Master’s degree program in

Digital Skills for Sustainable Societal Transitions

Academic Year 2023/2024

MASTER THESIS

Machine Learning to Optimize Energy Management in Energy Communities: Prediction and Scheduling of Energy Consumption and Renewable Energy Production.

Supervisors:

Prof. Mutani Guglielmina

Candidate:

Zhou Xuan

Abstract: This research aims to develop wind energy prediction models through machine learning techniques as part of a national-level Renewable Energy Community (REC) optimization planning platform. The study focuses on predicting wind energy production potential across urban districts by utilizing multi-source environmental variables, including regional topography, surface characteristics, and meteorological factors. Through the collection and preprocessing of relevant data, machine learning algorithms are applied to analyse the relationships between wind energy production, energy consumption, and input variables. Ultimately, the model will be integrated into the platform to generate detailed wind energy production forecasts, providing scientific evidence for planners to support optimal allocation of wind resources and energy management. This research provides technical support for the platform's wind energy module, contributing to the sustainable development of renewable energy communities.

Key words: Renewable Energy Community; energy platform; machine learning; wind energy; GIS; regression; sustainable.

1. Introduction

In the context of growing global energy demands and intensifying climate change threats, renewable energy development has become a crucial component of national energy policies worldwide. Renewable Energy Communities (REC) represent a key solution for achieving energy transition and addressing climate change challenges. (Ahmed, 2024) As an innovative model of collective energy management, RECs not only promote shared energy production and optimized utilization but also help reduce energy poverty, foster social inclusion, and stimulate economic revitalization in inland and rural areas.

The Italian government actively supports wind energy development through various policy measures. For instance, the National Energy Strategy (Strategia Energetica Nazionale) (Compagnucci, 2023) and the National Integrated Energy and Climate Plan (Piano Nazionale Integrato per l'Energia e il Clima) establish renewable energy development targets for 2030, providing corresponding financial incentives and technical support. These policies have encouraged participation from residents and local governments while promoting widespread adoption of technologies such as photovoltaic, wind energy, and biomass across different regions, particularly in areas facing energy poverty. (Tatti, 2023)

This research presents an innovative methodology that integrates machine learning, Geographic Information Systems (GIS), and multi-source data fusion techniques to assess wind energy potential in Italy. The approach ingeniously combines wind speed data from the Global Wind Atlas, Aeolian database, and PVGIS model to construct a comprehensive dataset, providing high-quality training samples for subsequent machine learning models. The study further simulates the power generation efficiency of various wind turbine models (such as Vestas V52 and V80-2.0) under specific conditions, evaluating generation potential across typical days, hours, months, and years. Through the integration of data from different sources and spatiotemporal scales, this methodology enables a thorough characterization of wind energy resource characteristics in the study area.

Furthermore, this paper thoroughly considers the influence of geographical elements, including terrain and surface characteristics, on wind energy resource distribution. By classifying different geographical regions of Italy (plains, hills, mountains, and coastal areas) and incorporating corresponding meteorological data, the methodology precisely describes variations in wind energy potential across different regions, reflecting the geographical concept of "regional differentiation." In the modeling process, this research employs advanced machine learning algorithms, such as deep neural networks and random forests, to establish complex non-linear relationships between geographical features and wind energy output. Compared to traditional physical models, machine learning models can automatically learn and extract key patterns hidden within massive datasets, demonstrating superior predictive capability and generalization performance.

To apply the prediction results to RES, this research also employs machine learning to analyse electricity consumption patterns across Italian regions. By comprehensively considering data on population, building distribution, and industrial distribution, the methodology can grasp the electricity demand characteristics of different regions, identifying key information such as peak usage periods and load distribution, thereby supporting power system supply balance, intelligent dispatch, and new energy integration.

The methodology developed in this research can be integrated into the national geographic portal platform and applied to nationwide REC planning, promoting optimal allocation of wind energy and other distributed energy resources, enhancing energy self-sufficiency and system resilience of RECs, and supporting Italy's energy transition and 2050 carbon neutrality goals. This research not only improves the accuracy of wind energy assessment but also provides robust technical support and decision-making basis for wind energy development in Italy and globally, advancing the construction of sustainable energy systems.

2. Case Study

Under the dual pressures of climate change and energy crisis, renewable energy development has become a crucial pathway for global energy transition. Italy actively responds to relevant international agreements and EU directives by formulating ambitious national energy transition plans. According to EU's "Fit for 55" targets (Council, 2024), Italy plans to increase renewable energy share to 32% by 2030. Against this backdrop, Italy strives to improve its energy system, making new progress in three major objectives: ensuring energy security, achieving environmental sustainability, and promoting energy equity.

Energy poverty refers to the phenomenon where low-income households struggle to afford basic energy services such as heating and cooling (González-Eguino, 2015). Italy's energy poverty situation shows distinct regional differences, with southern regions like Calabria and Puglia experiencing energy poverty rates exceeding 12%, while northern regions generally remain below 6%. The main cause of this north-south disparity lies in industrial structure imbalance, where the agriculture-dominated south struggles to provide sufficient employment, while the north benefits from strong industrial foundations and higher per capita income. (Bardazzi, Energy Research & Social Science)

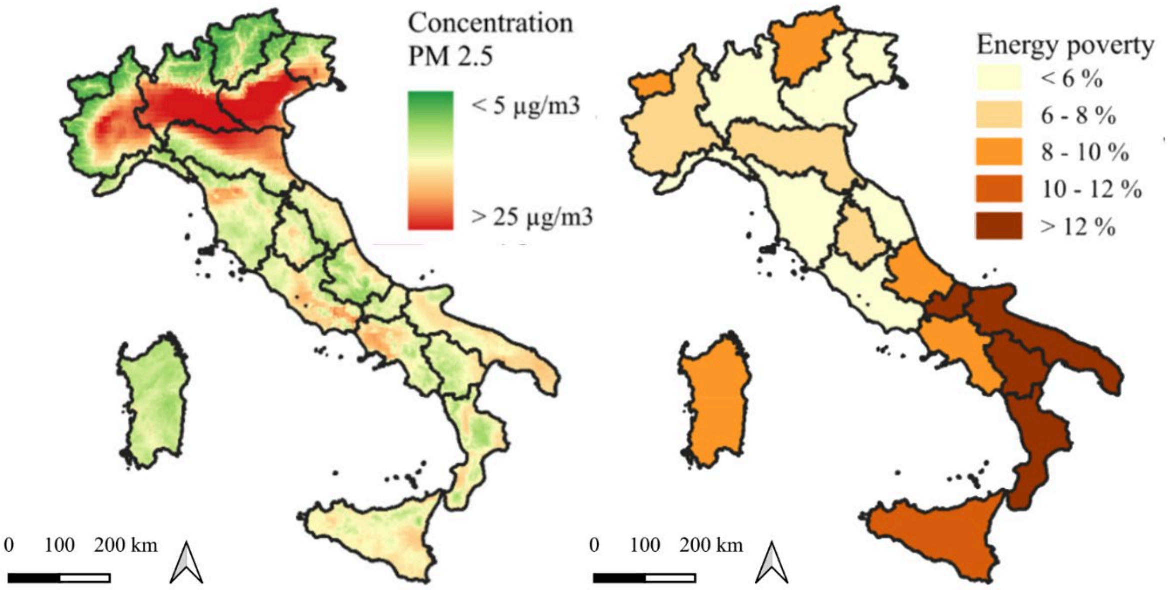


Figure 1. (a) PM2.5 annual average of Italy in 2019, (b) Energy poverty of Italy in 2022

Renewable energy promotion is viewed as one effective approach to mitigate energy poverty (Sovacool, 2012). First, distributed renewable energy systems (such as small-scale wind power and solar photovoltaic) can provide reliable energy supply to remote and impoverished areas, reducing dependence on traditional fossil fuels. Second, the construction and operation of renewable energy projects can create employment opportunities and promote local economic development. Furthermore, utilizing local renewable energy resources can reduce energy costs, improve energy affordability, and enhance community energy autonomy (REN21, 2021).

Wind energy, as a clean and sustainable form of energy, plays a crucial role in global energy transition. Wind power generation technology is mature with strong cost competitiveness, produces no greenhouse gas emissions, and is environmentally friendly (Gwec, 2021). Wind energy resources are abundant and widely distributed, suitable for large-scale development and utilization. With technological advancement and industry scale expansion, wind power costs have decreased significantly, becoming an important component in many countries' energy structure. The global wind power industry has achieved rapid development over the past decades. By the end of 2020, global cumulative wind power installed capacity reached 743 GW, an increase of 93 GW compared to 2019 (Joyce Lee, 2021).

2.1 Review and Current Status of Wind Power Generation in Italy

Italy possesses abundant wind energy resources owing to its long coastline and diverse topographical conditions. Since the first half of the 20th century, Italy began exploring wind energy resources for power generation. Due to technological limitations in materials and mechanical processing at the time, wind turbines were predominantly small-scale windmills, mainly distributed across southern islands and coastal regions for agricultural irrigation and civilian power supply. During this period, the Italian National Energy Agency initiated a series of wind power research and demonstration projects, some of which received EU funding. Modern commercial wind farms emerged from 1996. Benefiting from electricity price subsidies provided by the 1992 CIP6 decree, Italy's wind power installed capacity grew rapidly. In 1996, IVPC company constructed the first commercial wind farm near Benevento, with an installed capacity of 7.2MW (Pirazzi, 2005). By the end of 2022, the total number of wind turbines in Italy reaches about 7,450. The average installed wind turbine capacity is 3.3 MW and the average wind turbine capacity is 140 MW. (Greco, 2022) Compared to solar energy, wind power offers higher day-and-night generation stability, becoming the second-largest renewable energy source after solar power.

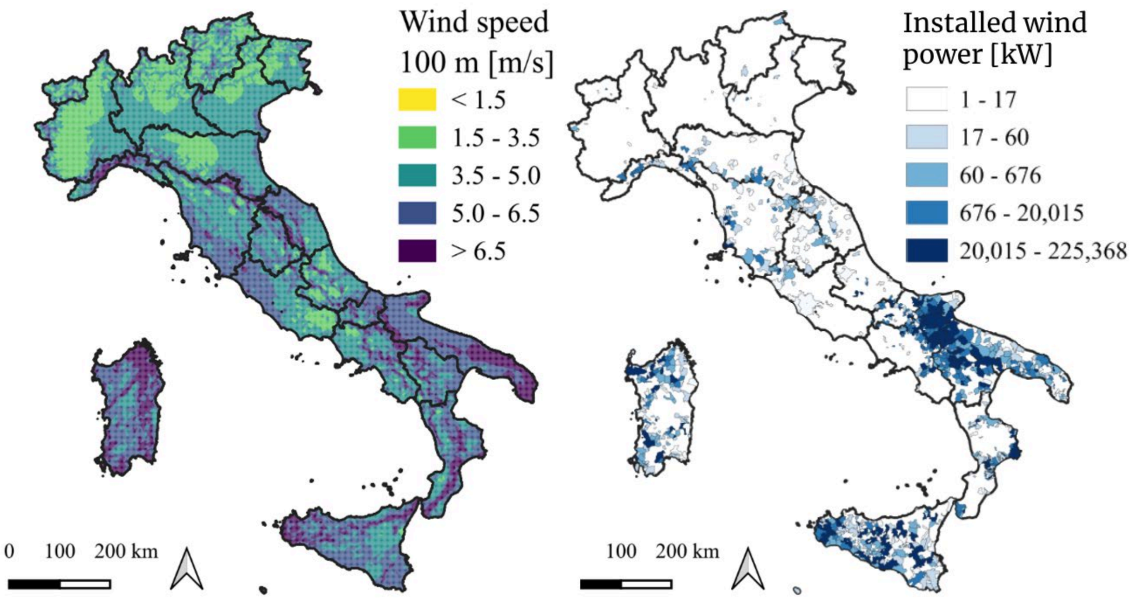


Figure 2. (a) Mean onshore wind speed at 100 m a.s.l. height, (b) Installed wind

generator power of Italy in 2021

Despite rapid development in Italy's wind energy resources and technological applications, their regional distribution shows significant imbalance. As shown in Figure 2, southern Italy experiences higher wind speeds (5.0-6.5 m/s and above), particularly in regions like Sardinia, Calabria, and Sicily, which have become the main concentration areas for Italy's wind power installed capacity. In contrast, northern regions have limited wind power development due to factors such as terrain and wind speed conditions. However, this concentrated development has led to issues with transmission grid capacity constraints and local energy balance. Additionally, since wind power generation is significantly affected by natural condition fluctuations, Italy still faces certain challenges in improving wind power stability. Future solutions may involve combining wind energy with other renewable sources and strengthening inter-regional energy coordination. Particularly within the energy community framework, the integration of wind power with solar and other distributed energy sources is viewed as an important solution for achieving carbon neutrality goals.

3. Methodology

This research aims to develop an integrated prediction model based on machine learning for assessing wind energy production potential across different regions of Italy, while predicting the corresponding energy consumption demand. On one hand, the model predicts wind energy production data for Italian municipalities by analysing a series of geographic, meteorological, and socioeconomic variables, such as elevation, vegetation coverage, distance from coastlines, and solar radiation index. On the other hand, the model thoroughly explores the spatiotemporal distribution patterns contained in energy consumption data.

To achieve this objective, this research adopts a data-driven modeling approach, using machine learning techniques to capture complex non-linear relationships, particularly in wind energy assessment. Unlike traditional physical models, the data-driven approach demonstrates greater flexibility in adapting to regional variations and improves prediction universality and accuracy by extracting useful features from large volumes of data.

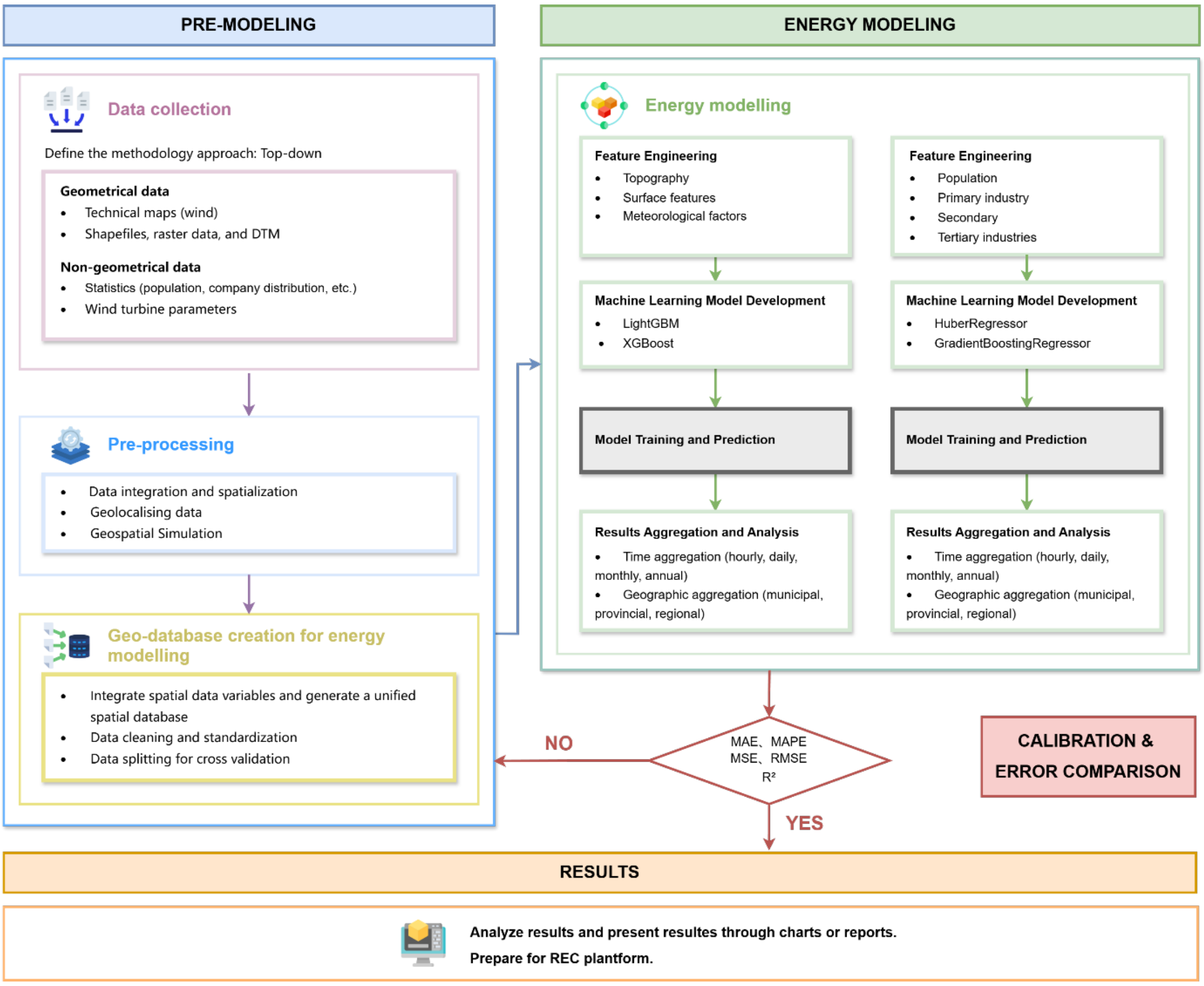


Figure 3. Methodological Flow Chart of This Research

The Figure 3 illustrates the research methodology, which consists of four main phases: pre-modeling, energy modeling, calibration and error assessment, and results and representation.

- **Pre-modeling:**

  The first phase encompasses data collection, preprocessing, and the creation of a geographic database. The data collection focuses on geometric data (such as technical maps, raster data, DTM) and non-geometric data (such as population distribution and wind turbine parameters). Subsequently, through data integration and spatial processing, geographic positioning and geographic simulation are integrated, and a unified spatial database is established using GIS technology. This phase emphasizes data cleaning, standardization, and the division of training and testing sets for cross-validation, preparing the foundational data for model development.

- **Energy Modeling:**

Through feature engineering, variables that significantly influence wind energy production are extracted and categorized into three types: terrain characteristics (such as slope and elevation), surface characteristics (such as land use, vegetation coverage, and building density), and meteorological factors (such as solar radiation). Subsequently, machine learning algorithms are employed to construct models for predicting wind energy production across different temporal scales (hourly, daily, monthly, yearly) and geographic regions (municipal, provincial, regional levels).

For energy consumption modeling, data primarily originates from ARERA. (ARERA, 2024) The baseline data consists of provincial energy consumption, which is mapped to the municipal level through multivariate linear correlation methods. Additionally, electricity consumption scenarios are determined for residential, primary, secondary, and tertiary sectors. (GSE, 2022) (Di Somma, 2018)

- **Calibration & Error Comparison:**

  Model performance evaluation and calibration follow the generation of prediction results. Multiple error metrics (such as RMSE, and R^2) compare model predictions to ensure high accuracy. When error metrics fall short of expected standards, the process returns to optimize the modeling phase until expected performance is achieved.

- **Results & Representation:**

  The final phase focuses on result analysis and presentation through charts, geospatial analysis, and time series analysis (hourly, daily, monthly, annual). This research ultimately provides robust data support for Renewable Energy Communities (REC) platform construction, helping quantify regional distribution of wind energy production and its potential impacts.

In the assessment of energy systems, the selection of quantification indicators directly impacts the accuracy and comprehensiveness of system performance evaluation. This research employs a comprehensive indicator system (Table 1) to evaluate the system across three dimensions: energy consumption, production, and economic performance. And it adopts a multi-level temporal scale assessment methodology, evaluating the energy system at hourly, daily, monthly, and annual levels. These indicators not only reflect the system's instantaneous operational status but also demonstrate its long-term operational effectiveness. (GSE, 2022) (Di Somma, 2018) (Mutani, 2021)

Table 1. Data collection information

Indicators	**Calculation Method**	**Significance**
Self-Consumption (SC)	min(C,P), where C is consumption and P is production	Reflects the actual efficiency of system energy utilization
Uncovered Demand	C-P (when C>P)	Characterizes the energy deficit in system supply
Over-Production	P-C (when P>C)	Quantifies excess energy production in the system
Self-consumption Index (SCI)	SC/P	Measures the effective utilization rate of energy production
Self-sufficiency Index (SSI)	SC/C	Evaluates the degree of demand satisfaction
Over-Production Index (OPI)	OP/P	Analyses the proportion of excess production
Energy Poverty Index (EPI)	Energy costs/Income	Assesses the economic feasibility of the system

Wind resource assessment demands high-quality, comprehensive data. This study collected multidisciplinary data covering natural geography and meteorology across the study area. Data sources span government departments, research institutions, and companies (Table 2).

To ensure data quality and consistency, this research implemented a standardized preprocessing workflow. This fundamental stage encompasses data cleaning, standardization, and spatial data alignment. Missing value imputation and spatial aggregation methods were employed to ensure compatibility and format uniformity across different data sources. For instance, wind speed and meteorological data were spatially aggregated by municipal boundaries to obtain median wind values, while population and building density data were averaged to reflect wind energy community potential across municipalities.

Table 2. Data collection information

Source	Data	Scale	Year
Copernicus	Corine Land Cover	Municipal	2018
Aeolian	Wind speed and direction data wind energy production	Municipal	2019
Global Wind Atlas	Wind variables - Wind Speed, Capacity Factor	Municipal	2019
ISTAT	Statistical Atlas of Municipalities	Municipal	2020
	Census of Population and Housing	Municipal	2011-21
	Census of Industry and Services	Municipal	2011
	General Census of Agriculture	Municipal	2010
PVGIS	Solar irradiation, diffuse / global ratio, irradiance, electricity production, temperature	Municipal	2020
Meteotest	Linke Turbidity Factor	Municipal	2019
Tinitaly	Digital Terrain Model - DTM 10 m	Municipal	2022

3.1 Place-based Approach

The place-based approach acts as the key methodology adopted in this research. This approach analyses from geographical and location-specific perspectives, fully considering multiple dimensions including environmental, social, economic, and policy factors, with specific locations at its core. In this research, this approach aims to explore complex relationships between wind energy potential and its influencing factors from a geographical perspective. Specifically, a location-specific model framework was established by analysing associations between geographical characteristics and wind energy production across different locations. Most notably, terrain significantly impacts wind resources. For instance, topographical features such as hills, valleys, cliffs, escarpments, and ridges affect wind patterns. Wind speeds accelerate near mountaintops and ridge peaks, while typically decreasing near ridge bases and valleys. This research utilizes GIS spatial analysis and mapping capabilities to visually present and assess spatial distribution of various assessment factors.

The research progresses from small to large scales, gradually revealing wind energy resource distribution characteristics and influencing factors across different spatial scales through comprehensive analysis at municipal, provincial, and regional levels. Municipalities serve as basic research units, with wind energy capacity calculations and predictions conducted at the municipal level. This bottom-up research approach requires collection of geographical characteristic data directly related to municipalities, including wind speed, air density, minimum and maximum elevation, forest coverage rates.

3.2 Process-driven Approach.

Process-driven methodology is a research and analysis approach based on explicit physical processes. In wind energy assessment, this traditional method primarily relies on meteorological, fluid dynamics, and statistical models, employing numerical simulation techniques to quantitatively evaluate wind resource potential. This approach uses core meteorological parameters such as wind speed and direction, utilizing long-term wind speed observational data (from weather stations or satellite data) to calculate average wind energy resource distribution in specific regions.

To implement a process-driven machine learning method, the primary task is constructing prediction models that incorporate multidimensional environmental characteristics including terrain, land use, and population density. This approach requires baseline datasets of energy production outputs from typical regions, obtained through classical wind energy calculations. Compared to traditional methods, process-driven machine learning approaches not only reveal the physical mechanisms of wind energy production but also provide more precise assessments of key wind energy parameters through data-driven methods. By establishing mapping relationships between wind energy and geographical characteristics, the model can directly output wind energy production (P) for target regions while comprehensively considering various environmental factors' effects on wind energy conversion efficiency.

Wind turbine power generation depends on several parameters, such as wind speed, wind direction, air density, and turbine characteristics: (Sohoni, 2016)

$$\boldsymbol{P = \frac{1}{2} \cdot \rho \cdot A \cdot CP(\lambda, \beta) \cdot v^3}$$

Where P is the wind power (W); ρ is the air density; A is the swept area; CP is the power coefficient, a function of the tip speed ratio λ and blade pitch angle β (considering Vestas V52 and Vestas V80-2.0); and v is the wind speed at hub height.

In practical applications, process-driven methods can construct complete computational chains for wind energy resource assessment through precise modeling of each parameter combined with numerical simulation techniques. The advantage of this method lies in its explicit physical mechanisms and traceable computational processes, providing reliable theoretical support for wind energy resource assessment. However, traditional process-driven methods often face challenges in computational efficiency and accuracy when simulating wind fields under complex terrain conditions, which has prompted researchers to explore hybrid methods combining machine learning techniques with traditional physical models.

Furthermore, process-driven methods provide a complete theoretical framework for wind energy prediction, giving model prediction results clear physical significance. This approach not only predicts wind energy production but also analyses various factors affecting wind energy conversion efficiency, providing scientific basis for wind farm site selection and operational optimization. By combining process-driven methods with modern data science techniques, model prediction accuracy and computational efficiency can be improved while maintaining physical mechanism correctness.

Historically, process-driven methods have played a crucial role in traditional wind energy development. However, with advances in data acquisition and processing technologies, data-driven methods have begun to emerge. This approach breaks free from reliance on complex physical process modeling, directly learning relationships between wind energy and its influencing factors through massive data, providing an efficient and flexible assessment method. The next section will specifically explore data-driven methods.

3.3 Data-Driven Approach

In the big data era, data-driven approaches have emerged as a crucial paradigm in wind energy resource assessment and electricity demand forecasting. Unlike traditional physical models, data-driven methods do not rely on calculations of complex physical processes but instead directly extract knowledge from massive empirical or simulated datasets to construct relationship models between wind energy potential, electrical load, and various influencing factors. The core of this paradigm lies in utilizing machine learning algorithms to achieve intelligent feature identification and prediction through the mining and fusion of multi-source heterogeneous data.

The theoretical foundation of data-driven methods stems from statistical learning theory and computer science. The fundamental concept is that energy production and consumption patterns are embedded within various meteorological, geographical, and socioeconomic data, and through analysis and learning of these data, the inherent laws of energy distribution and consumption can be revealed. From a methodological perspective, the data-driven approach encompasses several key steps:

- **Data Collection and Preprocessing:** On the production side, multi-source geospatial data including terrain, land use, and meteorological information are collected. On the demand side, socioeconomic indicators such as demographic data, building distribution data, household numbers, and industrial structure data are gathered. These raw data undergo cleaning, fusion, standardization, and other preprocessing operations to form structured training datasets.
- **Feature Engineering:** On the production end, feature variables related to wind energy potential, such as terrain roughness and solar radiation coefficients, are extracted or constructed from raw data. On the demand end, indicators reflecting energy consumption characteristics, such as building area by sector, are constructed. This process transforms implicit energy supply-demand patterns in the data into explicit, computable indicators. Simultaneously, features are screened to eliminate auto-correlated or strongly correlated features, avoiding data redundancy.
- **Model Training and Optimization:** Appropriate machine learning

algorithms are selected to construct both energy supply prediction models and demand prediction models. The supply model uses geographical and meteorological features as input to predict wind energy density and potential power generation; the demand model uses socioeconomic characteristics as input to predict energy consumption. Through techniques such as cross-validation and parameter tuning, the model's generalization performance is continuously optimized to improve prediction accuracy.

- **Model Validation and Application:** Independent test datasets are used to evaluate the predictive performance of both supply and demand models. The trained models, when applied to target municipalities, can generate not only high-resolution wind energy resource distribution maps but also energy demand forecasts. These bidirectional supply-demand prediction results provide more comprehensive decision support for regional energy planning.

Data-driven methods also face certain challenges. A primary issue is that model prediction performance heavily depends on the quality and representativeness of training data. If sample data contains bias or is incomplete, the trained model may fail to accurately reflect actual wind energy distribution patterns. Additionally, since data-driven models are typically "black box" models, their internal logic is difficult to interpret directly, which to some extent limits the interpretability and credibility of model results. To overcome these limitations, this research explores a hybrid modeling paradigm combining data-driven and process-driven approaches, striving to maximize the advantages of data intelligence while maintaining physical mechanism rationality.

3.3.1 Machine Learning Model Development

Wind energy resource assessment and energy consumption are essentially complex nonlinear regression problems. Traditional physical or statistical models struggle to accurately characterize the spatiotemporal distribution patterns of energy production and consumption. To improve prediction accuracy and resolution, this research introduces machine learning methods. Considering the nonlinear complexity of energy supply-demand relationships, this study employs machine learning models for prediction.

For wind energy resource assessment, we selected two ensemble learning algorithms based on Gradient Boosting Decision Trees (GBDT): LightGBM and XGBoost. The core concept of GBDT is to iteratively train a series of decision trees to gradually approximate the true mapping relationship. Each newly generated tree attempts to fit the prediction errors of previous trees, and through multiple iterations, the cumulative predictions continuously converge toward the true values.

LightGBM is an algorithm based on gradient boosting, implementing ensemble learning through multiple weak classifiers. As shown in Figure 4, its training process involves continuously adding decision trees, with each tree attempting to fit the prediction errors of previous trees, ultimately summing all trees' predictions to produce the model's output (Korstanje, 2021)。

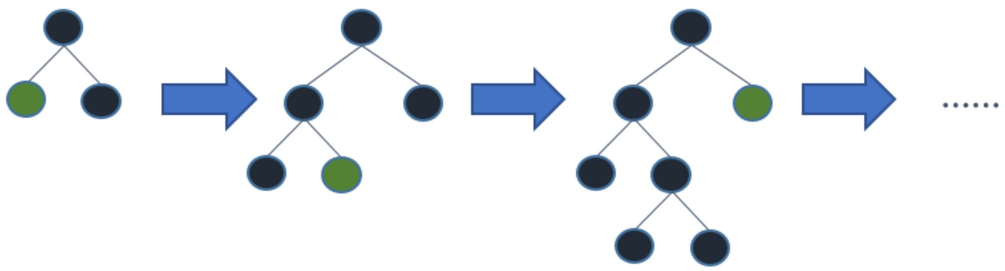


Figure 4. LightGBM trees grow by leaf pattern

At the algorithmic level, LightGBM employs histogram-based decision tree algorithms and a Leaf-wise tree growth strategy. While traditional GBDT algorithms require feature value pre-sorting when constructing decision trees, LightGBM discretizes continuous features into finite bins, significantly reducing time complexity. LightGBM's innovative Leaf-wise growth approach fundamentally differs from the Level-wise strategy used in traditional GBDT. The Leaf-wise strategy identifies and splits the leaf node with the maximum split gain during each iteration, theoretically converging to better accuracy faster than Level-wise approaches. However, it's worth noting that excessive growth may lead to overfitting, hence the algorithm introduces maximum depth limits as a regularization measure. Feature parallelism is another crucial characteristic of LightGBM, combining highly mutually exclusive features by calculating feature mutual exclusivity degrees, significantly reducing feature dimensionality without sacrificing accuracy. This optimization is particularly significant when handling multi-source heterogeneous data in wind energy prediction, such as Corine Land Cover land use data. During model training, LightGBM further enhances computational efficiency through Gradient One-Side Sampling (GOSS), retaining samples with large gradient absolute values while randomly sampling those with smaller gradients, reducing computation while maintaining accuracy.

In wind energy prediction: its efficient computational framework can handle large-scale historical data; second, its histogram algorithm and feature parallel strategy efficiently process multidimensional environmental features; then, its unique tree growth strategy and regularization mechanism help capture complex nonlinear relationships between wind speed and environmental factors. Research has shown that LightGBM provides higher prediction accuracy in wind energy prediction tasks compared to traditional machine learning methods, effectively extracting local data features and temporal characteristics to achieve accurate wind power prediction (Ren, 2022).

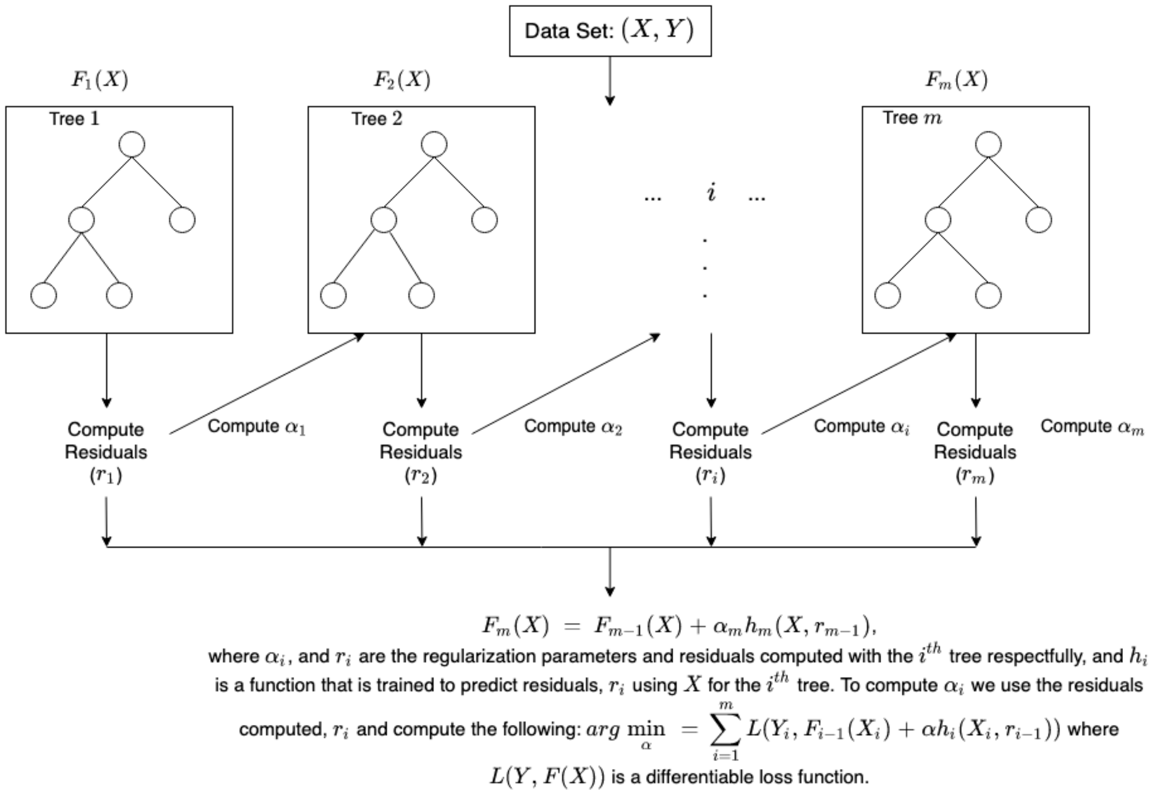


Figure 5. XGBoost model

XGBoost (eXtreme Gradient Boosting), developed by Tianqi Chen in 2014 (Chen, 2016), is another popular GBDT implementation. Compared to LightGBM.

As shown in Figure 5 (SageMaker, 2024), XGBoost adds regularization terms to the objective function and supports parallel computing, handling multiple tasks simultaneously for greatly improved efficiency. Additionally, XGBoost automatically handles missing values by finding optimal missing value split directions for each feature.

For energy consumption prediction, HuberRegressor and GradientBoostingRegressor algorithms were selected. Unlike classification trees, regression trees output continuous values rather than discrete class labels at leaf nodes. Regression trees determine each leaf node's output value by minimizing a loss function (such as mean squared error) to make sample predictions as close as possible to true values.

HuberRegressor employs the Huber loss function, which balances mean squared error and absolute error. For smaller prediction deviations, Huber loss approximates mean squared error and is insensitive to outliers; for larger deviations, it approaches absolute error, reducing outlier impact. This characteristic is particularly useful in energy consumption prediction, as electricity usage data may contain abnormal peaks or valleys caused by measurement errors or temporary events that don't represent overall data trends. Huber loss adaptively adjusts error contributions, maintaining mean squared error's efficiency while incorporating absolute error's robustness.

GradientBoostingRegressor represents the direct application of classical GBDT to regression tasks (scikit-learn, GradientBoostingRegressor, 2024). It follows Friedman's original formulation, using negative gradients as residual approximations, with each regression tree directly fitting previous trees' residuals. Compared to Newton method-based XGBoost, GradientBoostingRegressor's optimization process is more straightforward. It supports various loss functions, including Huber loss and quantile loss, and includes practical features like early stopping and feature importance evaluation, facilitating model tuning and feature selection. In energy consumption prediction, GradientBoostingRegressor has gained widespread application due to its concise efficiency and ease of parameter tuning.

4. Pre-modelling

Developing a machine learning model to accurately predict wind energy production across Italian municipalities first requires constructing high-quality training datasets. This dataset must contain two key components: wind energy production data from typical municipalities and their corresponding geographical feature data. Through this approach, it can enable machine learning models to understand the intrinsic relationships between geographical features and wind energy production.

Calculating wind energy production based on physical models is a complex process requiring consideration of multiple factors, including wind speed, air density, turbine characteristics, and terrain effects. Considering computational costs and data accessibility, this research adopted an approach based on typical municipalities. These typical municipalities were selected as representative samples based on terrain characteristics (plains, hills, mountains, and coastal areas). Through detailed wind energy calculations for these typical municipalities, we could obtain reliable training data without conducting time-consuming physical modeling for all municipalities.

Based on the physical model (as described in Section 3.2 process-driven approach), wind energy production calculations were performed for these typical municipalities across different time scales, including hourly, daily, monthly, and annual scales. These calculations fully considered turbine performance characteristics, terrain effects, and meteorological condition variations. This approach yielded a comprehensive training dataset incorporating temporal sequence features, providing rich data for subsequent machine learning models. It significantly reduced computational costs by focusing limited computational resources on representative samples. Additionally, this method allowed for more detailed analysis and validation of selected samples, ensuring training data quality. Finally, by ensuring typical municipalities covered the main geographical feature types of the study area, we could enhance the model's generalization capability.

This "typical sample calculation - machine learning generalization" strategy allows us to maintain prediction accuracy while dramatically improving computational efficiency. Through machine learning models, we can extend knowledge gained from typical municipalities to other regions without conducting tedious physical calculations. This approach not only improves wind energy assessment efficiency but also provides a feasible technical pathway for large-scale regional wind energy potential assessment.

4.1 Wind Turbine Technology

Before conducting wind energy production calculations, appropriate wind turbine models must be selected. According to the actual situation in the Italian wind power market, Vestas turbines hold the largest market share, with Vestas V52 (rated power 850 kW) and Vestas V90 (rated power 2,000 kW) accounting for 25.1% and 25.6% of installed capacity, respectively.

The Vestas V52 wind turbine features a 52-meter rotor diameter and maximum hub height of 86 meters. This medium-scale turbine is particularly suitable for deployment in areas with lower wind speeds or space constraints. Its cut-in wind speed is 4 m/s, with a cut-out wind speed of 25 m/s. The Vestas V90 represents a larger power class turbine, with a rotor diameter reaching 90 meters and maximum hub height of 105 meters. Despite its larger power capacity, this model maintains the same cut-in (4 m/s) and cut-out (25 m/s) wind speeds as the V52.

Calculation of Key Parameters

a. Wind Speed Time Series Construction:

The energy output of wind turbines primarily depends on wind speed. This research utilizes high-resolution wind speed data from the Global Wind Atlas as the foundational data source. This dataset employs advanced numerical simulation methods, combining terrain and surface characteristics to provide annual average wind speed data at 100 meters height. To improve temporal resolution, a time index method based on historical observational data was introduced. This method can be expressed as:

$$\boldsymbol{V_{hourly} = V_{annual} \cdot Index}$$

Where V_hourly represents hourly wind speed, V_annual is the annual average wind speed, and Index is a correction coefficient reflecting temporal variation characteristics.

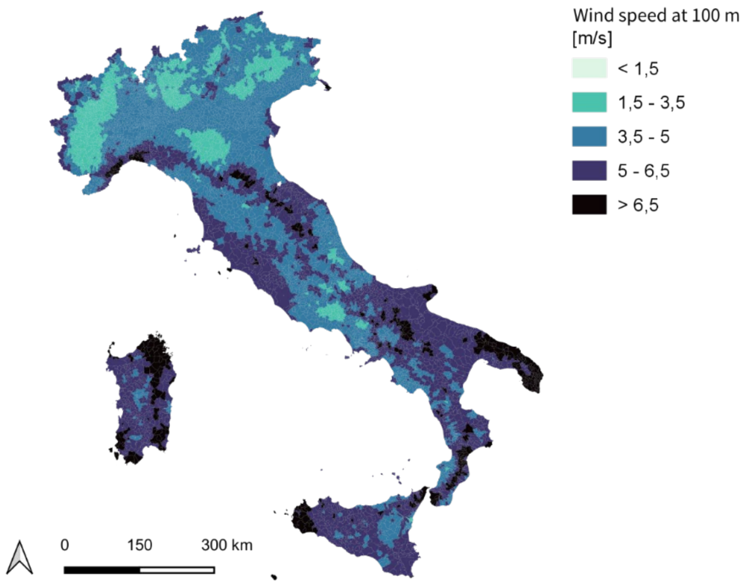


Figure 6. Annual average wind speed at 100 m above sea level in Italian municipalities

Analysis of typical municipalities' wind speeds revealed significant variations in wind patterns even within the same terrain classification. These differences manifest not only in absolute wind speed values but also in their temporal variation patterns. Through detailed analysis of eight representative municipalities, we discovered that spatiotemporal wind speed characteristics are far more complex than traditional classification methods would suggest.

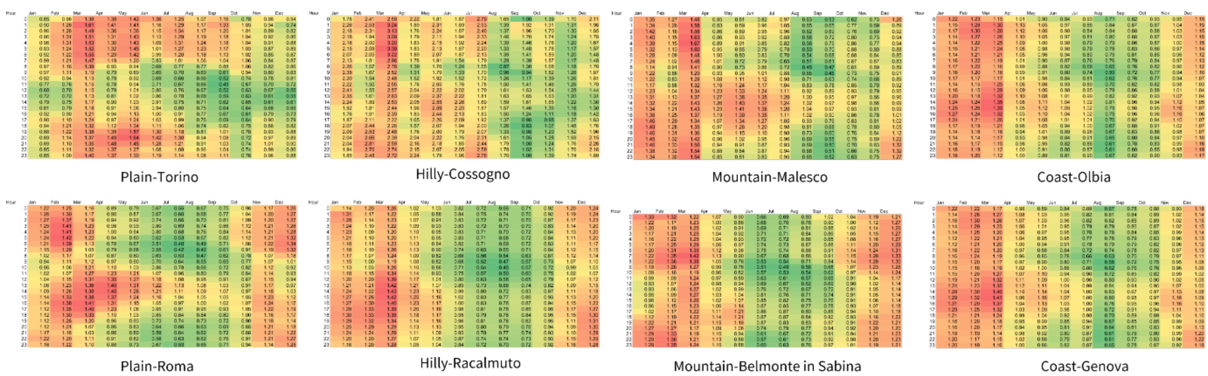


Figure 7. Wind speed trends in different regions

Taking plains regions as an example, despite both being in flat terrain areas, Torino and Roma display markedly different daily wind speed patterns. Turin exhibits stronger afternoon wind speed fluctuations during spring, with coefficients reaching 1.4-1.5, while Rome shows more moderate daily variations with generally smaller wind speed coefficient fluctuations.

The differences are even more pronounced in hilly regions. Cossogno's wind speeds can reach more than twice the annual average during spring afternoons, demonstrating strong seasonal and daily variation characteristics. In contrast, Racalmuto, though also located in a hilly region, shows relatively moderate wind speed variations, rarely exhibiting such significant peaks. These differences highlight the profound impact of local terrain features on wind speed patterns. Mountain regions' Malesco and Belmonte in Sabina similarly display distinct variation characteristics, while coastal areas' Olbia and Genoa are both influenced by sea-land breeze systems.

This indicates that simple terrain classification methods may not fully capture complex wind speed variation characteristics, emphasizing the need to consider more local factors when establishing prediction models. Traditional deterministic models might struggle to accurately describe such complex wind speed variation patterns, while data-driven machine learning methods show promise in providing more accurate predictions through learning these complex patterns.

b. Analysis of Air Density Spatiotemporal Characteristics:

In wind energy assessment, precise calculation of air density is crucial for result accuracy. High-resolution annual average air density distribution data at 100 meters height for the Italian region was obtained through the Global Wind Atlas.

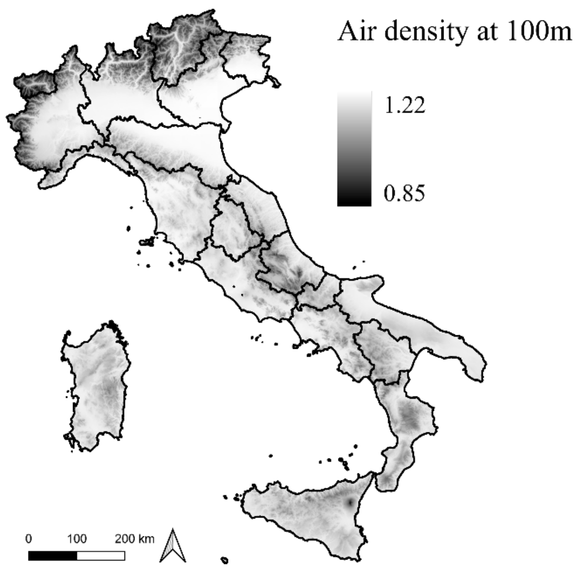


Figure 8. Air density at 100 m

However, for more precise modeling, monthly typical daily air density data is needed. To address this temporal scale conversion issue, we adopted a temperature correction method based on the ideal gas state equation.

The ideal gas state equation indicates that under relatively stable pressure conditions, air density is inversely proportional to temperature:

$$\boldsymbol{\rho}[\mathbf{kg}\cdot\mathbf{m}^{-3}]=\frac{\boldsymbol{\rho}_{\text{ref}}\cdot\boldsymbol{T}_{\text{K,ref}}}{\boldsymbol{T}_{\text{K}}}$$

Where ρref =1.29kg/m3 is the reference air density; TK,ref=273.15 is the reference temperature; and TK is the actual temperature.

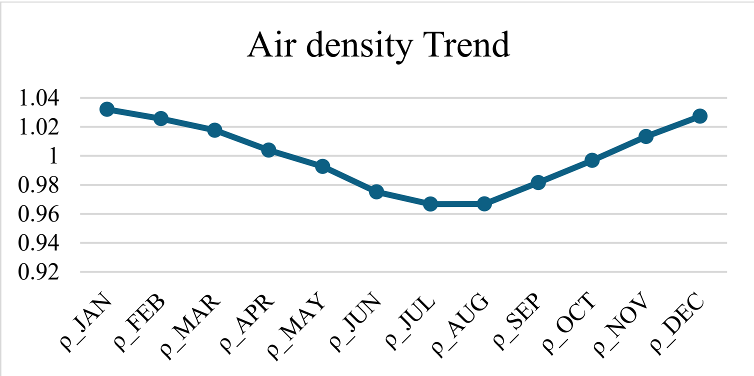


Figure 9. Air density trend

This theoretical relationship provides the foundation for converting annual average air density to monthly data. By analysing temperature observation records from 2016-2020 provided by the PVGIS system, we could determine monthly temperature variation characteristics relative to annual averages, deriving monthly air density variation patterns. Multiplying annual average air density values by monthly variation coefficients obtained in the first step yields estimated monthly air density values for each municipality.

c. Swept Area

The swept area A depends on the wind turbine diameter and is calculated as::

$$\mathbf{A}\ =\boldsymbol{\Pi}\cdot\left(\frac{\boldsymbol{diameter}}{\mathbf{2}}\right)^{\mathbf{2}}$$

d. Power Coefficient CP and Power Curve

The power coefficient (C_p) of wind turbines serves as a crucial indicator of wind energy conversion efficiency, exhibiting significant nonlinear characteristics that vary with wind speed. To accurately capture this nonlinear relationship, we implemented a machine learning-based power curve fitting method. This approach first uses gradient boosting regression to establish the mapping relationship between wind speed and power coefficient, followed by Savitzky-Golay filtering to smooth the fitted results and eliminate data noise.

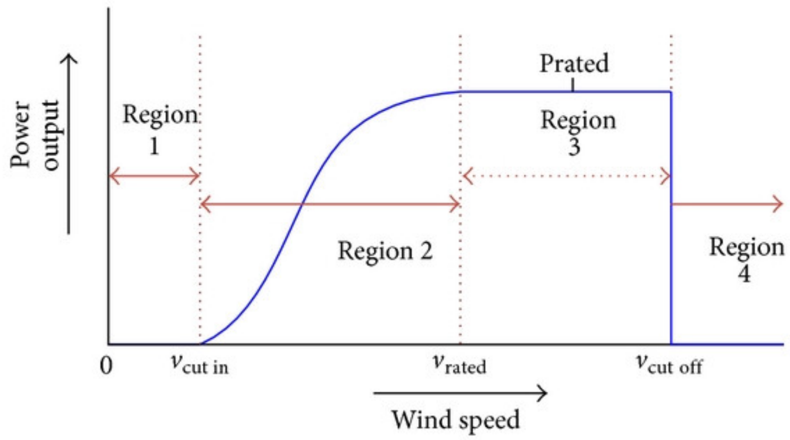


Figure 10. Typical power curve of a pitch-controlled wind turbine

The model achieved remarkably high precision levels for both Vestas V52 and V90 turbine models. The Vestas V52 showed excellent performance metrics with a mean square error (MSE) of 0.00005 and a coefficient of determination (R^2) of 0.9969. The Vestas V90 demonstrated even more impressive results, with an MSE of 0.00003 and R^2 reaching 0.9976. These metrics strongly validate the method's superiority in characterizing power curve features. When comparing the fitted curves with measured data, we found that the model accurately captures power coefficient variation trends, particularly in the critical range between cut-in and rated wind speeds, maintaining prediction errors within acceptable limits.

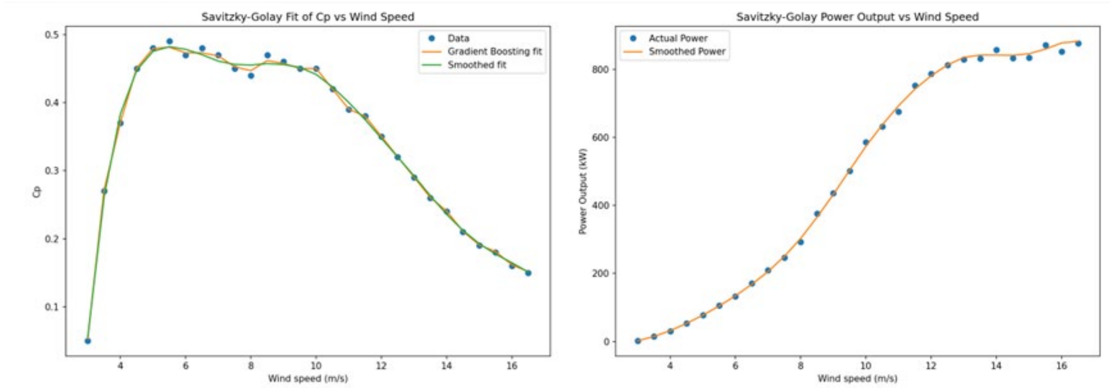


Figure 11. CP and P prediction results and trends of V52

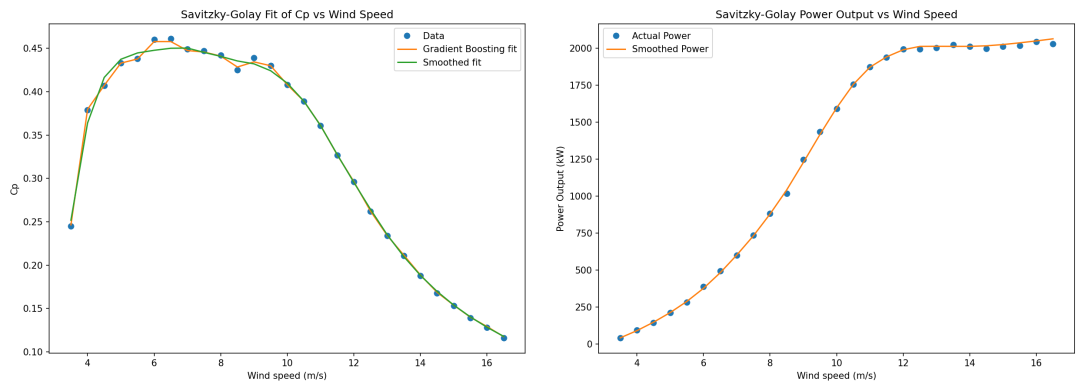


Figure 12. CP and P prediction results and trends of V80

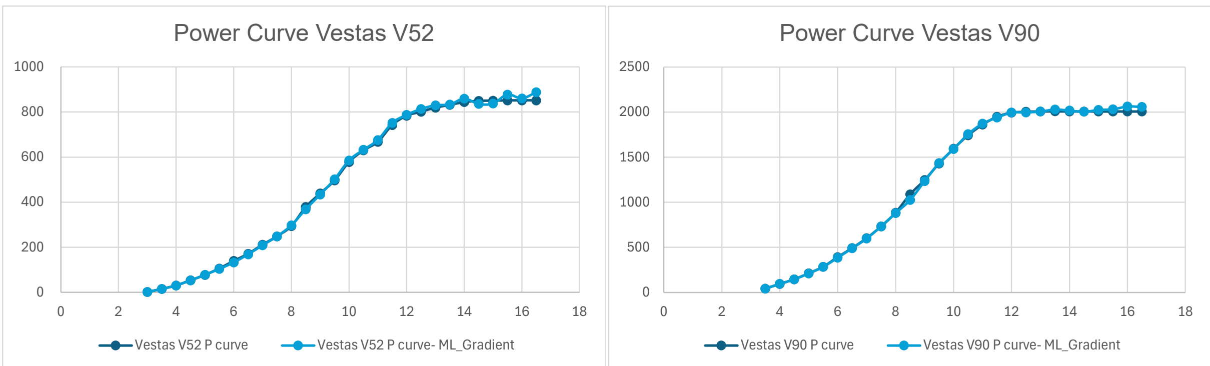


Figure 13. Comparison of predicted power curves and theoretical power curves for V52 and V80

Multi-scale Wind Energy Production Calculation

Based on the above parameter calculations, we established a multi-scale wind energy production calculation framework that spans from hourly to annual timescales. This framework employs a hierarchical calculation strategy to ensure consistency and reliability of results across different temporal scales. Using wind turbine power curves (wind speed v vs. power output P), the model can directly calculate hourly production. The calculation of wind energy output begins with the energy output of a single wind turbine. Starting from hourly outputs, it extends to daily output within 24 hours, then to typical monthly output. Monthly output is obtained by multiplying by the number of days in each month, and annual totals are calculated by summing all 12 months.

This hierarchical approach ensures that calculations at each time scale build upon and remain consistent with those at shorter time scales, providing a comprehensive view of wind energy production patterns. The method's strength lies in its ability to capture both short-term variations and long-term trends, making it particularly valuable for both operational planning and strategic decision-making in wind energy development.

Hourly power output

$$\boldsymbol{P_{hourly}} = \boldsymbol{f}(\boldsymbol{V_{hourly}})$$

f is the function derived from fitting the power curve, representing the relationship between wind speed and power output.

Hourly production:

$$\mathbf{E_{hourly\_wh}} = \boldsymbol{P_{hourly}} \cdot \boldsymbol{t} = \boldsymbol{P_{hourly}} \cdot \mathbf{1} = \boldsymbol{P_{hourly}}$$

$$\mathbf{E_{hourly\_kwh}} = \frac{\mathbf{E_{hourly_w}}}{\mathbf{1000}}$$

Daily production:

$$\mathbf{E_{daily\_kwh}} = \sum_{i=1}^{24} \boldsymbol{E_i} = \boldsymbol{E_1} + \boldsymbol{E_2} + \boldsymbol{E_3} + \boldsymbol{E_4} + \boldsymbol{E_5} \ldots\ldots + \boldsymbol{E_{24}}$$

Monthly production:

$$\mathbf{E_{monthly_{kwh}}} = \mathbf{E_{daily_{kwh}}} \cdot \boldsymbol{N_{day\_of\_month}}$$

Annual production:

$$\mathbf{E_{annual\_kwh}} = \sum_{i=1}^{12} \boldsymbol{E_i}$$

4.2 Wind Turbine Layout and Losses

Determining the number of wind turbines to be installed in each municipality is essential. The layout design of wind turbines is crucial in wind farm planning, significantly affecting overall power generation efficiency and economic benefits. The installation of wind turbines is constrained by factors such as world heritage sites, cultural heritage, nature reserves, Natura 2000 network, wetlands, bird areas, volcanic regions, lakes, rivers, woodlands, buildings, hazardous areas, and residential zones. These areas are identified using data from the Corine Land Cover, Natura 2000, and OpenStreetMap databases. These constraints determine the permissible locations for wind turbine installations (McKenna, 2022). In Italy, wind energy technology faces important constraints. On average, only 5% of the territory is suitable for wind energy technology, with a higher percentage in southern regions and islands.

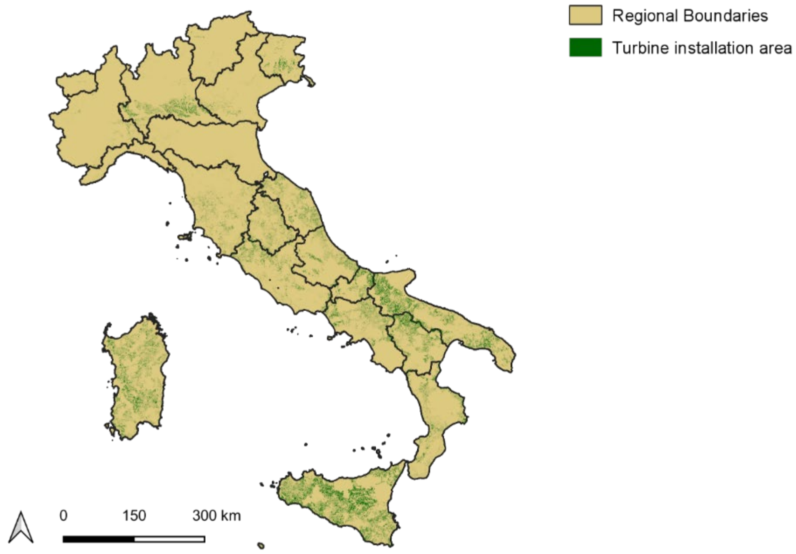


Figure 14. Areas in Italy where wind turbines can be installed

For individual turbines, only areas larger than 15 meters in length and width are considered. Given the V52 and V80 turbine models, the turbine tower is fixed on a concrete base approximately 15 meters in diameter, typically level with the surrounding ground. (Government, 2014) (Brian Tri, 2023)

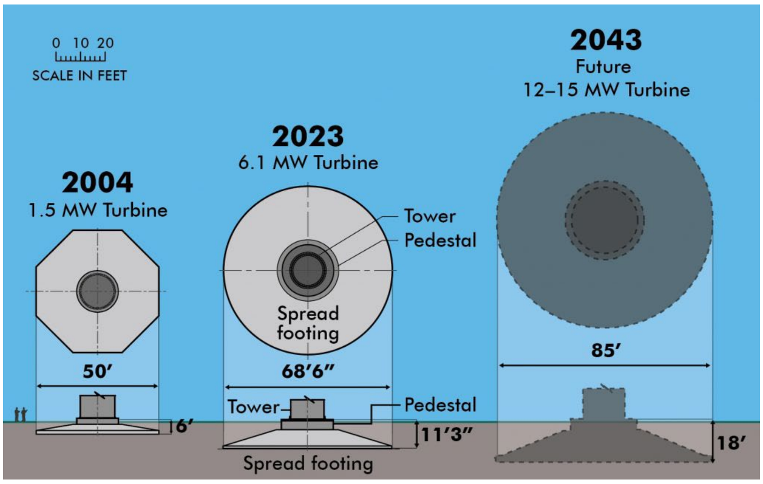


Figure 15. Required base diameters for wind turbines of different powers

For wind farms with multiple turbines, wake losses must be a key consideration. Wake losses refer to the impact of the wake generated by upstream turbine blades on downstream turbines, reducing the effective wind speed received and thereby decreasing power generation. The magnitude of wake losses depends on turbine spacing, arrangement, wind speed, and direction. Turbine spacing should be 5-8 rotor diameters in the main wind direction and 2-4 rotor diameters in the secondary wind direction (Gunecha, 2023) (Enewables, 2023)。 In this project, a spacing of 7 rotor diameters in the main direction and 3 in the secondary direction is adopted.

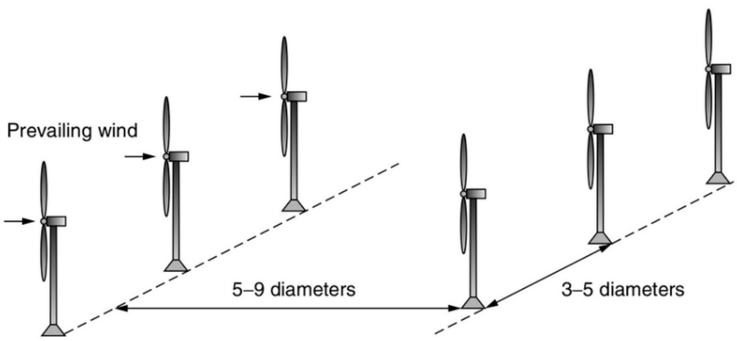


Figure 16. Wind generator layout distance

To accurately calculate the maximum number of turbines that can be installed in each municipality, a multi-constraint optimization model was developed. This model considers geographical constraints such as area, length, and width, while ensuring that the turbine layout meets spacing requirements.

Firstly, the maximum number of turbines based on area constraints is calculated. Given that each turbine occupies an area equivalent to 21 rotor diameters squared ($21D^2$), the total area of the region is divided by this unit area, rounding up to obtain the theoretical maximum number of turbines based on area. This step provides an area-based maximum turbine count without considering other constraints.

Secondly, length and width constraints are considered individually. For the length constraint, the height of the area is divided by 7 rotor diameters, rounding up to determine the maximum number of turbines that can be installed in the length direction. Similarly, for the width constraint, the width of the area is divided by 3 rotor diameters, rounding up to determine the maximum number of turbines that can be installed in the width direction. The product of the maximum turbine numbers in the length and width directions yields a matrix value representing the maximum number of turbines that can be arranged considering both length and width constraints.

Finally, the model compares the matrix constraint with the area constraint, selecting the smaller value as the final maximum number of turbines that can be installed. This step ensures that all constraint conditions are met. Once the number of wind turbines that can be installed in each municipality is determined, the annual, monthly, daily, and hourly wind power generation can be calculated accordingly.

So far, the number of wind turbines that can be installed in each municipality has been determined, and from this its annual, monthly, daily and hourly wind power production has been calculated.

5. Energy model

5.1 Feature Selection

This research integrates multi-source heterogeneous data to construct wind energy production and energy consumption models. Feature variables are divided into two main categories: wind energy production-related factors and energy consumption-related factors. We selected multiple environmental elements from terrain, land cover, meteorological, and socioeconomic data to establish a comprehensive wind power potential assessment indicator system.

Table 3. Factors Related to Wind Energy Production and Energy Consumption

Energy Model	**Class**	**Feature**	**Source**	**Data**	**Scale**	**Time**	**Year**
Wind Production	Topography	DEM	TINItaly	10 m-resolution Digital Elevation Models	Municipal	fixed	2022
		Slope	TINItaly	Slope map	Municipal	fixed	2022
		Hillshade	TINItaly	Shaded relief map	Municipal	fixed	2022
	Surface characteristics	Distance to Coastline	OpenStreet Map	Distance to coastline (km)	Municipal	fixed	2024
		Building Density	OpenStreet Map	Municipal building area and density	Municipal	fixed	2024
		Agriculture Coverage	Copernicus	Corine Land Cover - Land Use	Municipal	fixed	2018
		Forest Coverage	Copernicus	Corine Land Cover - Land Use	Municipal	fixed	2018
		Coast proportion	-	Proportion of coastal municipalities in the province	Provincial	fixed	2022
		Plain proportion	-	Proportion of plain municipalities in the province	Provincial	fixed	2022
	Meteorological factors	Temperature (Solar irradiation index)	PVGIS	Hourly solar radiation for typical days of the month for municipalities	Municipal	hourly	2019
		Air density	Global Wind Atlas	Annual average air density	Municipal	daily	unkown
Energy Consumption	Socioeconomic data	Population	ISTAT	Statistical Atlas of Municipalities Census of Population	Municipal	fixed	2021
		Industry and Services	ISTAT	Number of local units in each industry	Municipal	fixed	2011
		Agriculture Area	Copernicus	Corine Land Cover - Land Use	Municipal	fixed	2018

5.1.1 Wind Energy Production-Related Factors

To assess a region's wind energy production potential, it is necessary to analyse its wind energy endowment from multiple perspectives, including terrain, climate, and resources. As shown in Table 3, the feature variables selected to reflect wind energy production potential include terrain characteristics, surface characteristics, and meteorological factors.

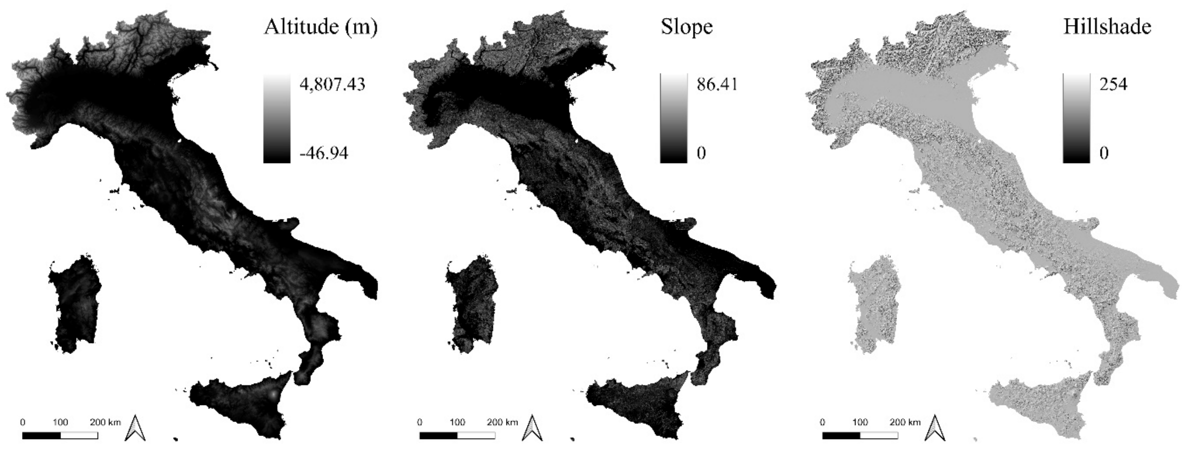


Figure 17. Altitude, slope, hill shade of Italy

Terrain Characteristics:

- Digital Elevation Model (DEM) Data: 10-meter resolution DEM data provided by TINITALY. Elevation is a crucial factor affecting wind speed - generally, higher elevations have lower surface roughness and often experience higher wind speeds.
- Slope: Terrain slope influences near-surface wind speed distribution. Gentle slopes facilitate stable wind direction and speed, while steep slopes tend to cause turbulent airflow.
- Hillshade: Terrain shadows characterize topographic relief from the perspective of slope aspect and gradient.

Surface Characteristics:

- Distance to Coastline: Sea-land thermal differences drive sea-land breezes, with coastal areas often possessing abundant wind energy resources. Using OpenStreetMap's coastline vector data, we calculated the distance from each municipal boundary to the coastline.
- Land Use/Land Cover: Different underlying surfaces have vastly different roughness, directly affecting near-surface wind speeds. This study focused on analysing CORINE land cover data from Copernicus, calculating Agriculture Coverage and Forest Coverage. Generally, open areas like farmland favor increased wind speeds, while forests tend to reduce surface wind speeds.
- Percent_coast%/Percent_plain%: To quantitatively analyse municipal terrain characteristics, we proposed a method using administrative division

information to characterize provincial terrain features. These indicators employ municipal-level geographical classification results, calculating the proportion of coastal and plain municipalities within each province. By analysing the proportion of different terrain types within each province, we can determine the surrounding topographic conditions of each municipality.

Meteorological Factors:

- Solar Irradiation Index: Solar radiation influences thermal wind circulation formation by altering ground-air temperature differences, affecting regional wind energy resource spatial distribution. We used hourly solar radiation data for typical days each month simulated by the PVGIS model.
- Air Density: Directly proportional to wind energy density, this is one of the parameters measuring wind energy resource abundance.

5.1.2 Wind Energy Consumption-Related Factors

Beyond wind energy resource endowment, wind power planning must consider local electricity demand and power consumption capacity. To analyse regional wind energy consumption potential, this research introduced socioeconomic indicators reflecting energy demand, such as population, industrial and commercial distribution, and agricultural scale, as shown in Table X. Industrial structure, in particular, directly impacts the scale and structure of regional industrial electricity demand. Through comprehensive analysis of population and the development scale of primary, secondary, and tertiary industries, we can quantitatively assess the foundation for regional wind power consumption.

- **Population:** Population size directly influences regional electricity demand and serves as an important reference for determining wind power development scale. This study used Municipal data from ISTAT 2021.
- **Industry and Services:** The secondary sector, especially electricity-intensive industrial sectors, constitutes the main body of power consumption. The thriving tertiary sector also brings substantial commercial and public service electricity demands. Analysing the number of industrial and commercial enterprises in each municipality helps estimate regional power load levels.
- **Agriculture Area:** To comprehensively assess the impact of industrial structure on regional electricity use, this study also included agricultural land

area indicators reflecting the primary sector's scale. Agricultural land area can, to some extent, reflect the agricultural sector's electricity demand.

5.2 Feature Data Preparation

The research uses data spanning multiple spatial and temporal scales, necessitating standardized integration through various spatial data processing methods. Spatially, the data ranges across multiple scales: from 10-meter resolution raster data to municipal-level statistics to provincial-scale data. Temporally, the data spans multiple years, including 2011 industrial and commercial data, 2018 land use data, and more.

Spatial interpolation methods primarily address administrative boundary changes over time. Italy's number of municipalities decreased from 8,092 in 2011 to 7,901 in 2023 due to administrative mergers and adjustments. When historical administrative regions (2011 municipalities) were merged or reorganized into new regions (2023 municipalities), we employed an area-based weight allocation method. Specifically, new municipalities received indicator values proportional to the area inherited from original municipalities. This method assumes that socioeconomic indicators (like number of business units) correlate with administrative area proportions.

Spatial aggregation methods handle scale conversions between different spatial resolutions. For high-resolution spatial data (like 10-meter resolution digital elevation models), we applied zonal statistics to calculate statistical characteristics within each 2023 municipality boundary. This approach preserved microscale terrain features while enabling integration with municipal-scale data.

Area-weighted methods process land use data. For 2018 agricultural land and forest cover data, we first identified land use type distributions within each 2023 municipal boundary, then calculated area proportions for each type.

For regional geographic characteristics, we employed a proxy indicator method based on administrative hierarchies. Each region's coastal and plain characteristics were represented through the proportion of municipality types within each province. Specifically, we calculated the ratio of municipalities classified as coastal cities and plain cities to the total number of municipalities in that province. This method effectively captures dominant regional geographic features, transforming municipal-scale classification information into quantitative indicators reflecting overall regional characteristics.

This data preprocessing methodology system achieved uniformity of multi-source heterogeneous data within the 2023 administrative framework, providing a standardized data foundation for subsequent energy modeling. Through appropriate spatial data processing methods, we maintained original spatial characteristics while ensuring comparability across different spatiotemporal scales.

5.3 Feature Engineering

Feature engineering, fundamental to machine learning, involves creating new features from existing data during the machine learning process. These new features help models better capture data patterns, thereby improving model performance.

Common feature engineering methods include: standardization and normalization of numerical features, encoding of categorical features, feature creation (such as calculating new features like means and standard deviations), and feature selection and elimination. Using the V80 wind turbine as an example, to optimize model input variables, this research evaluated relationships between these variables through correlation matrix analysis (Figure 18) to screen and reduce redundant features and decrease model complexity. The study conducted correlation analysis of wind energy production feature factors, retaining only information crucial for wind energy potential prediction. Below, we analyse and screen correlations for three types of variables: regional terrain, surface characteristics, and meteorological factors.

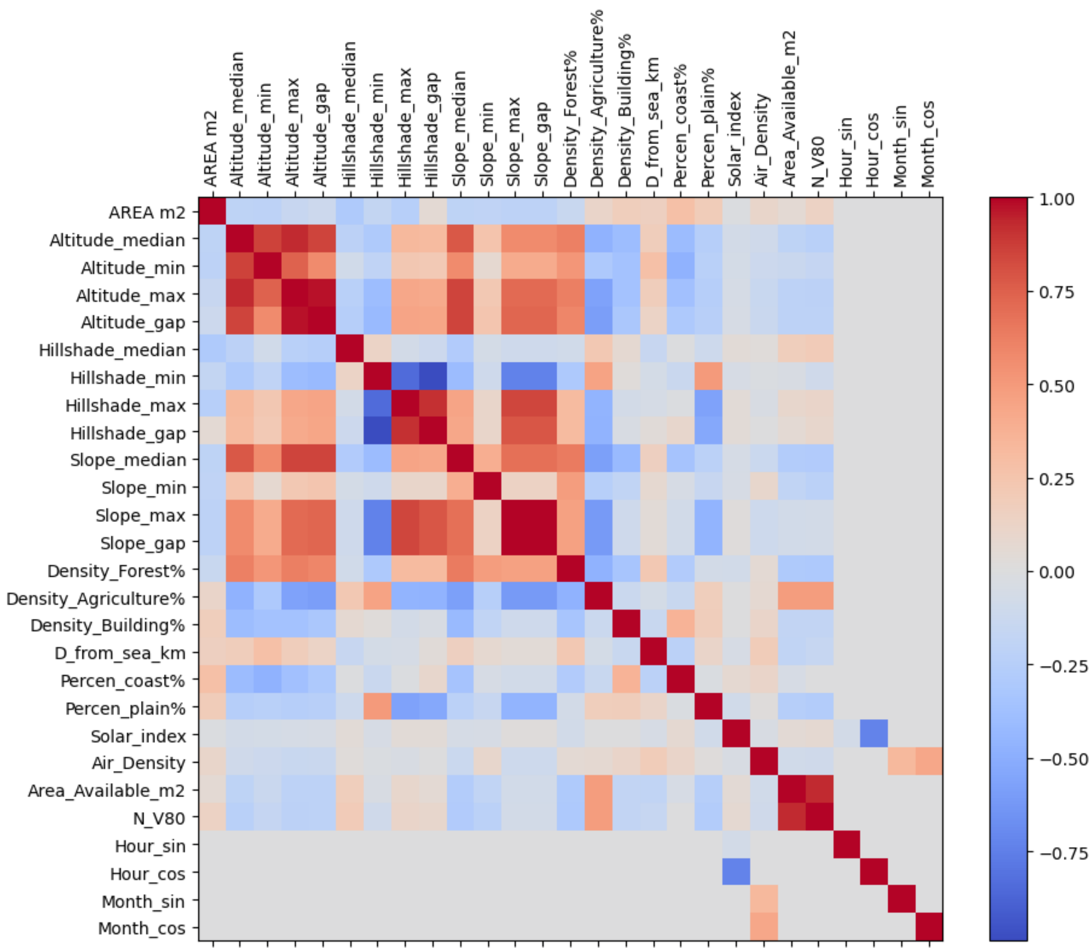


Figure 18. Correlation Matrix Between Features

Correlation Analysis of Terrain Features

This research calculated and analysed statistical characteristics of several key terrain indicators, including the minimum, maximum, median, and range (Gap) values for elevation (Altitude), slope, and terrain shadow index (Hillshade). These variables reflect both overall terrain trends and local relief characteristics.

Through correlation matrix analysis, we excluded redundant maximum and minimum variables, retaining median and range variables that separately reflect overall trends and local variations. This decision was made because the range (gap) and maximum/minimum values essentially describe the same extreme variation information. For example, Altitude_gap shows correlations approaching 1 with both Altitude_min and Altitude_max.

For elevation features, Altitude_median and Altitude_gap show high correlation, indicating that the median value sufficiently reflects regional relief amplitude. This correlation emerges from the clear continuity in elevation changes at regional scales. In complex terrain regions (such as the Alps and Apennines), Altitude_gap tends to be large and correlates with Altitude_median trends—higher median elevations typically correspond to larger elevation ranges. Conversely, in plains and lowlands (like the Po Valley), elevation variations are minimal, resulting in Altitude_gap values near zero that closely match median values.

In contrast, terrain shadow (Hillshade) and slope statistical values show lower correlations. Hillshade_median represents the median value of surface illumination conditions, influenced not only by slope but also by aspect, solar elevation angle, and azimuth. Hillshade_gap reflects extreme differences in regional illumination conditions, often caused by steep valleys or gentle plateaus. This complexity prevents linear correlation between Hillshade_median and Hillshade_gap.

Similarly, Slope_median and Slope_gap describe overall slope distribution trends and local relief intensity, respectively. While both are influenced by terrain changes, Italy's complex topography (including mountains, hills, and plains) means that slope medians and ranges don't necessarily correlate spatially. For instance, hilly regions might have low median slopes but high slope ranges due to contrasts between gullies and hillsides, while mountainous regions with steep but uniform slopes might show high medians but lower ranges.

Correlation Analysis of Surface Features

Surface features encompass land use characteristics (forest coverage, agricultural land proportion, and building density), regional spatial characteristics (distance to sea, coastal proportion, and plain proportion), and spatial constraints related to wind energy development (available area and number of turbines). Correlation matrix analysis reveals low correlations between most variables, indicating their ability to independently reflect different important dimensions in wind energy assessment. However, high correlation exists between available area (Area_Available_m2) and turbine numbers (N_V80/N_V52).

Land use features show strong independence. For example, forest coverage (Density_Forest%), agricultural land proportion (Density_Agriculture%), and building density (Density_Building%) show near-zero correlations, indicating they describe different dimensions of municipal land use characteristics. Different land use features have distinct surface roughness coefficients that affect wind speed distribution and turbulence characteristics. The surface roughness coefficient is a dimensionless parameter quantifying surface roughness effects on wind speed reduction. It's a function of roughness length (z_0), which represents the theoretical height where wind speed approaches zero under ideal conditions (measured in meters). The following table presents surface classification standards from European regulations (Eurocode, 2024):

Table 4. Terrain categories and terrain parameters

Terrain category		z_0 m	z_{min} m
0	Sea or coastal area exposed to the open sea	0,003	1
I	Lakes or flat and horizontal area with negligible vegetation and without obstacles	0,01	1
II	Area with low vegetation such as grass and isolated obstacles (trees, buildings) with separations of at least 20 obstacle heights	0,05	2
III	Area with regular cover of vegetation or buildings or with isolated obstacles with separations of maximum 20 obstacle heights (such as villages, suburban terrain, permanent forest)	0,3	5
IV	Area in which at least 15 % of the surface is covered with buildings and their average height exceeds 15 m	1,0	10
NOTE: The terrain categories are illustrated in A.1.			

According to surface classification standards, forests are classified as Type III terrain with a z_0 of approximately 0.3 meters. This significant roughness length indicates that forests substantially reduce wind speeds while increasing turbulence intensity (Porté-Agel, 2020). Areas with higher forest coverage experience stronger surface friction, resulting in more complex wind speed distributions.

Agricultural land typically falls under Type II terrain, with a z_0 of approximately 0.05 meters. Generally concentrated in plain regions, agricultural lands' lower roughness coefficient helps maintain higher wind speeds while producing weaker turbulent effects. Building density, a core characteristic of urbanized regions, is classified as Type IV terrain with a z_0 reaching 1.0 meters. Areas with high building density significantly enhance surface friction, producing the strongest wind speed reduction effects.

Regional spatial characteristics also demonstrate low correlations. Distance to sea (D_from_sea_km), coastal proportion (Percen_coast%), and plain proportion (Percen_plain%) each represent distinct aspects: regional geographic location, coastal characteristics, and terrain type distribution. Coastal regions, directly influenced by sea breezes, typically offer higher wind speed potential. Plain regions, with their gentle topography, feature more uniform wind speed distributions and help reduce turbulent effects.

Spatial constraint variables related to wind energy development include available area (Area_Available_m2) and number of turbines (N_V80/N_V52). The correlation matrix reveals high correlation between these variables, as mentioned in the process-driven approach - the number of turbines is a direct function of available area, determined by both available space and turbine spacing design. Retaining available area comprehensively reflects how regional terrain, land use, and environmental restrictions constrain wind energy development, making it more influential in wind energy assessment than turbine numbers.

Correlation Analysis of Meteorological Factors

This study selected Air Density, temporal features (Hour_sin, Hour_cos, Month_sin, Month_cos), and Solar Index as meteorological variables. Correlation matrix analysis reveals low correlations between these meteorological variables.

Air density directly determines the mass of air per unit volume, serving as a crucial parameter in calculating wind energy density. Higher air density typically corresponds to greater wind energy potential. Temporal features capture daily variations (hours) and seasonal patterns (months) through sine and cosine forms. Solar Index directly reflects local solar energy reception. While Solar Index doesn't directly represent temperature, it closely relates to surface heating processes, indirectly influencing changes in local thermal conditions and wind speed characteristics.

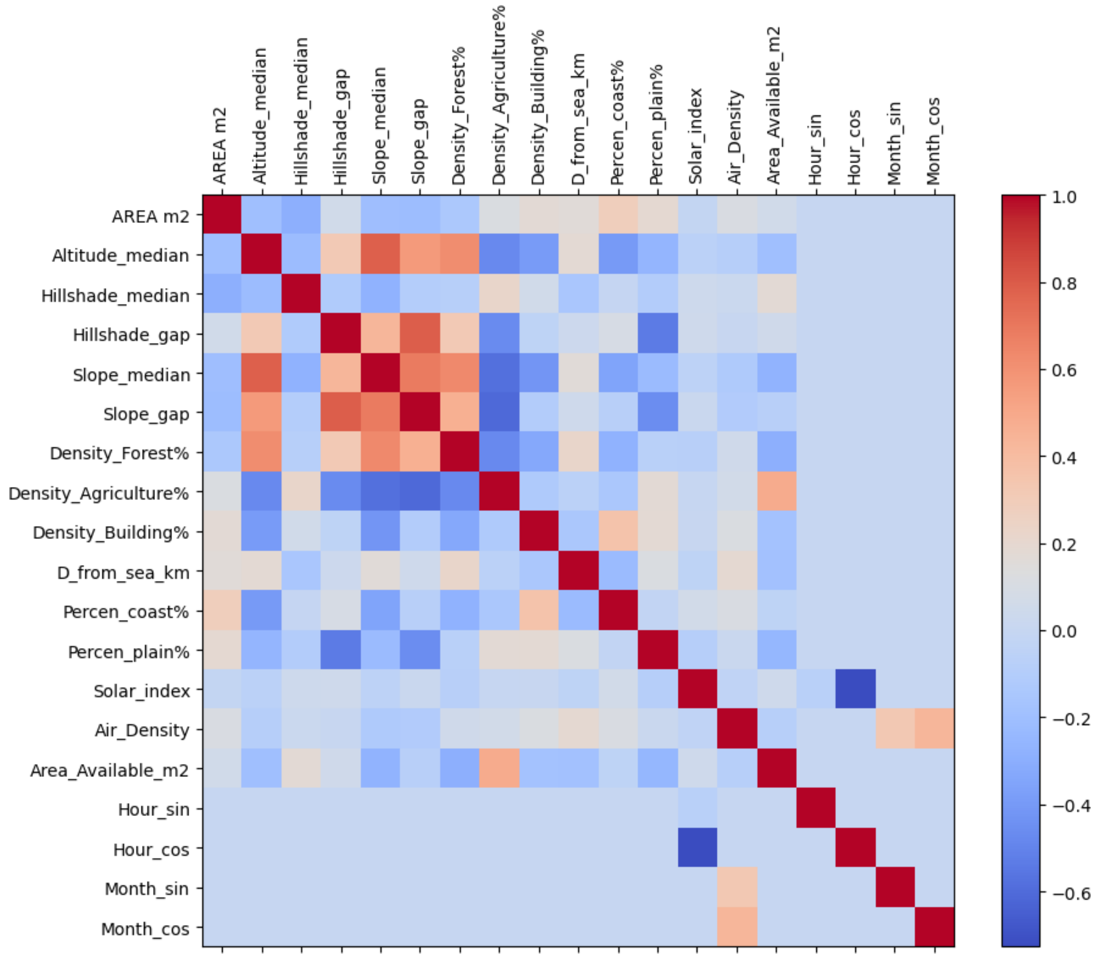


Figure 19. Correlation Matrix Between Features After Selection

Through systematic correlation matrix analysis, our study retained the following indicators that significantly contribute to wind energy assessment: Altitude_median, Altitude_gap, Slope_median, Slope_gap, Hillshade_median, Hillshade_gap, Density_Forest%, Density_Agriculture%, Density_Building%, D_from_sea_km, Percen_coast%, Percen_plain%, Area_Available_m2, Air_Density, Hour_sin, Hour_cos, Month_sin, Month_cos, and Solar Index. These filtered variables cover terrain, surface, and meteorological dimensions, optimizing model input features while preserving core information essential for wind energy assessment.

6. Results and Discussion

6.1 Energy Consumption:

Italy's energy consumption patterns demonstrate significant regional differences and industrial characteristics. In 2022, Italy's total national electricity consumption reached approximately 290 TWh, reflecting its energy demands as an industrialized nation. Looking at the industrial electricity consumption structure reveals several important patterns that help us understand Italy's economic landscape.

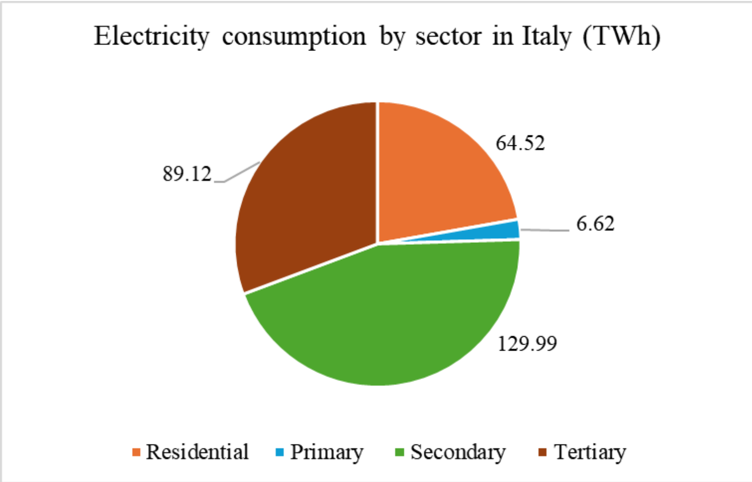


Figure 20. Proportion of energy consumption by sector in Italy

The secondary sector (industry) emerges as the largest electricity consumer, accounting for 45% of total consumption. This prominence underscores manufacturing's crucial role in supporting Italy's economy. The tertiary sector (commercial services) follows at 29%, while residential consumption represents 24%. The primary sector (agriculture) accounts for just 2% of total electricity consumption, indicating its relatively low energy intensity despite its cultural and economic importance.

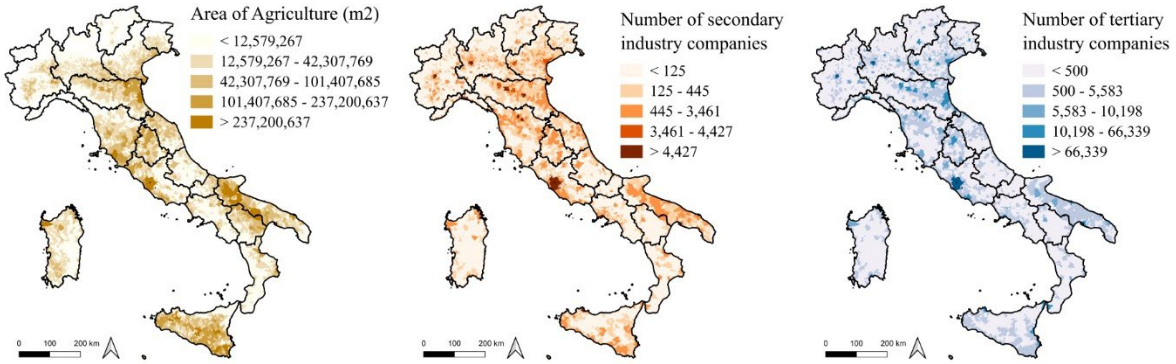


Figure 21. Agricultural area and secondary and tertiary industry distribution

According to industrial activity statistics (ISTAT 2011), analysis of 7,901 municipalities reveals Italy's highly diversified economic structure. Examining the first set of maps in Figure 21 shows distinct regional patterns in industrial distribution: Agricultural land distribution indicates that central and southern Italian plains possess the largest agricultural areas, where primary sector activities are most abundant. The northern regions and Lazio region show the highest density of secondary sector enterprises, forming distinct industrial clusters. Tertiary sector enterprises follow a similar distribution pattern, primarily concentrating in northern urban agglomerations.

The data reveals fascinating patterns of industrial participation across municipalities: 99.6% engage in secondary sector activities, and all municipalities have tertiary sector presence. Notably, 0.4% of municipalities rely entirely on the tertiary sector. However, industrial activity concentration varies significantly by terrain conditions:

- Coastal regions, though comprising only 648 municipalities, demonstrate the strongest industrial agglomeration effects. Each coastal municipality averages 319.5 secondary sector enterprises and 1,518.4 tertiary sector enterprises. This high-density distribution benefits from maritime logistics advantages and market opportunities provided by sea access.
- Plain regions, covering 44.3% of municipalities, serve as vital economic activity carriers. These areas show moderate enterprise density with an average of 192.5 secondary sector and 684.5 tertiary sector enterprises per municipality, reflecting how favorable plain topography supports industrial development.
- Mountainous areas, constrained by terrain, show relatively weak industrial activity with the lowest enterprise density, highlighting how geographical

limitations can impact economic development.

Regarding enterprise scale, Italy's industrial structure exhibits characteristics typical of small and medium-sized enterprise dominance. Secondary sector enterprises average 5.1 employees, while tertiary sector enterprises average 3.4 employees, indicating a granular economic fabric dominated by smaller business units.

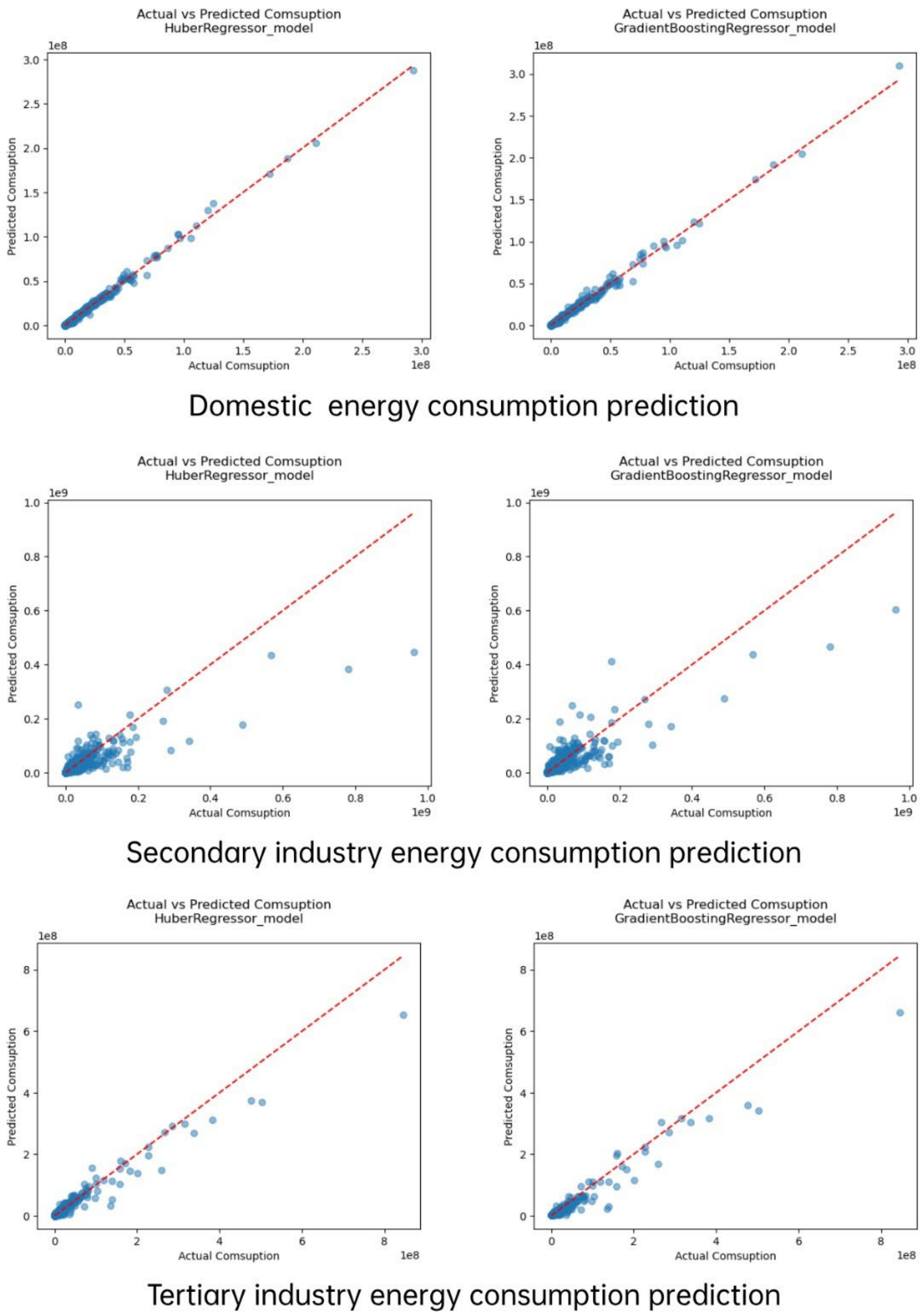


Figure 22. Energy consumption model performance

Based on socioeconomic characteristics, through running electricity demand prediction models for different user types, the research findings revealed the following patterns, as shown in Figure 22. The prediction models for residential and commercial electricity consumption demonstrated exceptional performance, with R^2 values exceeding 0.99, which can be attributed to the high correlation between the selected features and electricity consumption behaviors.

The residential electricity prediction primarily relies on population size, number of households, and building density. These indicators effectively capture the fundamental drivers of residential power usage. Meanwhile, commercial electricity prediction is based on commercial building area, number of business establishments, and employee count. These metrics effectively reflect the scale of commercial activities, as commercial electricity consumption patterns remain relatively stable, mainly centered around lighting, air conditioning, and electrical appliances.

In contrast, the industrial electricity prediction model showed less satisfactory results, with an R^2 value of only around 0.75. This model solely depends on industrial land area (A_B2_m2), number of enterprises (2_NUM_UNIT), and employee count (2_ADDETTI). This lower performance suggests that industrial electricity consumption patterns are more complex and would benefit from incorporating additional dimensional features such as industry type, production scale, and equipment energy consumption.

Regarding the primary sector, agricultural electricity prediction faces limitations in data collection. Currently, only two indicators are available: agricultural land area and agricultural building coverage. These metrics alone are insufficient to adequately reflect crucial factors in agricultural production, such as irrigation demands and mechanization levels. Due to this lack of essential feature data, it is presently not feasible to construct a reliable prediction model for agricultural electricity consumption.

Table 5. Energy consumption model performance

Sector	**Model**	**MAPE (%)**	**R^2**
Domestic	HuberRegressor	8.17	0.9952
	GradientBoostingRegressor	12.59	0.9928
Secondary	HuberRegressor	-	0.6982
	GradientBoostingRegressor	-	0.7586

Tertiary	HuberRegressor	28.82	0.938
	GradientBoostingRegressor	450.39	0.9293

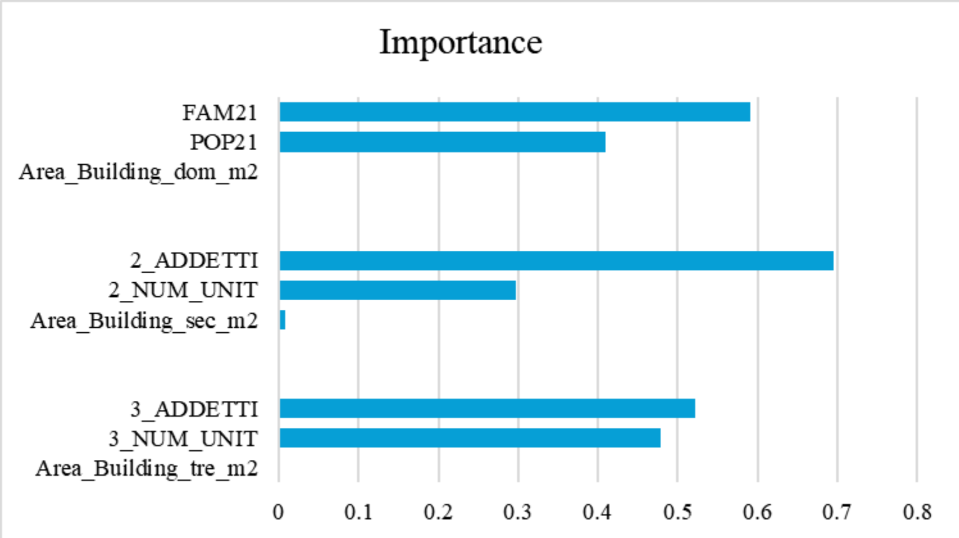


Figure 23. Feature importance of energy consumption

The importance distribution reveals that employment figures across sectors and population metrics consistently demonstrate the highest influence, reflecting several fundamental economic phenomena:

A comparative analysis of household units (FAM21) and population count (POP21) indicates that households emerge as more significant economic decision-making units. This pattern suggests that regional economic activity is predominantly driven by household-level consumption decisions rather than individual consumer behavior.

Within the secondary sector, the workforce indicator (2_ADDETTI) exhibits paramount importance (approximately 0.7). This high significance of industrial employment suggests that the region likely occupies a crucial position in the industrial value chain, potentially indicating a manufacturing-intensive zone. Conversely, the relatively lower importance of industrial establishment count (2_NUM_UNIT) suggests larger individual industrial enterprise scales, pointing toward high industrial concentration and the possible presence of major manufacturing entities.

The tertiary sector indicators display more balanced influence patterns: both service sector employment (3_ADDETTI) and establishment counts (3_NUM_UNIT) show similarly high importance levels. This balanced pattern reflects the characteristic diversification and dispersion within the service sector, where both the number of establishments and employment levels jointly influence sector development. This aligns with typical service sector characteristics: compared to industry, service establishments tend to be more numerous but smaller in scale.

An intriguing observation emerges from the consistently low importance of built-up area across all three sectors. This phenomenon illuminates a significant trend in modern economics: the weakening relationship between physical space requirements and economic output. This is particularly noteworthy in the secondary sector, potentially indicating the region's industrial evolution toward higher efficiency and technological sophistication, reducing dependence on physical space expansion.

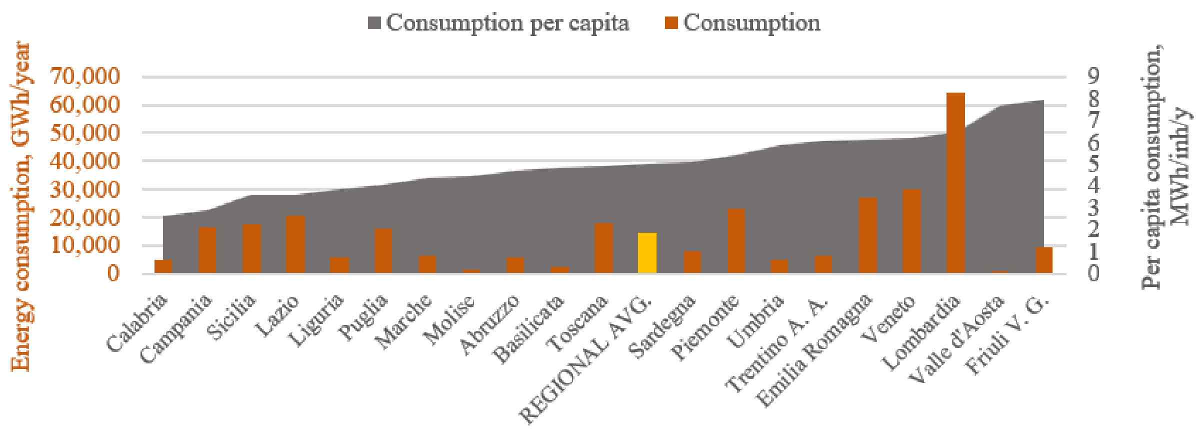


Figure 24. Electricity consumption by region and national average (2022)

From a regional distribution perspective, energy consumption exhibits distinct north-south disparities. The northern industrial zone, represented by the Lombardy region, consumes 64.3 TWh of electricity, accounting for 22.1% of the national total electricity consumption, with the secondary industry comprising a substantial 53.1% of this usage. These figures validate the northern region's status as Italy's industrial heartland. In contrast, economically less developed regions such as Valle d'Aosta and Molise consume merely 0.9 TWh and 1.3 TWh respectively, highlighting the imbalance in regional development.

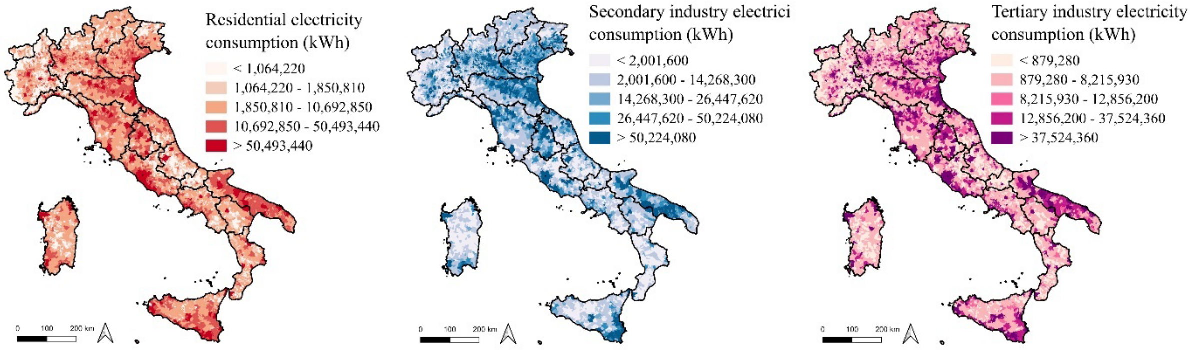


Figure 25. Distribution of Electricity Consumption Across Different Industries

Figure 25 shows the distinctive spatial distribution patterns of energy consumption across Italy, demonstrating unique geographical differentiation characteristics. Residential electricity consumption forms notable high-value centers in densely populated urban areas, particularly in metropolitan regions like Milan and Rome, reflecting the strong correlation between urbanization levels and residential power usage.

Industrial electricity consumption exhibits the most pronounced spatial concentration pattern. The northern industrial zones consistently record consumption levels overtaking 50 GWh, a highly concentrated distribution that closely aligns with the spatial clustering pattern of manufacturing enterprises, emphasizing the energy-intensive nature of industrial production.

The service sector's electricity consumption pattern displays spatial characteristics similar to the industrial sector, forming distinct high-value zones in major economic centers such as Rome, Milan, Turin, and Bari. It's particularly noteworthy that Puglia emerges as a heterogeneous region within the southern territory - due to its unique industrial structure and population distribution, it demonstrates relatively high levels across all electricity consumption indicators.

This spatial analysis effectively illustrates how economic activity, population density, and industrial development interact to shape regional energy consumption patterns across Italy. The clear north-south divide and the emergence of urban centers as major consumption hubs provide valuable insights into the relationship between economic development and energy usage patterns.

6.2 Wind Production

Figure 26 illustrates the daily wind power generation patterns across 24 hours during winter (January) and summer (July), revealing distinct seasonal variations in wind energy production.

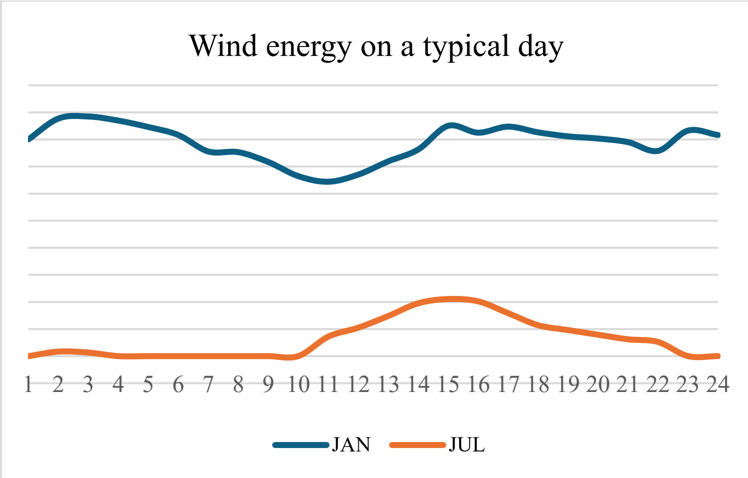


Figure 26. Wind energy on a typical day in winter andsummer

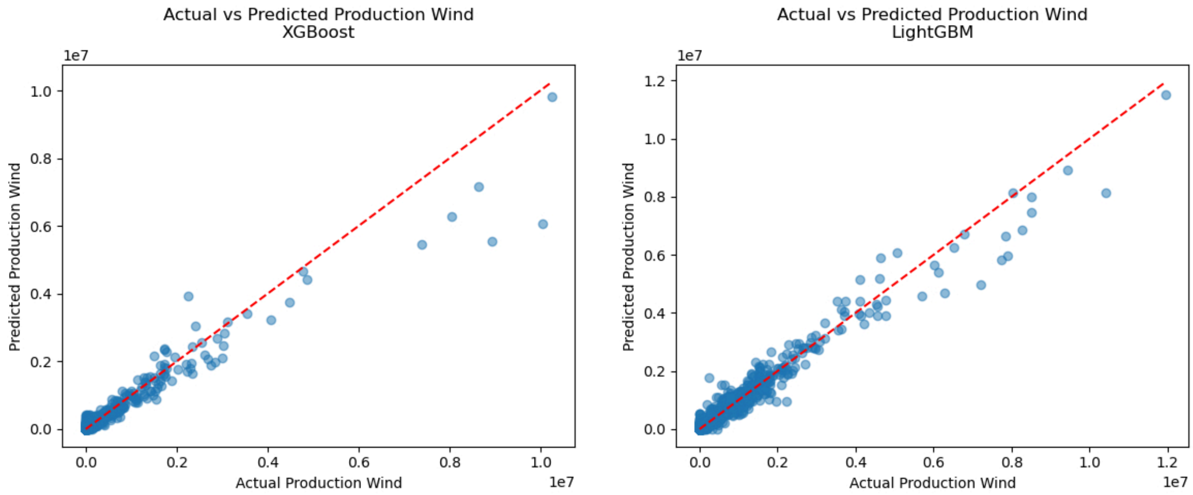
The winter pattern (blue line) demonstrates higher overall power generation but with notable fluctuations. During the early morning hours, wind power maintains elevated levels, providing substantial energy during the pre-dawn and early morning periods. However, a significant decline occurs around midday, creating a notable trough in power generation. This pattern suggests that winter winds are generally stronger but less stable throughout the day.

In contrast, the summer pattern (orange line) exhibits lower overall power generation but with more gradual variations throughout the day. The summer wind power reaches its peak during the afternoon hours, creating a gentle curve rather than the sharp fluctuations seen in winter. This smoother pattern indicates more consistent, though less powerful, wind conditions during summer days.

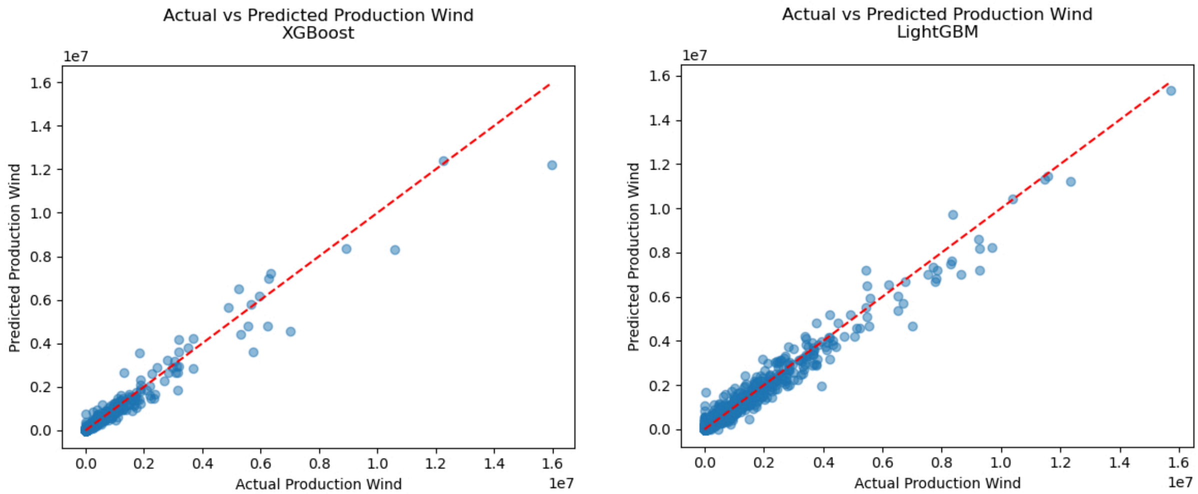
When analysing these patterns against typical electricity demand curves, several significant relationships emerge:

During winter months, wind power generation aligns favorably with morning (7-9 AM) and evening (6-10 PM) peak demand periods, helping meet increased energy needs for heating and lighting. However, the substantial midday decrease in wind power generation creates a potential supply gap that may require supplementation from other energy sources to maintain consistent power delivery.

The summer pattern shows an advantageous correlation between peak wind power generation (2-4 PM) and maximum air conditioning demand during the hottest part of the day. This natural synchronization helps meet increased cooling needs during peak temperature hours. However, the relatively lower wind power generation during morning and evening hours may necessitate additional energy sources to maintain stable power supply during these periods.



Scenario Vestas V52



Scenario Vestas V80

Figure 27. Wind production model performance

Table 6. Wind production model performance

Wind Prediction	**Model**	**R^2**
Vestas V80	XGBoost	0.9508
	LightGBM	0.9679
Vestas V52	XGBoost	0.9277
	LightGBM	0.9613

The wind production model results demonstrate high accuracy, with R^2 values higher than 0.95 for both the Vestas V52 and V80 turbine scenarios using LightGBM, confirming the model's effectiveness in predicting wind energy production. These outcomes highlight the reliability of machine learning techniques, such as LightGBM and XGBoost, in modeling complex wind energy generation patterns.

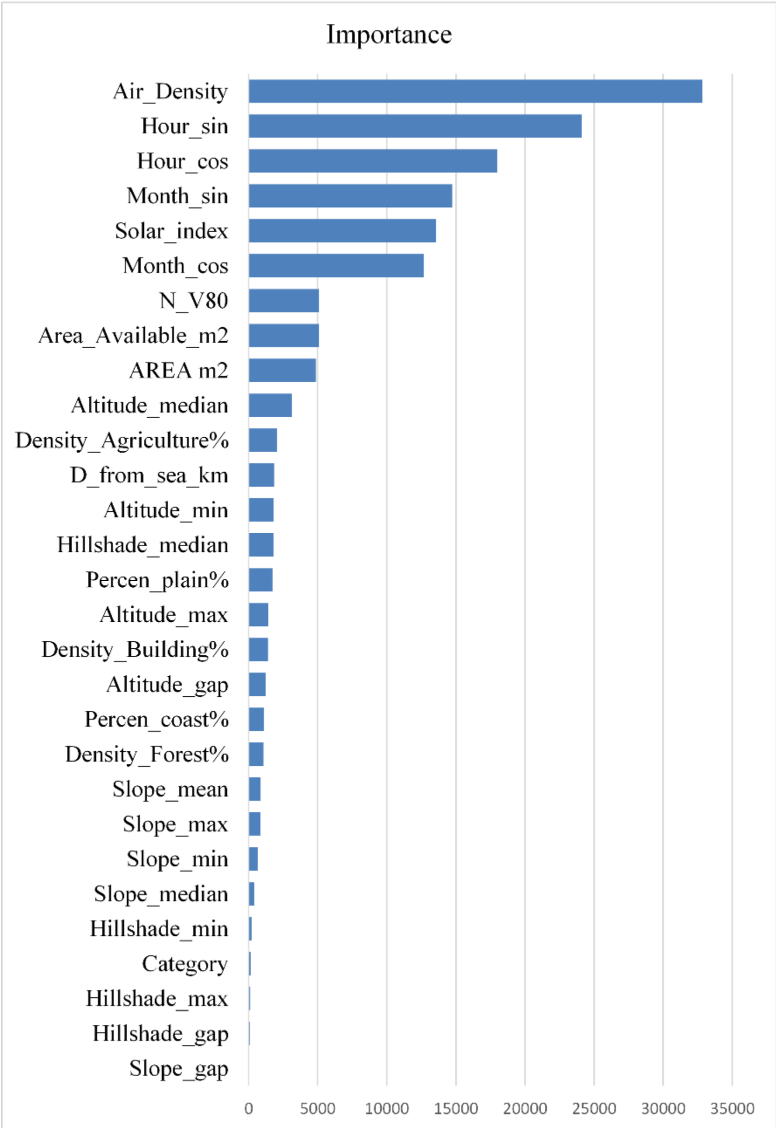


Figure 28.Feature importance of wind production

Figure 28 illustrates the relative importance ranking of various features in predicting wind energy production potential. The analysis reveals that

Air_Density emerges as the most significant predictor. This prominence aligns with physical principles, as air density directly influences wind turbine generation efficiency - higher density air carries more kinetic energy at the same wind speed, resulting in greater power generation. Air density varies with temperature, humidity, and altitude, making its spatiotemporal distribution crucial for accurate regional wind energy potential forecasts.

The next most significant features are Hour_sin and Hour_cos, which capture diurnal wind speed patterns. Wind speeds typically follow cyclical daily patterns, and these trigonometric transformations effectively model these non-linear relationships. These daily variations directly impact wind farm power output.

The temporal features Hour_sin, Hour_cos, Month_sin, and Month_cos collectively demonstrate high importance, reflecting both daily and annual wind speed periodicities. These cyclical patterns likely correlate with variations in solar radiation intensity and monsoon transitions. Notably, the model's feature ranking suggests that daily wind cycles carry more predictive weight than annual cycles, possibly because daily wind speed fluctuations tend to be more pronounced and have more immediate effects on short-term wind energy forecasts.

Solar radiation intensity also emerges as a significant feature, as its variations influence atmospheric thermal structure and consequently affect wind field distribution patterns.

Geographic features also play substantial roles. Variables such as Available_Area_m2, AREA_m2, and Altitude_median demonstrate meaningful impact on wind energy potential. Generally, larger available land area correlates with higher potential turbine capacity installation and overall wind resource abundance. Altitude influences local climate conditions and topographical effects, which in turn affect vertical wind speed distribution patterns.

Topographical factors such as distance from sea (D_from_sea) and terrain characteristics (Hillshade_median, Slope_median) show moderate importance. Near-surface wind speed gradients vary significantly, with slope affecting vertical wind distributions. Proximity to sea influences sea-land breeze circulation intensity.

Land use variables, including agricultural coverage (Density_Agriculture) and building density (Density_Building), rank lower in importance but remain relevant. These factors affect surface roughness, which influences near-surface wind speed distributions, justifying their inclusion in the model features.

In conclusion, this feature importance analysis provides valuable insights into wind energy potential determinants. The ranking aligns with physical mechanisms, where wind speed patterns - including their spatiotemporal distribution - emerge as primary factors. These patterns are primarily governed by meteorological conditions and topographical characteristics, while land cover types demonstrate relatively minor influence.

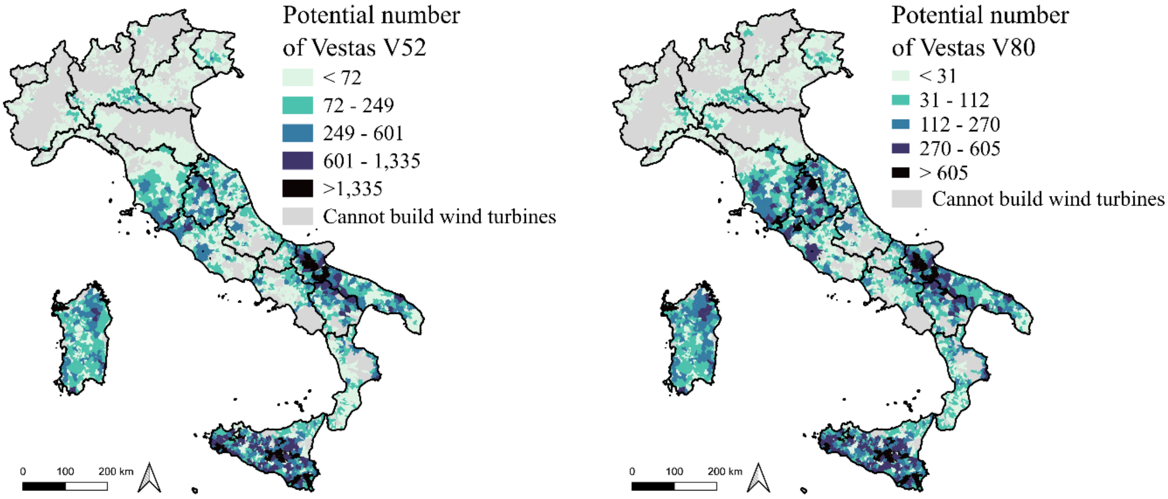


Figure 29. Potential number of Vestas V52 and V80

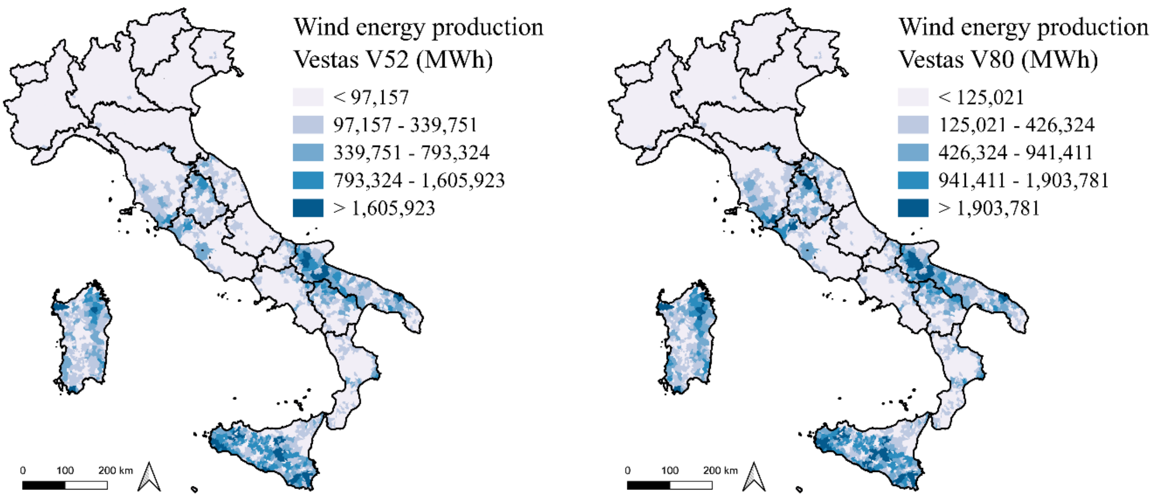


Figure 30. Future wind energy production of Vestas V52 and V80

Based on simulation analyses of different wind turbine models including Vestas V52, V80, and V90, research reveals significant geographical heterogeneity in Italy's wind energy resources. Data indicates that southern regions (Sardinia, Sicily) and certain central areas possess the highest wind power generation potential, with annual electricity generation reaching 366-532 TWh. The Puglia region along the Adriatic coast demonstrates exceptional generation potential, providing guidance for onshore wind farm development while highlighting offshore wind development prospects.

In contrast, northern regions show relatively lower wind potential due to Alpine terrain influence. This uneven resource distribution presents challenges for grid infrastructure, particularly regarding north-south power transmission. Research indicates only approximately 5% of national territory is suitable for wind facility deployment, predominantly concentrated in southern regions and islands.

In future scenario, wind energy production patterns demonstrate distinct regional variations between northern and southern regions. Compared to potential overcapacity issues associated with photovoltaic power generation, wind energy technology demonstrates superior alignment with actual user electricity demand patterns. Research data reveals significant regional variations in wind energy potential distribution. Specifically, southern regions exhibit high energy self-sufficiency rates but relatively low self-consumption rates due to their abundant wind resources, primarily attributed to excess local power generation. In contrast, northern regions demonstrate high self-consumption rates but lower energy self-sufficiency rates. Comprehensive data analysis shows that the Vestas V52 turbine model achieves an energy self-consumption rate of 90%, while the Vestas V80 reaches 80%. Regarding energy self-sufficiency rates, the Vestas V52 achieves 34%, while the Vestas V80 performs slightly better at 39%.

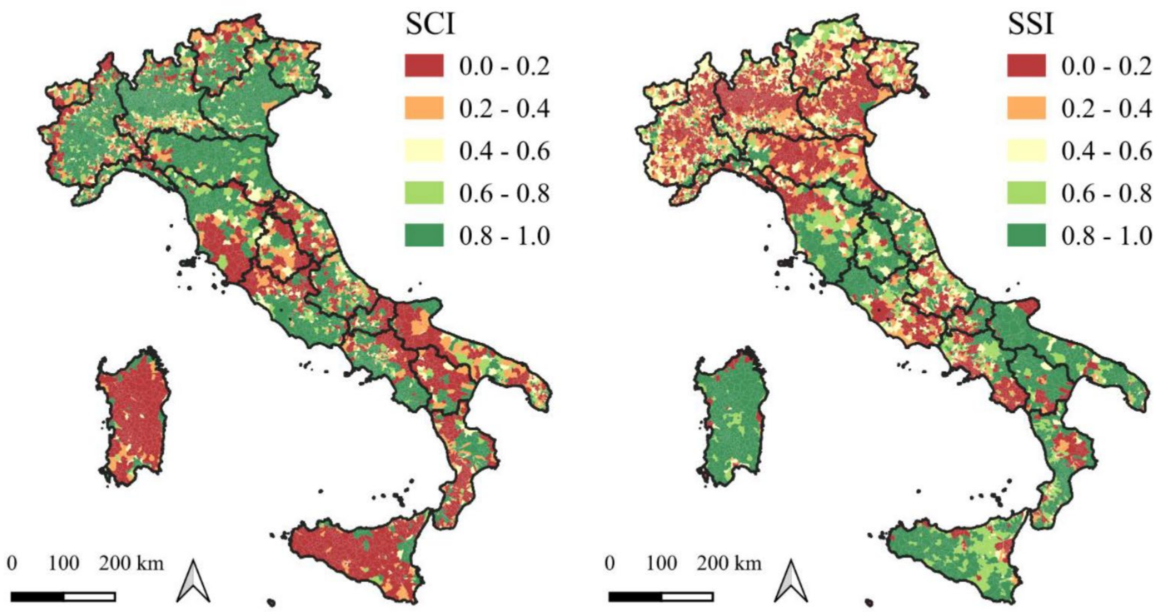


Figure 31. (a) Self-Consumption (SCI) and (b) Self-Sufficiency (SSI) indexes

for Scenario Vestas V52 (850 kW)

To fully utilize this resource endowment, the recommended strategies include strengthening north-south grid infrastructure, implementing targeted support policies in high-potential areas, and advancing diversified renewable energy development strategies in northern regions. These findings carry significant policy implications for Italy's energy structure transition.

7. Conclusions

This research provides an in-depth analysis of the spatial and temporal characteristics of wind energy production in Italy, revealing significant seasonal and regional disparities. Wind energy output is higher in winter but exhibits greater volatility. In contrast, summer wind energy production is generally lower but shows more gradual daily variations. Southern regions such as Sardinia and Sicily, as well as certain central areas, have the highest wind energy potential, while northern regions have relatively lower potential due to the influence of Alpine terrain.

Through simulation analyses of wind turbine models like the Vestas V52 and V80, it found significant differences in energy self-sufficiency and self-consumption rates across different regions. Southern regions exhibit high energy self-sufficiency but low self-consumption rates due to excess local wind resources. Northern regions show high self-consumption but lower self-sufficiency rates.

To fully harness wind energy resources, strengthening north-south grid infrastructure is recommended, particularly by optimizing northward transmission capacity. Additionally, targeted support policies should be implemented in high-potential areas, and diversified renewable energy development strategies should be advanced in northern regions. These measures have significant policy implications for facilitating Italy's energy transition.\.

References

Ahmed, S. &. (2024). Renewable Energy Communities: Towards a new sustainable model of energy production and sharing. Energy Strategy Reviews,, 101522.

ARERA. (2024). Retrieved from ARERA: https://www.arera.it/dati-e-statistiche

Bardazzi, R. B. (Energy Research & Social Science). To eat and not to heat? Energy poverty and income inequality in Italian regions. . 2021, 101946.

Brian Tri, M. j. (2023, Aug 14). Retrieved from Wind Systems Magazine: https://www.windsystemsmag.com/laying-the-foundation-for-wind-turbines-now-and-in-the-future/

Chen, T. &. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794.

Compagnucci, L. &. (2023). Il Piano Nazionale Integrato per l'Energia ed il Clima (PNIEC): Analisi e scenari per l'industria nazionale. In Policy Brief CiMET, 1-9.

Council, E. (2024, 4 12). Fit for 55. Retrieved from European Council- Council of the European Union: https://www.consilium.europa.eu/en/policies/fit-for-55/

Di Somma, M. B. (2018). Ottimizzazione economico–exergetica delle prestazioni di reti energetiche distribuite su scala distrettuale: applicazione ad un caso studio reale.

Enewables, V. (2023, Feb 1). Retrieved from Vector Enewables: https://www.vectorenewables.com/en/blog/do-you-know-what-the-wake-effect-is-in-a-wind-farm#:~:text=The%20%27wake%20effect%27%20is%20the,and%20even%20neighboring%20wind%20farms.

Eurocode. (2024, October 8). Roughness factor cr(z). Retrieved from Eurocode: https://eurocodes-tools.com/en/roughness-factor-crz/

González-Eguino, M. (2015). Energy poverty: An overview. Renewable and sustainable energy reviews, 377-385.

Government, S. (2014, May 28). Retrieved from Onshore wind turbines: planning advice: https://www.gov.scot/publications/onshore-wind-turbines-planning-advice/

Greco, L. (2022). Report 2022 - Italy. Italy: IEA Wind TCP.

GSE. (2022, 4 4). GRUPPI DI AUTOCONSUMATORI DI ENERGIA RINNOVABILE E COMUNITÀ DI ENERGIA RINNOVABILE. Retrieved from GSE: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.gse.it/documenti_site/Documenti%20GSE/Servizi%20per%20te/AUTOCONSUMO/Gruppi%20di%20autoconsumatori%20e%20comunita%20di%20energia%20rinnovabile/Regole%20e%20procedure/Autoconsumatori.pdf

Gunecha, S. (2023, Aug 11). Exploring Wake Losses in Offshore Wind. Retrieved from Empire Engineering: https://www.empireengineering.co.uk/exploring-wake-losses-in-offshore-wind/

Gwec, G. W. (2021). Global wind report 2021. Brussels, Belgium.: Global Wind Energy Council.

Joyce Lee, F. Z. (2021). Global Wind Report 2021. Brussels, Belgium : Global Wind Energy Council.

Korstanje, J. (2021). Gradient Boosting with XGBoost and LightGBM. Apress, Berkeley, CA.: Advanced Forecasting with Python.

McKenna, R. P. (2022). High-resolution large-scale onshore wind energy assessments: A review of potential definitions, methodologies and future research needs. Renewable Energy, 182, 659-684.

Mutani, G. S. (2021). Indicators and representation tools to measure the technical-economic feasibility of a renewable energy community. The case study of Villar Pellice (Italy). . nternational Journal of Sustainable Development and Planning, 16(1), 1-11.

Pirazzi, L. &. (2005). Wind power in Italy from the late 1970s to the present. In Proceedings of Solar World Congress, 34th ASES Annual Conference, and 30th National Passive Solar Conference.

Porté-Agel, F. B. (2020). Wind-turbine and wind-farm flows: a review. Boundary-layer meteorology, 1-59.

Ren, J. Y. (2022). A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism. . Energy Reports, 437-443.

REN21. (2021). Renewables 2021 global status report. REN21.

Ripple. (2022, Aug 18). Ripple. Retrieved from What makes up a wind turbine?: https://blog.rippleenergy.com/environment/what-makes-up-a-wind-turbine/

SageMaker, A. (2024). How the SageMaker XGBoost algorithm works. Retrieved from Amazon SageMaker.

scikit-learn. (2024). GradientBoostingRegressor. Retrieved from scikit-learn: https://scikit-learn.org/dev/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html

scikit-learn. (2024). HuberRegressor. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.HuberRegressor.html

scikit-learn. (2024). HuberRegressor. Retrieved from scikit-learn: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.HuberRegressor.html

Sohoni, V. G. (2016). A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems. ,. Journal of Energy, (1） 8519785.

Sovacool, B. K. (2012). The political economy of energy poverty: A review of key challenges. Energy for Sustainable Development, 272-282.

Tatti, A. F. (2023). The emerging trends of renewable energy communities' development in Italy. Sustainability, 15(8), 6792.

Wind – Renewables 2020 – Analysis. (2021). Retrieved from IEA-International Energy Agency.