Eskişehir Osmangazi Üniversitesi Müh.Mim.Fak.Dergisi C. XVIII, S.2, 2005
Eng.&Arch.Fac. Eskişehir Osmangazi University, Vol. XVIII, No: 2, 2005

STATISTICAL ANALYSIS OF WIND SPEED DATA

Veysel YILMAZ [1], Haydar ARAS [2], H.Eray ÇELİK [3]

ABSTRACT : *Wind speed is the most important parameter in the design and study of wind energy conversion devices. The energy which is obtained from wind is directly proportional with the cubic power of the wind speed. As the wind speed increases, the cost of the wind energy is reduced. In many studies in literature, it is assumed that the probability distribution related to wind speeds can be described by Weibull distribution, and it is accepted so without any statistical examination. In this study, the theoretical distributions of wind potentials fit to Weibull distribution for five different topographic situations from Turkey Wind Atlas are investigated and reported.*

KEYWORDS : *Wind Energy; Weibull Distribution; Statistical Tests.*

RÜZGAR HIZ VERİLERİNİN İSTATİSTİKSEL ANALİZİ

ÖZET : *Rüzgar enerjisi dönüşüm sistemlerinde rüzgar hızı en önemli parametrelerden biridir. Rüzgardan elde edilen enerji, rüzgar hızının küpüyle doğrudan orantılıdır. Rüzgar hızı arttıkça rüzgar enerjisi maliyetleri azalmaktadır. Literatürdeki pek çok çalışmada rüzgar hızı olasılık dağılımları hiç bir istatistik i test yapılmaksızın Weibull dağılımı olarak tanımlanmaktadır. Bu çalışmada Türkiye Rüzgar atlasında verilen beş farklı istasyondaki rüzgar hız verilerinin teorik dağılımının Weibull a uyup uymadığı araştırılmıştır.*

ANAHTAR KELİMELER : *Rüzgar Enerjisi; Weibull Dağılımı; İstatistiksel Testler.*

[1,3] Eskişehir Osmangazi Üniversitesi,Fen Edebiyat Fakültesi,İstatistik Bölümü,26480 ESKİŞEHİR
[2] Eskişehir Osmangazi Üniversitesi, Mühendislik Mimarlık Fakültesi, Makina Mühendisliği Bölümü, 26480 Batı Meşelik, ESKİŞEHİR

I.INTRODUCTION

As the world population grows rapidly, the energy demand increases proportionally. In spite of increasing energy demand, the limited reserves of fossil fuel, good quality energy, and productive usage of the produced energy becomes an important matter in all countries especially in developing countries such as Turkey [1]. Turkey has a considerably high level of renewable energy resources that can be utilized to satisfy a portion of the total energy demand [14]. In this study, statistical estimation techniques which are used in estimation of Weibull distribution parameters are quoted, then the wind speed distribution of five different regions fitted to Weibull distribution was examined using Chi-Square and Kolmogorov-Smirnov tests. Geographical coordinates for five different regions are given Table 1. The data for this study was obtained from Turkey wind atlas published by the General Directorate of Turkish State Meteorological Service and the General Directorate of Electrical Power Resources Survey Administration.

Table 1: Coordinate data for the five station

Station	Station Coordinates (Degree-Minute-Second)	Height of the sea level	Anemograph height
Afyon	41^0 01^I 59^{II} N 39^0 33^I 38^{II} E	1034m	13m
Van	38^0 28^I 14^{II} N 43^0 20^I 42^{II} E	1661m	10m
Sinop	42^0 01^I 51^{II} N 35^0 09^I 18^{II} E	32m	10m
Bozcaada	39^0 50^I 00^{II} N 26^0 04^I 25^{II} E	28m	10m
Silifke	30^0 22^I 58^{II} N 33^0 56^I 19^{II} E	15m	10m

Weibull distribution with two parameters is usually the used for probability distribution of wind speeds. It is generally accepted that measured wind data can be best characterized by Weibull distribution [3, 6-11, 13]. But in most studies fitting of data set to Weibull distribution was not examined. To determine if the frequency series fit to a theoretical distribution or which distribution fits better, can only be decided by statistical hypothesis tests.

II. METHODS FOR ESTIMATING THE PARAMETERS OF THE WEIBULL DISTRUBITION

The wind speed probability density function can be calculated as

$$f(v) = \left(\frac{k}{c}\right)\left(\frac{v_i}{c}\right)^{k-1} \exp\left[-\left(\frac{v_i}{c}\right)^{k}\right] \quad (1)$$

where f(v) is the probability of observing wind speed v, c is the Weibull scale parameter and k is the Weibull shape parameter. Basically, the scale parameter, *c*, indicates how windy a wind location under consideration is, whereas the shape parameter, *k*, indicates how peaked the wind distribution is [3, 10, 11]. Statistically the parameters of Weibull distribution can be derived by using various estimation techniques. Among those, some of the most widely used techniques are; Maximum Likelihood Estimation (MLE), Method of Moments (MOM) and, Least-Squares Method (LMS) related with graphical technique [4].

II.1. The Maximum Likelihood Method

Maximum Likelihood Technique, with many required features is the most widely used technique among parameter estimation techniques. The MLE method has many large sample properties that make it attractive for use. It is asymptotically consistent, which means that as the sample size gets larger, the estimates converge to the true values. It is asymptotically efficient, which means that for large samples, it produces the most precise estimates. It is also asymptotically unbiased, which means that for large samples, one expects to get the true value on average. The estimates themselves are normally distributed if the sample is large enough. These are all excellent large sample properties. Likelihood function for Weibull distribution with two parameters is as follows [10, 12],

$$L(c,k) = \prod_{i=1}^{n} f(v_i;\theta) \quad (2)$$

and using Eq (1) in Eq (2), we get

$$L(c,k)=\prod_{i=1}^{n} kc^{-k} v_i^{\,k-1} \exp(-c^{-k} v_i^{\,k}) \qquad (3)$$

If we take the natural logarithm of the likelihood function given by Eq (3)

$$InL(c,k)=\sum_{i=1}^{n}(Ink-kInc+(k-1)Inv_i-c^{-k}v_i^{\,k}) \qquad (4)$$

or

$$InL(c,k)=nInk-nkInc+(k-1)\sum_{i=1}^{n}Inv_i-c^{-k}\sum_{i=1}^{n}v_i^{\,k} \qquad (5)$$

is obtained. Then taking the derivative of the likelihood function (Eq.(5)) with respect to *c* and *k*, and set them to zero.

$$\frac{\partial InL(c,k)}{\partial c}=-nkc^{-1}+kc^{-(k+1)}\sum_{i=1}^{n}v_i^{\,k}=0 \qquad (6)$$

and

$$\frac{\partial \ln L(c,k)}{\partial k}=nk^{-1}-n\ln c+\sum_{i=1}^{n}\ln v_i-\sum(\ln\frac{v_i}{c})v_i^{\beta}c^{-k}=0 \qquad (7)$$

The shape factor *k* and the scale factor *c* are estimated using the following two equations:

$$\frac{\sum_{i=1}^{n}\ln v_i^k \ln v_i}{\sum_{i=1}^{n}v_i^k}-\frac{1}{k}=\frac{\sum_{i=1}^{n}\ln v_i}{n} \qquad (8)$$

and

$$k = \left(\frac{\sum_{i=1}^{n} v_i^k \ln(v_i)}{\sum_{i=1}^{n} v_i^k} - \frac{\sum_{i=1}^{n} \ln(v_i)}{n} \right)^{-1} \quad (9)$$

where v_i is the wind speed in time step *i* and *n* the number of nonzero wind speed data points.

Eq. (9) must be solved using iterative procedure, after which Eq. (10) can be solved explicitly. Care must be taken to apply Eq. (9) only to the nonzero wind speed data points.

$$c = \left(\frac{\sum_{i=1}^{n} v_i^{\,k}}{n} \right)^{1/k} \quad (10)$$

II.2. Least – Squares Method

This method is also called Graphical Method. With the help of this method the parameters are estimated with the regression line equation by using cumulative density function. The cumulative density function of Weibull distribution with two parameters can be written as,

$$F(v_i) = 1 - \exp(-c^{-k} v_i^{k}) \quad (11)$$

This function can be arranged as,

$$\{1 - F(v_i)\}^{-1} = \exp(c^{-k} v_i^{\,k}) \quad (12)$$

If we take the natural logarithm of Eq.(12),

$$-\ln\{1 - F(v_i)\} = (c^{-k} v_i^{\,k}) \quad (13)$$

and then retake the natural logarithm of Eq. (13), we get the following equation:

$$\ln[-\ln\{1-F(v_i)\}] = -k\ln c + k\ln v_i \quad (14)$$

In Eq. (14) represents a direct relationship between $\ln v_i$ and $\ln[-\ln\{1-F(v_i)\}]$ which should be minimized.

$$\sum_{i=1}^{n}\{\ln[-\ln(1-F(v_i)] - \ln[-\ln(1-E(F(v_i))]\}^2 \quad (15)$$

Parameters of Weibull distribution with two parameters are estimated by minimizing with Eq. (15). Parameters *c* and *k* intersects by using Eq. (16) and (17) as follows,

$$k = \frac{n\sum_{i=1}^{n}\ln v_i \ln[-\ln\{1-F(v_i)\}] - \sum_{i=1}^{n}\ln v_i \sum_{i=1}^{n}\ln[-\ln\{1-F(v_i)\}]}{n\sum_{i=1}^{n}\ln {v_i}^2 - \{\sum_{i=1}^{n}\ln v_i\}^2} \quad (16)$$

$$c = \exp\{\frac{k\sum_{i=1}^{n}\ln v_i - \sum_{i=1}^{n}\ln[-\ln\{1-F(v_i)\}]}{nk}\} \quad (17)$$

II.3. Method of Moments

Method of Moment is one of the oldest of the estimation methods. If there is a “j” parameter to estimate, when *j* universe moment connected with these parameters is equated to correlative sample moments, quality in “j”, which contains these parameters, is achieved. Estimation values of this equation, obtained on this “j” number, are found by solving the unknown parameters. In the estimation of the Weibull Distribution parameters with the method of Moments the first and second moments of distribution around zero are used. The j^{th} moment of the two-parametered Weibull Distribution around zero is as

$$E(V^j) = c^j\Gamma(\frac{j}{k}+1) \quad (18)$$

where Γ is the gamma function.

When equality number (18) is used, the first two moments around zero is as follows.

$$E(V) = c\ \Gamma(\frac{1}{k}+1) \quad (19)$$

$$\text{and}\ E(V^2) = c^2\Gamma(\frac{2}{k}+1) \quad (20)$$

Only one function which depends on the parameters of *j* is obtained by the division of the square of E(V) to E(V^2),

$$\overline{V} = c\Gamma(\frac{1}{k}+1) \quad (21)$$

and

$$\frac{\{\sum_{i=1}^{n} V_i\}^2}{n\sum_{i=1}^{n} V_i^2} = \frac{\{\Gamma(\frac{1}{k}+1)\}^2}{\Gamma(\frac{2}{k}+1)} \quad (22)$$

k Parameter is estimated by using Eq. (22) with standard iterative technique [6]. Then parameter *c* is estimated by placing k Eq. (23).

$$c = \frac{\overline{V}}{\Gamma(\frac{1}{k}+1)} \quad (23)$$

III. RESULTS & DISCUSSION

In this study to enable the probability distribution which estimates the potentials of the wind speed were determined for Afyon, Van, Sinop, Bozcaada, Silifke. With this purpose, wind speeds of five different regions in Turkey Wind Atlas have been recorded for 1989-1998 and its appropriateness to Weibull distribution has been studied with the help of statistical hypothesis test. *p-values* are often used in hypothesis tests where you either accept or reject a null hypothesis. The p-value represents the probability of making a Type 1 error, or rejecting the null hypothesis when it is true. A cutoff value often used is 0.05 that is, reject the null hypothesis when the p-value is less than 0.05. In many areas of research, the *p*-level of 0.05 is customarily treated as a "border-line acceptable" error level. According to results of statistical test, for the obtained data from stations Afyon and Sinop are appropriate theoretical models for Weibull distribution (Null hypothesis accept because p>α and D< D_α ; α=0.05). But for the other stations, it the test yielded not appropriate (Null hypothesis reject because p<α and D > D_α ; α=0.05). These results exposes that not all wind speed data fits Weibull distribution. Because of this, it is very important to examine the wind speed distribution with statistical tests before making investments to a region just looking at the wind speed potentials. Otherwise such an investment would not be a feasible one. If the distribution of wind speed was not determined correctly, obtained energy amount will not be as expected. Thus, it is very important to determine the probability distribution of wind speed potential correctly. Parameter estimations were done by using data set related to frequency distribution and cumulative frequency distribution which is given in Table 2 [5].

Table 2: Observed wind speed data in frequency distribution (O_i)

Speed Interval(m/s)	Stations				
	Afyon (1989-1998)	Van (1991-1998)	Sinop (1989-1998)	Bozcaada (1989-1998)	Silifke (1989-1998)
<1	277	131	458	23	66
2	399	232	306	33	288
3	210	138	156	23	486
4	74	110	59	38	121
5	29	95	17	65	31
6	9	89	3	69	7
7	2	86	1	107	0
8	0	40	0	110	0
9	0	26	0	114	0
11	0	36	0	173	0
13	0	12	0	146	0
15	0	3	0	68	0
17	0	1	0	20	0
>17	0	0	0	10	0

The results of the analysis for parameter estimation values in related distribution are given in Table 3. When Table 3 examined, parameter values obtained by three analytical methods (Rank Regression on X [RRX], Rank Regression on Y [RRY], MOM and MLE) can be seen.

Table 3: Shape and scale parameter estimation values of Weibull distributions according to RRX, RRY, MLE and MOM methods.

Station	Parameter Estimation	RRX Method	RRY Method	MLE Method	MOM Method
Afyon	c	1.9403	1.9372	2.5100	1.81401
	k	1.5916	1.5871	2.1631	1.52872
Van	c	4.0282	4.0221	4.4500	4.5003
	k	1.3823	1.3771	1.5902	1.46507
Sinop	c	1.4704	1.4672	2.1401	1.56121
	k	1.3147	1.3140	1.9721	1.39816
Bozcaada	c	8.9699	9.1000	10.174	13.5433
	k	2.1148	2.0244	2.5325	3.53856
Silifke	c	2.6701	2.6672	3.0974	1.93648
	k	2.6742	2.6494	3.2864	1.92098

Parameter estimations were inserted in equation (14) by using RRY. Estimated values can be seen in Table 4. LSM, or least sum of squares, regression requires that a straight line to be fitted to a set of data points, such that the sum of the squares the distance of the points to the fitted line is minimized.

Table 4: Estimated wind speed data in frequency distribution (E_i)

Speed Interval (m/s)	Stations				
	Afyon	Van	Sinop	Bozcaada	Silifke
<1	295	137	454	12	72
2	355	235	323	35	301
3	214	116	145	55	372
4	93	142	54	72	202
5	31	111	17	85	48
6	9	83	5	92	5
7	2	59	1	95	0
8	0	41	0	93	0
9	0	28	0	87	0
11	0	30	0	146	0
13	0	12	0	103	0
15	0	4	0	65	0
17	0	1	0	35	0
>17	0	0	0	11	0

This minimization can be performed in either the vertical or horizontal direction. If the regression is on the x-axis, then the line is fitted so that the horizontal deviations from the points to the line are minimized (RRX). If the regression is on the y-axis, then the line is fitted so that the vertical deviations from the points to the line are minimized (RRY). The rank regression estimation method is quite good for functions that can be linearized. As most of the distributions used in life data analysis are capable of being linearized. Further, this technique provides a good measure of the goodness-of-fit of the chosen distribution. Therefore, in this study RRY method was used. For data sets containing large quantities of suspended data points, MLE may be the preferable form of analysis. Observed and expected wind speed values of five different regions were given in Fig. 1a,b,c,d,e. Whether the measured wind speed for these five stations fitted well with the Weibull distribution or not, can be done only by using statistical tests. Null hypothesis, proving that "statistically there is no sensible difference between observed frequencies and expected frequencies", tested by Chi-Square and Kolmogorov - Smirnov tests. Formulas of testing statistics used in this test are given in equations 24 and 25,

$$\chi^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i} \tag{24}$$

where O_i is the observed frequency for each class, E_i is the frequency of each class estimated according to theoretical distribution and, r is the number of the class. Degree of freedom (df) of Chi-Square df is, r-1-m, where m is the number of parameters of concerned theoretical distribution.

$$D = \max \left| F_x(V) - S_T(V) \right| \tag{25}$$

Cumulative frequency of the distribution observed by Kolmogorov-Smirnov test based on the maximum absolute difference between ($F_x(V)$) and cumulative frequency of the estimated distribution ($S_T(V)$). Maximum absolute value of D among calculated D values, are compared with table values (D_α) at meaning level α which were prepared for Kolmogorov-Smirnov test. If D> D_α, the hypothesis saying that data comes from estimated distribution is rejected.

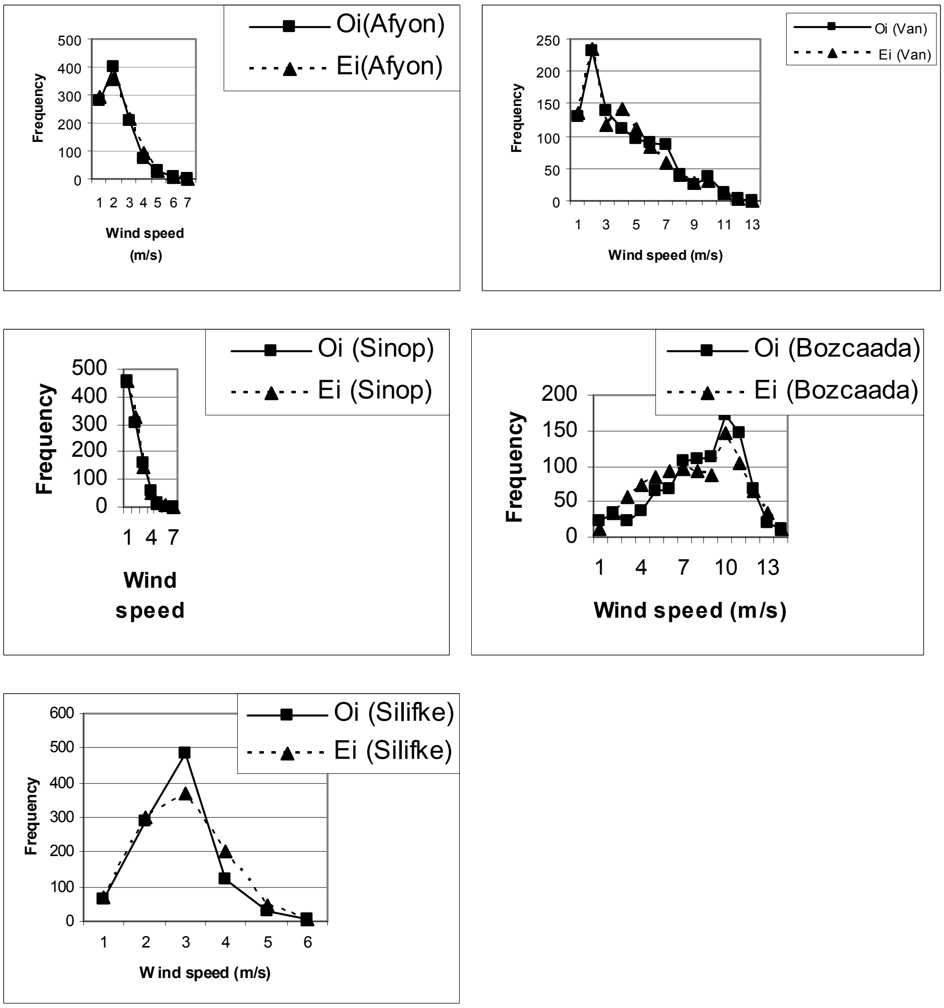

Figure1. Observed and estimated wind speeds for five stations (O_i : Observed frequency E_i : Estimated frequency)

The results of Chi-Square and Kolmogorov-Smirnov tests are given in Table 5.

Table 5: Test results of Chi-Square and Kolmogorov-Smirnov

Station	Chi-Square value and its meaningfulness *(α=0.05)*	D ve $D_\alpha \cong$ 0.04 compariso*(α=0.05)*
Afyon	10.64 ; df = 06 ; p = 0 .1002 Null *hypothesis* accept (p>*α)*	0.025<0.04 Null *hypothesis* accept
Van	28.40 ; df = 12 ; p = 0 .0049 Null *hypothesis* reject (p<*α)*	0.041>0.04 Null *hypothesis* reject
Sinop	03.03 ; df = 06 ; p = 0 .8054 Null *hypothesis* accept (p>*α)*	0.014<0.04 Null *hypothesis* accept
Bozca ada	97.93 ; df = 13 ; p = 0 .0000 Null *hypothesis* reject (p<*α)*	0.095>0.04 Null *hypothesis* reject
Silifke	75.29 ; df = 05 ; p= 0 .0000 Null *hypothesis* reject (p<*α)*	0.104>0.04 Null *hypothesis* reject

IV.CONCLUSIONS

Nowadays, wind energy is the most rapidly developing technology and energy source, and it is reusable. Due to its cleanness and reusability, there have been rapid developments made on transferring the wind energy systems to electric energy systems. Wind energy is an alternative clean energy source comparing to all other fossil originated energy sources which pollute the lower parts of the atmosphere [2]. As problems like environment pollution and supplying energy needs gets bigger, it is important to make use of a pollution free energy such as wind. Converting the wind energy to more widely used electrical energy can be done only with the wind turbines. The most important parameter of the power which was obtained from wind turbines is the wind speed. Due to this fact, it is very important to determine the probability distribution of wind speed values of an area correctly in which wind energy potential will be utilized. The wind potential of a region depends on various wind measurements of a location for years. To determine the wind energy potential of a region or field can be determined by correctly estimating the probability distribution.

REFERENCES

[1] Aras H, Condition and development of the cogeneration facilities based on autoproduction investment model in Turkey. Renewable and Sustainable Energy Reviews 2003; 7(6): 553-559.

[2] Aras H, Wind energy status and its assessment in Turkey. Renewable Energy 2003; 28(14): 2213-2220.

[3] Bivona S, Burlon R, Leone C. Hourly wind speed analysis in Sicily. Renewable Energy 2003 ;28: 1371-1385.

[4] Chen Zhenmin. Statistical inference about the shape parameter of the Weibull Distribution. Statistics and Probabilty Letters. 1997; 36: 85-90.

[5] Dündar C, Canbaz M, Akgün N, Ural G. Turkey wind atlas. Published by the General Directorate of Turkish State Meteorological Service and the General Directorate of Electrical Power Resources Survey Administration, Ankara, Turkey, 2002. [in Turkish]

[6] Dorvlo S.S. A. Estimating wind speed distribution. Energy Conversion & Management 2002; 43: 2311-2318.

[7] Gupta B.K. Weibull parameters for annual and monthly wind speed distributions for five locations in India. Solar Energy 1986; 37(6): 469-471.

[8] AL-Hasan M, Nigmatullin R.R. Idenfication of the generalized Weibull distribution in wind speed data by the eigencoordinates method. Renewable Energy 2003; 28: 93-110.

[9] Rehman S, HalawaniT.O, Husain T. Weibull parameters for wind speed distribution in Saudi Arabia. Solar Energy 1994;53(6): 473-479.

[10] Seguro J.V, Lambert T.W. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. Journal of Wind Engineering and Industrial Aerodynamics 2000; 85: 75-84.

[11] Weisser D. A Wind energy analysis of Grenada: an estimation using the "Weibull" density function. Renewable Energy 2003; 28: 1803-1812.

[12] Xie M, Tang Y, Goh T.N. A modified Weibull extension with bathtub-shaped failure rate function. Reliability Engineering and System Safety 2002;76: 279-285.

[13] Çelik A.N. A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. Renewable Energy 2004; 29 (4):593-604.

[14] Hepbasli A, Ozgener O. A review on the development of wind energy in Turkey. Renewable & Sustainable Energy Reviews 2004; 8(3): 257-276.

NOMENCLATURE

F (v_i)	:	Weibull Cumulative Distribution Function
D	:	Kolmogorov Smirnov test statistics
χ^2	:	Pearson Chi-Square test statistics
E_i	:	Estimated frequency
O_i	:	Observed frequency
k	:	Weibull Shape Parameter
$L(c,k)$	:	Likelihood function
v	:	Wind speed (m/s)

c	:	Weibull Scale Parameter (m/s)
n	:	The Total number of samples
p	:	The p-value represents the probability of making a Type 1 error, or rejecting the null hypothesis when it is true.
$f(v)$	:	Probability density function
m	:	Number of parameters
$\overline{V}$	:	Mean of wind speed
D_{α}	:	Kolmogorov Smirnov theoretical test statistics
α	:	Error level
$1-\alpha$	:	Confidence level
r	:	Number of the Class