2023

Developing a Standardized GIS Model Capable of Identifying Areas to Implement Wind Power Generation Infrastructure

Namidu Vishwanath De Silva
Minnesota State University, Mankato

Recommended Citation

De Silva, N. V. (2023). Developing a Standardized GIS Model Capable of Identifying Areas to Implement Wind Power Generation Infrastructure [Master’s alternative plan paper, Minnesota State University, Mankato]. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. https://cornerstone.lib.mnsu.edu/etds/1274/

Developing a Standardized GIS Model Capable of Identifying Areas to Implement Wind Power Generation Infrastructure

Namidu Vishwanath De Silva

An Alternative Plan Paper Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

In

Geographic Information Science

Minnesota State University, Mankato

Mankato, Minnesota

April 2023

Acknowledgement

This project took the encouragement and guidance of several individuals whose valuable input allowed me to effectively conduct this research. It would be truly impossible to individually thank each contributor, but this section will illustrate the main people and groups whose support and guidance were vital to this project. Firstly, I would like to thank my graduate advisor, Dr. Fei Yuan. Her efforts and encouragement were the main driving force behind my pursuit of this project and all the positive aspects that were derived from it. Also, I would like to extend gratitude to my APP committee members, Dr. Woo Jang and Dr. Rama Mohapatra for their advice. Additionally, would like to thank my graduate cohort at Minnesota State University, particularly Alex Murray and Jordan Kiesow, who constantly kept me motivated and in high spirits during the more strenuous parts of this project. I would also like to express my gratitude to my place of employment, Yes Energy, and my work team, the power infrastructure (PI) group. The members of the PI team have been some of the key advisors and sources of the energy industry and infrastructure knowledge for me throughout this project. The work and insight gained into the energy industry during my internship experience have allowed me to have a much-needed understanding of the energy industry and the best manner to conduct this research.

Finally, I would like to thank my family, for their constant support and motivation throughout my schooling experience. My mother and eldest sister were mainly my 2 greatest supporters and inspiration in not only pursuing my Master's degree in GIS but also this research project. Their constant reassurance allowed me to maintain my prowess, particularly in the direr times throughout this process. This research would not have been possible without my support system and all the people involved in it. I would like to express my sincerest gratitude to all the people who made this research possible and for all their support and encouragement.

Abstract

This project involves the use of multi-source GIS data, suitability analysis, and spatial modeling techniques to identify the most suitable areas for improving renewable energy infrastructures, primarily focusing on wind energy, within a given region. Previous literature conducted on the subject matter shows research being done in isolated areas or states. The primary objective of this project is to develop a GIS model and scripting tool that has the capability to test any input region to find the most suitable area within the given region that would be ideal for implementing new wind power infrastructure. The success rate of the model outputs was tested against pre-existing wind power infrastructure. The criteria used for the model were derived from consultations with energy industry experts and previous studies that have investigated building new wind power infrastructure in specified regions. Both the commercial ArcGIS Pro and open-source QGIS software were tested and compared. Results indicated an overall model accuracy of about 78% in both testing sites. The distance to transmission lines was found to be the most important factor affecting model performance. The Python scripting tool developed from the model however aims to be a universal tool that can be used in any region to find the most suitable areas for installing new wind power generation infrastructure.

Contents

Introduction

Energy is an important aspect of several decisions made by large entities such as companies or governments. The way this energy is provided for powering entire regions is of utmost importance. With increasing populations, migration, and demand for goods and services, the importance of generating power to fuel public endeavors is at an all-time high. Over the past decades, non-renewable power sources such as coal, petroleum, and natural gas have dominated the power market. Even today, much of the power generated to fuel most infrastructure is based on non-renewable fuel sources. This trend is predicted to gradually decline as some energy companies have started to research and implement more renewable power-producing infrastructure such as wind-based power (Jahangiri, 2016). This is an effort to increase carbon capture and reduce carbon footprints. The overall shift from non-renewables to renewable power generation would be difficult as the current power output from non-renewable power plants greatly outweighs the power output from renewable sources. Hence, building more renewable infrastructure, such as power plants, is vital to support the shift from non-renewables to renewable resource-based power. Building renewable power plants can prove more challenging as they generate power using natural features that are constantly changing and unstable over abbreviated time periods. In this context, finding regions that can be feasible for building renewable power plants will be the basis of this project.

The building of renewable power infrastructure is often credited to the need for power in an area and the availability of the fuel needed to generate the power. It is important to understand that building renewable power infrastructure is often difficult even though there is a significant need. Any powerplant relies on a specific type of fuel to operate. In terms of renewables, fuel can be a highly variable factor, particularly solar and wind power. A net power generation study conducted by the Energy Information Administration (EIA) for the 2022 year showed that 21.5% of all power generated in the United States (U.S.) comes from renewable power plants, but the hope is to improve this value by increasing the overall number of renewable power plants. The majority of the renewable power supplied to the grid is by using wind farms. This is due to wind farms being more reliable and providing a greater power output when compared to solar farms or any other renewable power plant. They can produce power throughout the day and can provide emergency power when there is a sudden influx in demand. However, wind farms also require large areas and can be difficult to build unless certain orographic and geographic features are present in the landscape. As increasingly more countries are planning to improve their wind power infrastructure, there is an increased amount of research on how to determine whether an area is suitable for implementing wind power and building wind farms. In the literature explored, each research article has its own way of determining if an area is suitable for wind power generation. Although most of the primary criteria are similar across the research articles, the values used to meet said criteria all vary from various parts of the world. Moreover, there is no general process on how to generate these suitable areas.

Hence, this project has developed a generalized mechanism that allows users to input values into the criteria and locate areas that can be used to implement new wind power generation infrastructure. A GIS model that integrates the analysis of various geophysical factors was created using ArcGIS Pro. This model not only allows users to determine whether a region has sufficient features to support wind power, but also allows users to find the areas within the region that can be used to develop wind farms based on the criteria and their respective input values. To further improve the usability of the model, a Python script was generated accordingly, which allows for the mechanism to be used on various GIS platforms as well.

Literature Review

The primary issue faced when trying to build new renewable power infrastructure is the feasibility of power output compared to non-renewable power generation infrastructure. Alongside the lower power output, renewable resource-based power generation infrastructure requires at least double the amount of area and resources to meet the power generation quantities of non-renewable resources (van Zalk and Behrens 2018). Several factors affect the installation of renewable power generation infrastructure, some of which include environmental, economic, and human intervention (Höfer 2016). The usage of geospatial data for analytical processes is a long-standing practice among energy companies. However, according to my literature review, each article studied used a GIS model specifically designed for the article's specific study areas. The development of a more standardized model that can be utilized over a variety of landscapes or areas is the basis of this project. The decisions made for implementing renewable infrastructure are a combination of geospatial data, performance models, and economic potential based on multiple factors. The ability to build any form of renewable power generation infrastructure can be attributed to a few primary aspects. These can further be subdivided into even more criterion that supports the infrastructure needs for renewable power generation described above.

The influencing factors behind power production are primarily based on the consumption levels of power itself (Supapo et al. 2021; Cox et al. 2018). The location of implementing new renewable power is further made difficult due to the importance of maintaining a balance between supply and demand. The proposed renewable power plants must have the ability to manage the demands of the region with minimal compromise. This is already questionable as renewable power plants already struggle to generate power outputs like non-renewable sources. The implementation of building any power generation infrastructure stems from two main concepts, the need for increased power, as mentioned earlier, and the abundance of resources or fuel to produce said power while also being eco-friendly (van Zalk and Behrens, 2018). Evidence gathered based on the above two factors shows that wind power will be the most suitable to assist with the shift towards more renewable power generation and ultimately the implementation of more renewable power infrastructure. Consequently, this project investigates how wind power generation infrastructure is built and creates a GIS-based tool that will allow a user to find areas for implementing new wind power generation infrastructure. As more renewable energy is in demand all over the globe, it is vital to find locations that are most viable for placing this infrastructure, including wind generation infrastructure.

This project will focus on the geophysical features that are involved in the siting process when new wind power generation projects are derived. The focus on the geophysical aspects stems from the socioeconomic and anthropogenic aspects of all, new power generation projects being highly difficult to predict. It is no surprise that some power generation projects often do not go beyond the permitting or pre-construction phases, as sudden changes in economic, business and government policy may potentially cause a halt in a project's progress or even the complete termination of the project itself. Power generation projects undergo heavy inspection and screening processes before even acquiring permits to begin construction. This timeline more than likely extends over a year and during said time, the company responsible for the project may undergo changes in project funding, face difficulties attaining land rights, meeting environmental legislation requirements or even changes in the government energy policies. These types of situations can all lead to the overall cancellation of a project and due to their ever-changing nature, it is virtually impossible to truly model and predict these situations. Hence this project aims to place the primary focus on the more stable aspects of power generation, the geophysical land features. Several entities over the past years have been increasing the development of onshore and offshore wind power projects because it is more reliable and has larger power output. The uncertainty in the ability of wind power plants to supply power adequately over different locations is why the location of the power infrastructure itself is considered highly important (Bunodiere and Lee 2020; Pakere 2022). To address this issue, the project aims to test the developed model and tool in two separate locations, which possess different ratios of renewable to nonrenewable power generation plants.

The way grid power currently operates is that the grid relies heavily on predictive models in order to identify and dispatch power to wherever it is needed. These predictive model outputs, coupled with the performance analysis of the pre-existing power generation infrastructure, can be utilized to gain an estimation of how much and where new wind generation power can be implemented (HuiBin et al., 2016). Having pre-existing renewable infrastructure will help in the implementation of new renewable power generation infrastructure, in the case of this project, wind power. The analysis taken by this project involves utilizing more physical and micro-level criteria to identify areas using a GIS model built in ArcGIS Pro 2.9, which was then converted to a Python scripting tool that would require minimal inputs to identify the most fitting areas in a region for the installation of new wind power generation infrastructure. The main criteria observed in the literature conducted on wind energy primarily include slope, regional windspeeds at a given height, proximity to urban and impervious surfaces, distance to major roads, and distance to transmission lines that transport power to and from the power grid (Atici et al. 2015). The mentioned criteria have been noted, across several research articles from around the world, to be the base criteria used to find ideal locations for building new wind power generation plants and infrastructure. Hence, these would be the primary criteria used for the development of the GIS model and script tool in this study. The basis for developing the model and scripting tool is to allow the siting to be as user-friendly as possible. In an effort to provide a model that can be utilized for several situations, this project will also aspire to develop and test the model in an open-source GIS platform and geospatial modelling environment. Apart from the development of the two wind energy siting mechanisms, an additional comparison analysis will be conducted in order to see if the model developed in this will perform optimally if the model is to be utilized in an open-source GIS data manipulation platform. Hence the model will also be developed in an open-source environment and tested using the same parameters as its more privatized software-based counterpart.

Study Area & Data Used

As mentioned earlier, in this study, the produced GIS model and geoprocessing tool were generated and tested in two different locations to assess its overall accuracy. The first iteration of the model was run in the northern state of Minnesota. Minnesota was chosen as this first region as it already possesses a significant wind power generation infrastructure, particularly in the southwestern corner of the state. This region is also smaller in size and will provide an ideal area for a newly developed model (Figure 1). The acquisition of data for Minnesota is also relatively simpler and due to the smaller size of the first study area, the model requires less processing power, so it can be executed with more optimized conditions and speed.

Once this first phase of testing is completed, the model will then be run on a much larger area to observe how it performs on a much larger and more robust dataset (Figure 1). This second phase of testing will be conducted on the entire land area covering the Electric Reliability Council of Texas (ERCOT). This is an organization or Independent System Operator (ISO) that is responsible for the management and upkeep of the entire Texas power grid. Its operational area covers most of Texas and due to the independent nature of the power grid in Texas, it provides an ideal environment for testing newer technology and power infrastructure. ERCOT's independent grid allows companies to host a variety of power generation infrastructure, renewables, or non-renewables, while still retaining the ability to meet the state's power demands most of the time. The independent grid has several drawbacks as well. The most prominent one is the inability to pull power from surrounding ISOs in case of emergencies such as the freeze that occurred in 2021 due to temperatures dropping below the standard operating temperatures of the power infrastructure. Having a large number of renewable power infrastructures that can continuously produce power in emergency events or provide supplemental power if the primary power generation plants were to temporarily fail in any event would be a positive impact of having more renewable power infrastructure (Shorabeh 2022). The shift towards building more renewables has also been greatly evident in ERCOT as it has the greatest amount of renewable power produced in the U.S.

The data acquired for this project came from a multitude of sources. Based on the criteria mentioned earlier, the datasets acquired for the model functioning include a Digital Elevation Model (DEM) of the study areas, land cover data, windspeed datasets, road network, transmission line network, and urban area boundaries. A 30-meter DEM of the entire U.S. was downloaded from the ArcGIS Online portal. The most recent available 2019 landcover raster for the entire was taken from the National Land Cover Database (NLCD). The DEM and landcover rasters were of the same spatial resolution. The road network was taken from the Department of Transportation database and was condensed to only the major road networks in the country. Transmission line data was acquired from Homeland Securities, Homeland Infrastructure Foundation Level Database (HIFLD). Only the larger and higher voltage transmission lines were acquired for this project. For the first stage of testing, the state boundary of Minnesota was acquired from the Minnesota Geospatial Commons. Mean wind speed data for the U.S. has been acquired from the National Renewable Energy Laboratory (NREL) wind energy database. This data came in the form of a raster depicting the windspeed of a region at a given height. Two wind speed rasters were downloaded from the NREL database, windspeed at 40 meters and 100 meters above the ground. Windspeed is a criterion of great importance for site selection for new wind power infrastructure, hence two datasets were acquired (Nasehi 2018). The majority of the data used for both stages of testing were similar except for the study area boundary. Once the model is completed, it aims to only require 2 inputs, the study area's boundary, and DEM, hence the majority of the data sets acquired in the first testing phase will be used for the second and subsequent testing phases. Models were developed and tested in both the ArcGIS Pro and QGIS software. After the models were developed, they were transferred to Python scripts and then converted to script tools.

Methodology

As mentioned earlier in this paper, this project was conducted using criteria derived from previous literature and from consultations with energy industry experts, to obtain the most suitable areas within the selected region for implementing new wind power generation plants. Models developed in previous research studies have more than often used raster-based datasets in order to generate the areas used for wind energy siting. The area outputs from these pre-existing models are generally single-factor, composite rasters in which the output areas are a result of the cell size of criteria raster datasets input for the data querying process (Szurek et al., 2014). Consequently, the final output raster areas will be heavily reliant on the input raster dataset cell sizes. This could potentially comprise the accuracy of the overall area output by the model. This was mainly done in order to possibly reduce the model's processing time and power requirements. Due to this project having a more generalized criterion and a combination of several research article findings, accuracy and precision are given a higher priority over processing time and speed. Hence it was decided that this model will operate with both raster and vector data, with the final area output from the model being a polygon feature class. This project was conducted in two phases, development, and testing, where the testing phase will be subdivided into two subphases based on the study area. The first phase involved developing the model using the geoprocessing tools and the Model Builder of ArcGIS Pro. The outputs from each of the geoprocessing tools were used in subsequent tools in the model. The model would then output a final area in the form of a polygon feature class showing the areas within the study area that would be ideal for building new wind power infrastructure. The accuracy of the model was assessed using pre-existing data regarding wind farms. The area output by the model was cross-referenced using the locational data for currently operating wind farms and wind turbines. Once the model proved to be operational in the study areas decided, it was converted to a Python script. The Python script was then modified into a geoprocessing script tool that can be applied by any user for their respective study areas. The methods and analysis conducted for each of the phases are outlined in the upcoming sections of this paper. The general steps of the method are illustrated in the flow diagram (Figure 2).

Model Development Phase in ArcGIS Pro

Based on the previous literature conducted on the site selection process for wind farms, a set of base criteria was developed. These base criteria will be the basis of the model and would lay the foundations for the geoprocessing tools to be used in the model. As the final output is to be a polygon feature class, the raster dataset would have to be queried first to only show the areas that adhere to their representative criteria and then be reclassified into binary rasters. The binary raster would then be converted to polygons for the final intersection of the suitable data layers.

The base criteria used for the preliminary development of the model include having an average windspeed greater than 15 miles per hour (7 meters per second) at a height of 40 meters above ground level (Parti et al., 2020). As mentioned earlier, the fuel for a power generation infrastructure is of great importance. The higher rate of wind speed in the region, the greater the amount of power generation by the turbines. As per research conducted by previous studies, the increase of windspeeds by simply 1-2 m/s can increase the voltage of the power output by a wind turbine by a factor of 2 to 3 (Parti et al., 2020). It is expected that each of the testing regions will have different wind speed rates due to changes in elevation, landscape, climate, and overall surface area of the region. Hence the expectation for the output for the required windspeed will differ heavily between the two test regions. Moreover, the ideal operation axis of a wind turbine is within 50 meters of elevation, with a buffer height of plus or minus 10 meters (Chamanehpour et al., 2017). Hence the acquired dataset for windspeeds of greater than 15 miles an hour at 40 meters of elevation will be most suitable for the purposes of this model. Similarly, the land area in which new wind power generation projects are supposed to be constructed can be a factor that could be the deciding factor between a project moving forward for construction or being canceled. In order to make the siting process done by the model more efficient, the landcover classes selected for the model are often deemed as landcovers in which wind power generation is feasible. This factor was confirmed by identifying the landcover classes used by pre-existing wind generation infrastructure and previous research. The land cover requirement for building wind farms must include, barren land, shrub/scrub lands, cultivated cropland, hay or pastures, herbaceous, developed open areas, and developed low intensity (Díaz-Cuevas 2018). The landcover raster is reclassified to only display the mentioned land covers and then converted to a binary raster which would then be converted to a polygon feature class for analysis. A similar approach is taken for identifying the required slope. The DEM of the study area is used to create a slope raster of the study area first. The slope raster is then queried to only show areas that are below 12 degrees of slope, as this is a common value used for slope in previous literature. The queried slope raster is then reclassified to a binary raster that shows the suitable slope areas, which are then converted to a polygon feature class for further usage. For finding the distance from transmission lines, major roads, and urban areas, there were two main options that produce the best results, the buffer tool, or the Euclidean distance tool. Both tools have the ability to create a buffer or a distance map from a feature input layer. This output would then be embedded and processed within the model, alongside its supplementary datasets in order to find the required distance to transmission lines and road networks. Previous models with similar characteristics to the model being developed in this research project have often used Euclidean distance due to its distance-to-feature calculation mechanism, as it produces a raster that shows the distance to features but also this distance is displayed in segments that can be used to represent certain distance intervals, similar to a distance decay from features. The Euclidean distance tool can be set to give an output raster that will have a maximum distance to features and a preset cell size for the raster. The smaller the cell size, the more accurate the output from the raster would be however a smaller cell size will also place immense strain on the model and will require a much greater processing time to generate an output. Comparatively, the buffer tool can produce a distance buffer that is more directly in line with the required distance for this research, but it does not produce an output that could be used for variable distances when testing the model. The usage of the buffer tool will potentially require the model or tool to be run several times with different buffer sizes in order to test which distances will be most suitable for the particular study area.

In order to identify which of the tools would be more reliable in order to gain the most accurate value for distance to features, both tools were run for the transmission line dataset. The transmission line dataset was first clipped to the study area and then input to the Euclidean distance and the buffer tools. Both tools used a threshold of 10,000 meters from transmission lines because this larger value would be ideal based on previous literature, and this would be one of the values used when testing the model's ability to handle large datasets. Both tools generated outputs that were compared against each other in order to find which produced the closest output to the required area output. Both outputs were compared by measuring the cross-sectional area from one edge of the output to its linear opposite edge. This process was conducted in several different areas of the tool outputs. Upon examination, the output from the Euclidean distance tool was observed to incorporate additional amounts of areas into its output resulting in a greater amount of area being deemed as suitable. This occurs as the accuracy of the Euclidean distance tool is determined by the spatial resolution of the input and output raster layers. In addition, it is designed to incorporate the areas that aren't considered as suitable as being completely suitable in order to build an edge that is most representative of the maximum distance to or from the feature, particularly in the edges of the output raster (see Figure 4). The cells on the edge of the Euclidean distance output raster were often only partially suitable but in order to have achieved a proper cell shape, the tool incorporates additional area into the output to form a square cell-based edge. Due to the Euclidean distance tool implementing additional area to the output and taking the upper limits of the output area, the total area from the output from the Euclidean distance tool was larger than the output from the buffer tool. This incorporation of the unnecessary area would be less than ideal as each cell is 1 square kilometer. A potential error in the suitable areas within a desired proximity to features is highly significant in infrastructure development, particularly in terms of the right of way and legally permitted building areas. When developing any form of energy infrastructure assets, a suitable area that is within proximity to certain physical features is not only important but also extremely costly, hence the addition of the increased, potentially unsuitable areas from the Euclidean distance tool would hypothetically compromise the model's overall functioning. Moreover, the output raster from the Euclidean distance tool is in the float raster data type while the majority of the model operates with vector-based datasets. Hence, the output raster will have to be post-processed to convert it to an integer-based raster, which would then have to be converted to a vector dataset in order to be used in the model. The conversion tools used for the aforementioned steps are the Int (Integer tool) and the Raster-To-Polygon conversion tool. These post-processing steps simply get the distance output to be in a format that can be utilized by the model, which increases the required processing time and power of the model. The increased tools and steps result in slower processing times and greater chances of errors or data corruption occurring while the model runs. Due to these issues, the buffer tool was deemed more suitable for the modeling process. Although the buffer tool's output is crude and does not provide an output with a distance decay measurement as the Euclidean distance tool, the output from the buffer does in fact provide a highly accurate depiction of the required area that is within proximity to the transmission line features. The buffer output is also a vector dataset in nature and can be directly implemented into the model without any post-processing. Thus, the buffer tool was to be used in the model being developed in this research project for proximity analysis.

The average distance from transmission lines used in previous literature ranges from 3000 meters to 10,000 meters (Pakere 2022). The buffer tool was incorporated to be used for the proximity analysis functions for the two transmission line distances. Once the buffer tools were implemented into the model, the model is to be run for both the mentioned distances to transmission lines. The urban areas feature also used a buffer tool, alongside the erase tool in order to identify the areas that are suitable for building new power infrastructure. The distance from urban areas was a more dynamic feature as wind farms are known to provide power to smaller towns; however, the larger metropolitan areas require the wind farms to be a certain distance away from the town as they damage the metropolitan area aesthetically. Hence, the urban areas feature was filtered to only include the larger metropolitan cities with populations greater than 50,000 population, while the area around the smaller, less populated urban areas was included in the model as a suitable area. The process was conducted using the "select by attribute" and "copy features" tools. This was conducted as previous literature often included smaller towns and urban areas within closer proximity to the wind farms. This is due to the wind farms often being a viable source of power for smaller towns thus having them in closer proximity was advantageous to the town's inhabitants. This was also the case when observing the pre-existing locations of wind farms with respect to urban area locations. Several wind farms were located within a 1-mile radius of these smaller towns and even some larger towns. Hence for the purpose of this model, a 1-mile (1600 meters) buffer distance away from the major metropolitan areas was to be used.

Once these criteria were derived, the model was produced. For each study area, the model is to output a feature class that would illustrate the areas ideal for building new wind power generation infrastructure (Figure 3). Using the locational data for currently operating wind farms, the ratio of wind farms located within the boundaries of the model's output areas will be an indicator of the model's overall performance. The model aims to attain between 75 to 100 percent of currently operating windfarms to be within the output area by the model. Once the model was formed, the testing phase of the project began.

Model Testing Phase: Minnesota

The primary testing phase of the model was conducted in Minnesota. The developed model was presented with the state boundary of Minnesota, alongside all required data mentioned earlier in this paper that fulfill the base criteria for the site selection process. The model was run using the mentioned criterion values for slope, windspeed, landcover and proximity to transmission lines, major roads, and larger cities. Once the initial output was generated, it was cross-referenced for accuracy using the locations of the pre-existing wind power infrastructure and plants. As the orographic feature criterion remains constant, anthropogenic features such as proximity to urban areas are subject to change over a variety of literature. The distance between transmission lines particularly was altered for different iterations of the model as the distance is inversely proportional to the amount of power being transported from the wind turbine to the grid (Pattanariyankool 2010) (Ghorbani 2017). Hence the model was re-run using different values of proximity for transmission lines, such as 3000 meters and 10000 meters from transmission lines.

The output areas were analyzed to determine the degree of variability between the area outputs from the model for each of the different values of proximity to anthropogenic features. Once each output was analyzed against each other, the accuracy of the models' outputs was tested using the locations of already operational wind farms. The number of wind turbines located within the specific areas predicted by the model was compared to the total number of wind turbines located within the entirety of the study area.

Model Testing Phase: ERCOT

Once the testing phase in Minnesota was complete, the model was re-run using ERCOT as a study area. Compared to the first study area, ERCOT covers a much larger surface area with a variety of land covers and features. This would push the model to process a much larger set of data and test its overall capabilities. The same processes used for the testing phase in Minnesota were applied in ERCOT. The outputs for this testing phase are predicted to be much larger than the outputs received from the first testing phase, mainly because ERCOT covers a larger land area and due to a large number of operational wind farms already present in ERCOT. Using the base criterion for the inputs mentioned earlier in this section and different values for the distance to transmission lines, the model gave three output areas located within ERCOT that can serve as ideal locations for building new wind power infrastructure. The output areas were examined for variation between the amount of area isolated from each of the different model outputs. Once the area variations were analyzed, the output areas were then tested for their overall accuracy.

The accuracy assessment was conducted using the ratio between the number of wind turbines within the area's output from the model against the total number of wind turbines in the study area. The accuracy assessment ratio between the Minnesota study area and the ERCOT study area was then compared against each other. This was done to assess the overall performance of the model when using a larger study area versus a smaller study area. The model's effectiveness with a larger area will determine its usability when given more custom study areas. The interpretation of the model outputs for both study areas is explained in greater detail in the Results and Discussion section of this paper. Once the model outputs were analyzed and the effectiveness of the model was assessed, the entire model was exported to a Python script to convert the model into a geoprocessing tool. This geoprocessing tool hopes to have minimal inputs to determine whether the given area will have ideal conditions for building new wind power infrastructure. The steps involved in the creation of the geoprocessing tool are outlined in the following subsection.

Creating the Geoprocessing tool

Using the Python script exported from the model, the model was converted to a geoprocessing tool. After examining previous literature, it can be stated that most of the orographic features used as criteria for identifying the most ideal areas for building wind power infrastructure remain within a certain range. This range does not have a large margin of variability hence the values for slope, landcover, wind speed, and distance to urban areas were left as the default value from the original model. However, as the distance from transmission lines tends to vary over a greater range, the input for these features will be input into the tool as per the requirements of the user. The model will ultimately be designed to have only four main inputs, namely the study area boundary, study area DEM, distance to transmission lines, and major roads, as well as one output area within the study site suitable for building wind power infrastructure.

The model script was input into the PyCharm IDLE for the editing process to be conducted. The script was configured as a Python 3 script as this version is more compatible with ArcGIS Pro version 2.9. Once the script was imported to PyCharm, the tool parameters were edited accordingly for the conversion process. The "GetParametersAsText(#)" syntax was used to give the script the ability to intake a variety of inputs for the input parameters mentioned earlier. As the outputs of certain tools will be utilized as the input for others, as outlined in the model, the need to have multiple inputs is greatly mitigated. The above-mentioned syntax was embedded into the script in lines or sections of the script where the user will be required to input either a source data layer or a value for the required parameter for 1 of the 6-base criteria. The input layers for the script tool included all the raw data files for each of the base criterion parameters in the model. A secondary set of the inputs via the above syntax was used to input the values for some of the user-defined parameters for the geoprocessing tools. In the effort to make this model effectively function in more than one environment, it must have the ability to allow the user to set the required parameters more easily for some of the criteria. This is due to certain regions of the world using different values for the base criterion. Hence in order for the model to perform in any region, these base criterion values must be defined and input by the user into the model and geoprocessing tool. The GetParametersAsText syntax was deemed as variables for each of the base criterion layer inputs and then the variables were embedded in the specific location within the script tool as inputs for the line of the script that is pertaining to the tool's input. Apart from using the syntax for loading raw data into the tool, the syntax was embedded within the raster calculator tools. This was done in order to allow the user to define the value used for the required windspeed, slope, and distance measure to physical features such as transmission lines or urban areas. The usage of the mentioned syntax in not only the raw data load but also in the individual sub-tools embedded in the final geoprocessing/script tool would allow the model to have the flexibility to be used in several different areas and with various values for the base criterion.

Once the base script was configured in a manner that can be used as a script tool, the user interface of the tool was built using ArcGIS Pro. A new script tool was created and fitted to use the base script from PyCharm to run the model on the back end. The model would be essentially run via the script tool with the input parameters. The user interface for the tool was designed to intake the study area boundary as a vector file, the study area DEM as a Raster, and the distance parameters as numeric values (Figure 14). Once the tool and user interface were developed, the base script from PyCharm was linked to the script tool so that it would be able to interact with ArcGIS Pro and run the model whenever the appropriate inputs were provided. Once the geoprocessing tool was set up, it was run using the same inputs for the Minnesota study area. The output of this tool was then compared to that provided by the stand-alone model in the initial testing phase.

QGIS Model development

A second factor that was taken into consideration, in efforts to make this model and geoprocessing tool as user-friendly as possible was to redevelop it in a more open-source environment. Due to ESRI ArcGIS being more expensive and firms not having to use it on a regular basis, most companies opt out of purchasing the ESRI platform. These entities often have in-house built geospatial mapping software or opt to use more open-source mapping platforms. The most commonly used of these open-source platforms is QGIS by the Open-Source Geospatial Foundation. Several companies that do not have the need or simply the resources to acquire the ESRI ArcGIS platform services, turn their attention to QGIS for providing geospatial analysis and solutions to clients. Being highly similar to ArcGIS, QGIS provides similar functionality for geospatial analysis and data processing functions. These functions, however, must be installed into QGIS either in the initial software installation process or in the form of add-ons toolsets called plugins, whereas most of the fundamental functionality tools in ArcGIS come pre-installed. In any case, the installation processes for these plugins are fairly unsophisticated, hence the increased usage of this open-source software. As a result, it was also decided that a QGIS version of the model/geoprocessing tool being built for this project will also be developed in the QGIS environment. The QGIS model was developed and tested in the Minnesota site. It was provided with the same raw data and criteria parameters as the ArcGIS model and script tool. Finally, the output provided by the QGIS model was compared to the output provided by the ArcGIS-based model and script tool in order to see if there are any large disparities between the model outputs and the wind energy siting areas provided by each one. With any kind of open-source software, it is questionable whether or not the platforms will have the capabilities to process large amounts of data and perform with the same amount of firmness as their more privatized counterparts. With regards to this issue, the QGIS platform is generally said to be capable of handling the same amount of data processing functionality as ArcGIS, if not more. The potential variation in processing time and overall performance are based on the back-end development programming languages. The ArcGIS platform is developed using .NET and C++, whereas QGIS was developed using Python and C++. The overall speed and performance were also tested as both these models are developed and run for this project.

The model development environment in QGIS is more robust and requires greater attention to detail than the environment in ArcGIS. The development of the QGIS requires the user to input the model variables as their data type. Users specify the types of data to be entered into the model by adding the raw data into the environment and having the software determine the data type. Each of the variables (model criterion raw data) was first added to the modelling environment and given the name of the base criterion to each of the variables was designated. Once these variables were added, the tools required to process the data in order to produce an output were added in a step-by-step process. Each of the base criterion's data processing steps was added and then the final outputs for the base criterion processes were combined in order to find the area that is most suitable for wind power generation. The majority of the raw data used for the ArcGIS model was directly implemented into the QGIS model. However, the landcover dataset was required to be pre-processed beforehand. This is primarily due to the landcover dataset being large with many different landcover types to be processed by the QGIS model all at once for the purposes of extracting out the study area and reclassifying the extracted land cover area within the study area that was most suitable for wind generation infrastructure development. Hence an already classified, binary, landcover raster was to be used in the QGIS model. The post-processed landcover raster was to be used in the QGIS model in order to maintain the model's data processing reliability. Another point of difference between the ArcGIS model and the QGIS model was the usage of the "Extract by mask layer" tool for clipping out the rasters according to the study areas provided by the user. The raster extraction tools used on both platforms were highly similar to the extract by mask tool in ArcGIS taking a little longer processing time as it effectively resamples the raster image based on the extracted output dimensions and rearranges the pixels. The clip tool in QGIS maintains the sample pixel arrangement as the original raster. All the tools necessary for the QGIS model were added with each of the tool outputs being the input for the proceeding tool wherever necessary, the final model was developed and ready for testing (Figure 16). The conditions or inputs for the base criterion parameters used for the QGIS model would simulate the parameters used for the 10000-meter iteration of the ArcGIS model in the same study area. This ensures a fair and unbiased test of the capabilities of both models.

The QGIS model, although designed as a model, operates, and behaves more like a geoprocessing tool. With the QGIS model not initially using the raw data and having variables for the raw data instead, QGIS builds a front-end interface for the user. This interface is developed based on the input variables and the outputs given by the model's processing tools. The user can also define what intermediate data can be saved by the model and which intermediate data to be removed once the model is run. Based on the requirements for the intermediate data, the model adds or removes more input information or variables to the model's user interface in order to help it add a path directory for saving data or values to the model's backend data processing mechanisms. Whenever the model is run, it opens a window with the automatically designed user interface, based on the input, output and processing criteria of the model (see Figure 17). The user is able to input the raw data into the model using this interface, and input the values needed for the data processing such as buffer distances or slope values and the interface also allows users to add data directory information to the model. Once all the required inputs are added via the interface, then only will the model be able to run and process the data. The model was run and with the provided output for the suitable land area for wind power generation, a comparison between the ArcGIS model and the QGIS model is to be undertaken. This comparison will be primarily based on the amount of land area that is covered by both models. Similarly, to the output provided by the ArcGIS model, the land area covered by the QGIS model output, and the quantity of pre-existing wind power infrastructure located within this output region will be the basis behind the comparison between the two models' outputs. Based on the values provided for these assessments, a judgment can be delivered which would state which platform would be most ideal for the model and for wind power generation site selection.

Results & Discussion

Minnesota Testing Phase Results & Discussion

The outputs from Minnesota were analyzed first. Most of the existing wind power infrastructure is concentrated in the southwestern corner of Minnesota. This is due to the presence of higher windspeed corridors within this region. The outputs of Minnesota were examined and compared to the pre-existing wind turbine areas. The 10,000-meter result for the Minnesota study area illustrated an area far greater than what was expected from the study area in question. The wind corridor towards the southwestern corner of the state was included but a larger portion of southern Minnesota, including the Minnesota River Valley and the south-eastern regions, were also captured by the model (Figure 5). Overall Minnesota has 2,690 wind farms within its land area, of this number, 2,118 were included in the region's output by the 10,000-meter transmission line iteration, the first iteration of the model. The first iteration managed to gain 78.7% of the wind farms in their respective output area (Figure 6). The model managed to achieve its overall accuracy requirement as the output area managed to identify the areas in the study site that can house wind power infrastructure while also remaining within the acceptable margin of error. The model was then rerun using the 3000-meter, second iteration, distances to transmission line. The output from the second iteration managed to include 818 wind farms from the study area (Figure 7). This is a much lower number of features located within the second iteration area as it only accounts for about 30% of the wind farms in the study area. The second iteration output areas were only able to capture 818 wind farms within their output areas. This value puts the second iteration at only a 30.4% success rate (Figure 8). Therefore, although the Euclidean distance of 3000 meters to transmission lines has been used in previous literature, it was not adopted in the study area used for the first testing phase of this project (Janke 2010; Díaz-Cuevas 2018). It can be said that the usage of 3000 meters distance to transmission lines would not be a suitable parameter for the study area for this specific testing phase. These circumstances and external factors can be used to account for the 30% of wind farms that did manage to be included within suitable areas such as varying distances to urban areas, road networks, land rights, or even sizes of wind farms. To account for these types of situations, the model's parameters can be changed according to the region's requirements when using the model in a different study area. These forms of issues can be expected by the model whenever testing in a different region. The model developed by this project aims to utilize a variety of parameters that can be dynamic and vary according to the user. The model aims to be user-friendly and support a variety of scenarios that can be used in different regions. The model can also be retrofitted to allow more parameters that can be implemented based on the user's needs.

Established by the output results, the first iteration parameters noticeably yield the most reliable outputs. Hence these parameters and iteration values were used for the second testing phase of the model as well. The overall results of the Minnesota study area were satisfactory regarding the model's ability to perform in reduced study areas. Once all the outputs from the Minnesota study area were assessed and the overall accuracy of the model with the first study site determined, the model was tested against a much larger study site - ERCOT. As mentioned earlier, ERCOT has a variety of features over a much larger area, which would be ideal for testing the overall capabilities of the model.

ERCOT Testing Phase Results & Discussion

Similar to the Minnesota output results, the outputs for ERCOT were also analyzed for variation in areas suitable for wind power development and the ratio of existing wind plants within output areas to total wind plants in the study area. The expectation of this testing phase is to be like the results of the Minnesota testing phase with the first iteration being the most accurate. The results for ERCOT took a long time as the big data processing required a large amount of computation. However, the outputs provided a thorough assessment of the model's overall performance.

Based on observations, it can be stated that most of the ERCOT region is suitable for implementing wind power generation. Most of the current wind power infrastructure in ERCOT is concentrated in the north or northwestern corner and in the southernmost regions of the study area. For all three iterations, except the distances to transmission lines, the remaining base criteria illustrated not only how ideal the study area would be for increased usage of wind power but also the model's ability to process big geospatial datasets. In terms of wind speed, 71.7% of the study area had acceptable wind speed at a height of 40 meters above ground level. For the slope criterion, 97.26% of the study area had slope values less than 12 degrees. This is to be expected as ERCOT has a relatively flat topography. The landcover requirement also provided a large area with 81.68% of the study area having suitable landcover for implementing new wind power generation infrastructure. Due to the inclusion of the smaller cities in the study area, the suitable distance from urban features covers a total of 95.18% of the study region. The output from the Euclidean distance tool for the road features shows that the majority of the study area has sufficient access to a major road. The slice tool was used to divide the distances provided by the Euclidean distance tool, using 10 geometric intervals. The geometric intervals were used as each class has the same number of corresponding values and the change between these intervals remains constant. The analysis used the first 5 of the intervals as these 5 intervals appear to include much of the existing wind farm within its area, with a reasonable margin of error. An exact area value could not be properly derived as the Euclidian distance tool outputs an area that extends beyond the bounds of the study area.

The final area outputs for each of the iterations in the ERCOT study area produced interesting results. The 10,000 meters distance to transmission lines output had a total area of 38,669,645.4 hectares. This area was spread throughout the study area, illustrating that much of the study had viable land areas suitable for building new power infrastructure (Figure 9). Many of the existing wind farms were also presented in the area output by the 10,000-meter iteration of the model. A total of 17,318 wind turbines are within the study area. Of the total wind turbines in the study area, 13,491 turbines exist within the output study area of the first iteration. Based on this information, the ratio of wind farms within the study area versus within the model output area was 3.095:4 and a percentage of 77.9% of the total wind farms within the estimated areas (Figure 10). This ratio came as an interesting value as upon visual inspection of the output, it appears as a vast majority of the wind power infrastructure lies within the model output area. After closer observation, however, the reasoning behind some of the wind farms remaining outside the expected output area appears to be caused by anomalies in one of the base criteria datasets. Errors present in the landcover dataset such as misinterpretations of the land cover classes, and their covered area caused uncertainties in the final output (Figure 11). Therefore, higher model accuracy can be achieved if more accurate landcover data is provided. For the second iteration, a 3,000-meter distance to transmission lines, a total of 7,915 wind turbines were discovered within the output area (Figure 12). This value only accounts for 45.7% of the total wind turbines within the ERCOT study area. This can be credited to the lower amount of output areas generated by the second iteration of the model (Figure 13). The loss in general areas can be attributed to the overall usage of the different values of distances to transmission lines. As mentioned in the previous section, usage of the 3000-meter distance to transmission lines is not unusual. However, these distance values often change throughout the literature and are based on different study areas (Pattanariyankool 2010). Hence, while the value of a 3000-meter distance to transmission lines wouldn't produce a viable result in this project's study areas, this value could potentially be used in other study areas using different parameters.

QGIS Model Testing Phase & Comparative analysis

Once the testing phases of the model in the ArcGIS Pro environment were completed, the QGIS model was tested using the same parameters. The model was redeveloped in the QGIS as illustrated in the methodology section of this paper. Once the datasets were uploaded onto the QGIS model and run. The processing of capabilities was particularly put to the test. However, observations of the model performance and capabilities of QGIS were not completely as predicted. Although predicted to perform at a faster pace than the ArcGIS model since the landcover raster was processed outside of the QGIS model, the QGIS model failed to completely finish a complete run for the Minnesota study region. The model was attempted to run several times, with each time the model only reaching partial completion after several hours of each run, indicating a limit on the amount of data or total features that QGIS can handle. This phenomenon greatly hindered the ability of this research to determine if the developed model would be effective in an open-source software platform or a more proprietary platform such as ArcGIS. However, the correctness of the model and whether it is capable of producing an accurate and usable wind energy site area output in the open-source platform could still be tested. In order to ease the data processing load on the QGIS model and allow it to run completely, it was decided to reduce the dimensions of the study area or test area that is to be input into the model. The reduced study area that was to be input into the model, was decided to be focused on the southern region of the original Minnesota study area. This will allow the model to process the raw data inputs to completion and produce a wind energy site area output. Once this output is generated, it was compared against the output provided by the ArcGIS model. The secondary study area is a rectangular polygon covering a large portion of the southern Minnesota region, it is 56,364.26 square kilometers (refer to Figure 15). It also has an evenly spread-out distribution of pre-existing wind power infrastructure; hence it provides an ideal environment for testing the two models against each other. Within this secondary study area, there are a total of 293 operational wind turbines. The majority of these wind turbines are often clustered together that are part of a wind farm. These wind farms provide power to factories or towns in their general proximity. The study area also has a sizeable number of individual or stand-alone wind turbines that aren’t associated with any larger windfarms. These stand-alone wind turbines are often used to power much smaller infrastructures such as a single building or farm. The output area of the ArcGIS model was redeveloped into the same study area of the QGIS model. The redeveloped area out from the ArcGIS model was compared to the output of the QGIS model first.

Once the covered area between the two model outputs was compared, they were compared against each other on their ability to account for pre-existing wind energy infrastructure. Theoretically, the amount of wind energy infrastructure covered by the area outputs by the two models with the same input datasets and processing tools should be similar.

The output generated by the QGIS model covers a total area of 41,918.2 square kilometers which accounts for 74% of the total area of the study area. The output generated by the ArcGIS model covered a total area of 35,311.19 square kilometers, accounting for 62% of the total area of the study area (see Figure 18). Using the output areas of both models, the number of wind infrastructures accounted for by each of the models' outputs was determined. The ArcGIS model output for the study area managed to account for a total of 224 wind turbines present within its area. This number is 77% of the total wind turbines in the region (Figure 19). In contrast, the QGIS model output area included a total of 281 wind turbines, accounting for 95% of the wind infrastructure in the study area. Based on the total area output and the number of wind turbines accounted for by the models' outputs, the QGIS model accounted for more of the wind infrastructure than the ArcGIS model for this study area, which is unexpected. The reasoning behind the QGIS model being able to perform a lot better than the ArcGIS model can be attributed to the manner in which the outputs were given. Upon visual examinations, the output provided by the QGIS model is more continuous and without any gaps leading it to be able to find more areas that are desirable for wind energy sites. The resulting polygon output is blockier in nature and a lot more generalized for representing the suitable siting area for wind generation. The output provided by the ArcGIS model, however, does have gaps within its own structure due to it being more refined in nature. The ArcGIS model output is more representative of the suitable area for wind energy development particularly as it represents the overall suitable area, directly down to its most minute and precise shape. Although more refined, the ArcGIS model output has several hundreds of smaller interlocked polygons representing wind energy siting areas whereas, as mentioned earlier, the QGIS model outputs an area that accounts for all these smaller areas or polygons and generates an output that is generalized and more representative of the suitable areas as a whole or as a singular, larger polygon area output. After further examination, it was found that the output of the QGIS model was generalized automatically based on the spatial resolution of the wind speed data layer (around 2000 m) while the output of the ArcGIS model was generated using the same spatial resolution as the 30-m DEM and landcover data layer. This would be the main reason for the QGIS model being able to incorporate a larger number of pre-existing wind infrastructures in an area.

Additionally, both models were unable to account for a certain number of wind turbines. The area outputs from both models omitted a general amount of wind turbines, hence a closer examination of the locations pertaining to these omitted turbines was conducted. For the turbines that were not accounted for by the model output areas, a closer look illustrated that these turbines are mostly individual or standalone turbines that are not part of a large wind energy generation project such as a wind farm. These standalone turbines aren’t often located within or closer to urban areas as they were built to provide power to only smaller infrastructures (Figure 20). These turbines are often built much closer to urban areas than turbines that are used for large scale power generation. They provide power to a single building or business. Some of these standalone turbines are built as part of research and development supporting infrastructure by research institutions or facilities for educational purposes. Ultimately, these types of turbine or wind energy projects are built using special permissions or permitting actions. As a result of these unaccountable circumstances, these types of turbines will not follow all the general standards used to build wind turbines for large scale power generation. This is believed to be the primary reason behind the certain number of wind turbines not being accounted for by the ArcGIS or QGIS model.

Upon examination of the models' outputs and the overall performance capabilities of both models themselves, certain conclusions can be drawn. Both models performed within an acceptable range of precision and accuracy. A general observation showed that the QGIS model with the generalized output was capable of performing at a much greater speed than the ArcGIS model but only on a smaller area. The ArcGIS model on the other hand is capable of producing a more refined output for wind power generation over a much larger area, however, it does so at a much slower speed. Unexpectedly, the generalized area output by the QGIS model was able to identify a greater percentage of wind infrastructure in the tested area. With both models being able to identify wind energy siting areas with an accuracy greater than 75%, both models can be viable options for usage for siting wind energy. When developing the model in the open-source environment, the primary ideology was to find out whether the model developed in this project would be capable of performing with the same prowess as the model in the more privatized software platform. Based on the information provided by this testing stage, it can be stated that both models are highly capable of providing a viable output that can be used for effective wind power generation infrastructure siting. In regard to which software is better for the task at hand is difficult to judge and will require several more testing, particularly in performance speed and data processing capabilities (utilizing larger study areas). However, the results indicated that there is a limit to the amount of data or total features that the open-source QGIS can handle. The final consensus that can be drawn from this project is that the usage of the model on their platform must be defined by the needs of the user and the scale of the project the user will attempt to utilize the model for. Both of the platforms have their own advantages and disadvantages in terms of utilizing the model developed in this project. It is the responsibility of the user, based on the project and their expectation or capabilities, to use whichever platform would best serve their purposes.

Conclusion

Based on the areas provided by the model and the geoprocessing tool, it can be stated that the model truly serves its developmental purpose. The parameters input to the model plays a large role in the behavior and as such the model's outputs for different variables may alter. Although several of the geophysical factors remain constant throughout the literature, more anthropogenic parameters often differ between study areas. One set of anthropogenic parameters cannot be declared for a second or third, hence further improvements to account for these kinds of changes would be required. The areas provided by the model, including the different values for distance to major transmission lines, are all within the acceptable degree of accuracy. The ability of the model to not only perform well in small study areas but also in much larger study areas with big data is truly the goal behind this project. The model was also capable of retaining its level of accuracy whilst processing the big data posed in the ERCOT study area.

The results generated by this project can be considered reliable. Based on the criteria used and the deliverables given, the model performed at an acceptable range of accuracy. Even though the model outputs were regarded as accurate, a certain degree of error was also observed and there are potential failures that could occur as there are certain areas within the study region that still would have wind power infrastructure. These areas of error are believed to be the result of more anthropogenic and specialty cases. The criteria used were created from combined results of different research articles around the world. To account for all forms of cases, testing the model using even more criteria and using overseas study areas would be the next stage in the development of models and scripting tools. It is difficult to attain over a 90% accuracy standard as several non-physical or non-geographic features also contribute to the building of wind farms over just suitable areas, such as land costs, capacity factors, and land rights. These are factors that cannot be entirely taken into account as they are highly dynamic and will require a much greater degree of computation. These factors will provide the basis for further improvements to the model itself and, ultimately, to enhanced geoprocessing tool. However, the overall research question pertaining to this project can be deemed as fulfilled. The model developed for this project was able to identify areas that would be highly suitable for the building and implementation of new wind power infrastructure.

As per the platform on which the model is to be utilized, the user must make the decision based on their needs. This project further illustrates the key advantages and disadvantages of using open-source software platforms compared to more privatized ones in the energy siting and modelling environments. The ArcGIS modelling environment produces a more refined output and can conduct the process for a larger study area that require more processing time. On the other hand, the QGIS model can process the data faster for smaller study areas but can only handle a limited amount of data and features. In addition, the QGIS model output may be less refined because it automatically resamples or generalizes the output using the data layer with the lowest spatial resolution. For the increase in processing power, the QGIS environment sacrifices the ability to completely process larger datasets. Hence, if a project requires a more generalized area for smaller study sites, the QGIS model would be the more viable option; however, if the project requires a much larger study area, the ArcGIS model would be preferable.

Prospects for Model Improvement & Expansion

With any model, there are always going to be areas in which they could be improved. The main aspects to be improved in this type of model are to make the model more user-friendly, expand capabilities and reduce processing time. For the model developed for this project, the expansion process would involve a more in-depth analysis of the base criterion and adding the ability to manipulate the base criterion itself. When expanding on the base criterion themselves, windspeed is a primary factor that can be potentially refined. As mentioned earlier, wind turbines have the ability to operate with windspeeds reaching 14 to 15 miles per hour at 40 to 50 meters above the ground. However, as windspeed increases, the rotors on the turbines begin to experience higher levels of atmospheric kinetic energy. This increased kinetic energy can result in turbulences occurring that would cause structural damage to the turbines (Miller & Kleidon, 2016). This upper limit of operation of a wind turbine is around 55 miles per hour or 24.5 meters per second. Moreover, the capacity factor of a turbine tends to define the overall power output of the turbine, irrespective of the wind speed. Hence, wind turbines are deactivated whenever the regional windspeed of an area approaches the upper limits of the turbines to avoid damaging the turbines. To improve the model for this project, a range of windspeeds can be set for the windspeed criterion by inputting a limiting factor on the desirable windspeed criterion. Proper siting practices for wind power infrastructure are further enhanced as the increased wind turbine activity in a region creates artificial turbulences that are capable of increasing a region’s surface temperature (Baidya Roy & Traiteur, 2010). To avoid these types of effects, strategically siting and placing wind turbines in a region is crucial. This could also be an additional factor that can be built into the models’ output analysis enhancements.

Moreover, as the model outputs these desirable areas for wind generation, the model can be expanded to further show how many wind turbines can be placed in this area. This is achieved by taking the average area required for constructing a wind turbine, its operational area radius and all its connection infrastructure, and dividing the area output from the model by the turbine's operational area. For the mentioned process, the ArcGIS Pro tool "Generate Grid From Area" tool can be used to develop the grid, and based on the number of grid cells produced, an area calculation aspect can also be built into the model, which would be able to divide the area output into equal area grid and based on the area provided, calculate how many wind turbines can be built in this area. Consequently, based on the number of potential wind turbines input into the gridded area, an estimate of the total power output and capacity factor can be derived. Another aspect that would considerably increase the model's capabilities would be adding the ability to conduct the energy siting process in the offshore environments. Offshore wind generation is an area of major research and development in several parts of the world. For offshore wind power site selection, the amount of data that needs to be processed is even greater. These datasets will include bathymetric data, tidal and ocean current data, ocean trade routes, ocean area rights information and weather datasets for the study area. The nature of the GIS model will allow for the addition of these types of datasets and their processing. Once these datasets are added and a beta version of the model is developed, the geoprocessing tool can also be redesigned to accommodate the offshore wind aspect.

The primary purpose of having the geoprocessing tool was to enable users who have an understanding of Python scripting and of the Arcpy libraries, to be able to add and remove aspects of the geoprocessing tool while improving the overall functionality of the model. The mentioned iterations of the model could be implemented in both the ArcGIS and QGIS environments. The purpose of having both environments is to allow this model and scripting tool to be as versatile and user-friendly as possible. This model should be available for usage by any entity that will require a basis for wind energy generation. The script tool will allow users to add newer geoprocessing tools into the overall functioning of the model, ultimately increasing its capabilities. These additions could hinder the processing time and performance of the model. However, these areas are to be tested when this model and the subsequent Python scripting tool undergo improvements or changes.

Bibliography

Atici, K. B., Simsek, A. B., Ulucan, A., & Tosun, M. U. (2015). A GIS-based Multiple Criteria Decision Analysis approach for wind power plant site selection. *Utilities Policy,* 37: 86-96.

Baidya Roy, S., & Traiteur, J. J. (2010). Impacts of wind farms on surface air temperatures. *Proceedings of the National Academy of Sciences*, 107(42), 17899–17904.

Bunodiere, A. and Lee, H. S. (2020). Renewable Energy Curtailment: Prediction Using a Logic-Based Forecasting Method and Mitigation Measures in Kyushu, Japan. *Energies*, 13(18).

Chamanehpour, E, Ahmadizadeh, S. R, & Akbarpour, A. (2017). Site selection of wind power plant using multi-criteria decision-making methods in GIS: A case study. *Computational Ecology and Software, 7*(2), 49-64.

Cox, Lopez, Watson, Grue, & Leisch, E. (2018). *Renewable Energy Data, Analysis, and Decisions: A Guide for Practitioners*. National Renewable Energy Laboratory.

Díaz-Cuevas, P. (2018). GIS-Based Methodology for Evaluating the Wind-Energy Potential of Territories: A Case Study from Andalusia (Spain). *Energies,* 11(10), DOI: 10.3390/en11102789

Ghorbani, A., Mehrjerdi, H., & Al-Emadi, N. A. (2017). Distance-differential protection of transmission lines connected to wind farms. International Journal of Electrical Power & *Energy Systems*, 89, 11–18.

Höfer, T., Sunak, Y., Siddique, H., & Madlener, R. (2016). Wind Farm siting using a spatial analytic hierarchy process approach: A case study of the städteregion aachen. *Applied Energy*, *163*, 222–243.

HuiBin, L., Wenxiao, Q., Qianqian, Y., Tianye, L., & Xiu, L. (2016). Research on the Short-term prediction model of wind power generation based on GIS. *China International Conference on Electricity Distribution*.

Jahangiri, M., Ghaderi, R., Haghani, A., & Nematollahi, O. (2016). Finding the best locations for establishment of solar-wind power stations in Middle East using GIS: A review. *Renewable and Sustainable Energy Reviews,* 66: 38-52.

Janke, J. R. (2010). Multicriteria GIS modeling of wind and solar farms in Colorado. *Renewable Energy*, 35(10): 2228-2234.

Krause, G. (2011). From Turbine to Wind Farms: Technical Requirements and Spin-Off Products. BoD – Books on Demand.

Miller, L. M., & Kleidon, A. (2016). Wind speed reductions by large-scale wind turbine deployments lower turbine efficiencies and set low generation limits. *Proceedings of the National Academy of Sciences*, 113(48), 13570–13575.

Nasehi, S, Karimi, S., and Jafari H.. (2018). Application of Fuzzy GIS and ANP for Wind Power Plant Site Selection in East Azerbaijan Province of Iran. *Computational Research Progress in Applied Science & Engineering*, 2 (3), 116-124

Pakere, I., Kacare, M., Grāvelsiņš, A., Freimanis, R., & Blumberga, A. (2022). Spatial analyses of Smart Energy System implementation through system dynamics and GIS modelling. Wind Power Case Study in Latvia. *Smart Energy*, *7*, 100081.

Parti, I. K., Mudiana, I. N., & Rasmini, N. W. (2020). Analysis of wind speed effect on voltage in wind power plant performance. Journal of Physics: Conference Series, 1450(1), 012132.

Pattanariyankool, S. and Lave, L. B. (2010). Optimizing transmission from distant wind farms. Energy Policy 38(6): 2806-2815.

Shorabeh, S. N., Firozjaei, H. K., Firozjaei, M. K., Jelokhani-Niaraki, M., Homaee, M., & Nematollahi, O. (2022). The site selection of wind energy power plant using GIS-multi-criteria evaluation from Economic Perspectives. *Renewable and Sustainable Energy Reviews*, *168*, 112778.

Supapo, K. R., Lozano, L., Tabañag, I. D., & Querikiol, E. M. (2021). A geospatial approach to energy planning in aid of just energy transition in small island communities in the Philippines. *Applied Sciences*, *11*(24), 11955.

Szurek, M., Blachowski, J., & Nowacka, A. (2014). GIS-based method for wind farm location multi-criteria analysis. *Mining Science*, *21*.

van Zalk, J. and Behrens, P. (2018). 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,* 123: 83-91.

Yunna, W., Yi-Sheng, Y., Tian-Tian, F., Li-Na, K., Wei, L., & Luo-Jie, F. (2013). Macro-site selection of wind/solar hybrid power station based on Ideal Matter-Element Model. *International Journal of Electrical Power & Energy Systems*, 50, 76–84.

List of Figures

Figure 14. The front interface for the Geoprocessing script tool in the ArcGIS Pro environment, developed using Arcpy and Pycharm IDLE. The raw data for the geoprocessing tools are inputted in the script using this interface. Whenever the script tool is modified, this interface must be redesigned for the user's needs.

Figure 15. shows the study area used to test the QGIS model. This study area focuses on the south, central region of Minnesota and has an even distribution of pre-existing wind power generation infrastructure giving the area ideal conditions for testing the QGIS model and comparing it to the output of the ArcGIS model (Generated using ArcGIS Pro).

Figure 16. Final model developed in QGIS for testing how the model will perform in an open-source environment (Generated using QGIS).

Figure 17. The front-end interface developed by QGIS. This interface is autogenerated when a model is created in the QGIS environment. The interface is based on the input parameters added to the model. Any changes occurring in the model are automatically reflected in this front-end interface. (Generated by QGIS).

Figure 18. The output provided by the QGIS model. The output was generated in the QGIS platform as a shapefile and transported to ArcGIS Pro for analysis. As seen, the QGIS model's output covers a large amount of the study area and accounts for a large amount of the existing wind power infrastructure.

Figure 19. The ArcGIS model's output for the secondary study area used for the comparison against the QGIS model. The output for the ArcGIS model also covers a large amount of the study area and visually accounts for a large number of the preexisting wind infrastructure.

Figure 20. A closer examination of the model output from the QGIS model illustrates the areas, particularly in the eastern regions, closer to urban areas where there are standalone wind turbines that didn't get accounted for by the models output. These are wind turbines built under specialty cases and do not follow the same base criterion as conventional wind turbines such as the ones used in this project.

Figure 1. Study area map showing the locations of Minnesota & ERCOT (Generated using ArcGIS Pro).

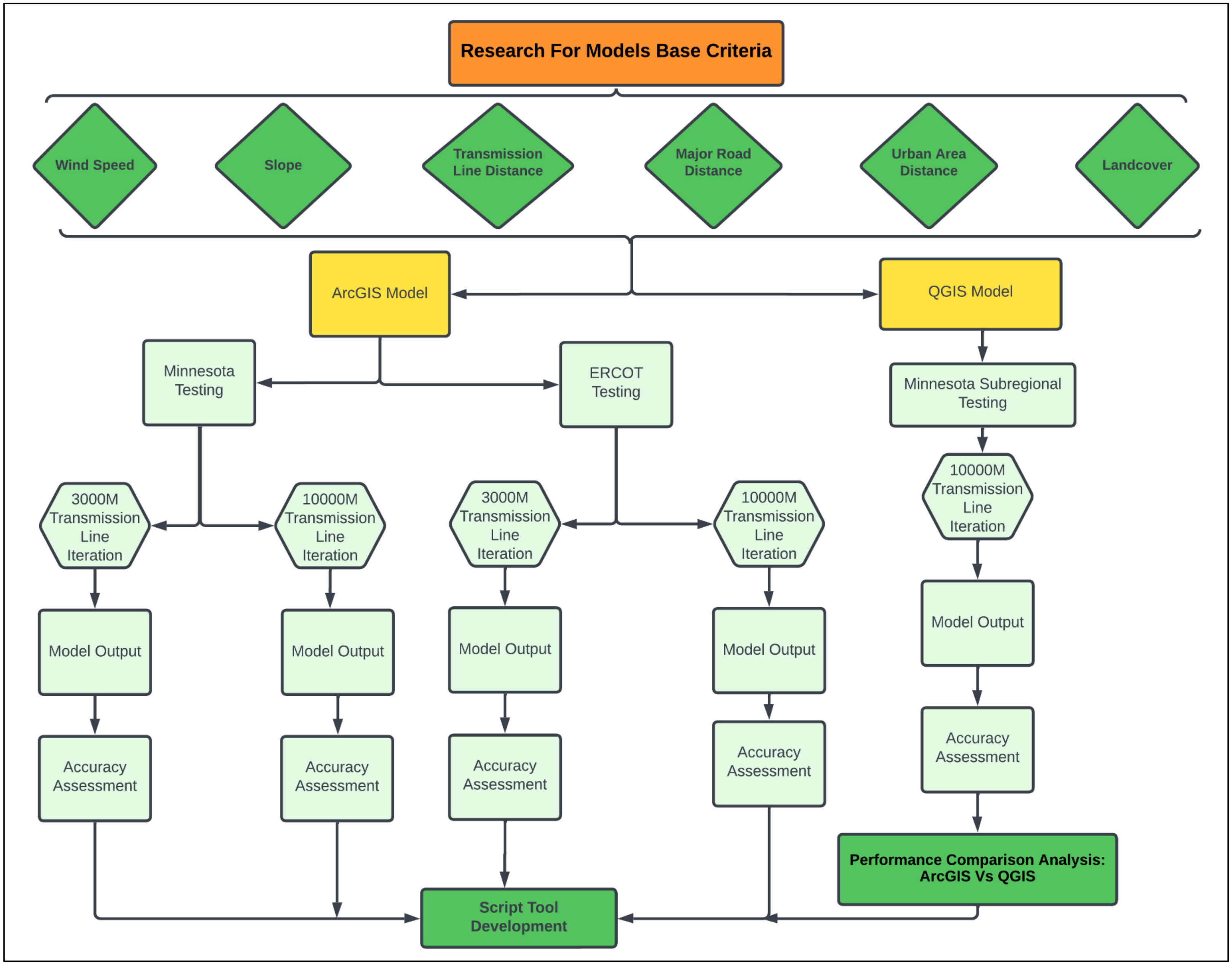


Figure 2. Flow chart showing the work flow conducted by this project for the wind energy.

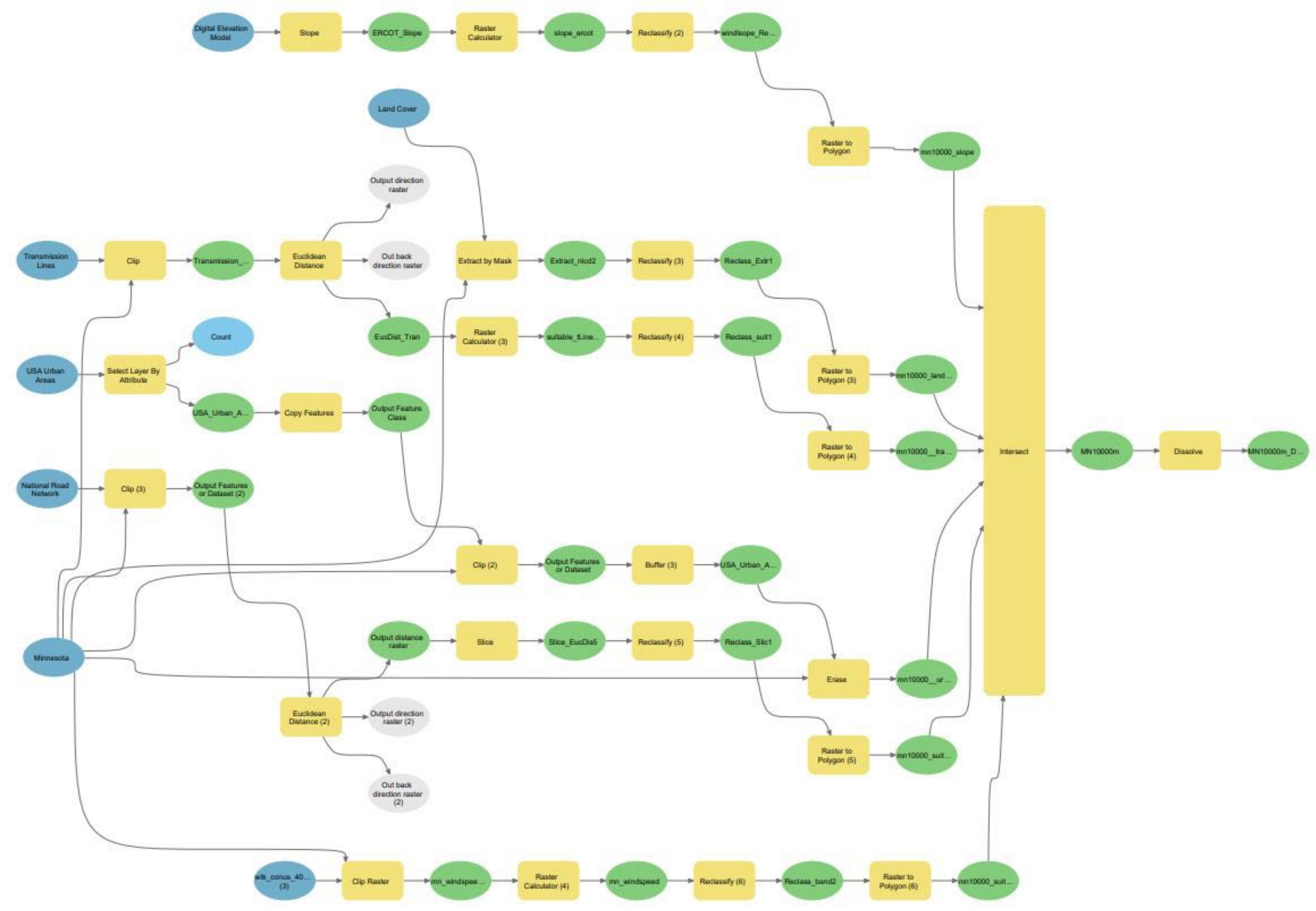


Figure 3. Final model developed for the project
(Generated using ArcGIS Pro)

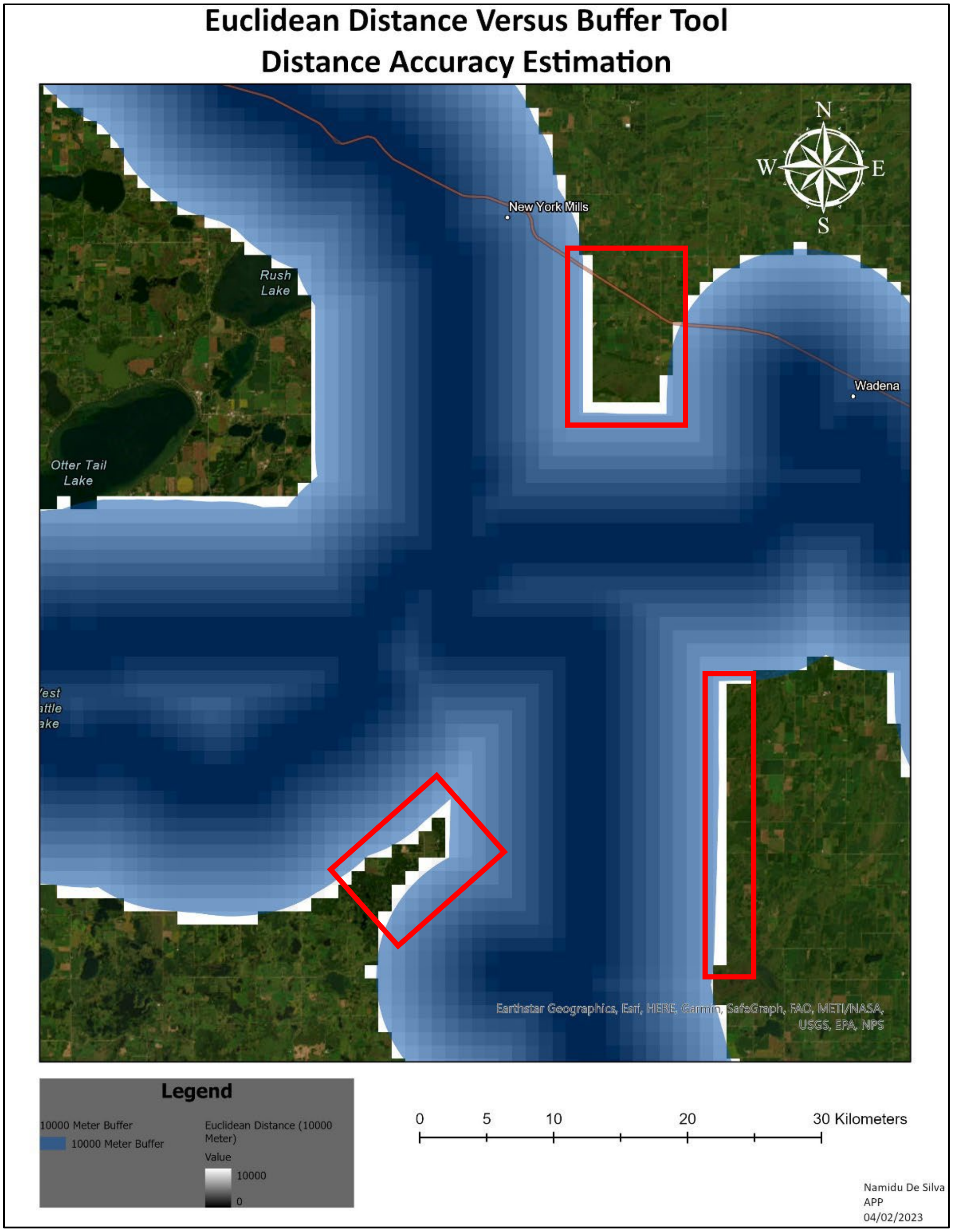


Figure 4. This map shows the area covered by the outputs provided by the Euclidean distance tool and the buffer tool, overlayed on top of each other. Some areas of the raster output from the Euclidean distance tool (highlighted in the red boxes) covers more areas than the buffer tool output but these areas aren't within the required bounds of the required base criterion. One of the raster cells has an area of 1 square kilometer and it multiple cells being part of the additional area, this extra area adds a significant area to the overall output. Hence the Euclidean distance tool output isn't as accurate as the Buffer tool output (Generated in ArcGIS Pro).

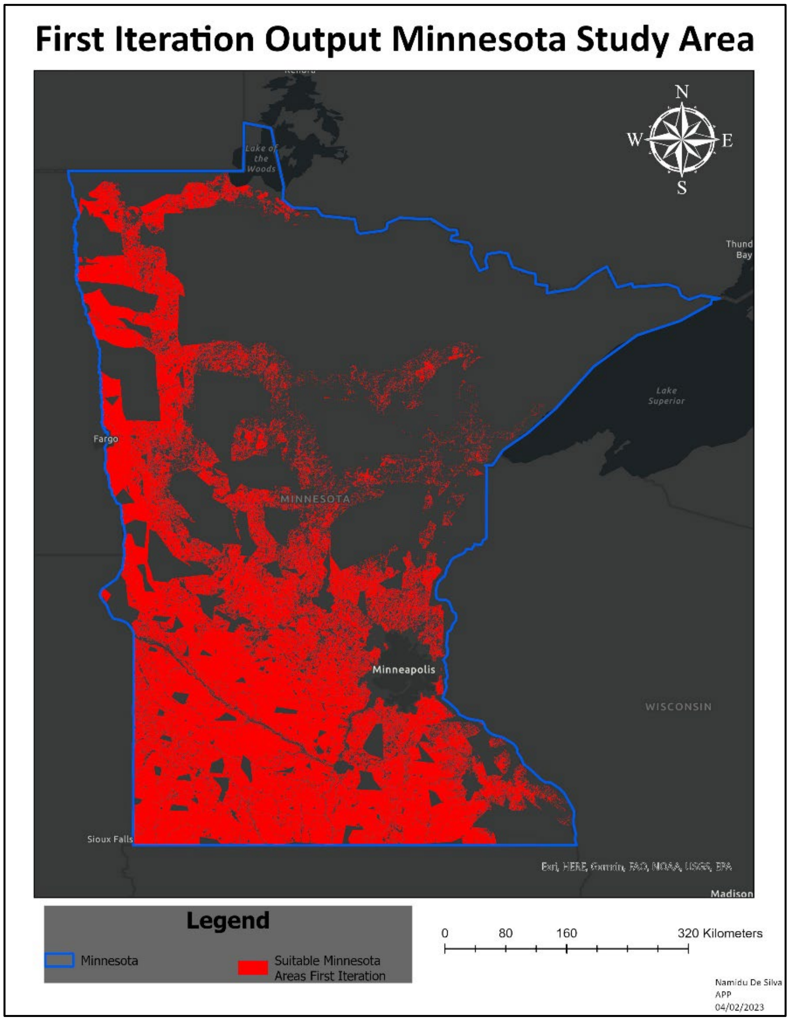


Figure 5. Minnesota study area map showing the output areas generated by the first iteration of the model (Generated using ArcGIS Pro).

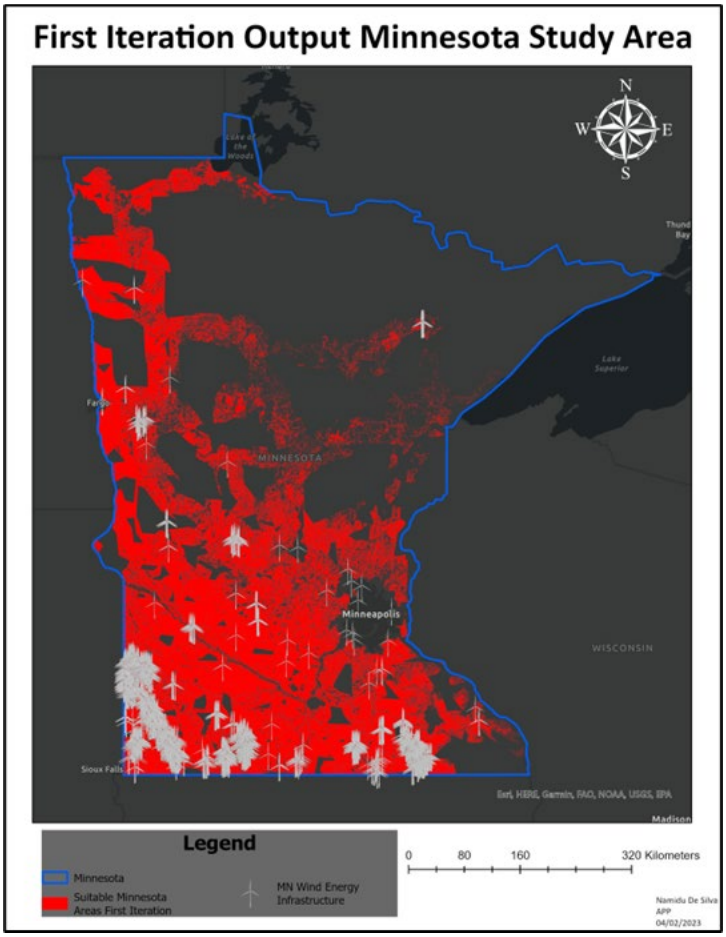


Figure 6. The output areas generated by the first iteration of the model in Minnesota and the distribution of pre-existing wind turbines with respect to the output areas. (Generated using ArcGIS Pro).

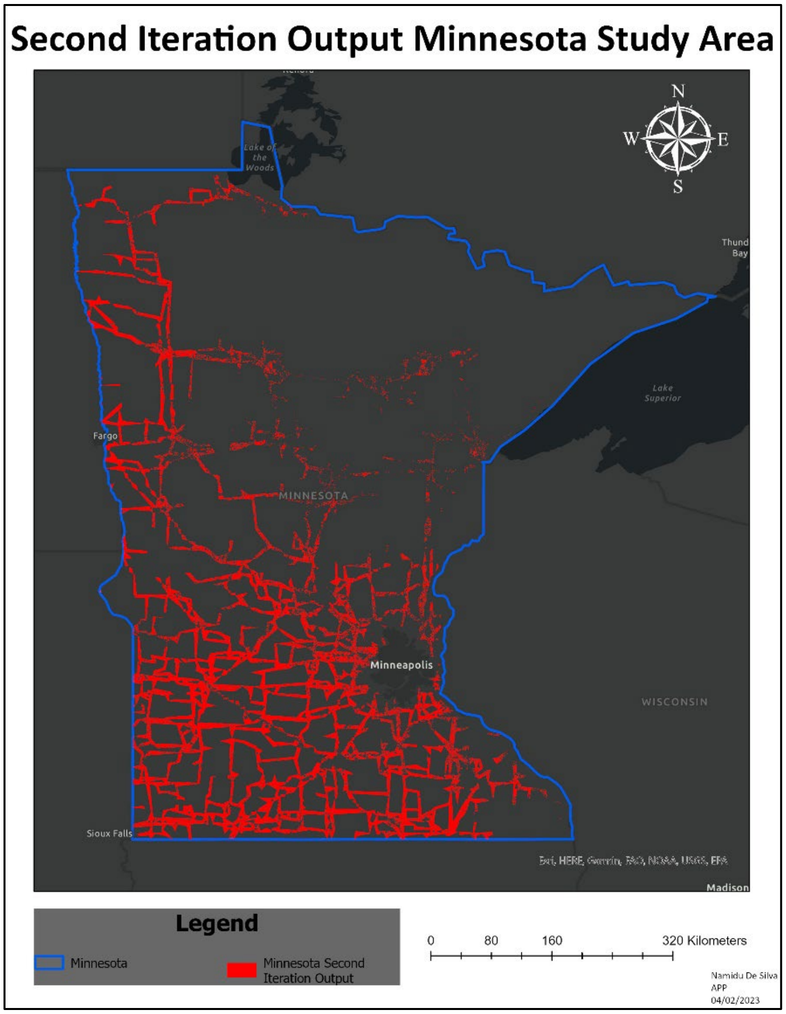


Figure 7. Minnesota study area map showing the output areas generated by the second iteration of the model (Generated using ArcGIS Pro).

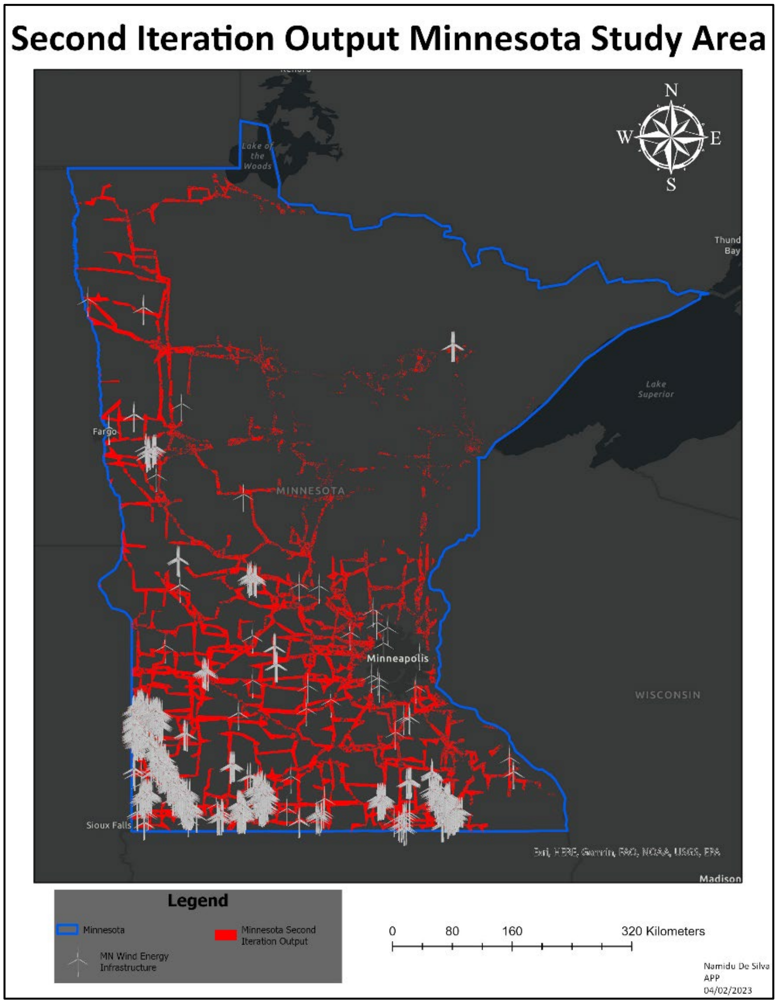


Figure 8. The output areas generated by the second iteration of the model in Minnesota and the distribution of pre-existing wind turbines with respect to the output areas. (Generated using ArcGIS Pro).

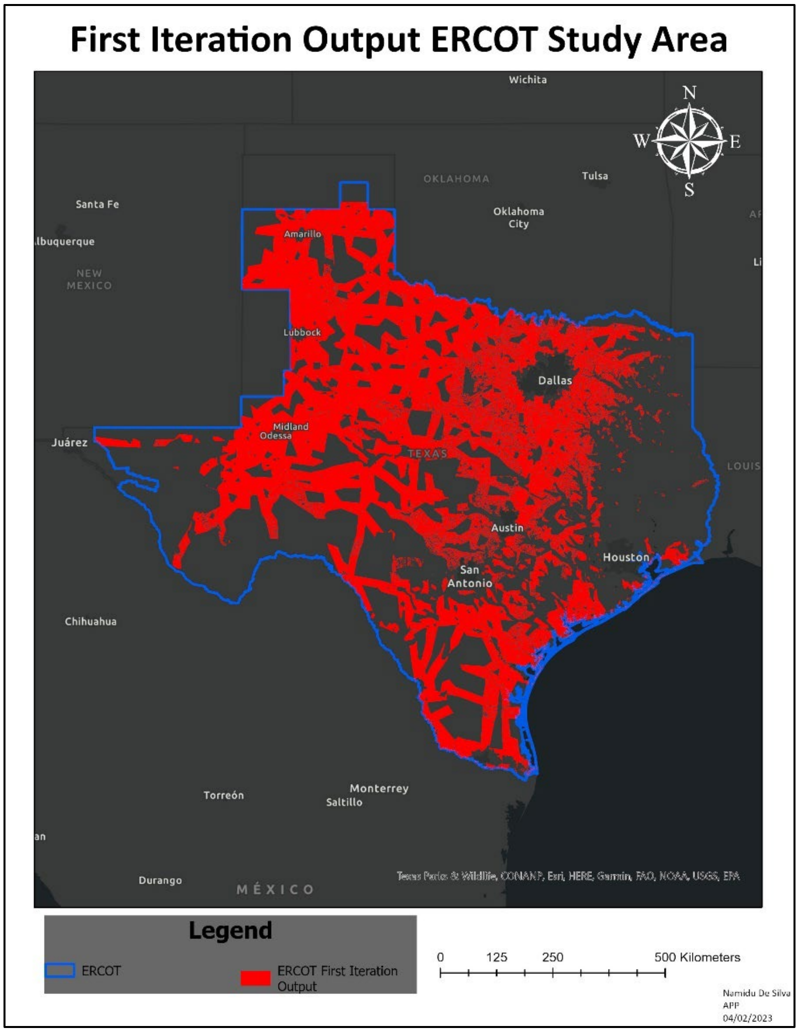


Figure 9. ERCOT study area map showing the output areas generated by the first iteration of the model (Generated using ArcGIS Pro).

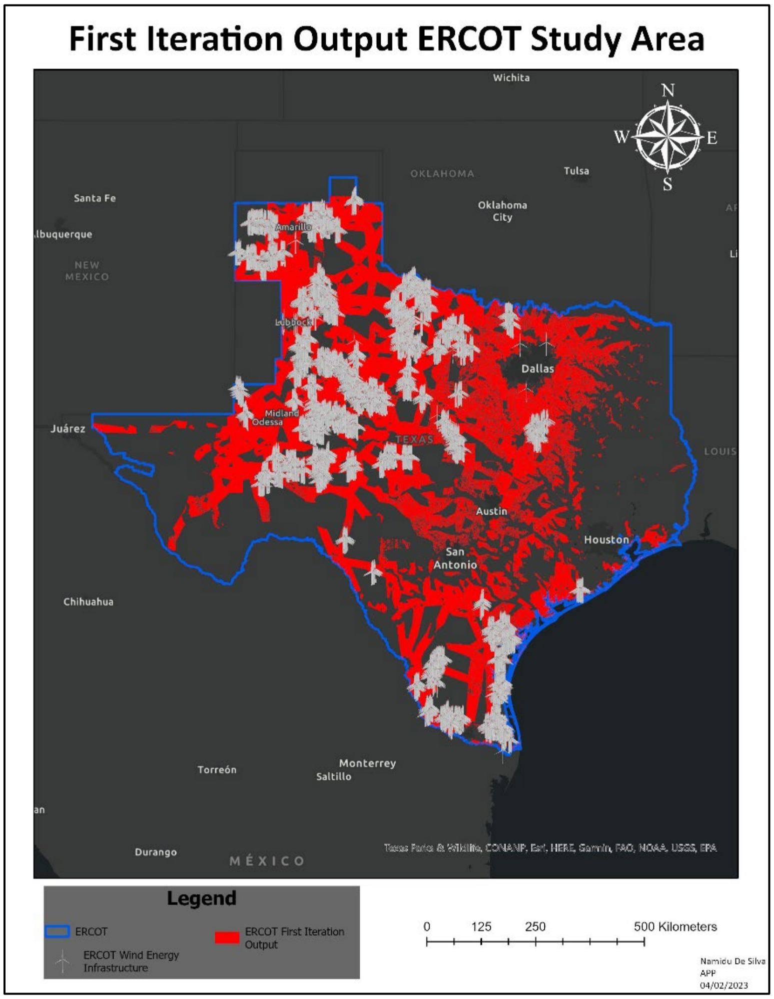


Figure 10. ERCOT study area map showing the output areas generated by the first iteration of the model and the distribution of operational wind farms in ERCOT (Generated using ArcGIS Pro).

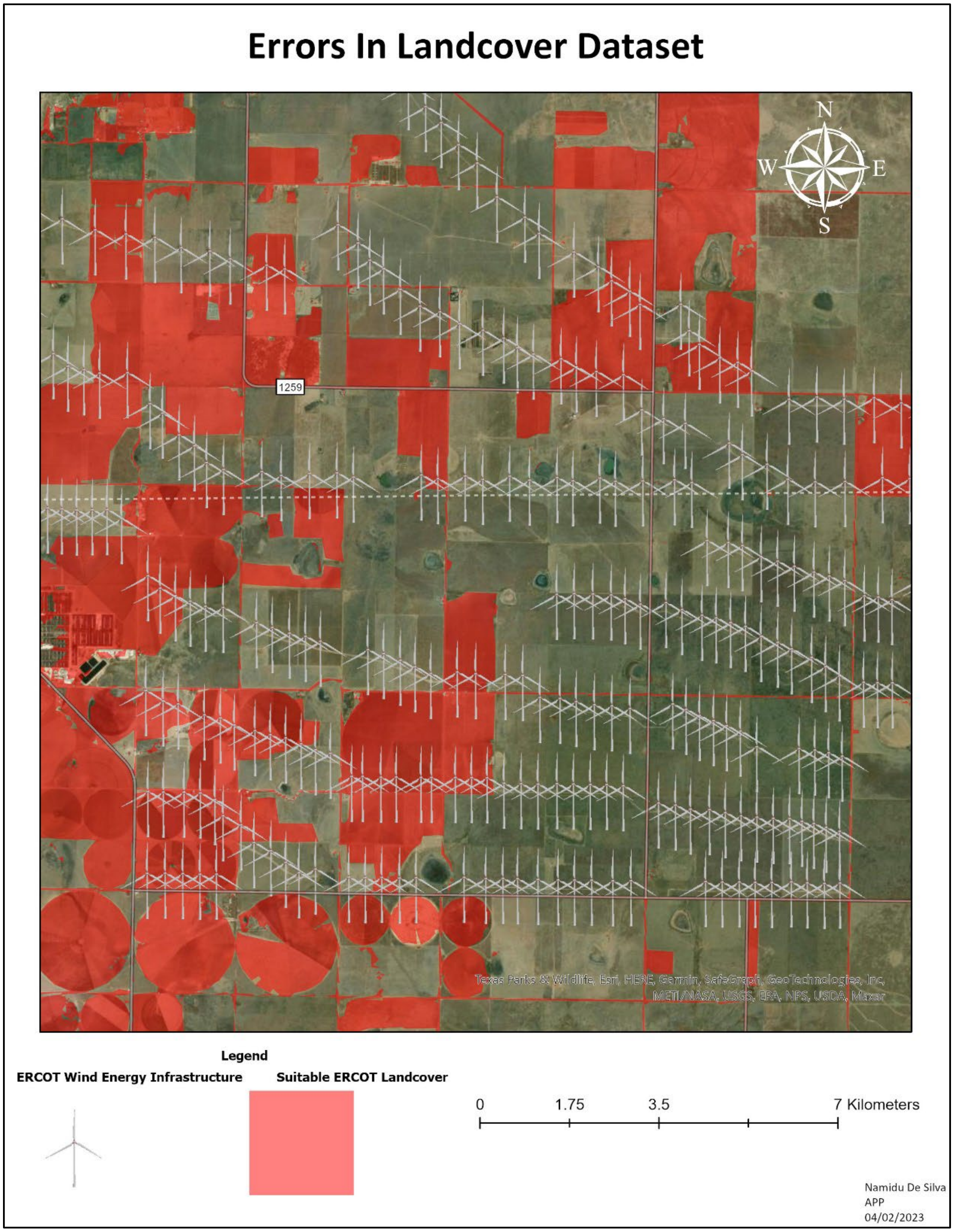


Figure 11. Errors in the suitable land cover area output from the suitable land cover selection stage of the model. Certain areas that should have been added to the output have been omitted. This error is also present in the Minnesota study area and its land cover output areas (Generated using ArcGIS Pro).

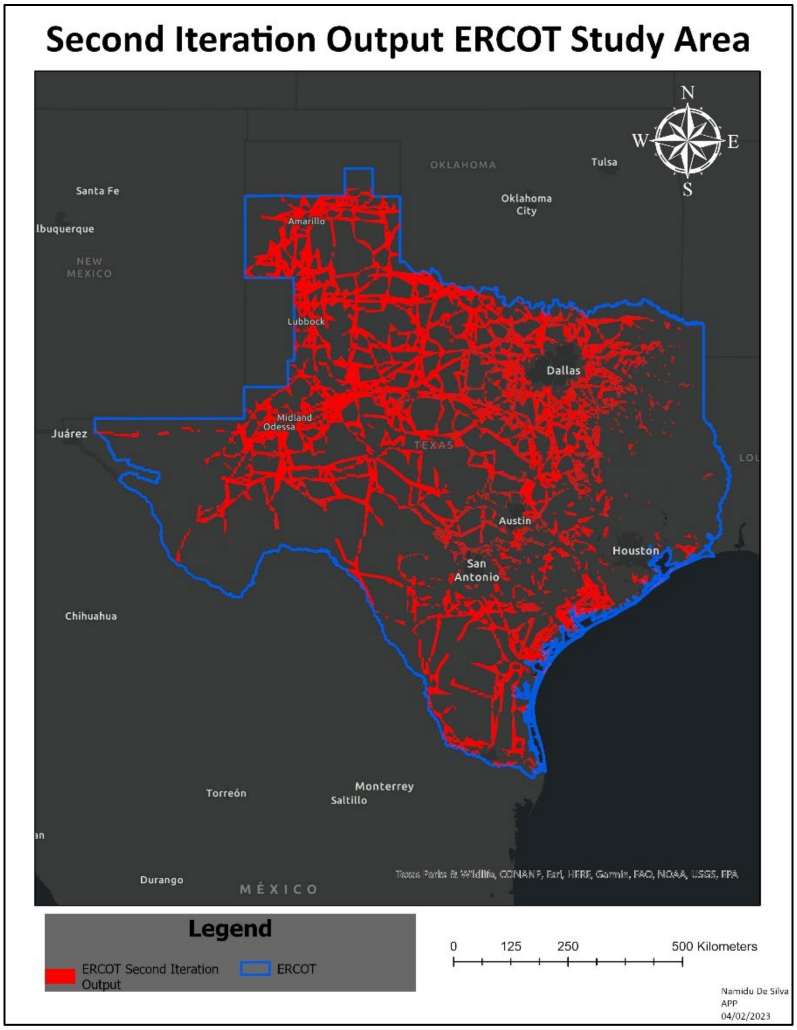


Figure 12. ERCOT study area map showing the output areas generated by the second iteration of the model (Generated using ArcGIS Pro).

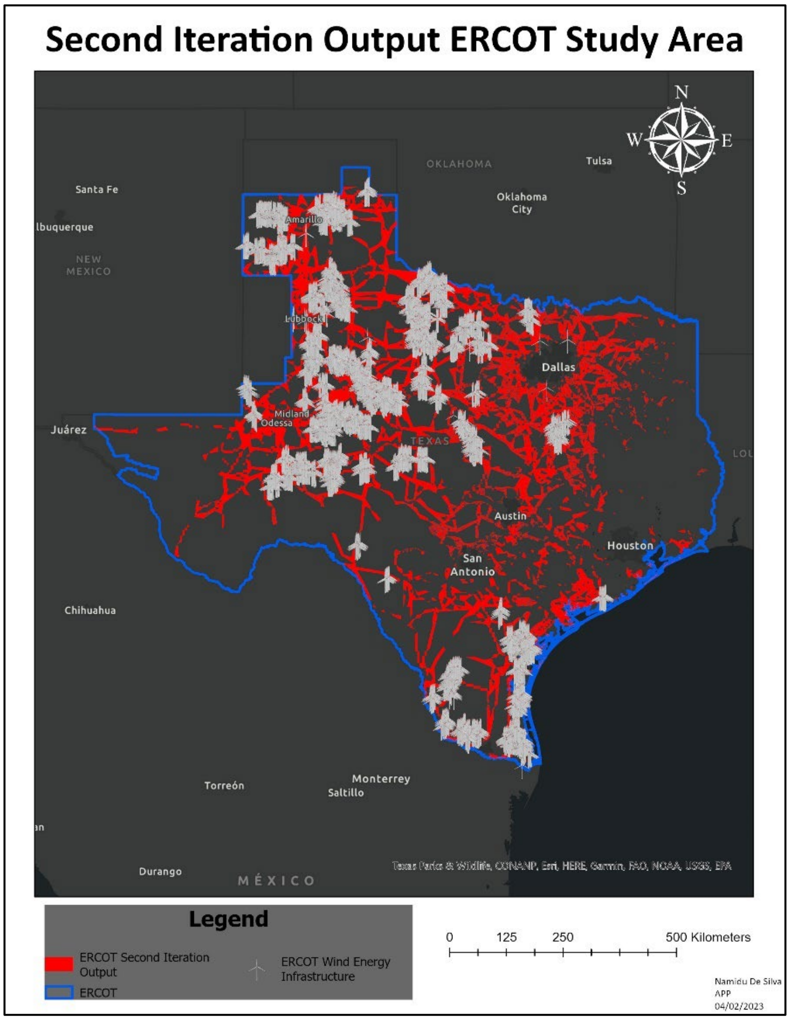


Figure 13. ERCOT study area map showing the output areas generated by the second iteration of the model and the distribution of operational wind farms in ERCOT (Generated using ArcGIS Pro).

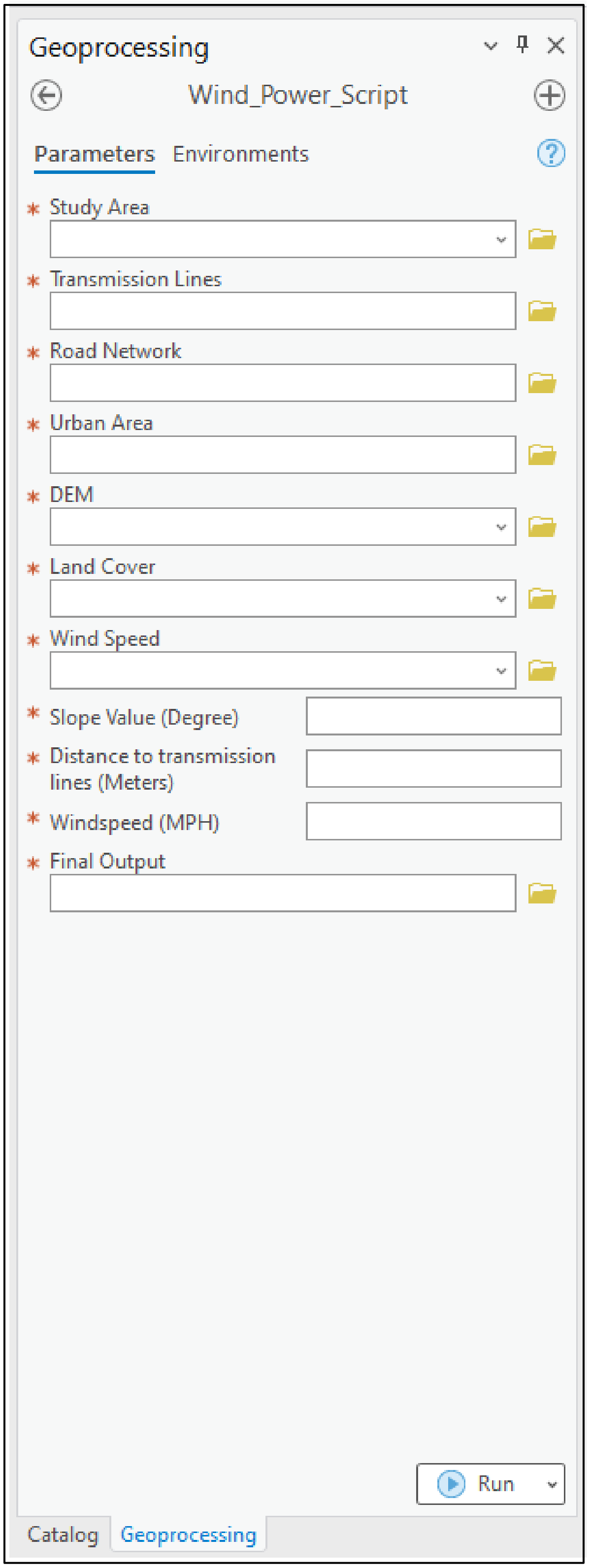


Figure 14. The front interface for the Geoprocessing script tool in the ArcGIS Pro environment, developed using Arcpy and Pycharm IDLE. The raw data for the geoprocessing tools are inputted in the script using this interface. Whenever the script tool is modified, this interface must be redesigned for the user's needs.

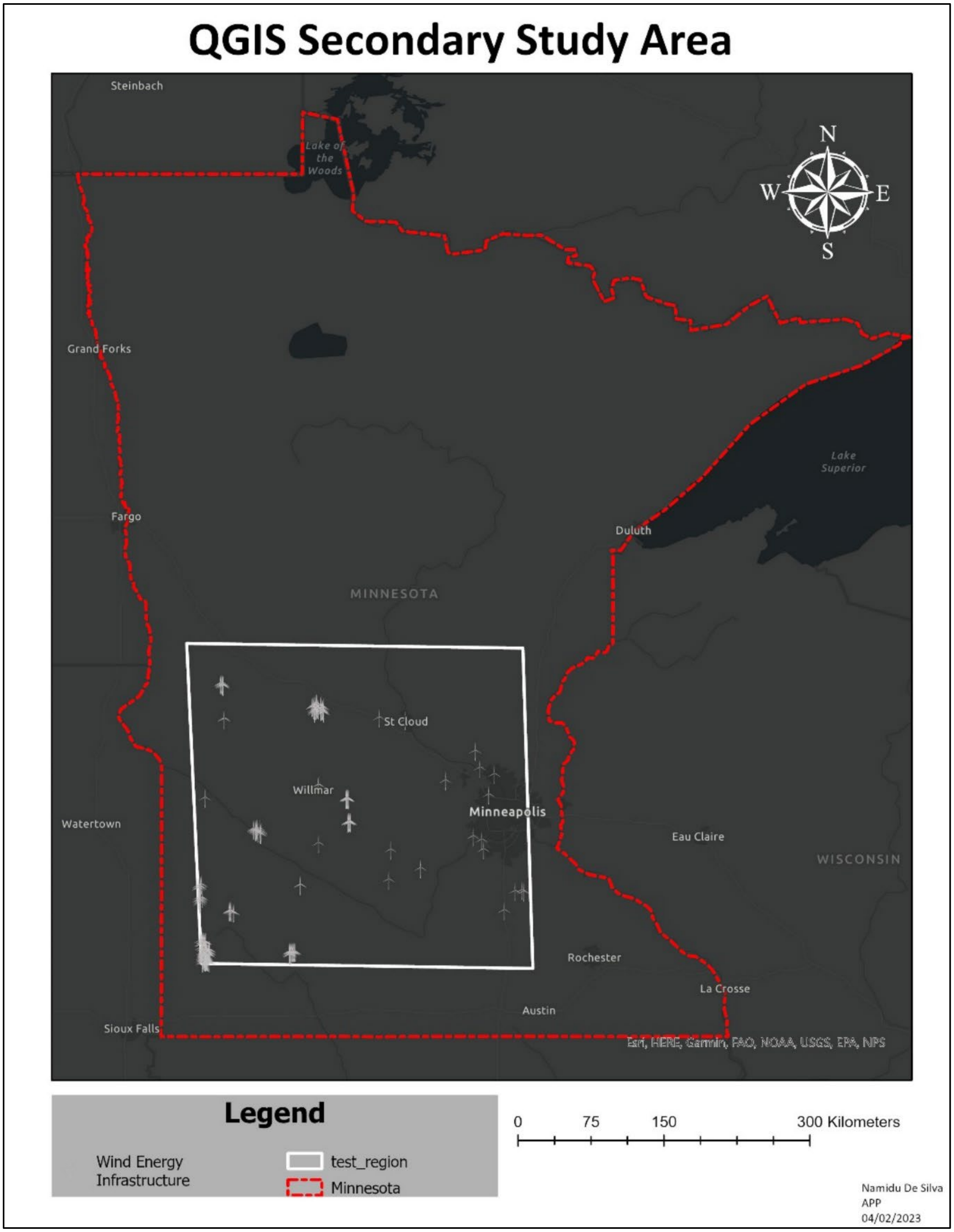


Figure 15. shows the study area used to test the QGIS model. This study area focuses on the south, central region of Minnesota and has an even distribution of pre-existing wind power generation infrastructure giving the area ideal conditions for testing the QGIS model and comparing it to the output of the ArcGIS model (Generated using ArcGIS Pro).

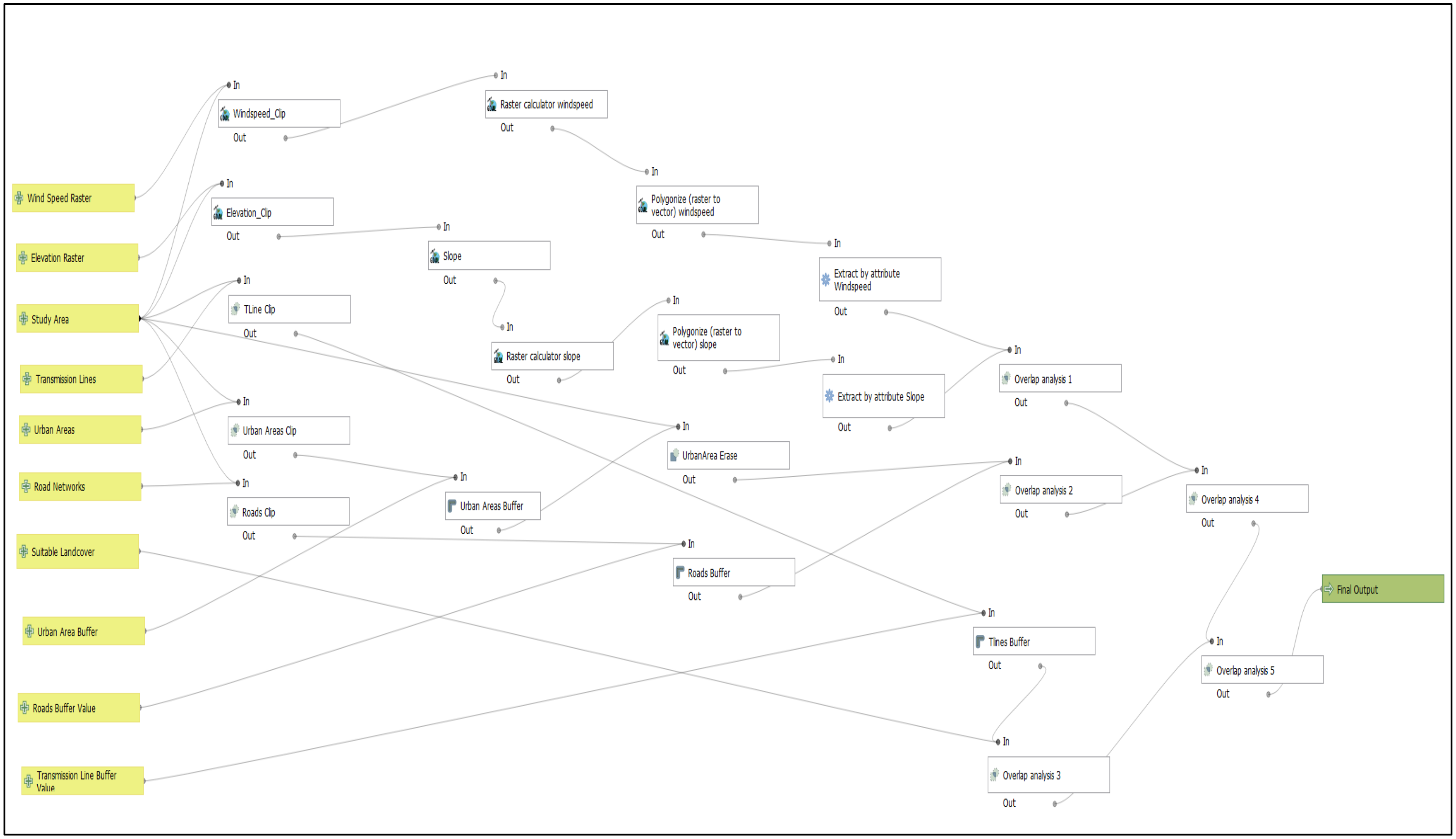
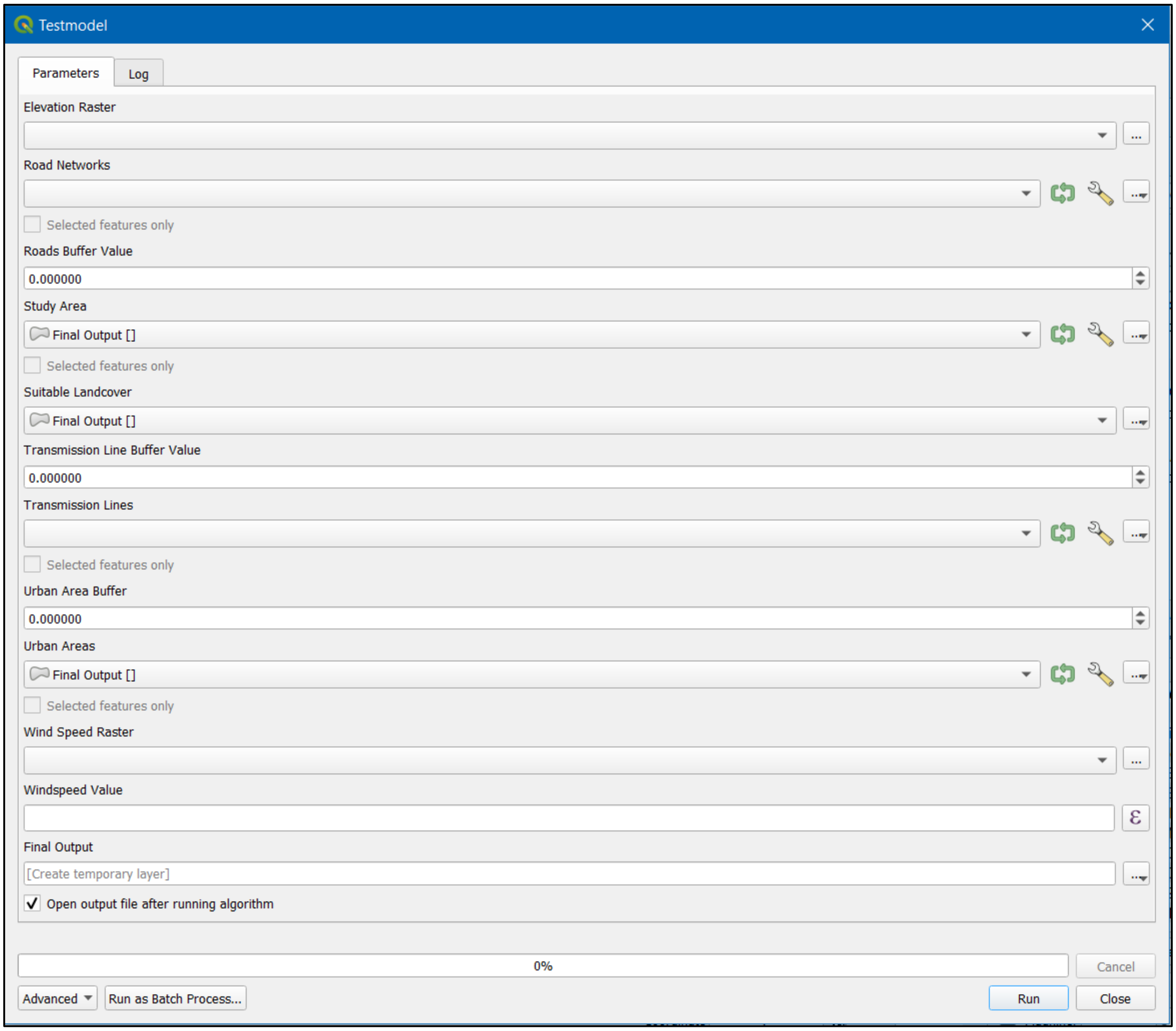


Figure 16. Final model developed in QGIS for testing how the model will perform in an open-source environment (Generated using QGIS).


Testmodel

Parameters | Log

Elevation Raster

Road Networks

☐ Selected features only

Roads Buffer Value

0.000000

Study Area

Final Output []

☐ Selected features only

Suitable Landcover

Final Output []

Transmission Line Buffer Value

0.000000

Transmission Lines

☐ Selected features only

Urban Area Buffer

0.000000

Urban Areas

Final Output []

☐ Selected features only

Wind Speed Raster

Windspeed Value

Final Output

[Create temporary layer]

☑ Open output file after running algorithm

0%

Figure 17. The front-end interface developed by QGIS. This interface is autogenerated when a model is created in the QGIS environment. The interface is based on the input parameters added to the model. Any changes occurring in the model are automatically reflected in this front-end interface (Generated by QGIS).

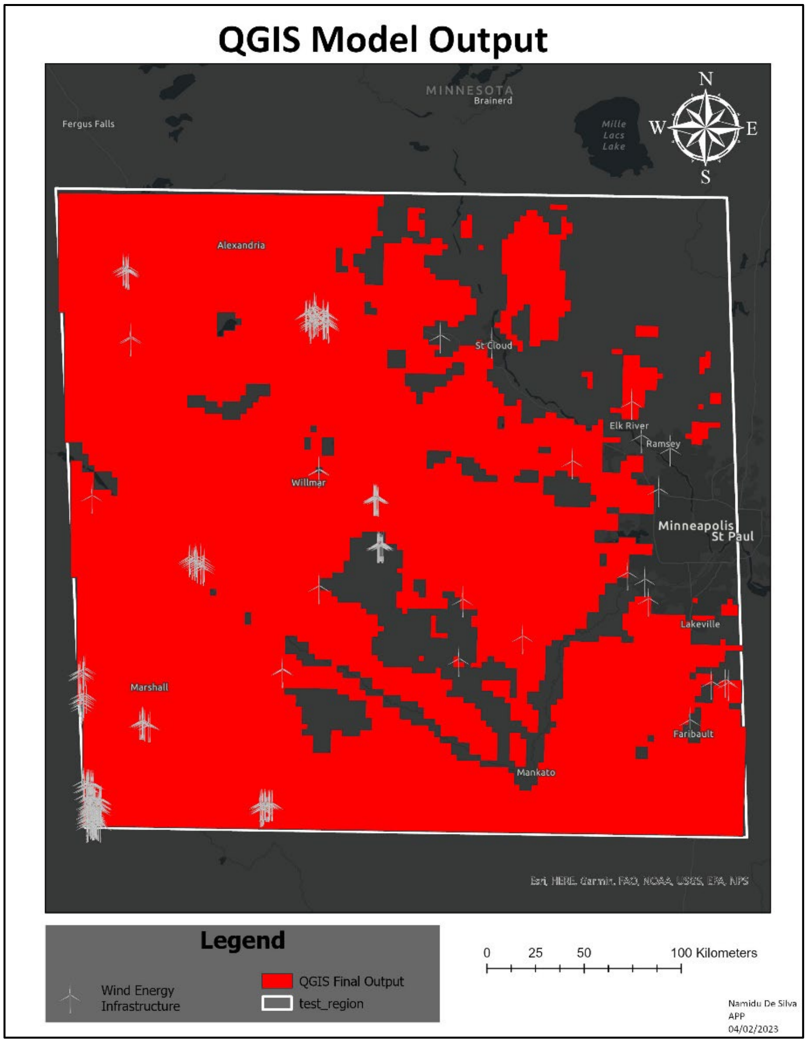


Figure 18. The output provided by the QGIS model. The output was generated in the QGIS platform as a shapefile and transported to ArcGIS Pro for analysis. As seen, the QGIS model's output covers a large amount of the study area and accounts for a large amount of the existing wind power infrastructure.

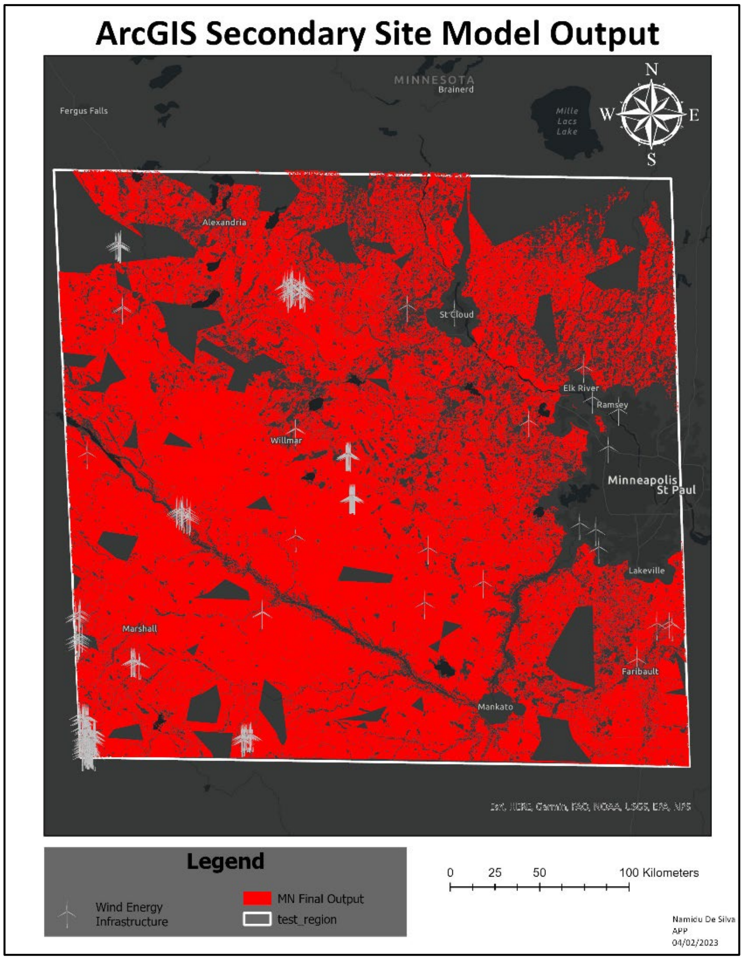


Figure 19. The ArcGIS model’s output for the secondary study area used for the comparison against the QGIS model. The output for the ArcGIS model also covers a large amount of the study area and visually accounts for a large number of the preexisting wind infrastructure.

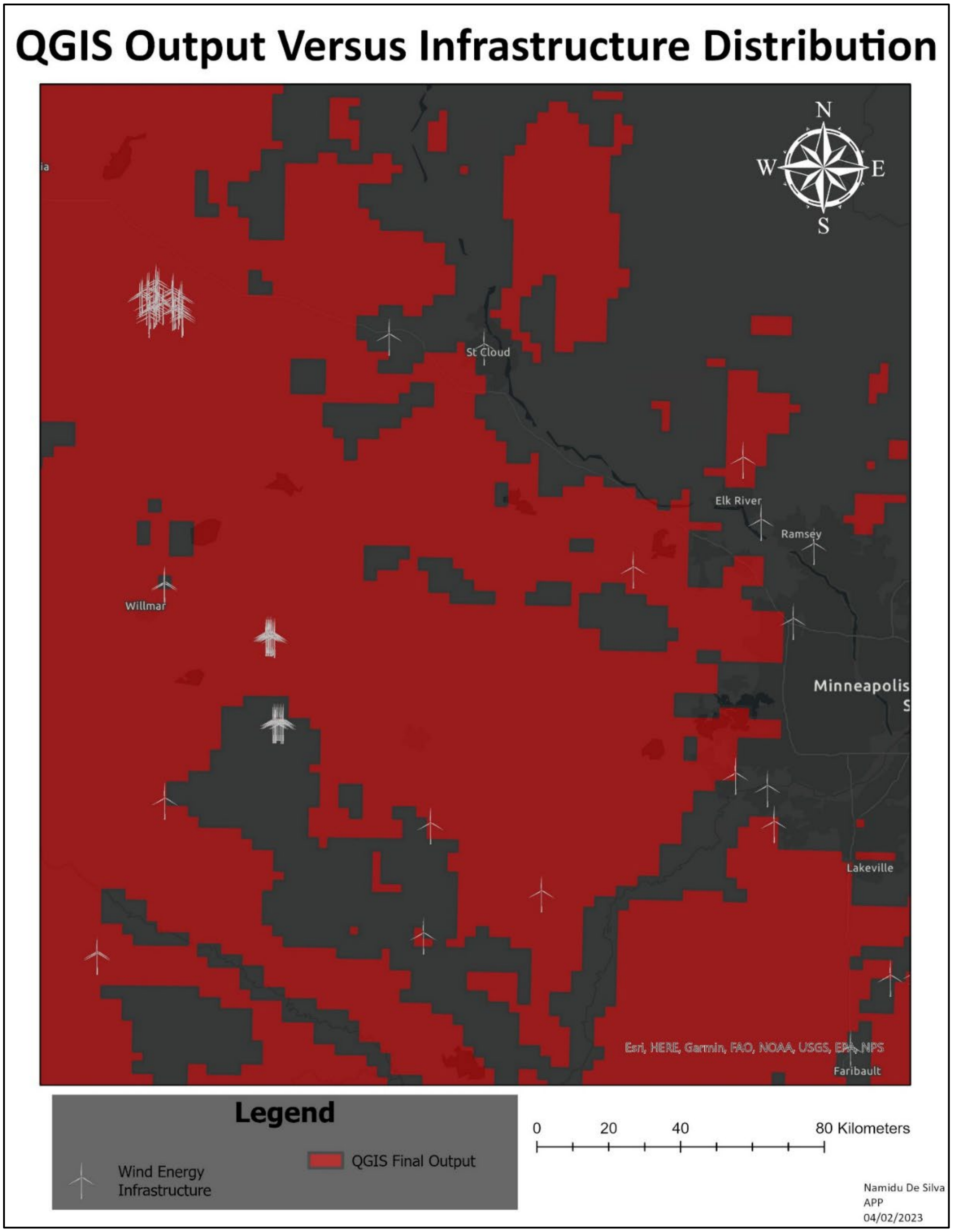


Figure 20. A closer examination of the model output from the QGIS model illustrates the areas, particularly in the eastern regions, closer to urban areas where there are standalone wind turbines that didn't get accounted for by the models output. These are wind turbines built under specialty cases and do not follow the same base criterion as conventional wind turbines such as the ones used in this project.