Wind_Farm_Site_Suitability_CONUS
Model Description and Instructions

Model Developer: Joshua "Jay" Wimhurst

Contact Details: *Email: Joshua.J.Wimhurst-1@ou.edu; Phone: (+1)405-255-2075*

Position: Graduate Student (ABD) at the University of Oklahoma, Norman, OK, United States.

Essential Reading to Understand the Model's Construction and Interpretation:

Insert links to the papers here once published.

Recommended Software: Python 3.9 (or higher) and ArcGIS Pro 3.0.0 (or higher).

Model Overview:

The objective of my doctoral dissertation was to build a logistic regression model that projects the most suitable places for constructing commercial wind farms across the Conterminous United States (CONUS) over the next 30 years. The model builds off of existing studies that have performed similar Wind Farm Site Suitability (WiFSS) modeling in other contexts (e.g., Mann et al., 2012; Foley, 2018; Harper et al., 2019) with this model introducing several additions to the overall approach:

1. Expansion of the predictors used in WiFSS modeling. This model introduces economic, social, and political predictors that have not been used in previous examples, such as registered state government lobbyists, Independent System Operators (ISOs), and retrainable workforces. This expansion is intended to capture a broader range of processes that impact wind farm siting decisions.

2. Incorporation of Receiver Operating Characteristics (ROCs) and confusion matrices to validate the model's performance. Previous WiFSS modeling studies have not validated the accuracy of their constructed suitability surfaces against the locations of existing commercial wind farms. Doing so is important because these wind farm locations are what are used to both train and test these models. This model uses ROCs to determine the probability threshold at which the model's ability to correctly classify locations as containing an existing wind farm maximizes. The model also uses confusion matrices to summarize the total number of correctly predicted wind farm locations in a given state and/or the CONUS, and to express the model's propensity for Type 1 (false positive) and Type 2 (false negative) classification errors.

3. The model is informed by an aggregated dataset prepared at 20 different grid cell resolutions, with the probabilistic occurrence of a wind farm evaluated within each grid cell. In order to meet end-user demands to investigate WiFSS over multiple locations and for multiple wind farm project sizes (in Megawatts), the predictors sourced for this model have been aggregated at a large number of grid cell sizes. Each grid cell size

represents a typical commercial wind farm capacity found in the CONUS based on capacities in the United States Wind Turbine Database (2022). The grid cells are hexagon-shaped in order to reduce the model's sensitivity to edge effects, and to provide a better approximation of radial distance from important physical features (e.g., nearest transmission line, nearest airport) than square grid cells can.

4. Provision of multiple predictor configurations. End-users have the option of trialing four different predictor combinations: *Full* (all predictors available), *No_Wind* (all predictors but wind speed), *Wind_Only* (wind speed is the only predictor), and *Reduced* (a refined predictor set that minimizes overfitting risk). These configurations allow end-users to investigate the predictors in which they are most interested in incorporating into the model, while also contrasting the configurations' propensity for (in)accurate predictions.

5. Incorporation of a cellular automata that advances the state of grid cells iteratively, thus representing a temporal component that allows the model to project grid cells most suited to acquire wind farms first, based on the probabilities computed by the logistic regression equation, model constraints, and neighborhood effects. This use of logistic regression-cellular automata follows examples frequently applied in urban expansion modeling (e.g., Shahbazian et al., 2019; Shu et al., 2020).

6. Integrating scenarios that modify the coefficients of the fitted logistic regression equation as the cellular automata is iterated. Doing so allows for a range of futures for the timing and location of wind farm installations to be produced when performing multiple model runs.

This model is ultimately designed with end-user utility in mind, such that anyone who downloads the model and its accompanying datasets can investigate (for a given state or the CONUS) what the most suitable locations for constructing commercial wind farms are, for their given project capacity and predictor configuration(s) of interest.

Outputs Generated by this Model:

All outputs generated by this model are summarized in the Console Output generated from its execution using Python (the repository contains LR_Equation_Example_Console_Output.pdf as an example of executing the logistic regression code, as well as CA_Model_Example_Console_Output.pdf produced by running the cellular automata code). These outputs are constructed for each of the four predictor configurations for which the user decides to run the model. The following information is contained within these console outputs:

- Predictors removed from the model due to the crucial assumptions of logistic regression. Application of a Box-Tidwell test to eliminate continuous predictors that do not possess a linear relationship with the logit of the dependent variable (Box and Tidwell, 1962). Variance Inflation Predictor (VIF) calculations to remove multicollinear predictors (Midi et al., 2010), i.e., predictors that do not have independent effects on whether a grid cell contains a commercial wind farm (VIF < 10 for retention in the model). Application of a Cook's Distance Test to remove grid cells that represent outliers in terms of their value for a single predictor (Martin and Pardo, 2009); a missing grid cell within a study domain is indicative of one that was removed by this test.

- Results from maximizing the model's median goodness-of-fit across 30 repeats of its calibration step. This goodness-of-fit is expressed in terms of log-likelihood ratios and McFadden's Adjusted Pseudo R-Squared statistics. From this goodness-of-fit computation, the obtained coefficients are re-expressed as odds ratios to generate odds ratio charts (OR > 1 for green bars, OR < 1 for red bars, whiskers represent the interquartile range of obtained ORs from the 30 calibration step repeats).

- ROC curves from 30 repeats of the model's validation step. The median and range of probability thresholds, and Area Under Curve statistics, are generated from each of these repeats. Confusion matrices using the same validation data are also constructed, showing the number and percentage of grid cells that are (in)correctly predicted as containing a commercial wind farm.

- Application of the model across all grid cells (having trained and tested the model) is expressed using boxplots, with 4 boxplots constructed showing the range of probabilities of each grid cell classified as true positive, false positive, true negative, and false negative. Statistically significant differences in the rank order of the true and false positive, and true and false negative, grid cells are assessed using a Mann-Whitney U-test.

- The probabilities computed for all grid cells are applied to the hexagonal grid cells to create the final WiFSS surface. The WiFSS surface can be viewed in ArcGIS Pro or any other GIS Map Software package. The WiFSS surface's accompanying attribute table contains 4 crucial fields:

    - The probability of each grid cell containing a commercial wind farm (*Probab*).

    - The classification of each grid cell as true/false positive or true/false negative (*Cell_State*).

    - The Getis-Ord (Getis and Ord, 1992) statistic computed for each grid cell to assess the statistical significance of clustering of the true positive and false positive grid cells (*GiZScore*, *GiPValue*).

- The values of the modified coefficients in each scenario are also provided, with each of the scenarios defined as follows (Coefficients do not change in the DEFAULT scenario. Percent changes in coefficients are fully customizable for the predictors of interest in the CUSTOM scenario. In all other scenarios, coefficients change by ±10% in each iteration):

1. DEFAULT: The coefficients of predictors remain constant and changes are driven solely by neighborhood effects.
2. CLIMATE_CHANGE: Temperature and wind speed increase, and bird and bat habitats are increasingly threatened.
3. DEMOGRAPHIC_CHANGES: Demographics that are statistically more supportive of wind energy projects comprise a greater amount of local populations.
4. SOCIOPOLITICAL_LANDSCAPE: Support for wind energy development among politicians and the electorate increases.
5. CHANGING_ENERGY_ECONOMIES: Older forms of energy generation age out, and a demand grows for green energy and green jobs.
6. NEW_INFRASTRUCTURE: Roads and transmission lines are built to support of new commercial wind farms.
7. NATURAL_AND_CULTURAL_PROTECTION: Protection of land that is historically, culturally, or environmentally significant is prioritized as commercial wind energy development continues.
8. URBAN_PROTECTION: Wind energy development continues at a distance set far enough away from industrial and domestic activities.
9. CUSTOM: A unique set of changes to coefficients can be made by the user.
10. NATIONWIDE: In model runs performed for the CONUS, predictors with effects at a nationwide level, such as legislation in effect, lobbyism, and land value can be implemented as an extra scenario.

- Quantity and Allocation Disagreement Index (QADI) tables that put numbers to the difference between maps produced between a Null (no predictors, intercept only) version of the model and each predictor configuration. These tables are a component of the model's sensitivity analysis, such that greater disagreement implies greater effect of the logistic regression equation on projections than the constraints or neighborhood effects.

Important Instructions and Caveats when Running this Model:

1. The two scripts are presented separately in case one wishes to only predict current WiFSS using the logistic regression equation (i.e., the *LR_Equation_Code.py* script). However, if one also wishes to project future wind farm locations, **the *LR_Equation_Code.py* script MUST be executed before the CA_Model_Code.py script, so that the fitted equation coefficients and intercept can be derived.**

2. Line 50 of the *LR_Equation_Code.py* script, and Line 42 of the CA_Model_Code.py script (see repository) should be filled in by the user prior to running the model. The relevant directories, as well as folders within that each directory for generating figures (Odds Ratio charts, Confusion Matrices, ROC curves, boxplots), coefficients and intercepts from fitting the model (needed for the Cellular Automata script) and WiFSS surfaces, should be set up before any attempt at model runs.

   a. Before running the *LR_Equation_Code.py* script, make sure folders named WiFSS_Surfaces, Coefficients, Intercepts and Figures are created, with those for Coefficients and Intercepts including sub-folders with the selected state (or CONUS) name.
   b. Before running the *CA_Model_Code.py* script, make sure folders named WiFSS_Surfaces, WiFSS_Future_Surfaces, Coefficients, Intercepts, QADI_Tables, Defined_Constraints, Defined_Neighborhoods, and Constraints_and_Neighborhood_Effects are created, with those for Coefficients Intercepts, and QADI_Tables including sub-folders with the selected state (or CONUS) name.

3. Ensure that the downloaded data for a given state and/or the CONUS have been extracted from their zipped folder and placed into a directory *LR_Equation_Code.py* script.

4. Ensure that all packages listed at the top of both the *LR_Equation_Code.py and CA_Model_Code.py* scripts have been installed. Of particular importance is the ArcPy package; this model was optimized for execution in the Spyder IDE (Python 3.9) sourced from ArcGIS Pro (Version 3.0.0). The following guide is useful for installing ArcPy: [Installing spyder IDE for ArcPro - Esri Community](#).

5. This model makes use of user inputs in order to inform the model of the study area (state or CONUS) grid cell resolution, predictor configurations, and the predictors to drop based on logistic regression's assumptions. Be sure to read the user input instructions in your Python IDE's console window carefully when starting a model run.

6. Other than the user inputs, the model is set up to print and save its console output to a PDF file. This console output is overwritten at the end of each model run, so make sure to rename the PDF file should you wish to retain it for later use. An example Console Output can be found in the repository.

7. This model cannot be executed over individual states that do not possess at least two grid cells that overlap with a commercial wind farm. The model requires grid cells of both classifications (i.e., wind farm and no wind farm) to exist in its training and testing data, meaning that state-level runs cannot be performed for the following 14 states: Alabama, Arkansas, Delaware, Connecticut, Florida, Georgia, Kentucky, Louisiana, Mississippi, New Jersey, Rhode Island, South Carolina, Tennessee, and Virginia. Representation of these states is only possible in CONUS runs of the model. **If wishing to project future wind farm locations for one of these 14 states using the *CA_Model_Code.py* script, the *LR_Equation_Code.py* script must first be executed using the CONUS.**

8. Fifteen of the model's 47 predictors can only be used in CONUS model runs, and not state-level model runs. These 15 predictors are as follows: *Cost_15_19, Farm_15_19, Prop_15_19, In_Tax_Cre, Tax_Prop, Tax_Sale, Numb_Incen, Rep_Wins, Interconn, Net_Meter, Renew_Port, Renew_Targ, Numb_Pols, Foss_Lobbs, Gree_Lobbs.* The datasets available for these predictors do not vary over a single state, meaning they can only be included in CONUS runs, otherwise logistic regression's assumption of independent assumptions is invalidated (le Cessie and van Houwelingen, 1994).

Data Sources for the Model's Predictors:

All data aggregated to the model's hexagonal grid cells have been obtained from publicly available sources. Below is a brief description of each predictor, its significance to the model, and its dataset source. Creative Commons licenses have been detailed where necessary.

Example Layout:

- **Predictor Name** (*Codename in the aggregated datasets*); description. *Role in model.*
  - Citation for dataset.
  - **NOTE** (if relevant).

Environmental Predictors:

- **Critical Habitats** (*Critical*); line and polygon shapefile of threatened and endangered species locations according to the United States Fish & Wildlife Service. *Risks of breeding and ecosystem fragmentation caused by wind turbines.*
  - Ritter Z. USFWS Threatened & Endangered Species Active Critical Habitat Report. United States Fish & Wildlife Service, Environmental Conservation Online System. 2023. https://ecos.fws.gov/ecp/report/table/critical-habitat.html.
- **Historical Landmarks** (*Historical*); polygon and point shapefile of landmarks, buildings, and locations protected under the National Historic Preservation Act. *Culturally and historically important locations are protected from development by federal law.*
  - Stutts M. National Register of Historic Places. National Park Services DataStore, Geospatial Dataset – (Code: 2210280). 2014. https://irma.nps.gov/DataStore/Reference/Profile/2210280.
- **Military Installations** (*Military*); polygon shapefile of land currently occupied for military bases and training operations. *Wind turbines present risks of radar signal disruption and aircraft collision.*
  - TIGER. Military Installation National Shapefile. United States Census Bureau, Data Catalog. 2019. https://catalog.data.gov/dataset/tiger-line-shapefile-2019-nation-u-s-military-installation-national-shapefile#sec-dates.

- **Mining Operations** (*Mining*); polygon and point shapefile of the locations of active and abandoned mines and pits. *Excavation of the ground may destabilize the foundations of an installed wind turbine.*
    - Horton JD, San Juan CA. Prospect- and mine-related features from U.S. Geological Survey 7.5- and 15-minute topographic quadrangle maps of the United States. United States Geological Survey, v8.0. 2022. https://doi.org/10.5066/F78W3CHG.
- **National Parks** (*Nat_Parks*); polygon shapefile of lands currently maintained and preserved by the US National Park Service. *These natural spaces are designated for protection from most forms of land development.*
    - National Park Services (NPS). Administrative Boundaries of National Park System Units – National Geospatial Data Asset (NGDA) NPS National Parks Dataset. National Park Services DataStore, Geospatial Dataset – (Code: 2225713). 2022. https://doi.org/10.57830/2225713.
- **Population Density by County** *(Dens_15_19*); annually averaged tabular data of county-level populations from 2015-2019. *Dense building of homes and settlements makes large wind farm projects difficult if not impossible.*
    - United States Census Bureau Vintage 2019 (USCBV). County Population Totals: 2010-2019 – United States. United States Census Bureau. 2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html.
- **Tribal Lands** (*Trib_Land*); polygon shapefile of lands under the jurisdiction of Native American, Native Hawaiian, and Alaskan Native tribes and people. *Financial barriers to development on tribal land may result from federal bureaucratic inefficiency.*
    - TIGER/Line Shapefiles. Current American Indian/Alaska Native/Native Hawaiian Areas National (AIANNH) National. United Census Bureau, Data Catalog. 2018. https://catalog.data.gov/dataset/tiger-line-shapefile-2018-nation-u-s-current-american-indian-alaska-native-native-hawaiian-area.
- **Wildlife Refuges** (*Wild_Refug*); polygon shapefile of tracts of land representing vulnerable natural ecosystems according to the United States Fish & Wildlife Service. *Risks of breeding and ecosystem fragmentation caused by wind turbines.*
    - Campbell J. FWS National Realty Tracts Simplified. United States Fish & Wildlife Service. 2022. https://gis-fws.opendata.arcgis.com/datasets/fws::fws-national-realty-tracts-simplified/about.

- **Average Elevation** (*Avg_Elevat*); raster file of height of the CONUS above/below sea level at a 1 arc-second (30-meter) resolution (in meters). *Wind speeds generally increase with height above the surface.*
    - United States Geological Survey (USGS). 3DEP Product Metadata – 1 arc-second resolution. USGS National Map Viewer Downloader. 2022. https://apps.nationalmap.gov/downloader/.
- **Average Temperature** (*Avg_Temp*); raster file of average temperature across the CONUS derived from a 30-year climate normal period (in degrees Celsius). *Warmer climates present a lower risk of wind turbine blade icing.*
    - University of Wyoming Department of Ecosystem Science (UWDES). United States Annual Temperature Raster. United States Geological Survey – ScienceBase Catalog. 2016. https://www.sciencebase.gov/catalog/item/57a26dd6e4b006cb45553f7a.
- **Average Wind Speed** (*Avg_Wind*); raster file of average wind speeds 80-meters above the ground across the CONUS (in meters per second). *Stronger winds mean a larger amount of energy that can be captured by a wind turbine.*
    - Draxl C, Clifton A, Hodge B-M, McCaa J. United States Wind Speed at 80-Meter above Surface Level. National Renewable Energy Laboratory – Wind Resource Maps and Data. 2017. https://www.nrel.gov/gis/wind-resource-maps.html.
- **Bat/Bird Habitat Range Count** (*Bat_Count; Bird_Count*); polygon shapefile collection of the habitat ranges of all bat/bird species observed natively over the CONUS. *Wind turbines present collision and habitat disruption risks (plus risk of barotrauma to bats specifically).*
    - United States Geological Survey (USGS). GAP Analysis Project (GAP) – Download Species Range and Predicted Habitat Data. United States Geological Survey – ScienceBase Catalog. 2018. https://gapanalysis.usgs.gov/apps/species-data-download/.
- **Proportion of Rugged Land** (*Prop_Rugg*), raster file of the proportion of land (in each cell of the raster used for *Avg_Elevat*) that possesses a land slope greater than 7%. *A lack of flat land makes construction large wind farms more expensive and physically challenging.*
    - United States Geological Survey (USGS). 3DEP Product Metadata – 1 arc-second resolution. USGS National Map Viewer Downloader. 2022. https://apps.nationalmap.gov/downloader/.
- **Proportion of Undevelopable Land** (*Undev_Land*), raster file of the proportion of land in each cell of the National Land Cover Database (2019) raster that can be classified as urban, forested, river, or wetland. *Wind farm construction on these land types is more difficult.*

- Dewitz J. National Land Cover Database (NLCD) 2019 Products. United States Geological Survey Data Release, v2.0. 2021. https://doi.org/10.5066/P9KZCM54.

Technological Predictors:

- **Nearest Airport** (*Near_Air*); point shapefile of the locations of airports of all sizes, passenger capacities, and functions across the CONUS. *Risk of disruptions to radar signal communication and collision with low-flying aircraft.*
  - United States Department of Transportation (USDoT). Aviation Facilities. Bureau of Transportation Statistics. 2022. https://data-usdot.opendata.arcgis.com/datasets/usdot::aviation-facilities/explore?location=11.590267%2C-1.633860%2C2.86.
- **Nearest Hospital** (*Near_Hosp*); point shapefile of the locations of all hospitals across the CONUS. *Some state governments legislate minimum setback distances from certain types of infrastructure.*
  - United States Geological Survey. USA Hospitals. Esri. 2022[b]. https://www.arcgis.com/home/item.html?id=f114757725a24d8d9ce203f61eaf8f75.
- **Nearest Major Road** (*Near_Roads*); line shapefile of the locations of all interstates and highways across the CONUS. *Constructing and maintaining wind farms is more difficult/expensive if roads must be built to access the construction site.*
  - United States Geological Survey. USGS Small-scale Dataset – 1:1,000,000-Scale Major Roads of the United States 201403 Shapefile. United States Geological Survey – ScienceBase Catalog. 2016. https://www.sciencebase.gov/catalog/item/581d052be4b08da350d524ce.
- **Nearest Major Transmission Line** (*Near_Trans*); line shapefile of the locations of all high-voltage (>69 kilovolt) electricity powerlines across the CONUS. *Connecting a wind farm to the electricity grid is more expensive without nearby existing transmission lines.*
  - HIFLD. Electric Power Transmission Lines. Homeland Infrastructure Foundation-Level Data (HIFLD). 2022. https://hifld-geoplatform.opendata.arcgis.com/datasets/electric-power-transmission-lines/explore.
- **Nearest School** (*Near_Hosp*); point shapefiles of the locations of all schools (public and private) across the CONUS. *Some state governments legislate minimum setback distances from certain types of infrastructure.*
  - HIFLD. Public Schools. Homeland Infrastructure Foundation-Level Data (HIFLD). 2022. https://hifld-geoplatform.opendata.arcgis.com/datasets/public-schools/explore.

    - National Center for Education Statistics (NCES). Private School Locations – Current. EDGE Open Data. 2021. https://data-nces.opendata.arcgis.com/datasets/nces::private-school-locations-current/explore.

- **Age of the Nearest (Non-Wind) Power Plant** (*Plant_Year*); point shapefile of the locations of all power stations that do not produce wind energy across the CONUS. *Older power plants are more likely to fail/be too costly to maintain versus a new wind farm.*
    - KTH Royal Institute of Technology in Stockholm (KTH-RITS). Global Power Plant Database. World Resources Institute. 2018. https://datasets.wri.org/dataset/globalpowerplantdatabase.
    - **NOTE:** Licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 United States License by World Resources Institute.
- **Age of the Nearest Wind Farm** (*Farm_Year*); point shapefile of the locations of all wind turbines across the CONUS. *Older existing wind farms suggest a more mature wind energy sector that is equipped to accommodate new wind energy projects.*
    - Hoen BD, Diffendorfer JE, Rand JT, Kramer LA, Garrity CP, Hunt HE. United States Wind Turbine Database. United States Geological Survey, American Clean Power Association, and Lawrence Berkeley National Laboratory, 2022. https://eerscmap.usgs.gov/uswtdb.
    - **NOTE:** This dataset is also used to derive *Wind_Turb*, the binary dependent variable of this logistic regression model, representing whether or not a hexagonal grid cell overlaps with a commercial wind farm.

Economic Predictors:

- **Electricity Cost** (*Cost_15_19*); annually averaged tabular data of state level electricity costs from 2015 to 2019. *Electricity obtained from wind farms generally sells at a lower wholesale price.*
    - Energy Information Administration (EIA). Electric Sales, Revenue, and Average Price – Average Annual Price. US Energy Information Administration. 2021. https://www.eia.gov/electricity/data/state/avgprice_annual.xlsx.
- **Independent System Operators** (*ISO_YN*); polygon shapefile of the areas served by independent system operators. *The active monitoring of electricity transmission and sales better guarantees sales of wind-derived electricity and minimizes wind turbine curtailment.*
    - HIFLD. Independent System Operators. Homeland Infrastructure Foundation-Level Data (HIFLD). 2022. https://hifld-geoplatform.opendata.arcgis.com/maps/geoplatform::independent-system-operators-1.

- **Nearest (Non-Wind) Power Plant** (*Near_Plant*); point shapefile of the locations of all power stations that do not produce wind energy across the CONUS. *A lack of existing nearby wind farms makes wind energy a candidate for creating jobs and selling cheap, local electricity.*
    - KTH Royal Institute of Technology in Stockholm (KTH-RITS). Global Power Plant Database. World Resources Institute. 2018. https://datasets.wri.org/dataset/globalpowerplantdatabase.
    - **NOTE:** Licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 United States License by World Resources Institute.
- **Farmland Value** (*Farm_15_19*); annually averaged tabular data of state-averaged farmland values from 2015 to 2019, according to the United States Department of Agriculture. *Cheap farmland makes wind energy development less economically attractive.*
    - United States Department of Agriculture (USDA). Agricultural Land Values – Annual Reports by the National Agricultural Statistics Service. https://usda.library.cornell.edu/concern/publications/pn89d6567?locale=es; 2023 [accessed 10 January 2023].
- **Property Value** (*Prop_15_19*); annually averaged tabular data of state-averaged property values from 2015 to 2019. *Cheap land for residential and urbanization land development makes wind energy development less economically attractive.*
    - Larson W, Shui J, Davis M, Oliner S. Working Paper 19-01: The Price of Residential Land for Counties, ZIP codes, and Census Tracts in the United States. Federal Housing Finance Agency. 2020. https://www.fhfa.gov/PolicyProgramsResearch/Research/Pages/wp1901.aspx.
- **Investment Tax Credits; Property Tax Exemptions; Sales Tax Abatements** (*In_Tax_Cre; Tax_Prop; Tax_Sale*); tabular data of compiled specific renewable energy and energy efficiency financial incentives that are either current or expired, as offered by state-level governments of utilities within a given state. *Wind energy is more affordable in an economy that offers financial supports for its development.*
    - NC Clean Energy Technology Center (NC-CETC). Database of State Incentives for Renewables & Efficiency (DSIRE). https://www.dsireusa.org/; 2023 [accessed 10 January 2023].
- **Total Number of Incentives** (*Numb_Incen*); tabular data of the number of all current and expired renewable energy and energy efficiency financial incentives offered by a state's government and/or electric utilities. *Wind energy is more affordable in an economy that offers financial supports for its development.*

- NC Clean Energy Technology Center (NC-CETC). Database of State Incentives for Renewables & Efficiency (DSIRE). https://www.dsireusa.org/; 2023 [accessed 10 January 2023].

Political Predictors:

- **Gubernatorial Election Results by State** *(Rep_Wins*); tabular data of the number of times a Republican candidate won each state's governor's election since 1976. *Generally, Republican administrations are less supportive of wind energy development.*
    - National Governors Association (NGA). Former Governors – NGA Archives. https://www.nga.org/former-governors/; 2023 [accessed 10 January 2023].
- **Presidential Election Results by County** (*Dem_Wins*); tabular data of the number of times a Democrat candidate won the popular vote in each states counties since 2000. *Generally, Democrat administrations are more supportive of wind energy development.*
    - MIT Election Data and Science Lab (MIT-EDSL). County Presidential Election Returns 2000-2020. Harvard Dataverse, v11. https://doi.org/10.7910/DVN/VOQCHQ.
- **Interconnection; Net Metering; Renewable Portfolio Standard; Size of Renewable Portfolio Standard Target** (*Interconn; Net_Meter; Renew_Port; Renew_Targ*); tabular data of compiled specific legislation that has been passed by a state's government. *More legislation in effect that supports wind energy development allows for faster sector growth.*
    - NC Clean Energy Technology Center (NC-CETC). Database of State Incentives for Renewables & Efficiency (DSIRE). https://www.dsireusa.org/; 2023 [accessed 10 January 2023].
- **Total Number of Statewide Legislative Pieces** (*Numb_Pols*); tabular data of the number of all pieces of legislation passed by a state's government. *More legislation in effect that supports wind energy development allows for faster sector growth.*
    - NC Clean Energy Technology Center (NC-CETC). Database of State Incentives for Renewables & Efficiency (DSIRE). https://www.dsireusa.org/; 2023 [accessed 10 January 2023].
- **Fossil Fuel Lobbies/Green Lobbies** (*Foss_Lobbs; Gree_Lobbs*); annually averaged tabular data of the number of oil and gas, mining, renewable energy, and pro-environment lobbies that have been registered with a state's government between 2015 and 2019. *Corporations and non-profits with a stake in the legislative process can influence political will in support/opposition of wind energy development.*
    - OpenSecrets. Follow The Money – Lobbyist Link. https://www.followthemoney.org/tools/lobbyist-link; 2023 [accessed 10 January 2023].

- **NOTE:** Licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License by OpenSecrets.

Social Predictors:

- **Employment Type by County** (*Type_15_19*); annually averaged tabular data of the percentage of people employed in utility, construction, and manufacturing jobs in each county between 2015 and 2019, according to the Bureau of Economic Analysis. *More people in these workforces may provide a larger workforce to retrain for work with wind energy.*
    - Bureau of Economic Analysis (BEA). Employment by County, Metro, and Other Areas – Interactive Data. Bureau of Economic Analysis. 2023. https://www.bea.gov/data/employment/employment-county-metro-and-other-areas.
- **Unemployment Rate by County** (*Unem_15_19*); annually averaged tabular data of the percentage of people that are unemployed between 2015 and 2019, according to the Bureau of Labor Statistics. *Greater opportunities exist to create more jobs in areas with high unemployment rates.*
    - Bureau of Labor Statistics (BLS). Local Area Unemployment Statistics – Annual Average Labor Force Data by County. US Bureau of Labor Statistics. 2023. https://www.bls.gov/lau/tables.htm.
- **Percent Female Population; Percent Hispanic Population; Percent of Population Under 25; Percent White Population** (*Fem_15_19; Hisp_15_19; Avg_25; Whit_15_19*); annually averaged tabular data of the percentage of people in each county that fell into these four demographics between 2015 and 2019, according to the United States Census Bureau. *Certain demographics may be correlated with support or opposition to wind energy development.*
    - United States Census Bureau (USCB). County Population by Characteristics: 2010-2019. United States Census Bureau. 2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html.
- **Public Support for Renewable Portfolio Standards** (*supp_2018*); tabular data of county-averaged attitudes toward Renewable Portfolio Standards in 2018. *Geographic differences in support for wind energy may correlate with amounts of existing wind energy installed.*
    - Marlon J, Neyens, L, Jefferson M, Howe P, Mildenberger M, Leiserowitz A. Yale Climate Opinion Maps 2018. Yale Program on Climate Change Communication. 2022. https://climatecommunication.yale.edu/visualizations-data/ycom-us/.