A Novel Approach for Large-Scale Wind Energy Potential Assessment

Tao Dai[1,2], Corinne D. Scown[1,2,3,4]*

[1]Life-Cycle, Economics and Agronomy Division, Joint BioEnergy Institute; Emeryville, CA 94608, USA.

[2]Biosciences Area, Lawrence Berkeley National Laboratory; Berkeley, CA 94720, USA.

[3]Energy Technologies Area, Lawrence Berkeley National Laboratory; Berkeley, CA 94720, USA.

[4]Energy and Biosciences Institute, University of California, Berkeley; Berkeley, CA 94720, USA.

*Corresponding author: taodai@lbl.gov

Abstract

Increasing wind energy generation is central to grid decarbonization, yet methods to estimate wind energy potential are not standardized, leading to inconsistencies and even skewed results. This study aims to improve the fidelity of wind energy potential estimates through an approach that integrates geospatial analysis and machine learning (i.e., Gaussian process regression). We demonstrate this approach to assess the spatial distribution of wind energy capacity potential in the Contiguous United States (CONUS). We find that the capacity-based power density ranges from 1.70 MW/km^2 (25th percentile) to 3.88 MW/km^2 (75th percentile) for existing wind farms in the CONUS. The value is lower in agricultural areas (2.73$\pm$0.02 MW/km^2, mean$\pm$95% confidence interval) and higher in other land cover types (3.30$\pm$0.03 MW/km^2). Notably, advancements in turbine manufacturing could reduce power density in areas with lower wind speeds by adopting low specific-power turbines, but improve power density in areas with higher wind speeds (>8.35 m/s at 120m above the ground), highlighting opportunities for repowering existing wind farms. Wind energy potential is shaped by wind resource quality and is regionally characterized by land cover and physical conditions, revealing significant capacity potential in the Great Plains and Upper Texas. The results indicate that areas previously identified as hot spots using existing approaches (e.g., the west of the Rocky Mountains) may have a limited capacity potential due to low wind resource quality. Improvements in methodology and capacity potential estimates in this study could serve as a new basis for future energy systems analysis and planning.

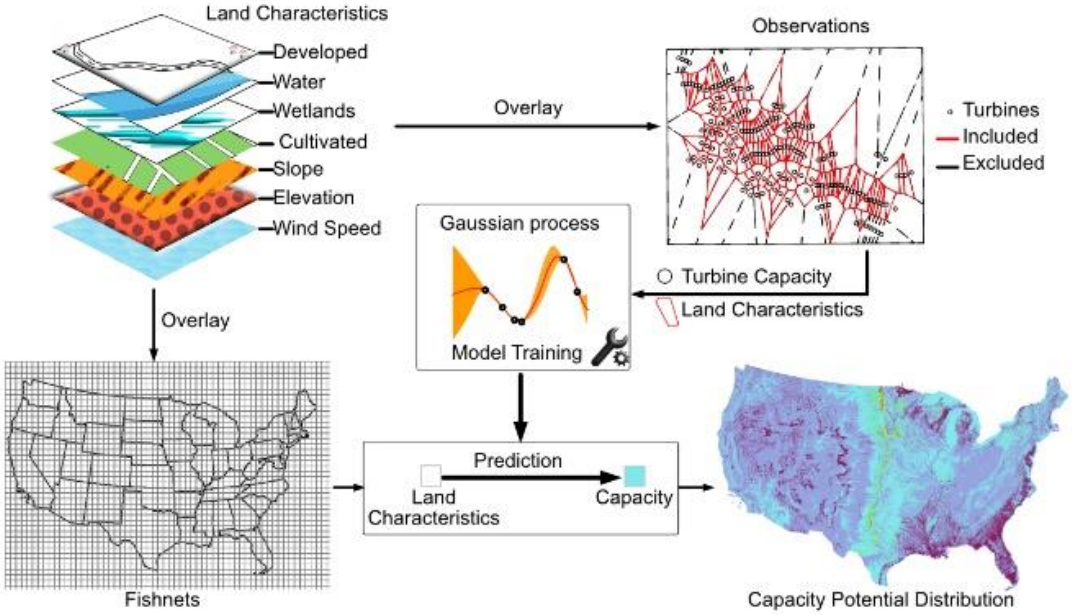
Highlights

- No standardized approaches exist for estimating large-scale wind energy potential
- A consistent wind energy power density quantification approach is developed
- A novel large-scale wind energy potential estimation approach is developed
- Wind energy distribution is shaped by wind resources rather than land availability
- Results of wind energy distribution could be a critical dataset for future studies

Keywords

Wind Energy; Power Density; Gaussian Process Regression; Renewable Energy; Sustainable Development; Machine Learning

Graphic Abstract



Nomenclature

List of abbreviations	
CONUS	The Contiguous United States
NLCD	National Land Cover Database
USWTDB	U.S. Wind Turbine Database
GPR	Gaussian process regression
PADUS	Protected Area Database in the United States
NREL	National Renewable Energy Laboratory
Symbols and units	
A	Area in square kilometers (km^2)
t_{cap}	Turbine capacity (MW)
PD_c	Capacity-based power density (MW/km^2)
$area$	Area as a predictor [km^2]
f_{water}	Fraction of water/snow (NLCD Class 11,12)
$f_{developed}$	Fraction of developed (NLCD Class 21-24)
$f_{cultivated}$	Fraction of cultivated crops (NLCD Class 82)
$f_{wetlands}$	Fraction of wetlands (NLCD Class 90,95)
$speed$	Median of annual average wind speed at 120 meters above the ground (m/s)
$elevation$	Mean elevation (m)
$slope$	Median of slope within the turbine area (degree)
$year$	Year of the project came to commercial use
m	Mean function of a Gaussian process regression
k	Covariance function
ϵ	Additive Gaussian noise
k_{base}	Kernel of the fundamental limitations of wind power potential
$K_{physical}$	Kernel of the physical limitations of wind power potential
$k_{landcover}$	Combination of land cover kernels
k_C	Constant kernel
k_{year}	Kernel of year
k_{area}	Kernel of available land
$k_{elevation}$	Kernel of elevation
k_{slope}	Kernel of slope
k_{speed}	Kernel of wind speed
$k_{developed}$	Kernel of the fraction of developed
$k_{wetlands}$	Kernel of the fraction of wetlands
k_{water}	Kernel of the fraction of water
$k_{cultivated}$	Kernel of the fraction of cultivated crops
k_n	A square exponential covariance function
σ_n	Output variance of a square exponential kernel
l_n	Length scale of a square exponential kernel

1. Introduction

Large-scale wind power stands as one of the most vital and cost-effective renewable energy solutions for transitioning away from fossil fuels and achieving a sustainable and climate-resilient energy future [1]. Increasing wind power supply is essential in all pathways to net carbon neutrality and a decarbonized grid in the United States [2–4]. The scaling-up of wind power requires land for siting additional turbines and constructing new infrastructure (e.g., access roads and transmission lines), which could have negative implications for wildlife habitat, biodiversity, and public health due to land conversion, land fragmentation, and noise [5–8]. Accurate predictions of location-specific wind energy potential and land requirements are key to optimizing energy systems planning and mitigating the associated community and ecosystem impacts [9–11]. This study focuses on improving the large-scale assessment of the capacity potential of wind energy, i.e., calculating how much wind power capacity can be built in a study scope at a continental or global level.

The capacity potential is commonly assessed as the product of the capacity-based power density (i.e., wind power that can be installed in a specific area in MW/km^2) and the area of suitable land (or the so-called geographical potential in km^2) [12–18]. This method is, however, theoretically not sound under current practices. Suitable land area has been determined in prior studies by excluding unsuitable lands, such as protected, developed, or physically prohibited areas (e.g., water bodies and high elevation or slope areas). However, when capacity-based power density is calculated based on observations of the spatial distribution of turbines within existing wind farms, the entire area of a wind farm would be treated as the denominator, which considers roads and other inaccessible or unsuitable areas that would be excluded in the calculation of the geographical potential [16,17,19,20]. As a result, the wind energy capacity potential is calculated based on a relatively underestimated capacity-based power density.

Wind power potential is also difficult to assess at a large scale because of the large spatial and temporal variations in capacity-based power density. Diffendofer and Compton manually digitized the land use by wind farms in the U.S. and found that the capacity-based power density can vary by >40 times, depending on the spatial scale of analysis [21]. More recent studies used geospatial analysis and obtained a capacity-based power density ranging from 3.1-7.6 MW/km^2 in the US and 6.2-46.9 MW/km^2 in Europe [20,22]. Developing a map of capacity-based power density may help address the spatial variations [23]. Temporally, studies have identified a declining trend in capacity-based power density and attributed this decline to the installation of wind turbines with a lower specific-power, which are theoretically of a lower capacity-based power density due to their larger rotor diameters requiring a larger spacing distance while providing relatively lower capacity [19,20,24].

Notably, existing measurements of capacity-based power density cannot be compared across studies due to inconsistencies in the definitions of the occupied land by wind farms. Early studies used the leased project area, which was later pointed out as an overestimation of land use [25–27], and geospatial analysis-based approaches such as defining a minimum bounding geometry of a wind farm, defining Thiessen polygons within a wind farm, or creating turbine rotor diameter-based buffering zones near turbines were widely applied as the increasing availability of spatially explicit datasets [19,28,29]. Limitations of existing geospatial analysis have also been recognized. For example, Dai et al. (2024) pointed out that the area of land determined by the minimum bounding geometry of a wind farm depends on the configurations of turbines (i.e., whether the turbines are formed as a single string, multiple strings, or irregularly distributed as a cluster) [30]. Enevoldsen and Jacobson argued that existing approaches are too simple to determine the land being actually used and proposed an automatic approach that quantifies the land use without

accounting for the space out of wind farms [22]. Our study proposed an approach that can automatically determine the turbines at the outer boundary of a wind farm, which ensures the consistency and replicability of land quantification across wind farms by avoiding the determination of a bounding geometry or a buffer distance near the outer turbines.

A more fundamental and unresolved question is: how can studies connect the measured capacity-based power density to the identified suitable land and complete the entire process of capacity potential assessment? One promising approach is to ignore the patterns of historical capacity-based power density and optimize the siting of a specific turbine model in the identified suitable lands [31]. This approach could obtain future-oriented, spatially explicit estimates of capacity potential and capacity-based power density by selecting turbines meeting future wind energy development needs. This approach still presents some challenges, including variations of turbine needs within suitable lands, land discontinuity, and data completeness. Suitable lands with different land characteristics (e.g., wind speed, land cover types, and slope) have a different need for turbine models (e.g., a turbine's specific-power) and a varying turbine siting pattern [21,32]. Regarding land continuity, Pett Ridge et al. (2023) found that requiring a contiguous area of 5 km^2 could remove one-third of the suitable land for wind energy development in the CONUS [4]. McKenna et al. identified the small patches of land created from the suitable land identification process as the primary uncertainty source for estimating wind energy potential in Germany [33]. Finally, obtaining and applying the siting constraints is also time-intensive and effort-intensive, especially for siting coordinates that could be regulated from a local level (municipality or township) to a state level. This challenges spatial analysis and may result in biased results due to an incomplete dataset [34].

Here, we developed a novel and systematic approach for capacity-based power density quantification and capacity potential assessment. In this workflow, first, the capacity-based power density is quantified by automatically selecting turbines bounded by other turbines, so uncertainty sources related to the land use at the outer boundary of wind farms are avoided [27]. Second, the amount of suitable land is treated implicitly, and the capacity potential in a given area is directly connected to land characteristics, including land cover types and other physical-based parameters, using a Gaussian process regression (GPR) model. Historical observations of capacity-based power density can then be utilized for projecting capacity potential assessment at a large scale, and land suitability conditions can be applied in postprocessing. A GPR model can adopt multiple predictors flexibly and obtain an uncertainty interval in the predictions, a unique feature among machine learning methods [35–37]. Based on a GPR model trained with data representing the siting characteristics of existing wind farms in the CONUS, we mapped the capacity potential of wind energy in the CONUS, which allows spatial explicit energy systems analysis associated with wind energy in future studies. Our analysis reveals how local land characteristics and technological advancements impact the capacity potential and shows the spatial distribution patterns of the capacity potential of wind power. The workflow provides a new approach toward the standardization of wind power potential estimation.

2. Material and methods

2.1 Quantifying Capacity-Based Power Density of Existing Wind Farms

The capacity-based power density values of existing wind farms are determined by generating Thiessen polygon-based zones from the inner areas of wind turbine clusters. Wind turbine locations are obtained from the U.S. Wind Turbine Database (USWTDB) [38]. We first clustered all the wind turbines using a density-based clustering approach with a search distance of 3 kilometers and a minimum number of turbines of 10 to guarantee the representativeness of samples (Figure 1a). The extent of land use by each turbine was

assigned by creating Theissen polygons per cluster. Each cluster of polygons was then dissolved into one polygon as the corresponding outer boundaries, and the Theissen polygons connected to the outer boundaries were removed. The remaining Thiessen polygons in the inner areas are then identified and used as samples. We determined 3 km as the search distance by examining the statistical distributions of the turbine capacity in the samples, aiming to include all isolated wind farms and to balance the involvement of land and the representatives of wind turbine capacity in areas with multiple wind farms, where the number of samples and the associated mean of turbine capacity both increase as the search distance increases due to the requirement of a larger spacing distance for turbines with a larger capacity (Figure 1b). Our approach tends to include wind farms with a cluster configuration or multiple arrays (>3) of parallel turbines; otherwise, turbines will be removed due to the connection to the outer boundary (Figure 1c). The area of the Thiessen polygon, A in km^2, is considered as the land use for the corresponding turbine, and the capacity-based power density, PD_c, is calculated as:

$$PD_c = \frac{t_{cap}}{A} \tag{1}$$

where t_{cap} is the turbine capacity in MW.

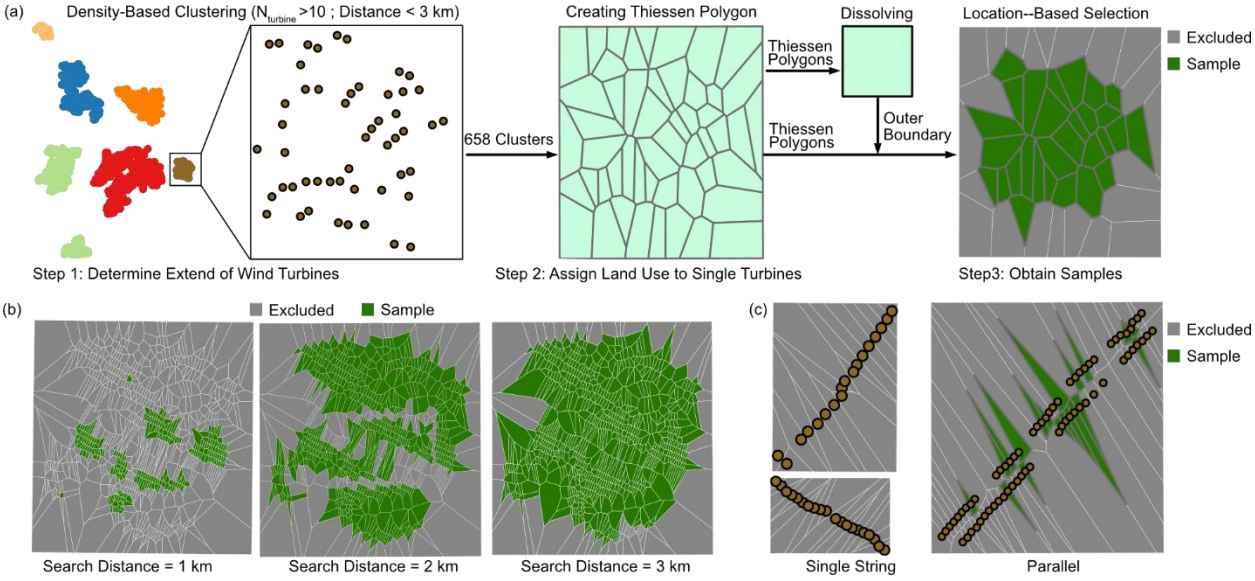


Figure 1. (a) The procedure to obtain turbine samples by excluding turbines located in the outer boundary of a wind farm based on geospatial analysis; (b) The impact of search distance in the density-based clustering step on the number and distribution of samples; (c) Turbines in wind farms with a single string and parallel configurations can be excluded from the analysis.

2.2 Examining Trends in Capacity-Based Power Density

We included observations of capacity-based power density ranging from the 5th to 95th percentile in our sample, considering a capacity-based power density of less than 0.6 MW/km^2 (5th percentile) may indicate that the land area is too large and has not been appropriately determined. A capacity-based power density larger than 8.7 MW/km^2 (95th percentile) can imply that the turbines are sited too close to be representative of modern wind power development, as turbines in wind farms before 2000 could be sited with spacing distance less than 3 times of the turbine diameter [30]. We examined how capacity-based power density was impacted by wind speed and physical constraints, including topographical slope, elevation, land cover

types, and turbine capacity. Turbine capacity was regarded as a proxy predictor of technological advancements. Each turbine's capacity was obtained by spatially joining the USWTDB with the Thiessen polygon layer. The rest of the predictors were obtained using a raster-based analysis starting with converting the Thiessen polygons to a raster map, in which the raster values were the turbine identities (i.e., the "caseID" in the USWTDB). The conversion process used the National Land Cover Dataset (NLCD) [39] series of land cover maps as reference layers of cell size (i.e., resolution), spatial reference and projection, processing extent, and a snap raster. Then, *Zonal Statistics as Table* in *ArcGIS Pro* is used to obtain the value of each of the predictors. Data sources for the wind speed, slope, and elevation are listed in Table 1. We merged the subgroups of land cover types in NLCD into nine land cover types (i.e., water, developed, barren, forest, shrub, herbaceous, pasture, cultivated and wetlands), which are high-level land cover class according to the Anderson Land Cover Classification System [40]. We assumed a three-year construction period, the estimated maximum time required to build a modern wind farm [41], before the wind farm came into commercial operation and used the corresponding land cover map to avoid the impact of turbine bases and access roads that can otherwise be categorized as "developed" in the NLCD maps (Table S1). Geospatial analysis technics (e.g., spatial join and *Zonal Statistics as Table*) are documented in the Spatial Analysis toolbox overview [42].

Table 1. Definition of predictors and the corresponding data sources

Predictor	Meaning	Data Source
$area$	Area	Calculated from Thiessen polygons
f_{water}	Fraction of water/snow (Class 11,12)	NLCD 2001-2019 [39]
$f_{developed}$	Fraction of developed (Class 21-24)	NLCD 2001-2019 [39]
$f_{cultivated}$	Fraction of cultivated crops (Class 82)	NLCD 2001-2019 [39]
$f_{wetlands}$	Fraction of wetlands (Class 90,95)	NLCD 2001-2019 [39]
$speed$	Median of annual average wind speed at 120 meters above the ground	Draxl et al. (2015) [43]
$elevation$	Mean elevation within the turbine area	Earth Resources Observation and Science Center [44]
$slope$	Median of slope within the turbine area	Calculated based on elevation data
$year$	Year of the project came to commercial use	US Wind Turbine Database [38]

2.3 Predicting Capacity Potential in the Contiguous United States

We used a GPR model with turbine capacity as the outcome variable and the land characteristics as predictors. A Gaussian process regression model can be written as

$$t_{cap} = GP(m, k) + \epsilon \tag{2}$$

$$\epsilon \sim N(0, \sigma^2) \tag{3}$$

where m is the mean function, which is usually assumed as 0, k is a covariance function, and ϵ is the additive Gaussian noise. A covariance function depicts the similarity between two wind turbine construction sites and indicates how the turbine capacity would change along with the differences in the predictors. The covariance function is constructed as:

$$k = k_{base} \cdot \left(K_{physical} + k_{landcover} + k_C\right) \tag{4}$$

where k_{base} is the kernel represents the fundamental limitations of wind power potential, which were constructed by year representing technological limitations and area of the available land:

$$k_{base} = k_{year} \cdot k_{area} \tag{5}$$

$k_{physical}$ represents a combination of physical constraints, including elevation, slope and wind speed, as

$$k_{physical} = k_{elevation} + k_{slope} + k_{speed} + k_{elevation} \cdot k_{slope} \cdot k_{speed} \quad (6)$$

The product of kernels indicates that a more significant similarity between two wind turbine sites occurs only when all the predictors are similar [45]. $k_{landcover}$ represents a combination of land cover kernels, as:

$$k_{landcover} = k_{developed} + k_{wetlands} + k_{water} + k_{cultivated} + k_{developed} * k_{cultivated} \quad (7)$$

where developed, wetlands, and water are land cover types that are physically not practical to build wind energy. Cultivated land represents lands with higher economic values compared to other land cover types such as shrubland and barren land. k_C is a constant kernel. We used the square exponential covariance function for all the predictors:

$$k_n(x, x') = \sigma_n^2 \exp\left(-\frac{|x - x'|}{l_n^2}\right) \quad (8)$$

where k denotes the covariance function and σ_n and l_n are the hyperparameters representing the output variance and length scale for each predictor.

Limited by computing resources, the samples were grouped by year and cluster and then randomly divided into a training set (20% of the turbines) and a test set (80% of the turbines). Python package *GPFlow* was used to solve the GPR model, and the hyperparameters were optimized using *SciPy* [46]. The model performance was measured by the coefficient of determination (i.e., R^2), the root mean square error (RMSE), an interpretation of the standard deviation of the prediction residuals, and the mean absolute error (MAE), a measure of the average absolute difference between the predicted values and the ground truth values.

Utilizing the trained GPR model to predict the capacity potential in the CONUS started with dividing the CONUS into fishnets with a size of 990 meters (33 times the resolution of NLCD). Each fishnet was ~0.98 km^2, close to the median area in the samples. The fishnets were then converted to a raster map with the NLCD layer as a reference (i.e., cell size, snap raster, processing extent, reference coordinate system, and snap raster) and the "ID" of each fishnet as the raster value. The values of the predictors are then obtained using *Zonal Statistics as Table* based on the dataset used in Table 1.

The predicted capacity potential included all the fishnets regardless of their suitability for wind energy development. We examined how the overall wind power potential is impacted by siting constraints with a scenario analysis of 5 configurations (i.e., *No-Protected-Area*, *Business-As-Usual*, *Conservative*, *Expanded-Siting*, and *Open-Access*). The scenarios were constructed by varying wind speed, elevation, slope, and fractions of water, developed, and wetlands in each fishnet polygon, reflecting the impact of siting ordinances, as shown in Table 2. Slope and elevation were included to examine the impact of the potential expansion of wind farms to areas with higher slopes or elevation. Specifically, predictor values in the *Business-As-Usual* scenario were selected by the 5th or 95th percentile in the samples. The *No-Protected-Area* scenario removed protected land, roadless areas, and the associated buffered areas described in Pett-Ridge et al. (2023) [4] (e.g., roadless lands [47], areas identified in the Protected Area Database in the United States (PADUS) which covers national parks, state parks, and other reserved lands [48]). The *Conservative* scenario assumed wind farms cannot be built near land cover types such as water, developed, and wetlands. The *Expanded-Siting* scenario set predictors to values determined in Lopez et al. (2021) [49], which allowed a lower wind speed, a large slope, and a higher elevation for wind turbine siting compared

to the *Business-As-Usual* scenario. The code for model training, performance examination, and prediction are in the Supplementary Material.

Table 2. Scenarios analyzed by varying the values of predictors. *P* is the percentile of the corresponding value of the predictor among the sampled turbines and *f* refers to the fraction of land cover in the specified categories (water, developed, and wetlands).

Scenario	Predictors											
	Min. WS*		Max. f_{water}		Max. $f_{developed}$		Max. $f_{wetlands}$		Max. Slope		Max. Elevation	
	Value (m/s)	*P*	Value (%)	*P*	Value (%)	*P*	Value (%)	*P*	Value (degree)	*P*	Value (meter)	*P*
No-Protected-Area	1.56		1		1		1		89.7		4119	
Business-As-Usual	7.07	5	0.8	95	8.3	95	5.1	95	8.09	97.5	2057	99
Conservative	7.07	5	0	0	0	0	0	0	8.09	97.5	2057	99
Expanded-Siting	1.72		0.8	95	8.3	95	5.1	95	14.0		2743	
Open-Access	1.72		2.5	97.5	9.8	97.5	9.8	97.5	28.3		2743	

*WS: Wind Speed

3. Results

3.1 Characteristics of Capacity-based Power Density

Our sampling approach obtained representative observations in terms of overall capacity, areas of development, and turbine capacity. The 41,381 observations have a total capacity of 88.2 GW, accounting for ~60% of the total installed capacity in the CONUS, with commercial use year ranges from 2004 to 2022 and turbine capacity from 0.065 to 5.6 MW, according to the USWTDB) [38]. The primary land cover types, i.e., the land cover with the largest share within the turbine area, are cultivated, herbaceous, and shrubland, which account for 55.2%, 21.6%, and 17.5% of total capacity, respectively (Figure 2a), and 58.3%, 20.2%, and 16.1% of total land use, respectively (Figure 2b). Turbines with a capacity of 2-3 MW have dominated among the newly commercially used turbines since 2015, accounting for 55.1% of all turbine capacity, followed by turbines with a capacity of 1-2 MW and 3-4 MW, accounting for 26.8% and 8.3%, respectively (Figure 2c), implying the significance of examining the dynamic patterns of wind energy.

The results show that the capacity-based power density is mainly impacted by land cover, wind speed, and turbine capacity, as shown in Figure 2d and Figure S1-S2. The capacity-based power density of sampled wind turbines is 2.99 ± 0.02 MW/km^2 (mean$\pm$95% confidence interval), concentrating between 1.70 MW/km^2 (25th percentile) and 3.88 MW/km^2 (75th percentile). Turbines in cultivated lands have a lower power density of 2.73 ± 0.02 MW/km^2 and turbines in herbaceous and shrubland have a power density of 3.26 ± 0.04 MW/km^2 and 3.37 ± 0.04 MW/km^2, respectively. The combined impact of wind speed and technological advancements on capacity-based power density also depends on land cover. In high wind speed areas (wind speed>8.35 m/s), increasing in turbine capacity is positively related to capacity-based power density, while in low wind speed areas (wind speed between 6.53 m/s and 7.1 m/s), a negative correlation is found, which may result from the popularity of the use of low specific-power turbines in these areas. Comparing turbines in herbaceous and shrubland areas with a wind speed of 7.1-8.35 m/s, the impact of turbine capacity is different, which may be due to turbine type choices. For the same wind speed classes in cultivated lands, capacity-based power density is positively related to turbine capacity, which may indicate redundancy in land use for turbines with a small capacity. Wind farms in cultivated lands are often divided into patches of land due to the existence of developed lands (e.g., roads for agricultural production), and these patches may install turbines with a larger rotor diameter without conflicting with any siting ordinances.

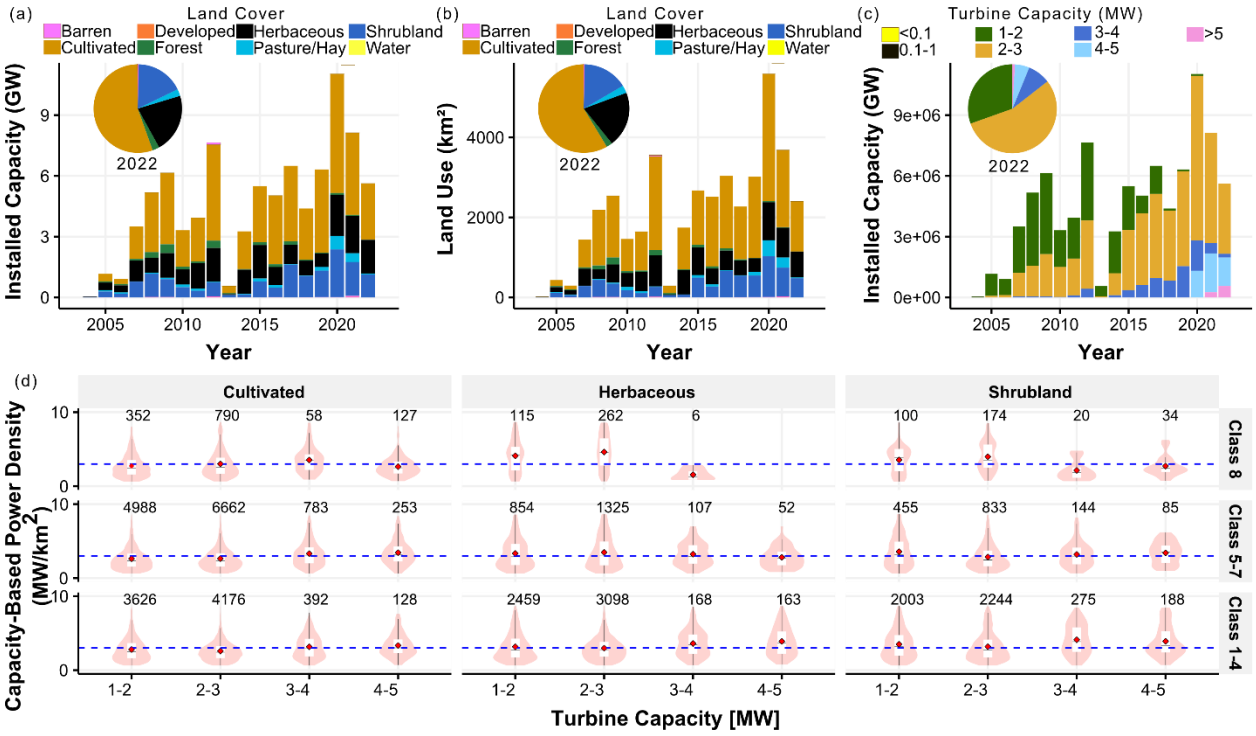


Figure 2. (a-c) Sampled observations show representativeness regarding total capacity, land cover, and turbine capacity. (d) The impact of turbine capacity and wind speed on capacity-based power density. Wind speed classes are determined by the wind speed 120 meters above the ground and are defined as follows: speed >8.35 m/s as Class 1-4, wind speed 7.1-8.35 m/s as Class 5-7, and wind speed 6.53-7.1 m/s as Class 8.

3.2 Patterns of Capacity Potential in the Contiguous United States

Our GPR model shows an overall high performance in predicting unknown observations, considering that the test set includes 4 times the observations in the training set. The model obtained an R^2 of 0.851, an RMSE of 0.261 MW and an MAE of 0.134 MW, showing a higher performance compared to the model in Harrison-Atlas et al. (2021) [23] which are 0.40 MW, 1.25 MW, and 1.02 MW, respectively, while fewer predictors have been used in our model. The prediction interval also represents a valid uncertainty estimate, with 94.8% of the values in the test set predicted within the 95% prediction interval. All the kernels show reasonable extension abilities, as shown in Table S2, in which all length scale values show a reasonably large magnitude.

The predicted wind power potential ranges from 1.66 MW to 6.19 MW per fishnet, with a mean prediction of 2.90 MW and a variation of 0.31 MW. The overall distribution of capacity potential in the CONUS is characterized by the wind speed distribution in the CONUS: lower capacity potential areas distributed in the West (e.g., the Cascade Range, the Great Basin, the Colorado Plateau, and the Rocky Mountains), East Texas and the Southeast (e.g., the Coastal Plains) (Figure 3a). Higher capacity potential areas are concentrated on the Great Plains and upper Texas. The model also successfully identified the impact of developed lands, water, and large slopes as barriers for siting wind turbines and predicted these areas with a lower wind power potential. Typically, the predictor of the fraction of water has helped capture the rivers, as shown in Figure 3b-3d. Figure 3c shows how large slope areas have been identified, and Figure

3d illustrates the effect of a combination of variations in land cover types (e.g., developed and cultivated), slope, and wind resources on wind power potential.

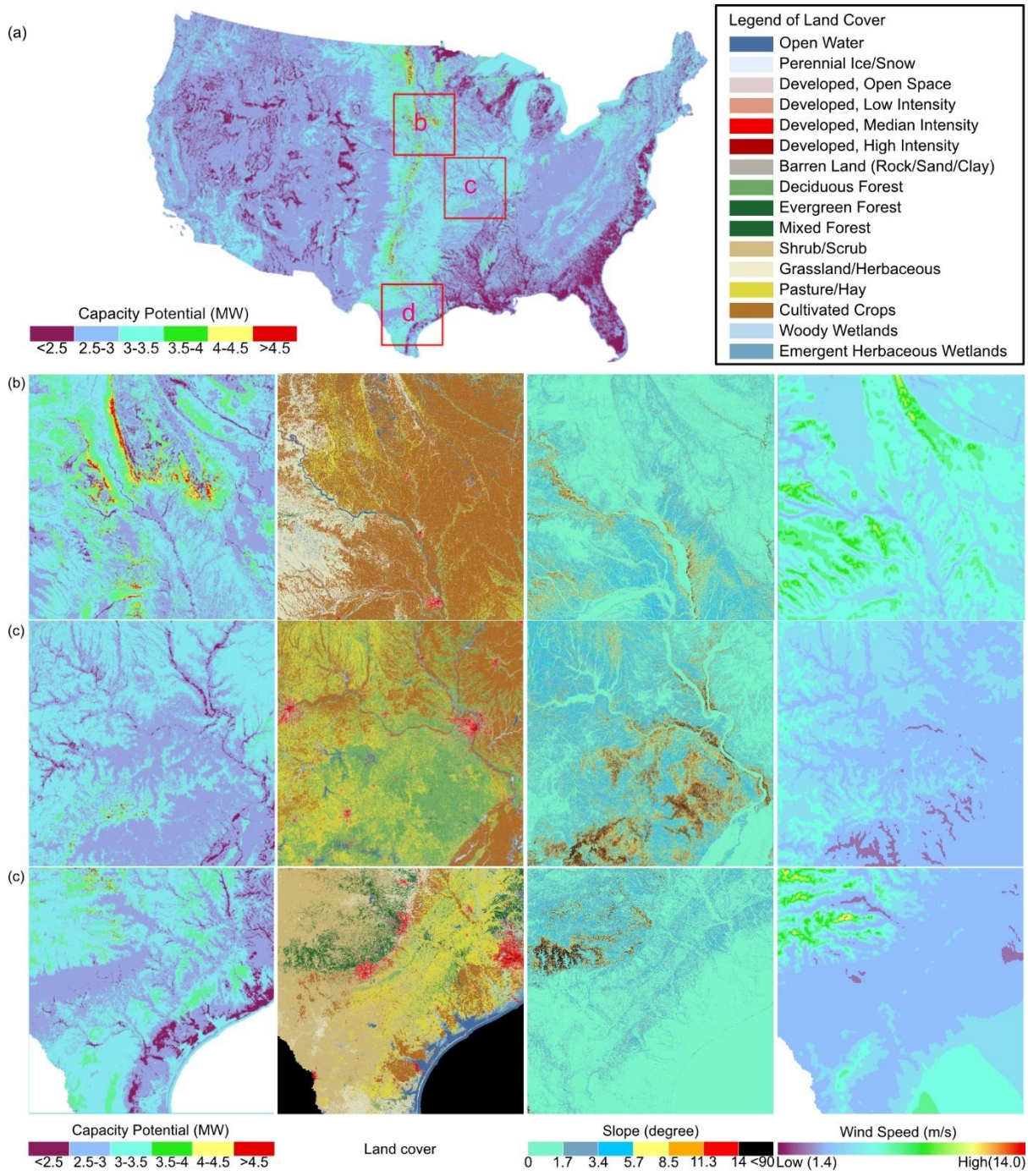


Figure 3. (a) The spatial distribution of predicted capacity potential in the CONUS; (b-d) Regional variations of capacity potential compared to land cover, slope, and wind speed variations.

Excluding all developed land from the total predicted capacity potential resulted in ~7.02 TW capacity potential and has a higher impact than excluding all water or excluding all wetlands, which resulted in 13.09 TW and 11.84 TW of total capacity potential, respectively, as shown in Figure 4a. The overall wind power

potential in the CONUS varies over an order of magnitude among the different scenarios, from 1.43 TW in the *Conservative* scenario to 15.73 TW in the *Open-Access* scenario (Figure 4b). Comparing the overall potential (Figure 3a) with the potential in the *open-access* scenario (Figure 4d), ~8 TW of wind power potential was excluded, which is due to the removal of protected land, water bodies in the Great Lakes, the East Coast and the Southeast Coast, and highly developed lands. The impact of changes in siting criteria on the high-capacity potential areas is relatively small. Additional limitations on land cover types of developed, water, and wetlands in the *Expanded-Siting* scenario excluded an additional ~5 TW of power potential in the Great Lakes areas, the Mississippi Delta, and the Southeast (Figure 4e). Increasing the lowest required wind speed, lowering the elevation and fraction of water, as in the *Business-As-Usual* (Figure 4f), excluded the majority of lands in the West and the Southeast. At last, a conservative siting that not allowing any wind farms near water, wetlands, and developed lands removes the capacity potential from lands in the Great Plains and the Central Lowland where developed land and cultivated lands are co-located, which represents ~4 TW of wind power potential (Figure 4g).

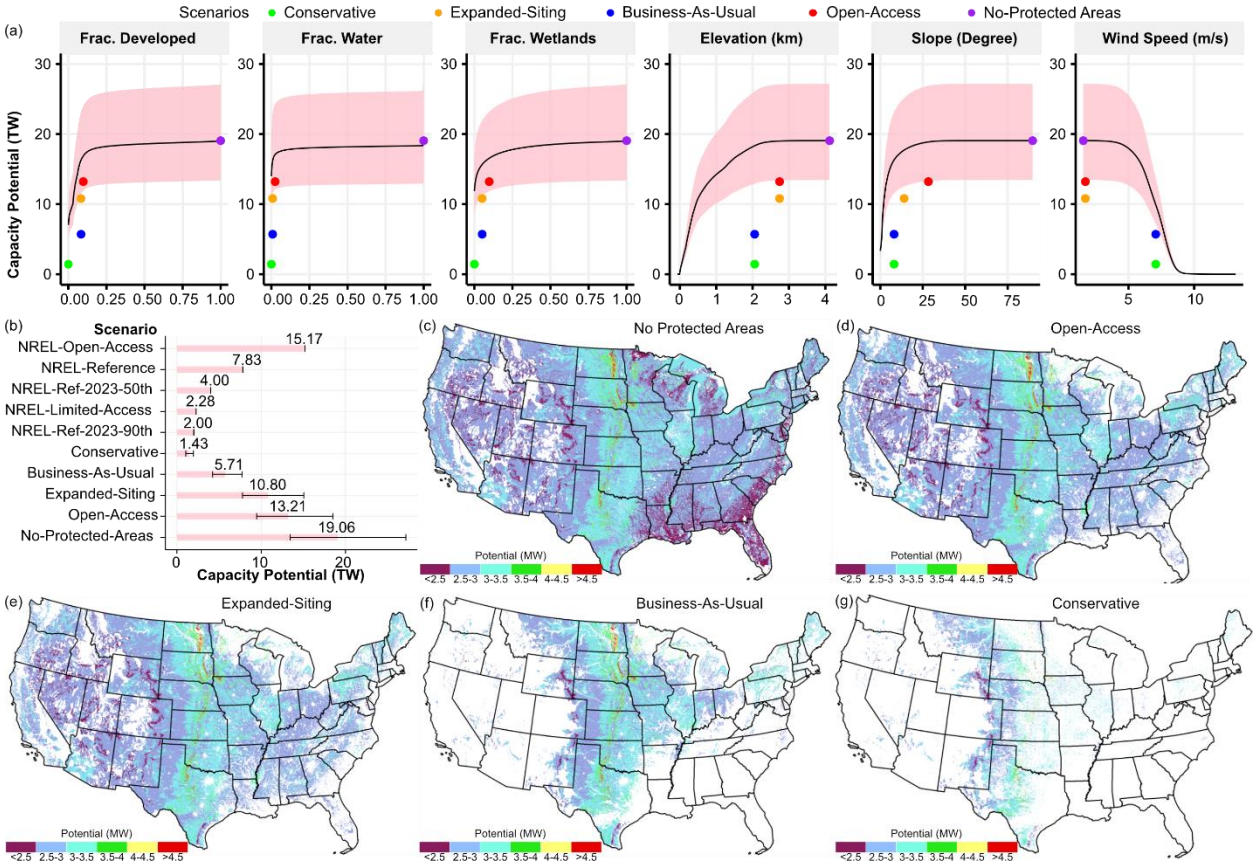


Figure 4. a. Results of local sensitivity analysis. Solid circles show the total capacity in each of the scenarios. "Frac" is an abbreviation of "Fraction". b. Results of capacity potential estimated under the designed scenarios. The results are compared with studies from the National Renewable Energy Laboratory (NREL) including Lopez et al. (2021) [49] and Lopez et al. (2023) [32]. c-g. The spatial distribution of capacity potential.

Compared to the spatial distribution of the capacity potential in Lopez et al. (2021) [49], which includes an open-access scenario (noted as *NREL-Open-Access* scenario), a reference scenario (noted as *NREL-Reference* scenario), and a limited access scenario (noted as *NREL-Limited-Access* scenario), the predicted

total, country-level capacity potential shows consistency for each pair of comparable scenarios. Results from Lopez et al. (2021) and its following study (i.e., Lopez et al. (2023) [32]) are representative datasets from the National Renewable Energy Laboratory (NREL) and have been widely used as a data source for the renewable energy-associated analysis (e.g., wind energy supply curve [50] and clean grid in the U.S. [2]). As shown in Figure 4b, for the more conservative estimates, the estimated capacity potential in the *Conservative* scenario and *Business-As-Usual* scenario is comparable to the *NREL-Limited-Access* scenario as well as the *NREL-Ref-2023-90th* scenario and the *NREL-Ref-2023-50th* scenario. The *NREL-Ref-2023-90th* and *NREL-Ref-2023-50th* are predictions obtained by applying the 90th percentile and 50th percentile siting ordinances in the U.S. (i.e., the regulated distance required to keep wind turbines from existing infrastructures such as roads and buildings) to the *NREL-Reference* scenario, respectively. In addition, the prediction in *NREL-Reference* is comparable to the results in *Expanded-Siting*, which both have no limits on wind speed, and *NREL-Open-Access* is comparable to the *Open-Access* scenario in this study.

Notably, the main discrepancies between estimates in this study and Lopez et al. (2021) [49] lie in the spatial distribution of capacity potential: our results highlighted the potential capacity of the Great Plains, while the *NREL-Reference* results show hot spots in the West, including the Great Basin, the west of the Rocky Mountains. This may reveal a shortcoming in the approach used in existing studies, which quantifies the wind energy capacity as the product of the area of suitable land and a spatial generic power density. The hot spots in NREL's studies are mainly areas with a lower development level, and thus, large areas of suitable land can be identified. A higher capacity potential can then be obtained using a spatially generic capacity-based power density (i.e., 3 MW/km^2). However, these areas are also characterized by a relatively lower wind speed compared to the east of the Rocky Mountains, so these areas should have a lower-than-average capacity-based power density. Results from optimizing turbine sitings by Lopez et al. (2023) [23] also show that the capacity potential in the Great Plains has been relatively underestimated by Lopez et al. (2021) [49]. Compared to Lopez et al. (2023) [23], our results show a lower prediction in the low wind speed areas in the Rocky Mountains (e.g., West New Mexico, West Colorado, and part of Wyoming).

Using a spatially generic power density could obtain a lower estimation than our approach's predictions and the achieved capacity in 2022, as shown in Figure 5. The intersect fishnets with a selection radius of 150 meters around existing turbines obtained nearly identical wind power capacity as the capacity came to commercial use in 2022, while with a search distance of 400 meters, the capacity from a power density-based approach reaches only ~1/3 the capacity. While this does not necessarily lead to a conclusion of which method is more accurate as we are comparing the achievable power potential with the technical capacity potential, the lower estimation of a power density-based approach can be explained as 1) the currently static and spatially generic power density cannot represent the power density level in 2022 due to technological improvement, especially in areas with good wind resource qualities, and 2) the power density has been underestimated due to the inconsistency of land definition in quantifying capacity-based power density and suitable land.

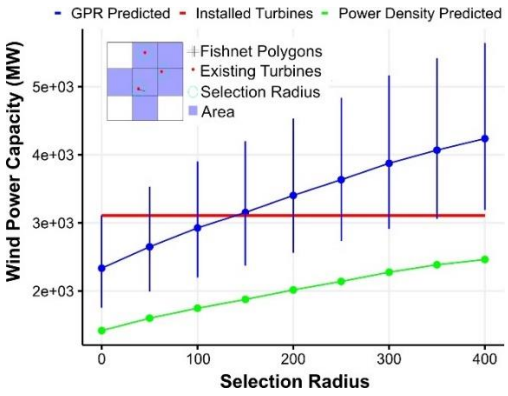


Figure 5. A comparison of wind capacity which came to commercial use in 2022, the GPR predicted capacity, and the capacity predicted with a power density-based approach. The estimates of the power density-based approach was obtained following Lopez et al. (2021) [49], assuming lands within 225 m of developed, water, and wetlands are unsuitable for wind energy development.

4. Discussion

The pattern that capacity-based power density is dependent on both wind speed and land cover type identified in this study indicates the temporal pattern of a declining trend of capacity-based power density due to technological advancements, as recognized in previous studies (e.g., Miller and Keith (2019) [19] and Harison-Atlas et al. (2021) [20], maybe not representative due to 1) overestimating the impact of low specific-power turbines and 2) inappropriate accounting of land use in agricultural areas by including all the lands within the wind farms. First, low specific-power turbines are designed for areas with a lower quality of wind resources to maximize wind energy capture and avoid excessive mechanical stress from high-speed wind. The trend of deployment of low specific-power turbines may be reversed when sites with a low wind speed and high accessibility to transmission lines or access roads are filling out. This is evidenced by the increasing trend of both the turbine's specific-power and the site's wind speed for wind farms built in 2019-2022 in the U.S. [32]. Second, the majority of existing turbines are sited within cultivated lands at a larger spacing distance than in other land types due to the existence of land co-uses. Using these lower estimates of capacity-based power density in agricultural areas may be misleading for future wind farms that are to be built in a different land cover type. Our results also show that there could be redundancies in existing wind farms in agricultural areas, i.e., larger wind turbines could potentially be installed in existing wind farms without using additional land: our results imply that repowering existing wind farms and installing turbines of 4-5 MW capacity could add ~70% (or ~30 GW) of capacity in areas with a wind speed class 1-4 and 70% (or ~18.8 GW) in agricultural areas with wind speed classes 5-7. Furthermore, the observations obtained in this study indicate that year alone is not a good predictor to identify the impact of turbine sizes on capacity-based power density due to mixed technological adaptations, the risk of oversimplifying the coupled impact of technological advancement and wind speed, and overestimating the impact of a lowering specific-power trend.

This study could help improve the understanding of wind farms' "land use impact", which has not yet formed a standardized categorization. We proposed to divide the land use impact of wind farms into visual impact, noise impact, directly impacted land, and operational land use, among which the vision impact, noise impact, and directly impacted land have an environmental consequence while the operational land use cannot be interpreted as from an environmental impact perspective. The operational land use can be quantified by the capacity-based power density but only shows the land requirement of turbine operation.

Mixing different aspects of land use by wind farms could overlook the extent of the actual impact [30,51]. Future studies could examine the relationships between vision, noise, direct impact land and operational land use and how local land characteristics help shape such a relationship.

We acknowledge the following limitations of this study from an energy analysis perspective in addition to the intrinsic limitations of the GPR method (e.g., the limited ability in expressiveness and extrapolation ability for complex models and the sensitivity of covariance function selection) [52]. First, this study uses the observations of achieved wind power to estimate the technical capacity potential, indicating a mismatch within the concepts. Whether the area of land assigned to the samples is closer to the achieved wind power potential or the technical potential seems to depend on land cover. In non-cultivated and flat areas, wind turbine sitings are close to an ideal siting configuration, resulting in the equivalence of technically required land and the achieved land use. In cultivated areas or where lands are limited by physical conditions (e.g., windspeed and slope), wind turbines are irregularly sited, making the land for achieved potential larger than for technical potential. We have used several approaches to bridge such a gap, including removing the Thiessen polygons at the outer boundary, considering data from the 5^{th} to 95^{th} percentile as valid samples, and a scenario analysis.

Second, our model tends to overestimate the impact of siting ordinances. We used the fraction of typical land cover types (e.g., developed, wetlands, and water) as a proxy of siting ordinances, whereas siting ordinances are not only dependent on location due to policy variations among different counties but also dependent on the turbine characteristics including turbine rotor diameter and tip height (i.e., the sum of rotor radius and turbine tower height). Our approach could not reflect these variations. Furthermore, using a fraction of the total area as a predictor aggregates the spatial information regarding the distribution of lands that may be impacted by siting ordinances. As the siting ordinance is usually smaller than the boundary length of the fishnet polygon (0.98 km), the impact of siting coordinates can be overestimated, especially in the scenario analysis.

Finally, data quality in the predictors impacts the accuracy of the results. As GPR is based on the covariance matrix of predictors, a higher level of details obtained in the training sample would help capture the features in each area's characteristics and improve prediction accuracy. We have mainly used the NLCD for land cover with a 30-meter resolution and have aggregated similar groups of land cover toward border and higher-level land cover classes. For example, the aggregated developed land comes from four sources (i.e., Developed-Open Space, Developed-Low Intensity, Developed-Medium Intensity, and Developed-High Intensity) [39], each shows a different level of coverage by impervious surfaces at 30-meter resolution, from less than 20% to more than 80%. Aggregating them has lost this information, but avoiding too much complexity also helps to prevent overfitting of the model. Additionally, the relatively low resolution of wind speed data (2,000 meters) can represent the overall distribution of wind resources in the CONUS but may miss some details in certain circumstances. We observed relatively high capacity potential in areas with both water bodies and cultivated land, which may implicitly show the impact of an increasing local wind speed on the capacity potential.

5. Conclusions

This work proposes a novel approach for estimating large-scale wind energy capacity potential that implicitly treats the amount of suitable land, ensuring consistency between observed capacity-based power density and predicted capacity potential. The proposed approach addresses the key limitations of existing methods, which include lack of a standardized definition of land use by wind farms, discrepancies between

the definition of suitable land and the land concept in the calculation of capacity-based power density, lack of spatial and temporal specificity, and lack of consistency in connecting measurements of capacity-based power density and capacity potential estimates. Applying this approach to the CONUS reveals that wind capacity potential is largely shaped by wind resource quality on a broad scale, while local land characteristics (e.g., land cover and physical conditions) also play an important role. The resulting spatially explicit capacity maps for the U.S. provide new insights for future energy systems analyses and have two important implications: first, studies using a conventional approach may have overestimated the capacity potential in less-developed areas, such as the west of Rocky Mountains, where large tracts of land are deemed suitable despite having relatively low wind resource quality. Second, contrary to the widely accepted notion of declining capacity-based power density, our results suggest that future wind farms could enhance performance in capacity-based power density through adopting advanced turbine technology in agricultural areas and by expanding into shrubland and herbaceous areas, as the most accessible locations are becoming fully utilized.

Future work could aim to refine capacity potential estimates and integrate our approach into the entire wind energy supply workflow. As this study focuses on methodological aspects and does not assess the suitability of specific locations, we do not consider whether particular land areas are suitable for wind energy development. Traditional land suitability conditions could be applied by either preprocessing or postprocessing, similar to the process demonstrated in the *No-Protected-Area* scenario. For instance, previous studies have addressed factors like existing infrastructure (e.g., transmission lines, railways, radars, and airports [49]) and up to township-level siting ordinances [34]. Furthermore, a complete framework for assessing electricity generation potential could also be developed by incorporating our approach with explicit capacity factors developed based on wind resource quality, in which a more comprehensive analysis considering the wind variability over time, extreme weather conditions, and turbulences is needed compared to using wind speed as a single predictor of wind resources. As our approach is not dependent on location (e.g., variations of siting ordinances across countries), the approach could be scaled up for global-level wind energy assessment. The approach could also be applied to other energy technology types such as photovoltaic with suitable datasets.

Acknowledgment

The authors thank the support of the U.S. Department of Energy (DOE) Office of Science and Office of Fossil Energy and Carbon Management under Contract No. DE-AC02-05CH11231. This work was enabled in part by the Joint BioEnergy Institute (https://www.jbei.org) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Competing interests: C.D.S. has a financial interest in Cyklos Materials.

Supplementary Material

The supplementary tables and code for model training, performance examination, and prediction are in the Supplementary Material.

References

[1] Wang F, Harindintwali JD, Yuan Z, Wang M, Wang F, Li S, et al. Technologies and perspectives for achieving carbon neutrality. Innov 2021;2:100180.

[2] Denholm P, Brown P, Cole W, Mai T, Serg B. Examining Supply-Side Options to Achieve 100 % Clean Electricity by 2035 Examining Supply-Side Options to Achieve 100 % Clean Electricity by 2035. Golden, CO: 2022.

[3] Jenkins JD, Mayfield EN, Larson ED, Pacala SW, Greig C. Mission net-zero America: The nation-building path to a prosperous, net-zero emissions economy. Joule 2021;5:2755–61. https://doi.org/10.1016/j.joule.2021.10.016.

[4] Pett-Ridge J, Ammar HZ, Aui A, Ashton M, Baker SE, Basso B, et al. Roads to Removal: Options for Carbon Dioxide Removal in the United States. Livermore, CA: 2023. https://doi.org/10.2172/2301853 LLNL-TR-852901.

[5] van Zalk J, Behrens P. The spatial extent of renewable and non-renewable power generation: A review and meta-analysis of power densities and their application in the U.S. Energy Policy 2018;123:83–91. https://doi.org/10.1016/j.enpol.2018.08.023.

[6] Luderer G, Pehl M, Arvesen A, Gibon T, Bodirsky BL, de Boer HS, et al. Environmental co-benefits and adverse side-effects of alternative power sector decarbonization strategies. Nat Commun 2019;10:1–13. https://doi.org/10.1038/s41467-019-13067-8.

[7] Turconi R, Boldrin A, Astrup T. Life cycle assessment (LCA) of electricity generation technologies: Overview, comparability and limitations. Renew Sustain Energy Rev 2013;28:555–65. https://doi.org/10.1016/j.rser.2013.08.013.

[8] Wolaver BD, Pierre JP, Labay BJ, LaDuc TJ, Duran CM, Ryberg WA, et al. An approach for evaluating changes in land-use from energy sprawl and other anthropogenic activities with implications for biotic resource management. Environ Earth Sci 2018 775 2018;77:1–14. https://doi.org/10.1007/S12665-018-7323-8.

[9] Rinne E, Holttinen H, Kiviluoma J, Rissanen S. Effects of turbine technology and land use on wind power resource potential. Nat Energy 2018;3:494–500. https://doi.org/10.1038/s41560-018-0137-9.

[10] Wu GC, Leslie E, Allen D, Sawyerr O, Cameron DR, Brand E, et al. Power of Place Land Conservation and Clean Energy Pathways for California 2019.

[11] Wu GC, Leslie E, Sawyerr O, Cameron DR, Brand E, Cohen B, et al. Low-impact land use pathways to deep decarbonization of electricity. Environ Res Lett 2020;15. https://doi.org/10.1088/1748-9326/ab87d1.

[12] Lopez A, Roberts B, Heimiller D, Blair N, Porro G. U.S. Renewable Energy Technical Potentials: A GIS-Based Analysis. Golden, Colorado: 2012.

[13] Feng J, Feng L, Wang J, King CW. Evaluation of the onshore wind energy potential in mainland China—Based on GIS modeling and EROI analysis. Resour Conserv Recycl 2020;152:104484. https://doi.org/10.1016/j.resconrec.2019.104484.

[14] McKenna R, Pfenninger S, Heinrichs H, Schmidt J, Staffell I, Bauer C, et al. High-resolution large-scale onshore wind energy assessments: A review of potential definitions, methodologies and future research needs. Renew Energy 2022;182:659–84.

[15] World Energy Council. New renewable energy resources: a guide to the future. Kogan Page; 1994.

[16] Hoogwijk M, de Vries B, Turkenburg W. Assessment of the global and regional geographical, technical and economic potential of onshore wind energy. Energy Econ 2004;26:889–919. https://doi.org/10.1016/j.eneco.2004.04.016.

[17] Zhou Y, Luckow P, Smith SJ, Clarke L. Evaluation of global onshore wind energy potential and generation costs. Environ Sci Technol 2012;46:7857–64. https://doi.org/10.1021/es204706m.

[18] Smil V. Power density: a key to understanding energy sources and uses. 2015.

[19] Miller LM, Keith DW. Addendum: Observation-based solar and wind power capacity factors and power densities (Environ. Res. Lett. (2018) 13 (104008) DOI: 10.1088/1748-9326/aae102). Environ Res Lett 2019;14. https://doi.org/10.1088/1748-9326/ab12a2.

[20] Harrison-Atlas D, Lopez A, Lantz E. Dynamic land use implications of rapidly expanding and evolving wind power deployment. Environ Res Lett 2022;17:044064. https://doi.org/10.1088/1748-9326/ac5f2c.

[21] Diffendorfer JE, Compton RW. Land cover and topography affect the land transformation caused by wind facilities. PLoS One 2014;9:e88914. https://doi.org/10.1371/journal.pone.0088914.

[22] Enevoldsen P, Jacobson MZ. Data investigation of installed and output power densities of onshore and offshore wind turbines worldwide. Energy Sustain Dev 2021;60:40–51. https://doi.org/10.1016/j.esd.2020.11.004.

[23] Harrison-Atlas D, Maclaurin G, Lantz E. Spatially-Explicit Prediction of Capacity Density Advances Geographic Characterization of Wind Power Technical Potential. Energies 2021;14:3609. https://doi.org/10.3390/en14123609.

[24] Leslie E, Pascale A, Jenkins JD. Princeton's Net-Zero America Study Annex D: Solar and Wind Generation Transitions. 2021.

[25] Strickland D, Johnson D. Overview of What We Know About Avian/Wind Interaction. Natl. Wind Coord. Collab. Wildl. Workgr. Res. Meet. VI, 2006.

[26] Bureau of Land Management. Final Programmatic Environmental Impact Statement on Wind Energy Development on BLM Administered Land in the Western United States. 2005.

[27] Denholm P, Hand M, Jackson M, Ong S. Land Use Requirements of Modern Wind Power Plants in the United States. vol. Technical. 2009. https://doi.org/10.2172/964608.

[28] Walker BL, Neubaum MA, Goforth SR, Flenner MM. Quantifying habitat loss and modification from recent expansion of energy infrastructure in an isolated, peripheral greater sage-grouse population. J Environ Manage 2020;255. https://doi.org/10.1016/j.jenvman.2019.109819.

[29] Diffendorfer JE, Dorning MA, Keen JR, Kramer LA, Taylor R V. Supplementary Information: Geographic context affects the landscape change and fragmentation caused by wind energy facilities. PeerJ 2019;2019:e7129. https://doi.org/10.7717/PEERJ.7129/SUPP-1.

[30] Dai T, Jose Valanarasu JM, Zhao Y, Zheng S, Sun Y, Patel VM, et al. Land Resources for Wind Energy Development Requires Regionalized Characterizations. Environ Sci Technol 2024;58:5014–23. https://doi.org/10.1021/acs.est.3c07908.

[31] Lopez A, Stanley PJ, Roberts O, Williams T, Pinchuk P, Buster G, et al. Estimating National-Scale Wind Potential Using Spatially Explicit Turbine Layout Optimization. 2023.

[32] Wiser R, Bollinger M, Hoen B, Millstein D, Rand J, Barbose G, et al. Land-Based Wind Market Report: 2023 Edition. Berkeley, CA: 2023.

[33] McKenna R, Hollnaicher S, Fichtner W. Cost-potential curves for onshore wind energy: A high-resolution analysis for Germany. Appl Energy 2014;115:103–15.

[34] Lopez A, Cole W, Sergi B, Levine A, Carey J, Mangan C, et al. Impact of siting ordinances on land availability for wind and solar development. Nat Energy 2023;8:1034–1043. https://doi.org/10.1038/s41560-023-01319-3.

[35] Dai T, Jordaan SM, Wemhoff AP. Gaussian Process Regression as a Replicable, Streamlined Approach to Inventory and Uncertainty Analysis in Life Cycle Assessment. Environ Sci Technol 2022;56:3821–9. https://doi.org/10.1021/acs.est.1c04252.

[36] Rahman A, Smith AD. Predicting fuel consumption for commercial buildings with machine learning algorithms. Energy Build 2017;152:341–58. https://doi.org/10.1016/j.enbuild.2017.07.017.

[37] McElreath R. Statistical Rethinking. Second. CRC Press; 2021. https://doi.org/10.1080/09332480.2017.1302722.

[38] Rand JT, Kramer LA, Garrity CP, Hoen BD, Diffendorfer JE, Hunt HE, et al. A continuously updated, geospatially rectified database of utility-scale wind turbines in the United States. Sci Data 2020;7:1–12. https://doi.org/10.1038/s41597-020-0353-6.

[39] U. S. Department of the Interior, U. S. Geological Survey. National Land Cover Database. Multi-Resolution L Charact Consort 2015. https://www.usgs.gov/centers/eros/science/national-land-cover-database?qt-science_center_objects=0#qt-science_center_objects (accessed July 12, 2021).

[40] Anderson JR, Hardy EE, Roach JT, Witmer RE. A Land Use And Land Cover Classification System For Use With Remote Sensor Data. Washington DC: 1976.

[41] U.S. Department of Energy. Land-Based Wind Energy Economic Development Guide. 2022.

[42] ESRI. An overview of the Spatial Analyst toolbox n.d. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/an-overview-of-the-spatial-analyst-toolbox.htm.

[43] Draxl C, Hodge BM, Clifton A, McCaa J. Overview and Meteorological Validation of the Wind Integration National Dataset Toolkit. Golden, CO: 2015.

[44] Earth Resources Observation and Science Center (EROS). LANDFIRE 2020 Elevation (Elev) CONUS. Sioux Falls, SD: 2022.

[45] Duvenaud DK. Automatic Model Construction with Gaussian Processes. University of Cambridge, 2014.

[46] Matthews AGDG, Van Der Wilk M, Nickson T, Fujii K, Boukouvalas A, León-Villagrá P, et al. GPflow: A Gaussian Process Library using TensorFlow. J Mach Learn Res 2017;18:1–6.

[47] USDA-Forest Service. Roadless Areas: 2001 Roadless Rule n.d. https://data.fs.usda.gov/geodata/edw/datasets.php.

[48] U.S. Geological Survey (USGS) Gap Analysis Project (GAP). Protected Areas Database of the United States (PAD-US) 3.0: U.S. Geological Survey data release 2022. https://doi.org/https://doi.org/10.5066/P9Q9LQ4B.

[49] Lopez A, Mai T, Lantz E, Harrison-Atlas D, Williams T, Maclaurin G. Land use and turbine technology influences on wind potential in the United States. Energy 2021;223:120044. https://doi.org/10.1016/j.energy.2021.120044.

[50] National Renewable Energy Laboratory. Geospatial Data Science: Wind Supply Curves n.d.

https://www.nrel.gov/gis/wind-supply-curves.html.

[51] Palmer JF. Deconstructing viewshed analysis makes it possible to construct a useful visual impact map for wind projects. Landsc Urban Plan 2022;225:104423. https://doi.org/10.1016/j.landurbplan.2022.104423.

[52] Rasmussen CE., Williams CKI. Gaussian Processes for Machine Learning. vol. 7. Massachusetts Institute of Technology; 2006.