasets representing e.g. wetlands, critical habitat for threatened and endangered species, bat habitat, intact habitat (areas of critical environmental concern, high integrity grasslands, priority conservation areas), focal bird habitat (sage grouse, whopping crane) [29]; see table 1 for links to more methodological details. Additionally, we developed spatial lag layers using kernel density estimation based on the locations of existing projects to include as a predictor variable and account for spatial autocorrelation or non-stationarity.

To more accurately represent of solar resource quality in terms of electricity generation potential, we calculated solar PV capacity factors in the system advisory model (SAM). We used the following assumptions for single axis tracking systems: ground mount single-axis tracking configuration, DC/AC Ratio = 1.4, average annual soiling losses = 5%, module mismatch losses = 2%, diode and connection losses = 0.5%, DC wiring losses = 1%, AC wiring losses = 1%, availability losses = 1%, degradation = 0.5% per year.

Although additional factors could have been considered [30], we attempted to balance comprehensiveness with preserving statistical power, given our sampling approach, which treats each power plant project as a single sample instead of analyzing individual wind turbines or census blocks. Sparse datasets with a large number of predictor variables relative to the sample size is prone to over-parameterization and overfitting. All data layers were preprocessed and standardized into 250-meter resolution rasters with a consistent coordinate reference system for alignment.

To assess regional and community disparities as a example use of the siting probability surfaces, we used tract-level identification of disadvantaged communities (DAC) using the CEJST developed by the White House U.S. Council on Environmental Quality [31]. CEJST defines DACs based on socioeconomic, environmental, and health vulnerabilities. However, because CEJST relies on 2010 U.S. Census tract boundaries and associated tract IDs, we used a dataset [31] that maps these 2010 tract boundaries to their updated 2021 equivalents.

4.2. Solar and wind power plant sampling

Our response variable consisted of a roughly balanced sample of existing onshore wind and utility-scale solar PV power plants and pseudo-absence locations. For both wind and solar projects, we included only projects built in or after 2017 to maintain temporal alignment with transmission and substation infrastructure data, which are from 2016, and other siting criteria datasets (e.g. land use land cover, social vulnerability indicators) (table 1).

We used the U.S. large-scale solar photovoltaic database (USPVDB) [39] dataset for solar plants (containing operational year and nameplate capacity) with all power plant locations (reported in latitude and longitude) to select only projects with nameplate capacity greater than or equal to 1 MW and operational year from 2017 to 2023.

Table 1. Siting factors evaluated for utility-scale solar and wind.

Category	Dataset name	Source	Description	Data type/ resolution
Renewable resource	Solar PV capacity factors	NREL System advisory model [32]	Point locations of estimated annual average capacity factors for fixed tilt solar PV calculated using SAM (see text for calculation assumptions)	CSV with geographic coordinates/10 km
Renewable resource	Onshore wind capacity factors	NREL ReV model [33]	Point locations of simulated wind speeds and estimated annual average capacity factors of quality wind resource areas in the U.S.	CSV with geographic coordinates/2 km
Technical	Slope	CGIAR	Calculated slope in percentage from SRTM digital elevation model—Resampled 250 m SRTM 90 m Digital Elevation Database v4.1	Raster/250 m
Technical	Electric substations	Homeland Infrastructure Foundation Level Database (HIFLD) from 2017	Substations are considered facilities and equipment that switch, transform, or regulate electric power at voltages equal to, or greater than, 110 kV	Shapefile
Technical	Transmission lines	Homeland Infrastructure Foundation Level Database (HIFLD) from 2017	Transmission lines with voltages between 110 kV to 765 kV	Shapefile
Technical	Planned substation and transmission lines	Hitachi Velocity Suite	This is a proprietary dataset that is only available to subscribers	Shapefile
Technical	Roads	Open Street Map through Geofabrik [34]	Downloaded in 2022; selected the following types of roads: 'motorway', 'motorway_link', 'trunk', 'trunk_link', 'primary', 'primary_link', 'secondary', 'secondary_link'	Shapefile
Environmental	Environmental sensitivity	The Nature Conservancy [18]	Environmental sensitivity scoring system based on data categorizations from the Power of Place—National and Power of Place West studies [2]	Raster/250 m
Environmental	Land use and land cover type: agricultural, forested, grassland, developed, shrubland, riparian, other	LANDFIRE (2016) [35]	Existing Vegetation Type (EVT) 'EVT_PHYS' field	Raster

(Continued.)

Table 1. (Continued.)

Political	Renewable Portfolio Standard or Target	Database of State Incentives for Renewables & Efficiency (NC Clean Energy Technology Center, 2022) [36]	Used most near-term RRS target	CSV
Socio-economic	Population density	ORNL Landscan [37]	Persons per km^2	Raster/1 km
Socio-economic	Land acquisition cost	Nolte (2020) [38]	Estimates of fair market value of private lands in the contiguous United States	Raster/1 km
Socio-economic	Percentage living below poverty	Social Vulnerability Index (SVI) (2018) from U.S. Centers for Disease Control [37]	Percentage of persons below poverty estimate, 2014-2018 ACS (using the metric,'EP_POV')	Shapefile
Socio-economic	Percentage minority	Social Vulnerability Index (SVI) from U.S. Centers for Disease Control (Centers for Disease Control, 2021) [37]	Percentage minority (all persons except white, non Hispanic) estimate, 2014-2018 ACS (using the metric, 'EP_MINRTY')	Shapefile
Socio-economic	Percentage unemployed	Social Vulnerability Index (SVI) from U.S. Centers for Disease Control (Centers for Disease Control, 2021) [37]	Unemployment Rate estimate, 2014-2018 ACS (using the metric,'EP_UNEMP')	Shapefile
Spatial clustering	Spatial lag	U.S. Large-Scale Solar Photovoltaic Database (USPVDB) [39] & United States Wind Turbine Database (USWTDB) [40]	Created based on the locations of existing projects	Raster/250 m

We relied on the United States Wind Turbine Database (USWTDB) v. 6.1.0 [40] dataset, filtered for projects with operational start year between 2017 and 2023 and with project installed capacity greater than or equal to 10 MW. Additionally, we excluded projects consisting of only a single turbine. For turbines with unknown project names, we visually inspected the locations of turbines and manually grouped turbines together to form projects based on reasonable proximity. We then generated wind farm project footprints in the form of polygons by first buffering each turbine point location by 500 m and then applying a convex hull algorithm on the unioned buffered locations. We also manually reassigned turbines as needed when visual inspection showed overlap of the occupied turbine area with another project's footprint. The total area of each project was calculated in order to generate area-equivalent pseudo-absence footprints.

To generate pseudo-absence locations and polygons, we first buffered the polygons of all existing solar and wind projects (not filtered by size or operational year) by 500 m for solar and 2000 m for wind. We then converted these to raster and masked out (or 'erased') the buffered existing project locations from a published site suitability map (Candidate Project Areas) for both solar PV and onshore wind from the Power of Place—National study [18]. Methods for generating the CPA dataset are described in detail in Wu *et al* [2]. Briefly, we generated these minimally suitable areas by applying thresholds for a comprehensive set of siting criteria (e.g. for wind: slope 20%; capacity factor less than 30%; no siting in legally protected areas, airports and landing strips, military areas, or areas prohibited by zoning ordinances [6]). The goal of this study was to understand among all the factors in places considered by developers, which factors are the most important and whether these factors differ regionally. Sampling outside of these minimally suitable areas, as other studies have done, may inflate the influence of certain siting characteristics. For example the effect of capacity factor may be inflated, since such an approach would introduce pseudo-absence samples that have even lower capacity factors than existing project locations, which further reduces the overlap between the two presence and absence distributions. However, we know that such locations are not economically viable locations for utility-scale wind and thus should not be considered an 'absence' in terms of developer preference. As a result, we believe that while no national model of techno-economical suitability will be perfect as it would need to be updated regularly to reflect changing conditions, we believe sampling within minimally suitable areas reduces bias in the results.

Using this site suitability raster with existing projects removed, we then randomly sampled the same number of pseudo-absence locations as there are existing solar and wind projects in our presence sample within each state and used the existing projects' land areas to generate land-area equivalent paired pseudo-absence polygons by buffering the points within each state. This balanced sampling of 'presence' and 'pseudo-absence' locations is recommended for generalized linear models as well as boosted regression trees and random forest, as per [19]. Barbet-Massin *et al* 2012 does not provide guidance about the sampling area of each location since species distribution modeling uses point locations, but we chose to maintain balance of project size. For states without existing projects, we selected 10 samples using the average polygon area for the respective technology within each state to ensure geographic representativeness, as per the recommendation to stratify sampling [19]. We bootstrapped subsequent analyses by repeating the above steps to generate 10 random sets of pseudo-absence locations. Given our sample of approximately 200–500 pseudo-absence locations, between 5 to 12 replicates are recommended to achieve optimal model performance [19].

To generate the model input dataset, we then used zonal statistics to calculate the average value (or majority value for categorical dummy variables) for each presence/existing and pseudo-absence location (polygon converted to raster) across all predictor variables.

4.3. Model assumptions and checks, evaluating model performance, and generating prediction surfaces

We ran a logistic regression after scaling variables/features and compared the model results of the 10 different datasets of the pseudo-absence locations and reported the estimated odds ratios of predictors across the 10 sets for sampling robust checks. We chose the pseudo-absence set with the median performance among the 10 sets using the Akaike information criterion for further analysis. We additionally ran lasso regression, random forest, and XGBoost models to understand sensitivity to model choice. All models were built in R [41] using the caret package [42]).

Method selection—Logistic regression (generalized linear regression) was chosen because it is commonly used in binary classification problems and performs reasonably well across a wide range of applications [15, 43]. Lasso regression was chosen for its ability to automate variable/feature selection and thus address multi-collinearity, which must be ex-ante addressed when using logistic regression [43]. Due to the limitations that come with meeting linearity and the non-multi-collinearity requirements with these two more conventional statistical learning methods, we also chose to run a decision tree approach—random forest—that is well suited to handle a large number of possibly non-linear or collinear variables/features given that it bootstraps both data selection and feature selection when generating decision trees. We complemented the random forest model with XGBoost, a tree-based algorithm designed to efficiently capture non-linear relationships in binary data. This approach is particularly well-suited for handling sparse data, as seen in solar and wind siting outcomes.

Variable importance—Variable importance was determined independently for each method. In logistic regression, it was based on the p-value (or the absolute value of the t-statistic). For scaled logistic regression and lasso regression models, variable importance corresponds to the value of the coefficient, with larger coefficients indicating greater importance. In random forest and XGBoost, it is measured as the average difference in prediction accuracy across all trees when the model was run with and without each variable; larger differences indicate greater variable importance [42].

Global model prediction—Using the best fitting model for each method, we generated prediction surfaces showing the probability (0 to 1) of observing a solar or wind power plant in each grid cell. However, instead of using the model to predict on the same set of historic data (variables in table 1), we generated a new raster stack by replacing the historic substation and transmission line locations with existing and planned substations and transmission lines. This produces probabilities for future siting occurrence that consider changing landscapes.

Random slope model—To explore the regional differences in siting criteria, we chose three of the most important predictors identified in our analysis for each technology. Using a random slope model (generalized linear mixed model using a logit link function), we estimated regional coefficients for the selected predictors, allowing for variability in their influence across regions (groups). Regions include West, Mountain-West, Midwest, Texas, South, and Northeast. Random slope models reveal how each predictor's impact varies across regions by allowing each region to have its own relationship between the predictor variables and the siting outcome. For example, factors that are highly significant in one region may play a lesser role in another. This approach provides insights into spatial heterogeneity and the localized effects of siting factors on project siting outcomes.

Acknowledgments

GCW received funding from the David H Smith Conservation Research Fellowship administered by the Society for Conservation Biology with funds from the Cedar Tree Foundation. We wish to thank Eric O'Shaughnessy, Ben Hoen, Joseph Rand, and Salma Elmallah for invaluable feedback on the early sets of results. We also wish to thank Joseph Fargione for feedback on the manuscript, Tamma Carleton for advice on statistical regressions, Christoph Nolte for assistance interpreting land acquisition cost data, Ciara Hovis for preprocessing the planned transmission and substation data, Emily Leslie for providing the wind and solar PV site suitability maps used in the Power of Place—National study, and Niklas Greissbaum for help with data downloads.

Data availability statement

The resulting geospatial data underlying figure 3 (prediction surfaces) are available through the following repository: https://doi.org/10.5281/zenodo.19686722 [44].

Supporting figures available at https://doi.org/10.1088/1748-9326/ae5faa/data1.

Code and data availability statement

All code to preprocess the variables, run the analysis, and generate the figures are available in the following repository: https://github.com/spatialClimateSolutions/energySiting_analysis. The following publicly available dashboard with interactive figures and webmaps allows users to examine the results in more detail: US wind and solar siting dashboard. Data for running the analysis are all publicly available via the links provided in table 1, unless otherwise indicated.

Author contributions

Grace C Wu 0000-0002-8290-119X
Conceptualization (lead), Data curation (equal), Formal analysis (equal), Methodology (lead), Project administration (equal), Validation (supporting), Visualization (supporting), Writing – original draft (lead), Writing – review & editing (lead)

Yohan Min 0000-0002-8001-4124
Conceptualization (supporting), Formal analysis (lead), Methodology (equal), Validation (equal), Visualization (lead), Writing – original draft (equal), Writing – review & editing (equal)

Ranjit Deshmukh 0000-0002-5593-675X
Conceptualization (supporting), Methodology (supporting), Project administration (equal), Supervision (supporting), Writing – review & editing (equal)

Paloma Cartwright 0000-0002-9556-4065
Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (supporting), Visualization (supporting), Writing – review & editing (supporting)

Joseph DeCesaro
Data curation (equal), Formal analysis (equal), Methodology (supporting), Visualization (supporting), Writing – review & editing (supporting)

Daniel Kerstan
Data curation (equal), Formal analysis (equal), Methodology (supporting), Visualization (supporting), Writing – review & editing (supporting)

Desik Somasundaram 0000-0002-1053-8280
Data curation (equal), Formal analysis (equal), Methodology (supporting), Visualization (supporting), Writing – review & editing (supporting)

Henry Strecker 0009-0001-4012-5209
Data curation (equal), Formal analysis (equal), Methodology (supporting), Writing – review & editing (supporting)

References

[1] Williams J H, Jones R A, Haley B, Kwok G, Hargreaves J, Farbes J and Torn M S 2021 Carbon-neutral pathways for the united states *AGU Adv.* **2** e2020AV000284

[2] Wu G C *et al* 2023 Minimizing habitat conflicts in meeting net-zero energy targets in the western United States *Proc. Natl Acad. Sci.* **120** e2204098120

[3] van de Ven D-J, Capellan-Peréz I, Arto I, Cazcarro I, de Castro C, Patel P and Gonzalez-Eguino M 2021 The potential land requirements and related land use change emissions of solar energy *Sci. Rep.* **11** 2907

[4] Susskind L, Chun J, Gant A, Hodgkins C, Cohen J and Lohmar S 2022 Sources of opposition to renewable energy projects in the United States *Energy Policy* **165** 112922

[5] Rand J and Hoen B 2017 Thirty years of North American wind energy acceptance research: what have we learned? *Energy Res. Soc. Sci.* **29** 135–48

[6] Lopez A, Cole W, Sergi B, Levine A, Carey J, Mangan C, Mai T, Williams T, Pinchuk P and Gu J 2023 Impact of siting ordinances on land availability for wind and solar development *Nat. Energy* **8** 1034–43

[7] Rand J, Strauss R, Gorman W, Seel J, Kemp J M, Jeong S, Robson D and Wiser R 2023 Queued Up: characteristics of power plants seeking transmission interconnection as of the end of 2022 (Lawrence Berkeley National Lab (LBNL)) *Technical Report* (available at: https://emp.lbl.gov/publications/queued-characteristics-power-plants-1)

[8] Lopez A, Mai T, Lantz E, Harrison-Atlas D, Williams T and Maclaurin G 2021 Land use and turbine technology influences on wind potential in the United States *Energy* **223** 120044

[9] Wang H-W, Dodd A and Ko Y 2022 Resolving the conflict of greens: a GIS-based and participatory least-conflict sitting framework for solar energy development in southwest Taiwan *Renew. Energy* **197** 879–92

[10] Mann D, Lant C and Schoof J 2012 Using map algebra to explain and project spatial patterns of wind energy development in Iowa *Appl. Geogr.* **34** 219–29

[11] Petrov A N and Wessling J M 2015 Utilization of machine-learning algorithms for wind turbine site suitability modeling in Iowa, USA *Wind Energy* **18** 713–27

[12] Lück L and Moser A 2018 Combining machine learning and multi criteria decision analysis modeling regulatory, economic and social influences on wind turbine allocation *2018 15th Int. Conf. on the European Energy Market (EEM)* pp 1–5 (available at: https://ieeexplore.ieee.org/document/8470016)

[13] Sun Y, Zhu D, Li Y, Wang R and Ma R 2023 Spatial modelling the location choice of large-scale solar photovoltaic power plants: application of interpretable machine learning techniques and the national inventory *Energy Convers. Manage.* **289** 117198

[14] O'Shaughnessy E, Wiser R, Hoen B, Rand J and Elmallah S 2022 Drivers and energy justice implications of renewable energy project siting in the United States *J. Environ. Policy Plan.* **0** 1–15

[15] Wimhurst J J, Greene J S and Koch J 2023 Predicting commercial wind farm site suitability in the conterminous United States using a logistic regression model *Appl. Energy* **352** 121880

[16] Tao L, Hayashi K, Gyeltshen S and Shimoyama Y 2025 Spatial assessment of utility-scale solar photovoltaic siting potential using machine learning approaches: a case study in Aichi prefecture, Japan *Appl. Energy* **383** 125329

[17] Sun Y, Li Y, Wang R and Ma R 2024 Modelling potential land suitability of large-scale wind energy development using explainable machine learning techniques: applications for China, USA and EU *Energy Convers. Manage.* **302** 118131

[18] Jones W, Leslie H, Fargione J, Hiltibran K, Cohen and Williams 2023 Power of place - national: excutive summary the nature conservancy *Technical Report* (available at: www.nature.org/content/dam/tnc/nature/en/documents/f/i/FINAL_TNC_Power_of_Place_National_Executive_Summary_5.28.25_Update.pdf)

[19] Barbet-Massin M, Jiguet F, Albert C H and Thuiller W 2012 Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* **3** 327–38

[20] Wimhurst J J and Greene J S 2023 Using logistic regression-cellular automate to project future sites for commercial wind energy development *Appl. Geogr.* **159** 103070

[21] Electric Reliability Council of Texas 2018 Report on existing and potential electric system constraints and needs (Electric Reliability Council of Texas (ERCOT)) *Technical Report*

[22] Meijer J, Steinfeld G, Vollmer L and Dörenkämper M 2024 The global blockage effect of a wind farm cluster - an LES study *J. Phys.: Conf. Ser.* **2767** 092093

[23] Wiser R *et al* 2024 Land-based wind market report: 2024 edn *Technical Report* (Lawrence Berkeley National Laboratory (LBNL))

[24] Weisse E and Bhat S 2026 Wind, solar power face a common foe - creative local governments (available at: www.usatoday.com/story/news/nation/2026/02/21/restrictions-wind-solar-energy-bans-setbacks-government/85952104007/)

[25] Rand J, Manderlink N, Gorman W, Wiser R, Seel J, Kemp J M, Jeong S and Kahrl F 2024 Queued Up: characteristics of power plants seeking transmission interconnection as of the end of 2023 Lawrence Berkeley National Lab (LBNL) *Technical Report* (available at: https://emp.lbl.gov/sites/default/files/2025-08/LBNL_Ix_Queue_Data_File_thru2023_v2.xlsx)

[26] Lopez A, Roberts B, Heimiller D, Blair N and Porro G 2012 U.S. renewable energy technical potentials: a GIS-based analysis *Technical Report* (National Renewable Energy Laboratory)

[27] Wu G C, Leslie E, Sawyerr O, Cameron D R, Brand E, Cohen B, Allen D, Ochoa M and Olson A 2020 Low-impact land use pathways to deep decarbonization of electricity *Environ. Res. Lett.* **15** 074044

[28] California Public Utilities Commission [CPUC] 2009 Renewable energy transmission initiative (RETI) phase 1B *Technical Report* (available at: https://energyarchive.ca.gov/reti/documents/index.html)
[29] Jones W, Leslie H, Fargione J, Hiltibran K, Cohen, and Williams Power of place - national - environmental and social impact score datasets 2023 (available at: https://zenodo.org/records/7878144#.ZFmRJ-xKjc8)
[30] Wimhurst J J, Nsude C C and Greene J S 2023 Standardizing the factors used in wind farm site suitability models: a review *Heliyon* **9** e15903
[31] Min Y 2024 Disadvantaged communities in the U.S (available at: https://zenodo.org/records/13381832)
[32] N R E L SAM photovoltaic models - system advisor Model - SAM 2024 (available at: https://sam.nrel.gov/photovoltaic.html)
[33] Maclaurin G, Grue N, Lopez A, Heimiller D, Rossol M, Buster G and Williams T 2021 The renewable energy potential (reV) model: a geospatial platform for technical potential and supply curve modeling (National Renewable Energy Laboratory (NREL)) *Technical Report* (available at: www.osti.gov/biblio/1563140)
[34] GmbH G Geofabrik download server 2024 (available at: https://download.geofabrik.de/)
[35] U S D A Full extent downloads | landfire 2024 (available at: https://landfire.gov/data/FullExtentDownloads)
[36] D S I R E Database of state incentives for renewables & efficiency® 2024 (available at: www.dsireusa.org/)
[37] C D C Social vulnerability index 2024 (available at: www.atsdr.cdc.gov/place-health/php/svi/index.html)
[38] Nolte C 2020 High-resolution land value maps reveal underestimation of conservation costs in the United States *Proc. Natl Acad. Sci.* **117** 29577–83
[39] Fujita K S, Ancona Z H, Kramer L A, Straka M M, Gautreau T E, Garrity C P, Robson D, Diffendorfer J J E and Hoen B 2024 United states large-scale solar photovoltaic database (ver. 2.0, August 2024) (available at: www.sciencebase.gov/catalog/item/66707f69d34e89718fa3f82f)
[40] Hoen B, Diffendorfer J J E, Rand J, Kramer L A, Garrity C P and Hunt H United states wind turbine database 2025 (available at: www.sciencebase.gov/catalog/item/57bdfd8fe4b03fd6b7df5ff9)
[41] R Core Team 2022 R: a language and environment for statistical computing (R Foundation for Statistical Computing) (available at: https://www.R-project.org)
[42] Kuhn M 2008 Building Predictive Models in R Using the caret Package *J. Stat. Softw.* **28** 1–26
[43] James G, Witten D, Hastie T and Tibshirani R 2022 *An Introduction to Statistical Learning With Applications in R* (*Springer Texts in Statistics*) 2nd edn (Springer) (https://doi.org/10.1080/24754269.2021.1980261)
[44] Wu G, Min Y, Deshmukh R, Paloma C, DeCesaro J, Kerstan D, Somasundaram D and Strecker H 2026 Wind and solar siting prediction surfaces for the United States *Zenodo* (https://doi.org/10.5281/zenodo.19686722)