Resource Assessment for Distributed Wind Energy: An Evaluation of Best-Practice Methods in the Continental US

Caleb Phillips[1], Dmitry Duplyakin[1], Lindsay Sheridan[2], Jenna Ruzekowicz[3] Matthew Nelson[4], Dimitrios Fytanidis[5], Rod Linn[4], Rao Kotamarthi[5], Heidi Tinnesand[1]

[1]National Renewable Energy Laboratory, Golden, Colorado, USA
[2]Pacific Northwest National Laboratory, Richland, Washington, USA
[3]Stanford University, Stanford, California, USA
[4]Los Alamos National Laboratory, Los Alamos, New Mexico, USA
[5]Argonne National Laboratory, Argonne, Illinois, USA

Correspondence: caleb.phillips@nrel.gov

Abstract. Current wind resources within the United States (US) indicate a potential to profitably install nearly 1,400 gigawatts of distributed wind (DW) capacity. This amount is equivalent to over half of the United States' current energy demand from electricity, making it enough to power millions of homes and businesses and replace countless fossil fuel-based generating plants. Despite the potential growth of DW in the US, deployments are presently hindered by a lack of confidence in resource estimation methods. One potential challenge is that smaller-scale turbines, with hub heights of 40 meters or less, are disproportionately impacted by obstacles such as buildings and vegetation. These obstacles may produce complex wake effects, best modeled with high-fidelity complex fluid dynamics (CFD) models that are too computationally expensive to use for routine siting and resource assessment. Thus, installers today make use of heuristics and simple equations to approximate the impact of obstacles while also leveraging long-term resource data from commercial or publicly available atmospheric models. This study evaluates these historical and commonly used methods alongside new lower-order obstacle models produced from CFD simulations and measurement-based bias correction. The preliminary results from this study show the importance of taking care in the choice and application of mesoscale atmospheric models and the significant value of bias correction using measurements from nearby meteorological towers. Detailed obstacle modeling provides only modest additional gains in performance and, in some cases, can add error, especially at sites where turbines have already been located to avoid obvious impact from upwind obstacles. These findings reinforce the importance of collecting *in situ* measurements and suggest that obstacle models may be better applied in practice to automated or computer-aided siting, rather than in economic wind resource assessments.

1. Introduction

Distributed energy resources (DERs) provide energy solutions to communities, individuals, and regions at or near the location of energy use. In the realm of wind energy, distributed wind (DW) involves 10- to 60-meter hub-height turbines appropriate for deployment in a broad range of application areas including residences, farms, industrial sites, and campuses. Recent studies have found a significant

opportunity to utilize distributed wind to meet US energy needs [1]. Despite this significant opportunity, adoption of such technology is hindered by the cost of projects, confidence in the underlying technology and resources, and availability of incentives. A lack of confidence in the technology is propelled by limited availability of accessible energy production predictions. While their utility-scale counterparts can afford in-depth wind resource site assessments, DW relies on less computationally expensive models and survey methods to make estimations of the available resources. Due to the financial limits of DW siting operations, a higher dependence is found on the results of such models and estimations. However, because of the presence of obstacles and vegetation closer to the hub height of the turbine, modeling the resource is significantly more difficult.

This study looks at methods for modeling obstacle impacts on resource assessment and siting. Models determined to be usable for this application include: (1) classic simplified empirical models well utilized by industry (e.g., [2]), (2) new reduced-order models developed from machine learning and complex fluid dynamics (CFD) simulations [3], and (3) urban dispersion models adapted for use in wind turbine siting [4]. A prior study with EAZ Wind in the Northern Netherlands found newer obstacle models tended to outperform legacy methods and that data-informed approaches, including *a priori* bias correction, resulted in the largest performance gains [5]. The present study significantly extends that work to develop best practices and recommendations in a broad range of topographies, environments, and resource conditions in the continental US. To analyze each approach, this study leverages data from existing installations in collaboration with Bergey Windpower. The chosen turbine sites represent a spatially diverse cohort, with detailed data available from multiple years of continuous operation.

Table 1. Summary of data sources

	Data Source	Variables	Locations	Spatial Resolution	Data Points	Duration
Atmos. Model	WIND Toolkit [8]	Wind speed, direction, temperature, pressure	19	2 km	1.1 M for sites, 613 k for bias	7 years (2007–2013)
Observational Data	Bergey Wind Power Production and Excel 10 Power Curve	Power (kW)	19	N/A	67 k (daily), 6.4 M (5-minute)	13 years, beginning in 2010
	Northern Power Systems Nacelle Anemometer	Wind speed and direction	131	N/A	63.5 million	Varies by site between 2010 and 2023
	USGS 3DEP Lidar Point Cloud [11]	Intensity (elevation)	1-km radius around each turbine	0.34–1.4 m	Billions	Collection date varies by site 2017–2021
	USGS 3DEP Topobathymetric Digital Elevation Model [11]	Elevation	1-km radius around each turbine	Derived from lidar	Billions	Collection date varies by site 2017–2021

	3DBuildings.com Vector Building Data [10]	Polygonal buildings	1-km box around each turbine	N/A	2,463 objects, 17,063 vertices	Downloaded March 2023
GIS Data	Google Satellite Orthoimagery	Polygonal buildings, computed heights	Obstacles within 100-m radius	+/- 5 m	187	Coded March–April 2023

2. Data

Table 1 provides a summary of the data sets used in this study grouping by data category. In the following subsections, we will discuss each data set in detail.

2.1 Reanalysis Mesoscale Atmospheric Data

Wind source data from mesoscale atmospheric models are essential to both drive the obstacle deficit models and for understanding the long-term wind resource and temporal (seasonal and interannual) variability. Prior studies have shown significant differences in the performance of both publicly available and commercial models for DW siting and resource assessment [6]. In this study, we use and evaluate the popular, well understood, and publicly available Wind Integration National Dataset (WIND) Toolkit (WTK). Studies have shown that these data are best used in tandem with vertical and spatial interpolation, and when augmented with *a priori* bias correction from fitted measurements [7]. Thus, we use these methods in tandem with WTK.

The WTK provides wind resource data at relatively high horizontal (2-km), vertical (generally every 20 m up to 200 m), and temporal (5-min) resolution. The WTK meteorological data set, a product collaboration between the National Renewable Energy Laboratory and 3Tier, was created using the Weather Research and Forecasting (WRF) model version 3.4.1. The WRF model was initialized and forced at the boundaries using the European Centre for Medium-Range Weather Forecasts Interim Reanalysis data set with the model terrain, roughness, and soil properties sourced from the U.S. Geological Survey (USGS) GTOPO30 data. The model physics included the Noah land surface model, the Yonsei University boundary layer parameterization, and topographic wind enhancement [8].

2.2 Production Data from Turbines

Production data for the Bergey Wind Power Excel 10 turbines are accessed through a portal created by APRS World, LLC [9]. The Bergey production collection includes daily turbine generation data (kWh energy) along with turbine diagnostics that provide insight into turbine availability and fault occurrences. Beginning in July 2018, the Bergey energy generation and fault timeseries switch to 10-second resolution. Data used in this study are aligned to the corresponding highest-resolution wind resource data available, hourly for the WTK.

Bergey turbines report various error codes including curtailment for grid conditions ("soft grid"), reset, manual stop, and general fault. We exclude those days from this study with >1% error or curtailment, requiring no cumulative error greater than 15 minutes. We choose to include those days with 0 kW production but no reported errors, which we assume are days where the wind is below the cut-in. The Excel 10 turbines used in this study have a maximum power output of 12.5 kW (at approximately 16 m/s) and a cut-in wind speed of 2.5 m/s.
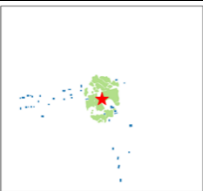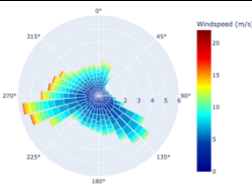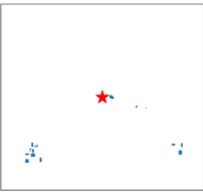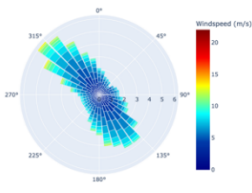
2.3 Obstacle Data
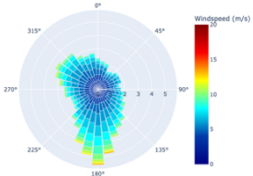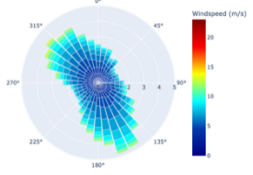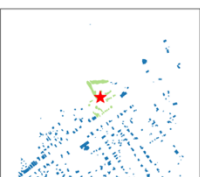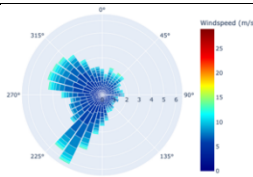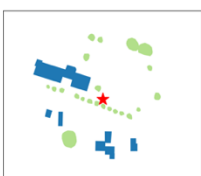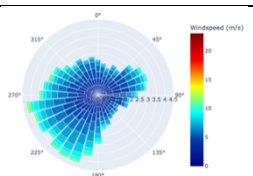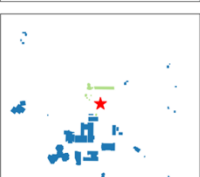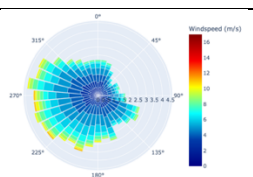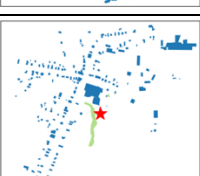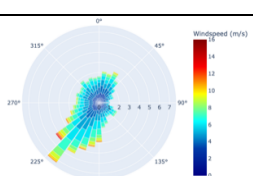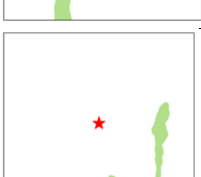
To curate obstacle data for our models to use for each site, we combined a commercial building footprint data set from 3dBuildings.com with height data calculated using recently released USGS 3DEP lidar point cloud data [10,11]. The process for creating the final obstacle model for each site is as follows:

1. **Footprints**: Building footprints are obtained for a 1 km x 1 km bounding box around each turbine using a commercial service. These data are downloaded as a GeoJSON file using the API.
2. **Lidar:** The USGS 3DES lidar data index is used to identify lidar data availability at the turbine location. If this data is at least QL2 or better (approximately 0.7-m accuracy), we download the corresponding metadata and point cloud data from the associated server location. All tiles whose extent overlaps with a 1-km diameter circle around the turbine are downloaded.
3. **DEM:** Using the same index, raster topobathymetric Digital Elevation Model (DEM) data are downloaded. These data are derived from the lidar data and serve as an estimate surface model from which the obstacles in the point cloud lidar data can be separated.
4. **Annotation:** Google satellite orthoimagery is used to identify missing obstacles from the building footprint files, and these are manually annotated. In addition, large vegetative objects (trees, shrubs, hedgerows, and forested areas) are added at this time. These manual annotations are done on a best-effort basis. Polygons are drawn around each apparent obstacle interactively using the QGIS tool within a 100-m diameter area around each turbine [12].
5. **Zonal Statistics:** Heights of all obstacles are calculated using the lidar data, which is rasterized, aligned with the DEM and zonal statistics are calculated for each polygon (building or vegetation footprint). As our models assume constant height obstacles, we utilize the median point cloud height for buildings and the maximum point cloud height above the surface for trees.

While this method works well, there are boundary cases that are worth mentioning. Because orthoimagery is from the current year (2023) and the lidar data may have been collect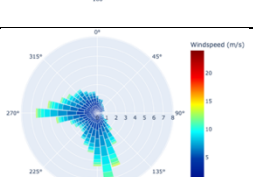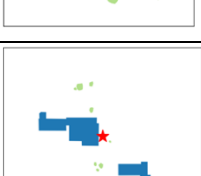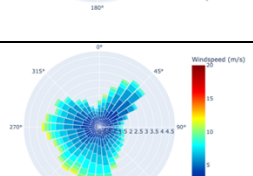ed years earlier, and at a different year for each site, there is the possibility that obstacles will be identified that are not present at the time of lidar scanning. In practice, any obstacle with a calculated height less than 2 m is excluded. Relatedly, because each GIS data source is collected at a singular point in time, it is impossible to fully understand how changes in obstacles over the course of several years may have impacted a turbine differently, e.g., construction, demolition, or simply the growth of trees and vegetation. We ignore this possible impact, assuming that point-in-time obstacle assessment is close enough for practical purposes and represents a realistic use case.

Table 2. Summary of Validation Sites providing wind rose and simplified obstacle maps. The obstacle maps show vegetation in green and built obstacles (buildings primarily) in blue. The location of the turbine is at the center of each image.

ID	State	Hub Height (m)	Land Use	Obstacles Within 500-m Radius of Turbine	Obstacles Within 100-m Radius of Turbine	Wind Rose
t034	NY	31	Forest			
t041	IA	37	Cropland			

t133	IL	37	Cropland			
t139	IA	31	Cropland			
t140	NY	37	Cropland and suburban residential			
t169	IN	37	Suburban residential			
t170	MI	24	Forest, cropland, and rural residential			
t182	NY	37	Forest, cropland, and rural residential			
t192	VT	43	Cropland and forest			
t207	IL	37	Suburban residential			

3. Methods

The design of our validation experiment is as follows:

1. **Select candidate validation sites** in representative locales and obtain obstacle data. For this study we identified 10 sites with sufficient data, metadata, and representing a diverse cohort for analyses. Details including obstacles, wind rose, and terrain type are provided in Table 2.
2. **Gather data from the WTK** and perform vertical and spatial interpolation as appropriate following the methodology in [7].
3. **Bias-correct the WTK data** using site-proximal measurement data. We evaluate scenarios both with and without bias correction.
4. **Assess obstacle impacts** with each model using resource data inputs (wind speeds and directions) and obstacle data. We evaluate scenarios both with and without these models.

3.1 Bias Correction

Bias correction is the process of *a priori* adjustment of the input wind resource data using data from nearby meteorological tower data that overlaps temporally with the resource data. To provide a bias correction for our sites, we consulted with Northern Power Systems (NPS) to utilize a data set of 93 nacelle-mounted anemometers (see Table 1) having at least one year of continuous data. These data are broadly distributed across the central and eastern US, Alaska, and Hawaii with a temporal resolution of 10 minutes and height of 30–37 m. For each validation site, we select those NPS data within a 200-km radius. The availability of data for bias correction is a significant factor in the selection of validation sites, limiting our study to the 10 sites selected. While we have evaluated the inclusion of meteorological data from diverse sites (including airports), we believe these data to be best-in-class, both for broad spatial distribution and consistency in instrument and measurement quality.

We follow the same method of multivariate least squares linear bias correction proposed and evaluated in our prior work in the Northern Netherlands [5]. This involves fitting a multiple linear regression with the following form (Equation 1):

$$w_{obs} = x_0 + w_{wtk}x_1 + d_{wtk}x_2 + hx_3 + mx_4,$$

Equation 1: Multivariate least squares linear bias correction.

Where w_{obs} is the observed wind speed at the meteorological tower, w_{wtk} is the WTK estimate for the wind speed, d_{wtk} is the direction of the model data in degrees, h is the hour of the day (0-23), and m is the month of the year (1-12). Values for the coefficients x_0, x_1, x_2, x_3, and x_4 are fitted with least squares regression. We have found in repeated experiments that the wind speed and direction contribute the greatest information to this fit. However, the hour and month do slightly improve the fidelity of the fit overall. Similarly, nonlinear and nonparametric methods have not shown a meaningful performance improvement compared to this linear approach. For each meteorological tower, we fit this model using spatially and vertically interpolated WTK data at that site. Finally, we apply this model to wind resource data at the site of the nearby wind turbine producing the bias corrected estimates.

3.2 Obstacle Models

We evaluate four obstacle models in this study—three from the literature and two of novel design:
1. **Perera:** This classic model provides a closed form equation for the deficit in velocity that occurs behind a thin (in wind direction), infinite-length obstacle of a given porosity such as a fence or hedgerow. This model, designed in 1981 from wind tunnel experiments, was tailored to the scenario in which the wind was perpendicular to the length of the given obstacle [2].
2. **Shelter and Shelter+:** In addition to the classic Perera model, a variety of extensions exist, one of which is the SHELTER model proposed as part of the WaSP toolkit [13]. This model allows for finite obstacle lengths to be added, which can better model buildings and vegetation. Shelter+ improves Shelter by allowing rotation in the simplified obstacles. Within this study, implementation of these models was performed following the literature descriptions.

3. **PILOWF:** As a potential improvement using more modern methods, we developed the Physics-Informed Low Order Wake Flow (PILOWF) model for predicting wake characteristics behind buildings. To train the model, we used an extensive data set obtained from Reynolds-averaged Navier-Stokes (RANS) simulations, focusing on the flow structure in the wake of buildings [3].
4. **QUIC:** The Quick Urban & Industrial Complex (QUIC) dispersion modeling system was designed to compute wind fields in dense built-up urban areas [4]. QUIC can run on a laptop with one simulation, requiring only seconds to minutes, and has been demonstrated to predict wind fields that were comparable to CFD modeling results for practical applications [14]. To support DW applications, we improved QUIC with a diffusive wake model that extends both laterally from the sides and vertically above the top of the building using machine-learning techniques on time-averaged high-fidelity Large Eddy Simulation.

Each obstacle model is evaluated with four different sets of obstacles: (1) bldgsonly: all buildings within 1 km, (2) bldgsonly_100m: all buildings within 100 m, (2) treesasbldgs_100m: buildings and vegetation within 100 m, (3) treesasbldgs: all buildings within 1km and vegetation within 100 m.

3.3. Error Metrics

To support consistent and comparable evaluation of model performance across diverse sites, we propose the nomenclature in Table 3. The metric, *f,* selected here is normalized to daily production, allowing comparability between sites with different performance. We also require error metrics that can be averaged without canceling error. Thus, we define the following quantities:

Table 3: Variable Definitions and Error Metrics

Name	Notation	Description
Predicted energy	e	Daily energy production estimates obtained by running input wind speeds through the turbine's power curve and aggregating hourly.
Energy measurements	e^*	Daily energy production observed. These values are treated as "ground truth" for validation purposes.
Performance factor, f	$f = \dfrac{\sum e}{\sum e^*}$	Energy estimates scaled by measurements for a given time frame. $f = 1$ would occur when estimates perfectly match observations. Values below 1 indicate a net underestimation, whereas values above 1 suggest overestimation of the wind speed (resource).
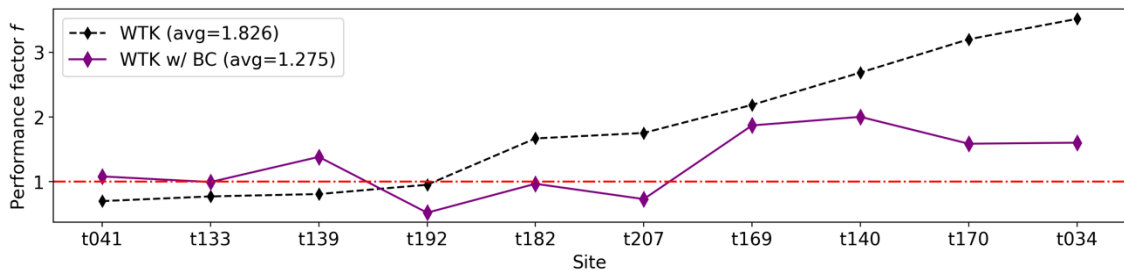
Figure 1: Baseline performance for WTK at each site, with and without bias correction. Bias correction shows improvement in all but two cases and an average improvement of 55%.

4. Results

To evaluate the performance of the models, we begin by defining the baseline performance of the mesoscale (WTK) data alone and with the addition of bias correction. Figure 1 shows these results. For the sites studied, the uncorrected WTK shows a positive bias (overestimation) for most sites. Applying
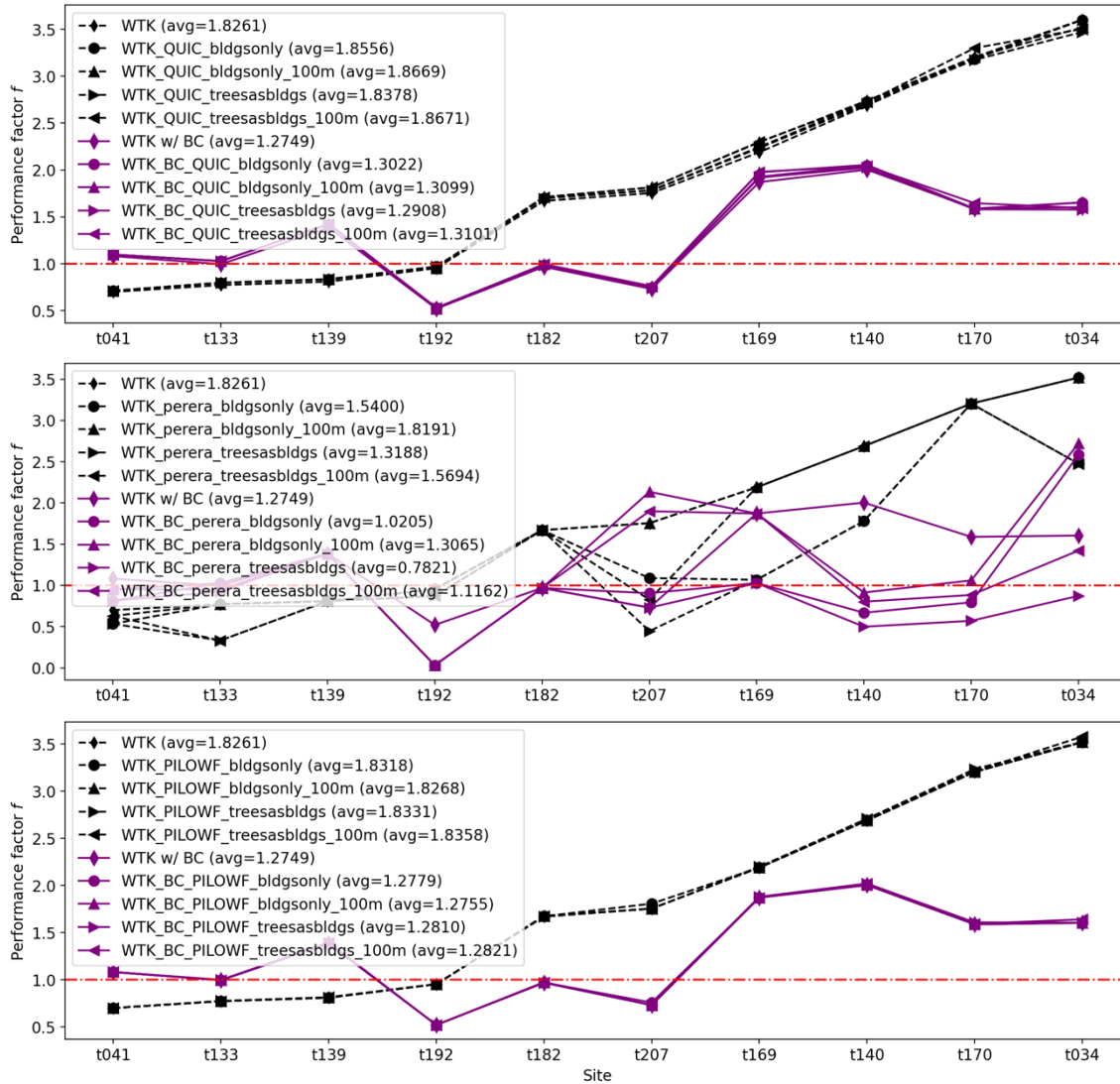
Figure 2: Performance for the obstacle models studied: (top) QUIC, (middle) Perera, and (bottom) PILOWF.

bias correction improves the accuracy of the WTK at nearly all sites. Notably, sites t034 and t140 are coastal sites—the former on the coast of Lake Ontario, the latter on the Atlantic seashore. At these sites, the WTK significantly overestimates windspeed, presumably due to the grid including the water, but bias correction significantly improves the result. Similarly, t170 and t169 are sites in suburban areas and all happen to be on college campuses with appreciable surrounding infrastructure. These results suggest that on average a 55% improvement in performance may be achieved with bias correction, and at some sites (e.g., t034 or t170) the gain can be much greater (e.g., 150-200%).

Figure 2 provides results for the addition of obstacle modeling in overall performance for each of our sites, and Table 4 provides aggregate results for the bias-corrected case. For brevity, Shelter and Shelter+ have been excluded but show performance very similar to Perera. We can see that the more advanced obstacle models are very conservative in their calculation of velocity deficit, suggesting very minor adjustments that make little impact on the overall performance. The Perera family of models make larger adjustments, improving the performance for some sites, but worsening it for others. While the Perera model is the best-performing overall, we suspect that this may be due to it having benefitted from the typical overestimate of the wind resource in the source WTK data, whereas the use of a different mesoscopic data may lead it to underestimate in these same conditions.

Table 4: Results for all obstacle models with bias correction.

Wind Source	Obstacle Model	Obstacle Group	Avg. Performance Factor f
Bias-Corrected WTK	None	none	1.27
	PILOWF	bldgsonly	1.28
		bldgsonly_100m	1.28
		treesasbldgs	1.28
		treesasbldgs_100m	1.28
	QUIC	bldgsonly	1.3
		bldgsonly_100m	1.31
		treesasbldgs	1.29
		treesasbldgs_100m	1.31
	Perera	bldgsonly	1.02
		bldgsonly_100m	1.31
		treesasbldgs	0.78
		treesasbldgs_100m	1.12
	Shelter	bldgsonly	1.35
		bldgsonly_100m	1.36
		treesasbldgs	1.25
		treesasbldgs_100m	1.29
	Shelter+	bldgsonly	1.36
		bldgsonly_100m	1.37
		treesasbldgs	1.3
		treesasbldgs_100m	1.33

Evaluating the choice of which obstacles to include does not lead to appreciable differences for most of the models. Indeed, the PILOWF model shows almost identical performance, regardless of which obstacles are included, and does worsen the performance on average when vegetation is included. QUIC behaves similarly, making only very minor adjustments, yet worsening the performance overall. We suspect this may be due to the QUIC model overemphasizing the diffusive wake (wind speed acceleration over obstacles) and this combining with the starting overestimation of the WTK. QUIC may perform better with resource data exhibiting lower starting positive bias. Shelter and Shelter+ appear to make more cautious adjustments compared to the baseline Perera model, but do not make a significant improvement in doing so. As mentioned above, the performance of Perera is best on average, but the high degree of variability in performance with differing obstacle input sets suggest caution should be advised when using this model in practice. Simply put, the Perera model reduces the wind speed proportionally to the number of obstacles included, which works here due to the positive bias of the WTK but may not in other scenarios.

5. Conclusions

In this study, we performed a first-of-its-kind evaluation of current best-effort methods to estimate performance of distributed wind turbines in realistic environments in the continental US. Despite the observation that smaller turbines are disproportionately impacted by wake turbulence from terrestrial obstacles, our findings show that much greater gains in performance estimation can be obtained by improving atmospheric, mesoscopic inputs, as compared to modeling obstacle impacts. In alignment

with this finding, measurement-based bias correction appears to offer consistent gain in performance estimation and siting. Including detailed obstacle models, however—whether modern and advanced, or staid examples from the literature—appears to decrease accuracy as often as it improves it and demonstrates a concerning sensitivity to the size and composition of the obstacle specifications used. Taken together, these results suggest necessary caution in application of these models, additional work to understand where they are best applied, and renewed and ongoing investment in measurement.

Although this study is the most comprehensive of its kind, there are limitations that merit discussion. All turbines used in this validation were presumably sited to minimize impact from surrounding obstacles. Because of this, we can assume the apparent impact due to obstacles is minimized *a priori,* putting the obstacle models at a disadvantage in creating a meaningful impact. Secondly, our study data set has some limitations in terms of spatial diversity due to our inability to locate suitable turbines or meteorological towers—this is especially true for the western US. Lastly, we chose to use the WTK due to its popularity for applications in this area and high fidelity despite it being a 7-year data set (not 20, as is industry standard). While we cannot practically address the first concern, we intend to address the latter two in future work by incorporating additional bias correction data and comparative re-analyses data sets (e.g., WTK-LED, WTK's forthcoming successor, and ERA5).

Based on these findings, installers of distributed wind turbines are best advised to combine the highest-accuracy long-term wind resource data they can obtain with statistical bias correction using either nearby meteorological tower data or, ideally, short-term measurement deployment. If no nearby observations exist to perform bias correction, it is recommended to consult multiple wind resource data sets to obtain a range of estimates. Based on our findings here, we cannot advocate for one obstacle model over another, but we advise comparing multiple models if choosing to include obstacle assessment in resource estimation. In future work, we expect to further evaluate obstacle models for the purpose of automated turbine siting and sizing, which may be an application better suited to their abilities.

References

[1] McCabe K, Prasanna A, Lockshin J, et al 2022 Distributed Wind Energy Futures Study. *National Renewable Energy Laboratory* NREL/TP-7A40-82519 Golden, CO

[2] Perera MDAES 1981. *Journal of Wind Engineering and Industrial Aerodynamics*, 9, 93-104

[3] Fytanidis DK, Maulik R, Balakrishnan R, Kotamarthi R 2021, A physics-informed data-driven low order model for the wind velocity deficit at the wake of isolated buildings *Argonne National Laboratory*, Report ANL-21/24.

[4] Brown M, Gowardhan A, Nelson M, Williams M, Pardyjak E 2013 *International Journal of Environment and Pollution* 52 263–287

[5] Phillips C, Sheridan L, Conry P, Fytanidis DK, Duplyakin D et al. 2022 *Wind Energy Science* 7 1153–1169

[6] Sheridan LM, Phillips C, Orrell AC, Berg LK, Tinnesand H, et al. 2022 *Wind Energy Science* 7 659-676

[7] Duplyakin D, Zisman S, Phillips C, Tinnesand H 2021 Bias Characterization, Vertical Interpolation, and Horizontal Interpolation for Distributed Wind Siting Using Mesoscale Wind Resource Estimates *National Renewable Energy Laboratory* NREL/TP-2C00-78412

[8] Draxl C, Clifton A, Hodge BM, McCaa J 2015 *Applied Energy* 151 355-366

[9] APRS World http://www.aprsworld.com/

[10] 3DBuildings https://3dbuildings.com/

[11] USGS 3DEP Product Metadata https://www.usgs.gov/

[12] QGIS Association http://www.qgis.org

[13] Astroup P, Larsen SE 1999 WAsP Engineering Flow Model for Wind over Land and Sea. *Riso National Laboratory* Roskilde, Denmark

[14] Neophytou M, Gowardhan A, Brown M 2011 *International Journal of Wind Engineering and Industrial Aerodynamics* 99 4 357-368