CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2024.319

Assessing wind energy exploitation potential in several regions of Viet Nam using Kernel density estimation model

Tin Trung Chau[1], Tuan Ngoc Nguyen[2], and Ton Duc Do[1*]

[1]Department of Robotics and Mechatronics, School of Engineering and Digital Sciences, Nazarbayev University, Astana, Kazakhstan

[2]Faculty of Information Technology, University of Economics Ho Chi Minh City - Vinh Long Campus, Viet Nam

**Corresponding author (doduc.ton@nu.edu.kz)*

Article info.

Received 1 Jul 2024
Revised 4 Sep 2024
Accepted 6 Oct 2024

Keywords

Bandwidth selection, kernel density estimation, non-parametric distribution, wind energy, wind speed distribution

ABSTRACT

This article analyzes and assesses the potential for wind energy exploitation in six regions of Viet Nam. The wind speed data are used to construct wind speed probability distributions (WSPDs) based on kernel density estimation (KDE). The KDE distribution, with six bandwidth selection methods, is implemented to generate probability density functions (PDFs) for each region's data to describe wind speed characteristics. The statistical tests Cramér-Von Mises (CvM), Anderson-Darling (A-D), and Kolmogorov-Smirnov (K-S) are applied to evaluate the PDFs' goodness-of-fit performance. The analysis results present the KDE distribution using the least-squares cross-validation (LSCV), and the Scott bandwidth selection method has outstanding fitting performance. Based on these PDF distributions, the wind turbine (WT) power curve is used to estimate and predict the amount of electricity that can be produced. This study also proposes a reliable method for wind power output planning based on wind speed that can be universally applied.

1. INTRODUCTION

Energy is an indispensable need for human life. Currently, most energy sources are produced from fossil fuels. Using fossil fuels causes many negative environmental effects and is the leading cause of the greenhouse effect (Farghali et al., 2023). Today's urgent need is to develop a type of clean energy that can replace energy from fossil fuels to protect the environment. Many studies are being conducted on clean energy, in which wind energy is highly appreciated and promising in becoming a suitable alternative energy source (Khaloie et al., 2020). Wind energy is a clean, sustainable, and widely developed renewable energy source. Wind power is receiving strong government attention and development. In 2023, 117 GW of new wind power capacity will be integrated into the grid, bringing total wind output to more than 1 TW globally, representing a growth of about 10% compared to 2022 (IRENA, 2024). In Viet Nam, wind power is also very interested in developing. According to statistics from GWEC, Viet Nam has installed about 4.6 GW of wind power capacity. The government aims to reach 11.8 GW by 2025 and 36 GW by 2030 (Global Wind Report 2024 - Global Wind Energy Council, n.d.).

In the field of wind energy, Wind speed plays a key role in wind generators, and it is proportional to the power generated by wind turbines (WTs). However, wind speed has the characteristics of fluctuations and random changes (Masseran, 2016). Therefore, mastering wind energy's characteristics is essential to optimizing its exploitation. Because wind speeds fluctuate randomly, statistical probability methods can be applied to analyze their characteristics. The probability density function (PDF) is a prominent method in the field of statistics. Specifically, PDFs describe the properties of random variables in terms of density, distribution shape, and other related aspects (Elliott et al., 2004). On that basis, PDFs are also significant in determining the optimal power output for WTs or wind farms. Therefore, PDF analysis is also significant in estimating output power for strategizing and operating wind farms.

Many studies have deployed PDFs to describe WSPD, and many distribution functions have been applied, such as Weibull (Azad et al., 2014), Gamma (Aries et al., 2018), Gumble (Kang et al., 2015), and Rayleigh (Bidaoui et al., 2019) distribution. These distributions are also combined together to form mixed distributions to describe multimodal WSPDs, typically mixed distributions Weibull-Weibull (Carta & Ramírez, 2007), Weibull-Gamma (Ouarda et al., 2015), Gamma-Weibull (Chang, 2011). These distributions are parametric because they depend on skewness, location, and shape. To apply parametric distributions, it is necessary to choose appropriate functions and parameters, which also pose many challenges. On the other hand, parametric distributions depend on the shape of the functions and, therefore, cannot be a good description for distributions with shapes other than the general one (Shi et al., 2021).

In contrast, the Kernel Density Estimate (KDE) distribution is known as a nonparametric distribution that can overcome the above challenges. The KDE method estimates the PDF directly from data without assuming the underlying distribution (Qin et al., 2011). This helps describe distributions that do not conform to known distributions or when flexibility in the distribution structure is needed. This method smooths data points using kernel functions, most commonly Gaussian functions (Zhang et al., 2013). The important point in the KDE method is the bandwidth parameter. Bandwidth, also known as the smoothing parameter, is a factor that determines the smoothness of the PDF estimate by affecting the width of the kernel. Many studies have been conducted to select the optimal bandwidth for KDE distribution, including Sheather-Jones Plug-in (SJPI) (Sheather & Jones, 1991), Silverman's Rule of Thumb (SROT) (Harpole et al., 2014), and Least Square Cross-validation (LSCV) (Demir, 2018). Each approach has certain advantages and disadvantages.

With its outstanding advantages, the KDE distribution can be applied to describe WSPD, especially in complex wind regimes. However, not many studies have been conducted using KDE for WSPD distribution. Furthermore, relevant research does not exist in Viet Nam. Therefore, this research was implemented with the following main contributions:

KDE models characterize wind speed distributions (WSPD) using six bandwidth selection methods, including LSCV, Oversmooth, Scott, Sheather-Jones, Silverman, and Normal Scale. To determine the optimal distribution, statistical tests such as CvM, K-S, and A-D, we evaluated the goodness-of-fit of the distributions.

Wind speed data of 6 regions in 2015 in Viet Nam are used to generate the WSPD. The highly suitable WPSD model is then applied to a specific turbine to estimate the corresponding power output.

The research is deployed in sections: section 2 describes the wind speed data set. Section 3 presents the estimation of WSPD using KDE according to bandwidth selection methods. Section 4 implements the appropriate goodness-of-fit tests for every PDF model. Section 5 presents and analyzes the performance of KDE models and estimated output. Conclusions are drawn in section 4.

2. MATERIALS AND METHOD

2.1. Wind speed data set description

The wind speed data set was taken from the project “Establishment of Legal Framework and Technical Assistance to Grid Connected Wind Power Development in Viet Nam”. Wind speed data was collected every 30 minutes over a 1-year period (2015) in six regions, including EaPhe (12°43' 53.280" N - 108° 21' 35.520" E), Da Loan (11° 34' 28.140" N - 108° 22' 5.700" E), Ea Drang (13° 13' 19.740" N - 108° 12' 45.540" E), Ia Der (13° 59' 4.010" N - 107° 56' 16.300" E), Kon Dong (14° 2' 20.570" N - 108° 16' 11.200" E), and My Thanh (20° 33' 00" N - 105° 32' 00" E), shown in Figure 1. Stations The anemometer is located at a height of 80 meters, which is usually the ideal height for installing WTs. Wind speed data is collected every second (1 Hz), then wind speed is averaged over 30 minutes. At each location, 17520 wind speed samples were collected. Table 1 presents basic information on wind density distribution, such as average wind speed, skewness, kurtosis, and standard deviation. Figure 2 presents the histograms of the wind speed data set in six regions.

Table 1. The basic parameters of probability density distribution for each region

	Ea Phe	Da Loan	Ea Drang	Ia Der	Kon Dong	My Thanh
Mean	5.0336	4.9393	4.3620	5.4674	5.5385	5.4655
Skewness	0.1192	0.5129	0.4770	0.3469	0.3463	0.5239
Kurtosis	2.5969	3.1561	3.0340	2.6589	2.5871	2.5492
Standard Deviation	2.3593	2.3496	2.0128	2.3402	2.9447	3.0904

In all regions, the mean wind speed is quite similar, from 4.3 m/s to 5.5 m/s. Positive skewness shows that the wind distribution is asymmetrical but does not vary, and most wind speeds have small wind speed values. The kurtosis value of the distribution is close to 3, indicating that this data distribution almost has a kurtosis equivalent to the normal distribution (Ea Drang). The standard deviation represents the wind speed data dispersion. My Thanh has the highest standard deviation (3.0904), showing that the wind speed variation in this area is the largest. Meanwhile, Ea Drang has the lowest standard deviation (2.0128), indicating the least variability in wind speed.

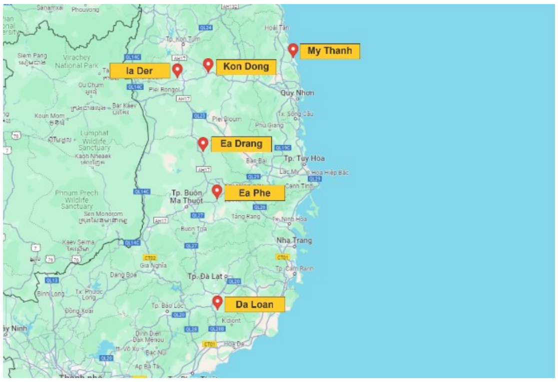

Figure 1. Location of 6 wind speed measuring stations by region

2.2. WSPD based on KDE distribution

KDE is an outstanding nonparametric method for accurately estimating probability distributions. Because it does not depend on general function shapes, KDE is especially suitable for describing WSPDs with multimodal distributions and complex shapes. The formula for KDE to estimate the PDF distribution as follows (Silverman, 2018).

$$\hat{f}(x) = nh\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \quad (1)$$

Where $\hat{f}(x)$ presents the estimated density at the point x, x_i are the data points at i^{th}, n denotes the number of data points, *h* is the bandwidth, and *G* is the kernel function.

The Gaussian kernel function used in this research is a commonly and widely used kernel:

$$K\left(\frac{x-x_i}{h}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-x_i)^2}{2h^2}\right) \quad (2)$$

Choosing the right bandwidth is arguably more crucial than choosing the kernel (Zhang et al., 2013). Bandwidth can be selected based on experience or through trial and error. Besides that, bandwidth also has optimal selection methods for each data set.

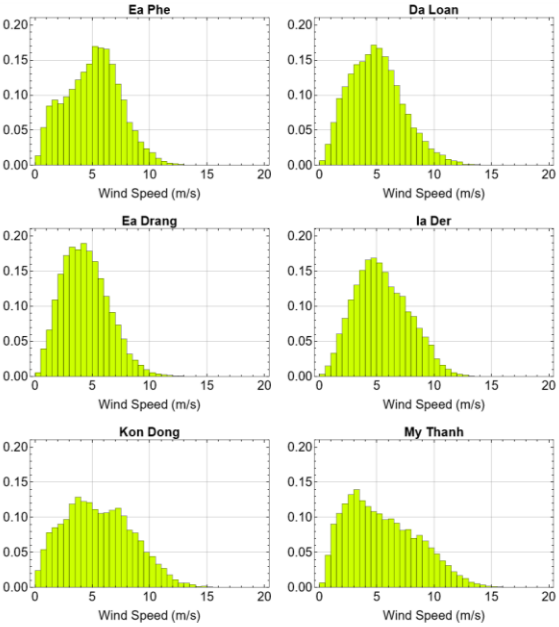

Figure 2. Wind speed histogram for each region

2.3. Bandwidth selection methods

Bandwidth is the crucial factor affecting the smoothness of KDE distribution. Many proposed bandwidth selection methods are typically mentioned in the study (Demir, 2018) proposing the LSCV method. The LSCV method finds the optimal bandwidth by minimizing the loss function based on the squared error between the estimated and actual density. The loss function is presented:

$$L(h) = \int\left[\hat{f}_h(x) - \hat{f}(x)\right]^2 dx \quad (3)$$

where $\hat{f}_h(x)$ is estimated probability density with bandwidth \$h\$ and actual probability density. The optimal bandwidth h_{LSCV} is the bandwidth for which the loss function is the smallest.

The Oversmooth method (Terrell, 1990) estimates large bandwidths to create smoother trend densities that help clarify overall trends in noisy data. However, a large bandwidth can cause a loss of important details, making it unsuitable for data with many complex details and fluctuations. Oversmooth uses a simple formula to estimate bandwidth *h*:

$$h_{OS} = An^{-\frac{1}{5}} \tag{4}$$

Where h_{OS} is the Oversmooth bandwidth, *A* is a constant related to the distribution type and kernel used, and *n* number of data samples.

Scott's Rule (Scott, 1992) is a simple and commonly used method for bandwidth selection in KDE. It calculates the bandwidth based on the standard deviation and the sample size of the data, providing a reasonable balance between bias and variance. While it is easy to compute and works well for data that follows a normal distribution, it may not perform adequately with multimodal or highly skewed datasets. The assumption of normality limits its applicability in diverse data scenarios. The bandwidth h_{SC} is calculated as:

$$h_{SC} = 1.06\sigma n^{-\frac{1}{5}} \tag{5}$$

where σ is the standard deviation, and *n* is the sample size.

The Sheather-Jones method selects the bandwidth by minimizing the expected Kullback-Leibler distance. This method typically provides accurate density estimates, making it suitable for multimodal and skewed data distributions. However, similar to LSCV, it is computationally intensive and requires numerical optimization methods. Due to its large computational load, it can be challenging to apply to large datasets. The specific formula and procedure for calculating information bandwidth using this method are presented in the study (Sheather & Jones, 1991).

Silverman's Rule (Harpole et al., 2014) is a widely used rule of thumb for bandwidth selection in KDE. It calculates the bandwidth related to the standard deviation and the interquartile range of the data, providing a simple and effective method for unimodal and symmetric data distributions. While easy to calculate, Silverman's Rule may not perform well with multimodal or highly skewed data, as it also assumes normality. Its simplicity and general applicability make it a popular choice, though it may not always yield the best results for all data types. The bandwidth h_{SR} is calculated as:

$$h_{SR} = 0.9\min\left(\sigma, \frac{IQR}{1.34}\right)n^{-\frac{1}{5}} \tag{6}$$

where σ denotes the standard deviation, *IQR* refer to the interquartile range, and *n* is the number of data points.

The Normal Scale method (Harpole et al., 2014) is commonly used to select bandwidth for KDE distribution. This method uses the basic characteristics of the Normal distribution to estimate the optimal bandwidth. Bandwidth h_{NS} is calculated by:

$$h_{NS} = \sigma\left(\frac{4}{3n}\right)^{\frac{1}{5}} \tag{7}$$

Where σ denotes the standard deviation, and n is the sample size.

In this study, each bandwidth selection method has its advantages and disadvantages. The LSCV method provides high accuracy by optimizing the error but is computationally complex. Oversmooth smooths the data but can lose important details. The Scott method is simple and provides a good balance between bias and variance but is limited when the data are not normal. Sheather-Jones is suitable for complex data but is computationally expensive. Silverman and Normal Scale are computationally easy and efficient for simple data but are not suitable for complex distributions.

2.4. Goodness-of-fit

The goodness-of-fit models describing WSPD is evaluated through statistical tests, including the Cramér-von Mises (CvM), Anderson-Darling (A-D), and Kolmogorov-Smirnov (K-S). Each test offers a distinct evaluation method, providing unique insights into the distribution models' accuracy. Utilizing all statistical tests ensures a comprehensive and detailed assessment, enhancing the reliability of the distribution models' performance.

2.4.1. The Cramér–von Mises test

The CvM test measures the squared difference between the empirical and estimated cumulative distribution functions (CDFs) over the entire data range. The CvM test (Darling, 1957) is defined as:

$$CvM = n\omega^2 \tag{8}$$

Where $\omega^2 = \int_{-\infty}^{+\infty}\left[F(x)-F^*(x)\right]^2 dF^*(x)$ with *F* and *F** are the empirical and estimated distribution CDFs, respectively. The CvM test is appropriate for wind speed analysis due to its sensitivity to variations across all data values.

2.4.2. The Kolmogorov–Smirnov test

The K-S test (Darling, 1957) calculates the maximum difference between the empirical and estimated CDF. The K-S statistic is defined as:

$$D = \max_{1\le i\le N}\left(F_{(x_i)} - \frac{i-1}{N}, \frac{i}{N} - F_{(x_i)}\right) \tag{9}$$

Where $F(x) = \frac{1}{N}\left[n(i)\right], N$ is the sample size, and x_i are the sample values sorted in ascending order. Here, $n(i)$ is the count of observations less than x_i

2.4.3. The Anderson–Darling test

The A-D test (Corder & Foreman, 2011) enhances the K-S test to determine if an empirical CDF is suitable with an estimated CDF. It focuses more on deviations at the distribution's tails, making it more effective at evaluating the goodness of fit for extreme wind speed values.

$$AD = -N - \frac{1}{N}\sum_{i=1}^{N}(2i-1)\times\left\{\ln\left[1-F\left(x_{n-(i+1)}+\ln F(x_i)\right)\right]\right\} \tag{10}$$

Where: N is the sample size, and x_i is the wind speeds in the i^{th}.

Statistical tests provide detailed assessments of WSPD in all wind speed ranges. Each test provides a distinct perspective, allowing for a comprehensive evaluation of the AKDE distribution's performance.

In these tests, lower test statistic values indicate a better fit. The P-value, set at 0.05, helps determine significance. A *P-value ≤ 0.05* means the model does not fit the data, leading to rejecting the null hypothesis. A *P-value > 0.05* suggests the model reasonably fits the data (Di Leo & Sardanelli, 2020). This approach ensures robust conclusions about the model's ability to describe WSPD.

3. RESULTS AND DISCUSSION

3.1. WSDP modeling using KDE

The bandwidth values and computation times (in seconds) for six KDE distribution methods corresponding to each wind speed dataset are presented in Tables 2 and 3. The variation in bandwidth values between methods and regions reflects the diversity in the structure and nature of the wind data.

Table 2. Bandwidth values for KDE models across different regions

Kernel	Ea Phe	Da Loan	Ea Drang	Ia Der	Kon Dong	My Thanh
LSCV	0.15408	0.30576	0.26587	0.25042	0.18597	0.27138
Oversmooth	0.38238	0.38081	0.32621	0.37928	0.47724	0.50086
Scott	0.33426	0.33288	0.28516	0.33154	0.41719	0.43782
SheatherJones	0.24077	0.30163	0.26913	0.32259	0.277	0.26597
Silverman	0.30083	0.2996	0.25664	0.29839	0.37547	0.39405
NormalScale	0.35405	0.3526	0.30205	0.35118	0.44189	0.46376

Table 3. Computation time (s) for bandwidth estimation of KDE models across different regions

Kernel	Ea Phe	Da Loan	Ea Drang	Ia Der	Kon Dong	My Thanh
LSCV	0.420	0.129	0.079	0.077	0.077	0.080
Oversmooth	0.005	0.005	0.014	0.009	0.009	0.009
Scott	0.004	0.005	0.010	0.007	0.007	0.008
SheatherJones	193.6	163.8	173.0	167.9	178.2	260.4
Silverman	0.056	0.063	0.039	0.044	0.048	0.067
NormalScale	0.007	0.015	0.010	0.007	0.008	0.013

The computation times for bandwidth estimation using different KDE methods vary significantly across the six wind speed datasets. Notably, the Sheather-Jones method exhibits the longest computation times, particularly in the My Thanh region, where it reaches 260.4 seconds. In contrast, methods such as Oversmooth and Scott consistently show fast performance, with computation times typically below 0.01 seconds. This brings attention to the trade-off between accuracy and computational efficiency for various KDE methods, which is influenced by the characteristics of the dataset.

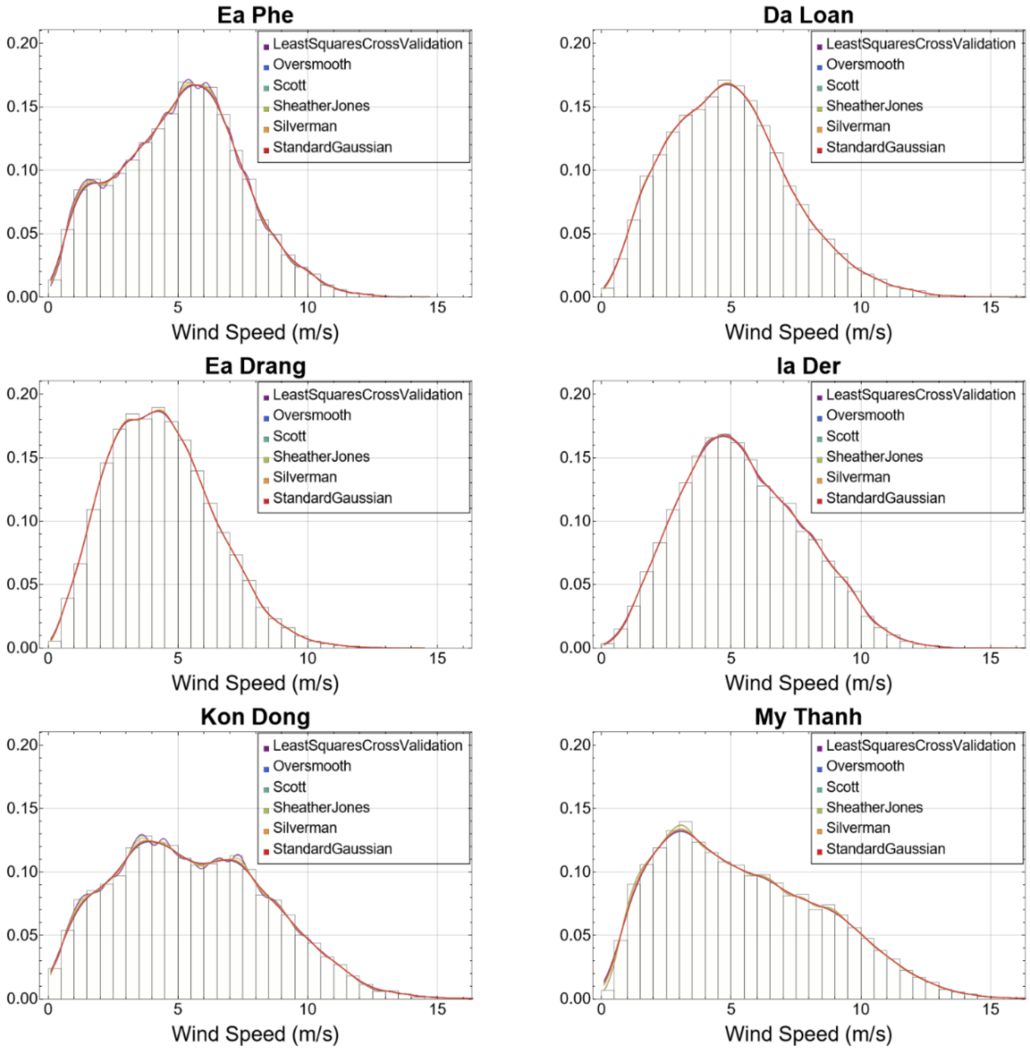

Figure 3. Fitting curve of KDE distribution according to each bandwidth selection method

Figure 3 presents the curves of the KDE distribution according to each bandwidth selection method. For the EaPha and Kon Dong regions, bandwidth values differ between distributions. The LSCV method has a small bandwidth, so it closely follows the fluctuations of the details. On the contrary, the bandwidth values of the remaining methods are relatively large and similar, giving smoother and more general curves. In these two regions, the wind speed distribution has strong fluctuations, and the distribution has many details. Therefore, a small bandwidth value is suitable to handle fluctuations and capture these details.

The Da Loan, Ea Drang, Ia Der, and My Thanh regions have quite similar KDE distribution curves compared to the methods. Although the wind speed data sets in these regions do not have a general distribution shape, they also do not have much variation in distribution.

3.2. Evaluation of the goodness-of-fit performance

The goodness of fit of the distributions must be evaluated not only through images but also based on specific data. Evaluation tests, including CvM, K-S, and A-D, were implemented, and the lower the statistical value results, the more suitable the distribution is. Besides, the P value is also used if a *P-value > 0.05* recommends the model reasonably fits the data.

3.2.1. CvM test result

The CvM test results are presented in Table 4 the LSCV, Oversmooth, Scott, and SheatherJones methods are highly suitable. They have a *P-value > 0.05*, representing no significant difference between the actual and the estimated distribution. The Ea Phe, Ia Der, Kon Dong, and My Thanh regions are highly suitable for all methods.

Table 4. CvM test results for each region

	Ea Phe		**Da Loan**		**Ea Drang**		**Ia Der**		**Kon Dong**		**My Thanh**	
Kernel	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value
LSCV	0.2387	0.2035	0.3473	0.1000	0.4298	0.0604	0.0225	0.9939	0.1220	0.4874	0.1905	0.2868
Oversmooth	0.2837	0.1504	0.3458	0.1010	0.4282	0.0609	0.0449	0.9066	0.1217	0.4886	0.1898	0.2881
Scott	0.2643	0.1710	0.3025	0.1330	0.3852	0.0791	0.0313	0.9716	0.1133	0.5237	0.1709	0.3317
SheatherJones	0.2588	0.1775	0.3443	0.1019	0.4267	0.0615	0.0560	0.8393	0.1214	0.4897	0.1892	0.2895
Silverman	0.2431	0.1975	0.4791	0.0450	0.5610	0.0280	0.0342	0.9603	0.1455	0.4035	0.2461	0.1935
NormalScale	0.2667	0.1683	0.5261	0.0342	0.6078	0.0214	0.0302	0.9755	0.1535	0.3790	0.2654	0.1697

3.2.2. K-S test result

Table 5 presents the K-S test results of the KDE distribution models describing the WSPD for the corresponding regions. In two-thirds of the regions, the KDE distribution with LSCV and Scott bandwidth selection methods suits most wind data sets, while the remaining methods have poorer goodness-of-fit performance. In this test, the Ia Der and Kon Dong regions are a good fit for all KDE distributions.

Table 5. K-S test results for each region

	Ea Phe		**Da Loan**		**Ea Drang**		**Ia Der**		**Kon Dong**		**My Thanh**	
Kernel	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value
LSCV	0.0097	0.0737	0.0107	0.0370	0.0121	0.0115	0.0028	0.9990	0.0075	0.2719	0.0090	0.1192
Oversmooth	0.0114	0.0214	0.0107	0.0375	0.0121	0.0117	0.0036	0.9795	0.0075	0.2741	0.0090	0.1200
Scott	0.0109	0.0300	0.0102	0.0515	0.0118	0.0155	0.0031	0.9961	0.0071	0.3445	0.0086	0.1517
SheatherJones	0.0108	0.0339	0.0106	0.0380	0.0121	0.0118	0.0039	0.9531	0.0075	0.2760	0.0089	0.1208
Silverman	0.0101	0.0573	0.0120	0.0127	0.0133	0.0041	0.0032	0.9941	0.0084	0.1724	0.0113	0.0227
NormalScale	0.0110	0.0283	0.0125	0.0087	0.0139	0.0023	0.0031	0.9966	0.0086	0.1522	0.0120	0.0131

3.2.3. A-D test result

The AD test results are presented in Table 6. The KDE with Scott bandwidth selection method shows the best fit for 5 out of 6 regions. The LSCV, Oversmooth, and SheatherJones methods determine the appropriate bandwidth to distribute KDE in 2 out of 3 regions. In contrast, the Normal Scale and Silverman methods show the worst performance, only suitable for 1 of the two regions. Da Loan and Ea Drang are two areas where most KDE distributions are unsuitable.

Remark: From the overall results of 4 statistical tests, the KDE distribution with the LSCV and Scott bandwidth selection methods has the highest goodness of fit. These two KDE distributions are suitable for most regions, so the KDE PDF results are reliable enough to be applied to estimating wind power output.

Table 6. A-D test results for each region

	Ea Phe		Da Loan		Ea Drang		Ia Der		Kon Dong		My Thanh	
Kernel	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value
LSCV	1.4388	0.1918	2.9044	0.0306	3.3345	0.0186	0.2593	0.9653	1.1889	0.2718	2.4260	0.0542
Oversmooth	2.2471	0.0675	2.8825	0.0314	3.3128	0.0190	0.5467	0.6997	1.1823	0.2744	2.4067	0.0555
Scott	1.9073	0.1033	2.2473	0.0674	2.6932	0.0393	0.3721	0.8759	0.9900	0.3626	1.8380	0.1130
SheatherJones	1.8063	0.1177	2.8613	0.0322	3.2918	0.0195	0.6897	0.5675	1.1759	0.2769	2.3884	0.0567
Silverman	1.5162	0.1727	4.7113	0.0040	5.1182	0.0025	0.4096	0.8392	1.7151	0.1325	3.9030	0.0097
NormalScale	1.9499	0.0979	5.3241	0.0020	5.7298	0.0013	0.3578	0.8893	1.8889	0.1058	4.3750	0.0057

3.3. Wind power output calculation

Wind speed directly affects the power generated by WT. Each WT has its own power curve, Figure 4 shows the power curve of WT Enercon E82/2300. The power curve presents the system between wind speed (m/s) and output power (kW) of the WT. WT goes through four main zones: zone 1 when the wind speed is less than v_{cutin} , the turbine is inactive, and the output power is 0. Zone 2, when the wind speed is than v_{cutin} to v_{rated} , the output power gradually increases with the wind speed. Zone 3, when the wind speed is from than v_{rated} to v_{cutout} , the turbine reaches its rated capacity and does not increase further. In zone 4, when the wind speed exceeds v_{cutout} , the turbine stops working to avoid damage.

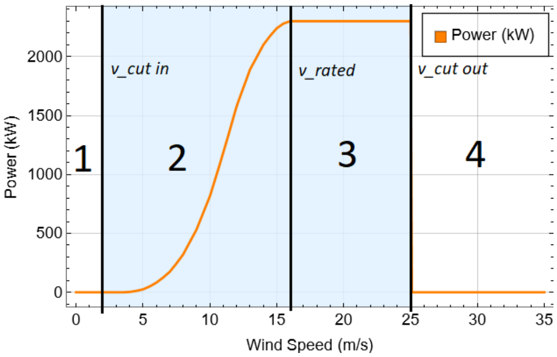

Figure 4. Enercon E82/2300 wind turbine power curve

Accurately estimating the power output of each turbine is important for making sound design decisions, planning wind farm construction, or integrating into the power system. This is done by accurately estimating wind speed variations using reliable and mathematically rigorous probabilistic models. The WT power curve is combined with the PDF curve to estimate the output through the expression:

$$P_{output} = t\int_{x_{cut-in}}^{x_{cut-out}} P(x)\hat{f}(x)dx \qquad (11)$$

where P_{output} is output power, $P(x)$ the power curve of WT, and $\hat{f}(x)$ estimated PDF WSPD.

The two KDE distributions with the LSCV and Scott bandwidth selection methods have the highest goodness of fit with the selected data sets deployed to estimate the power output. The fit curves of these two distributions are combined with the power curve of WT. The estimated output electricity output results are showed in Table 7.

Table 6 shows each region's total energy production (kWh) over a year with the Enercon E82/2300 wind turbine model based on the KDE LSCV and Scott distributions. The difference between the two estimation methods is not large, only ranging from a few kWh to about 20 kWh in different areas. This presents that both methods give relatively consistent results. My Thanh has the highest energy production of both methods, with LSCV at 2042.62 kWh and Scott at 2056.38 kWh. Ea Drang has the lowest energy production in both methods, with LSCV at 560.81 kWh and Scott at 561.90 kWh. The output estimation results show that Kon Don and My Thanh are the two regions with the best potential for exploiting wind energy. This place has a high average wind speed, and high wind speed points have high density.

Table 7. The total energy produced (kWh) over the course of a year for each region with the wind turbine model Enercon E82/2300

Kernel	Ea Phe	Da Loan	Ea Drang	Ia Der	Kon Dong	My Thanh
LSCV	1021.41	1035.77	560.806	1343.47	1893.09	2042.62
Scott	1033.19	1037.87	561.898	1350.38	1911.55	2056.38

4. CONCLUSION

This study evaluated the potential for wind energy exploitation in six different regions of Viet Nam using the KDE model. Among the six applied bandwidth selection methods, the LSCV and Scott methods performed superiorly in describing the wind speed distribution in these areas.

The KDE model has proven effective in capturing the complex and multiple nature of wind speed data. Therefore, it provides a reliable PDF for estimating wind power output.

The analysis shows significant differences in wind speeds across the six regions, highlighting the importance of assessing local wind speeds for effective wind energy planning. The KDE model is tested with statistical tests, including CvM, K-S, and A-D, confirming the robustness and reliability of the estimated PDF.

Power generation estimates based on KDE distributions for the Enercon E82/2300 WT model highlight significant wind energy potential, especially in areas like My Thanh and Kon Dong. These findings support strategic decisions regarding wind farm development, turbine placement optimization, and operational efficiency.

This study provides a robust methodological framework for assessing wind energy potential, taking advantage of the flexibility and accuracy of the KDE model. The method outlined herein can be widely applied to other regions with different wind regimes, contributing to the overall goal of sustainable and efficient wind energy exploitation. Future research could enhance the model's prediction accuracy by integrating additional environmental factors and exploring more advanced KDE techniques.

REFERENCES

IRENA. (2024). *The Global Atlas for Renewable Energy: A decade in the making, International Renewable Energy Agency, Abu Dhabi*. www.irena.org/Publications

Global Wind Report 2024 - Global Wind Energy Council. (n.d.). Retrieved October 17, 2024, from https://gwec.net/global-wind-report-2024/

Aries, N., Boudia, S. M., & Ounis, H. (2018). Deep assessment of wind speed distribution models: A case study of four sites in Algeria. *Energy Convers. Manag.*, *155*, 78–90.

Azad, A. K., Rasul, M. G., Alam, M. M., Uddin, S. M. A., & Mondal, S. K. (2014). Analysis of wind energy conversion system using Weibull distribution. *Procedia Eng.*, *90*, 725–732.

Bidaoui, H., Abbassi, I. El, & Bouardi Abdelmajid El and Darcherif, A. (2019). Wind speed data analysis using Weibull and Rayleigh distribution functions, case study: Five cities Northern Morocco. *Procedia Manuf.*, *32*, 786–793.

Carta, J. A., & Ramírez, P. (2007). Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions. *Renew. Energy*, *32*(3), 518–531.

Chang, T. P. (2011). Estimation of wind energy potential using different probability density functions. *Appl. Energy*, *88*(5), 1848–1856.

Corder, G. W., & Foreman, D. I. (2011). Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach. *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*, 1–536. https://doi.org/10.1002/9781118165881

Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramer-von Mises Tests. *The Annals of Mathematical Statistics*, *28*(4), 823–838. http://www.jstor.org/stable/2237048

Demir, S. (2018). Adaptive kernel density estimation with generalized least square cross-validation. *Hacet. J. Math. Stat.*, *48*(3).

Di Leo, G., & Sardanelli, F. (2020). Statistical significance: p value, 0.05 threshold, and applications to radiomics-reasons for a conservative approach. *Eur. Radiol. Exp.*, *4*(1), 18.

Elliott, D., Schwartz, M., & Scott, G. (2004). Wind Resource Base. In *Encyclopedia of Energy* (pp. 465–479). Elsevier.

Farghali, M., Osman, A. I., Chen, Z., Abdelhaleem, A., Ihara, I., Mohamed, I. M. A., Yap, P. S., & Rooney, D. W. (2023). Social, environmental, and economic consequences of integrating renewable energies in the electricity sector: A review. *Environmental Chemistry Letters*, *21*(3), 1381–1418. https://doi.org/10.1007/S10311-023-01587-1

Harpole, J. K., Woods, C. M., Rodebaugh, T. L., Levinson, C. A., & Lenze, E. J. (2014). How bandwidth selection algorithms impact exploratory data analysis using kernel density estimation. *Psychol. Methods*, *19*(3), 428–443.

Kang, D., Ko, K., & Huh, J. (2015). Determination of extreme wind values using the Gumbel distribution. *Energy (Oxf.)*, *86*, 51–58.

Khaloie, H., Abdollahi, A., Shafie-khah, M., Anvari-Moghaddam, A., Nojavan, S., Siano, P., & Catalão, J. P. S. (2020). Coordinated wind-thermal-energy storage offering strategy in energy and spinning reserve markets using a multi-stage model. *Applied Energy*, *259*, 114168. https://doi.org/https://doi.org/10.1016/j.apenergy.2019.114168

Masseran, N. (2016). Modeling the fluctuations of wind speed data by considering their mean and volatility effects. *Renew. Sustain. Energy Rev.*, *54*, 777–784.

Ouarda, T. B. M. J., Charron, C., Shin, J.-Y., Marpu, P. R., Al-Mandoos, A. H., Al-Tamimi, M. H., Ghedira, H., & Al Hosary, T. N. (2015). Probability distributions of wind speed in the UAE. *Energy Convers. Manag.*, *93*, 414–434.

Qin, Z., Li, W., & Xiong, X. (2011). Estimating wind speed probability distribution using kernel density method. *Electric Power Syst. Res.*, *81*(12), 2139–2146.

Scott, D. W. (1992). *Multivariate density estimation*. John Wiley & Sons.

Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Series B Stat. Methodol.*, *53*(3), 683–690.

Shi, H., Dong, Z., Xiao, N., & Huang, Q. (2021). Wind speed distributions used in wind energy assessment: A review. *Front. Energy Res.*, *9*.

Silverman, B. W. (2018). Density estimation: For statistics and data analysis. *Density Estimation: For Statistics and Data Analysis*, 1–175. https://doi.org/10.1201/9781315140919/DENSITY-ESTIMATION-STATISTICS-DATA-ANALYSIS-BERNARD-SILVERMAN

Terrell, G. R. (1990). The maximal smoothing principle in density estimation. *J. Am. Stat. Assoc.*, *85*(410), 470.

Zhang, J., Chowdhury, S., Messac, A., & Castillo, L. (2013). A Multivariate and Multimodal Wind Distribution model. *Renew. Energy*, *51*, 436–447.