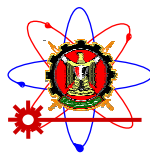


Military Technical College
Kobry Elkobbah,
Cairo, Egypt
May 16-18,2006



3rd International
Conference on Engineering
Mathematics and Physics
(ICMEP-3)

Using Genetic Programming to Estimate WECS Output Power

Said A. Gouda, Ph. D.
Department of Mathematics
Air-Defense College, Egypt

Abstract: A statistical technique to characterize wind speed data for use in wind energy conversion system (WECS) is presented. We start by examining the frequency distribution of long-term data. The histogram is generated for observed wind speed for a particular hour over a month for a number of years. It is fitted to two distributions (Weibull, and β). Genetic Programming (GP) also used to fit the distribution. Two goodness-of-fit criteria are employed in checking the best fit. These are Chi-square (χ^2), and Kolmogorov-Smirnov (KS). Long-term data from International Logan Airport, Boston, MA, USA are analyzed. It is found that the GP fits the long-term hourly data best.

I. INTRODUCTION

There are a number of WECS analysis models that have been used for wide range of design and economic analyses, and policy studies. These models provide a broad set of capabilities to be used for characterizing wind turbine technologies and systems. Some of these capabilities include modeling of turbine characteristics; power conditioning characteristics; plant level characteristics; operation and maintenance characteristics; and site-specific characteristics. One factor that is missing from this long list of characteristics is the characterization of weather data. It is known that the output power is strongly dependent on wind speed. This paper deals with characterizing only the wind data for use in WECS design and analysis models.

The technique of characterizing the wind speed starts with examining the frequency distribution of the data [1-5]. This distribution provides clues concerning the probable distribution (s) that may fit the data. Subsequently, the data are analyzed to check for the best fit among several distributions. A number of goodness-of-fit tests are performed in the process of choosing the best distribution. The distribution parameters are then calculated for the available data. Long-term data Logan airport, Boston, MA, USA was used to test our technique of wind characterization [4].

II. BACKGROUND INFORMATION

There are numerous publications on WECS models for evaluating electricity generation. The great majority of these have a deterministic approach. Thus, these models generally use the known value for wind speed to determine the system power output. In a few cases, however, the probabilistic nature of

the variability in wind speeds has been considered in connection with WECS performance evaluation models. Several of these studies will be discussed here.

It is possible to get different probability distributions for wind speed. The probability distribution parameters will vary from one hour to another. Thus it is believed that, if long-term field data were available for a site, then the approach presented in this paper will give better prediction of a WECS performance.

In the following section we discuss how the wind data are analyzed to determine probability distribution functions that can be utilized in WECS performance analysis models.

III. WIND SPEED HISTOGRAMS

In order to characterize the wind data for various times during the day, long-term observations of wind speed for the same hour of the day (for a large number of days) are examined. Histograms are plotted to check the type of variations encountered in the data for the same hour on different days. In Fig. 1, four such histograms are presented. These are due to wind data for the same hour of the day in a particular month for 30 years. Plots are shown for 12 noon readings in March, June, September and December. These histograms are plotted such that the full spectrum of wind data is represented. For example in Fig. 1-d we see the midpoint wind speed of 4 m/s has a frequency of 67. That implies there are 67 observations between 3 and 5 whose midpoint is 4. Accordingly the midpoint wind speed of 10 m/s represents the 40 data points between 9 and 11. It should be noted that we have shown the data for four seasons as samples only. Data for other hours and months were also examined and found to follow similar trends.

A look at these histograms clearly shows the variations of the wind speeds for the same hour in different seasons. Shapes of these histograms, however, suggest that their relationships can be exploited to fit some form of frequency distribution. This would be an improvement over the usual approach of predicting the performance of WECSs by directly using the observed data for a set of hours. The technique of fitting distributions is discussed in the following section.

IV. FITTING DISTRIBUTIONS

The appropriate probability distribution model needed to describe a random phenomenon, like the wind speed, is difficult to discern. There are, however, occasions when the required distribution is determined empirically. Alternatively an assumed distribution is accepted or rejected on the basis of observed data. Histograms often provide clues for initially selecting such distributions. This is the approach we have taken for fitting a distribution to observed wind data.

Histograms shown in Fig. 1 suggest that wind data, for the same hour over a number of days, would have a form of bimodal probability distribution function.

Furthermore, it is generally seen that, there would be a concentration of observations at the lower end of the spectrum, and another concentration at the higher end of the spectrum. There are of course seasonal and climatologically factors that will determine the exact nature of the frequency distributions. An examination of the histograms in Fig. 1 suggests that superposition of two known probability distribution functions would closely replicate these frequency distributions. Accordingly, the speed measurements are separated into two groups such that a known frequency distribution would fit each group of data. This grouping is done visually by looking at the histograms in Fig. 1. Sometimes it may be necessary to group some of the speed measurements (for the same hour) into three parts for a better

fit. Then the distribution functions are weighted according to break points. For example, if we determine that the first 1/3 of the speed observation have a probability distribution function of F(x) and the other 2/3 have a distribution function of G(x), then the entire data set would be represented by:

$$P\{p \leq x\} = [F(x)/3] + [2G(x)/3] \tag{1}$$

where

$$F(x) = P\{x \geq m\}, \text{ and } G(x) = P\{m \geq x\}$$

where x=m is the break point

Three most likely probability distributions are tested for the best fit. These are Weibull and β , using two measures of fit then compares the predicted and observed class frequencies. These are Chi-square (χ^2), and KS test. All statistics of fit are computed using estimated parameters of the distributions with the data as a whole.

In the following section we discuss briefly the distributions that have been considered for fitting the wind speed data. Two distributions chosen to fit the wind speed data will now be discussed.

A. Weibull distribution

The probability density function is given by [1]

$$f(x) = \left(\frac{k}{c}\right)\left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right], 0 < x < \infty \tag{2}$$

where c = scale factor and k = shape factor. This is a two-parameter distribution and has been used to characterize also wind speed data.

B. β distribution.

A probability distribution appropriate for a random variable where values are bounded, between finite limits 0 and 1, is the β distribution. Its density function is given by [1]

$$f(x) = \left(\frac{\Gamma(k)\Gamma(c)}{\Gamma(k+c)}\right) \cdot x^{k-1}(1-x)^{c-1}, 0 < x < 1 \tag{3}$$

where k and c are the parameters of the distribution.

Two goodness-of-fit statistics are checked for each distribution in order to choose parameters for the probability density function that fit the available wind data best. These statistics are given here [2].

C. Chi-square Test [1]

$$\chi^2 = \sum_{i=1}^n \frac{(P_{oi} - P_{ci})^2}{P_{ci}} \tag{4}$$

where

n = sample size,

P_{oi} = observed probability for class i ,

P_{ci} = computed probability for class i .

D. KS Test [1]

$$d = i \max |S_{oi} - F_{ci}| \quad (5)$$

where

S_{oi} = observed cumulative probability in class i,

F_{ci} = computed cumulative probability in class i.

E. Genetic Programming

The genetic programming GP, as proposed by John Koza [6], offers a tree structure for the usual chromosome. Instead of coding the solutions of the problem as a String, it is coded in GP as a tree.

The elements of the tree are called nodes .the nodes are classified into non-terminal nodes and terminal nodes.

- Non-terminal nodes (function nodes) consumes one or more inputs and provide one output, an example of function nodes are $\{+,-,*,./\}$
- Terminal nodes represent external inputs or constants. These are the leaves of the expression tree

The structure of the tree can be as shown in Fig 2.

The evolution in GP proceeds in similar way to standard GA i.e. an initial population is generated at random and each individual is evaluated its fitness value, then it is evolved by means of genetic operators as follows:

Selection: pairs of parent trees are selected based on its fitness for reproduction [6].

Crossover: another method of creating new members. This process mimics sexual reproduction in organisms. In sexual reproduction, combing elements of the parents' genes creates a new set of genes. In computer breeding, two population members exchange pieces of themselves to create two new programs. These new programs are then inserted into the population [6].

Mutation: works in much the same way as natural mutation does. In the program, a population member receives a random alteration. This process is similar to the random mutation of a DNA molecule in an organism. In this way, several mutations of surviving population members are created and inserted into the population.

These standard operators and other operators continue to work in cycles until either a predefined maximum number of generations are reached or a certain criteria are achieved. Fig 2 shows an example of the Tree Structure for a mathematical formula. GP utilizes computer programs concept, so that form, length and/or complexity of a possible solution has not to be specified in advance. Additionally, computer programs offer variability and dynamic, that exceeds that of fixed length bit-fields, conditional execution, and computations on variables of different types, iterations and recursions. Last, but not least, the functions used in these computer programs may be implemented in a certain ways that help solving the problem [6].

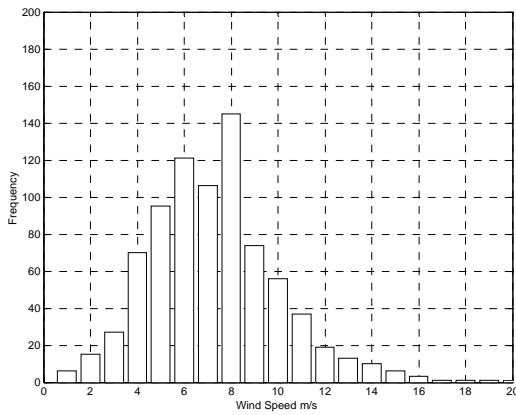
V. RESULTS AND DISCUSSION

Results obtained by applying the two goodness-of-fit criteria mentined in equation (4) and (5) to determine the best distribution fit (out of the three examined) for the data are presented and discussed in this section. The data are separately fitted to three distributions, GP, Weibull and β , all of these distributions were explained in the previous section

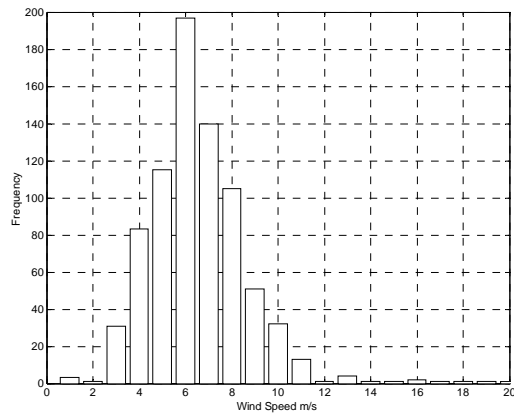
In Tables 1-2, we have summarized the results of the distribution fit based on Matlab statistical toolbox, for the data, in the order the test statistics came out, with 1 being the best. The goodness-of-fit tests applied were: (i) χ^2 test, (ii) KS test.

VI. CONCLUSION

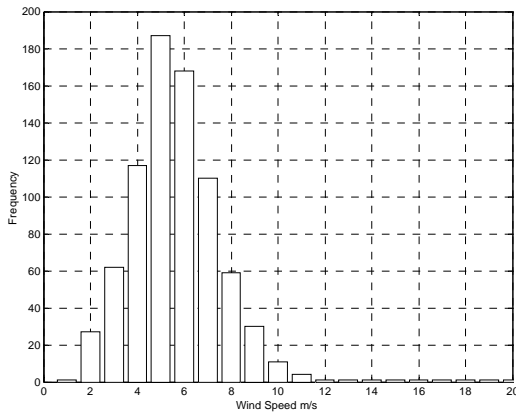
We conclude that the GP distribution represents the long-term hourly wind speed best. Both weibull and beta distribution parameters can easily be determined using the MATLAB statistics toolbox. GP can fit distributions better than weibull and beta function.



(a) March, 12 noon



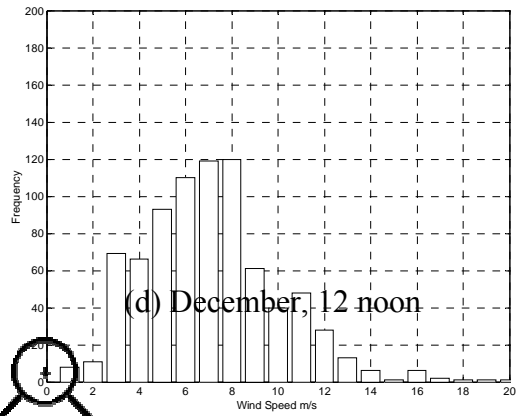
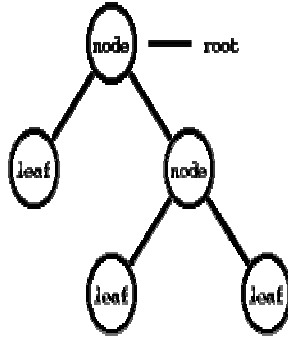
(c) September, 12 noon



(b) June, 12 noon

Fig.

1. Histograms for wind speeds



(d) December, 12 noon

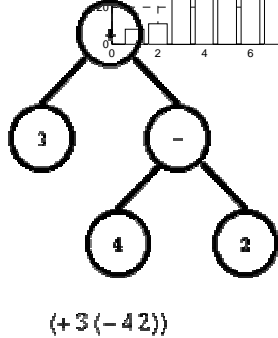


Fig. 2. GP Chromosome and tree

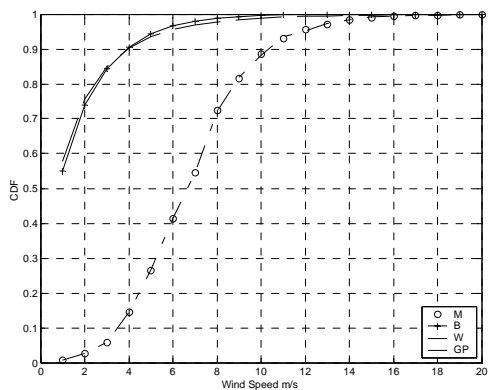
Order	3	6	9	12
17	-1.35e13	6.57e11	-8.15e12	-4.09e12
16	8.62e13	-4.16e12	5.20e13	2.65e13
15	-2.55e14	1.21e13	-1.53e14	-7.96e13
14	4.64e14	-2.19e13	2.79e14	1.47e14
13	-5.80e14	2.72e13	-3.50e14	-1.86e14
12	5.30e14	-2.46e13	3.20e14	1.72e14
11	-3.65e14	1.68e13	-2.21e14	-1.20e14
10	1.94e14	-8.83e12	1.17e14	6.48e13
9	-8.04e13	3.62e12	-4.88e13	-2.71e13
8	2.61e13	-1.1648e12	1.59e13	8.89e12
7	-6.63e12	2.93e11	-4.05e12	-2.28e12
6	1.31e12	-5.77e10	8.06e11	4.54e11
5	-1.99e11	8.75e09	-1.23e11	-6.97e10
4	2.30e10	-1.00e09	1.43e10	8.08e09
3	-1.95e09	8.57e07	-1.22e09	-6.88e08
2	1.17e08	-5.16e06	7.39e07	4.13e07

1	-4.62e06	205727	-2.94e06	-1.63e06
0	106207	-4771.56	68044.1	37592.9
KS	0.0	0.0	0.0	0.0
χ^2	0.0	0.0	0.0	0.0

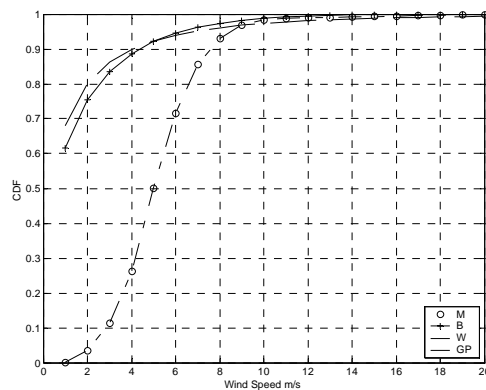
Dist	Month	Distribution Parameters		Goodness-of-fit-test	
		C	K	KS	χ^2
Beta	3	11.34	0.59	0.40	0.83
	6	7.07	0.37	0.25	0.66
	9	7.46	0.39	0.30	0.68
	12	11.40	0.59	0.40	0.82
Weibull	3	9.90	0.72	0.35	0.78
	6	6.33	0.50	0.25	0.55
	9	6.58	0.52	0.30	0.58
	12	9.89	0.72	0.35	0.78

Table 1: Distribution Fit Results

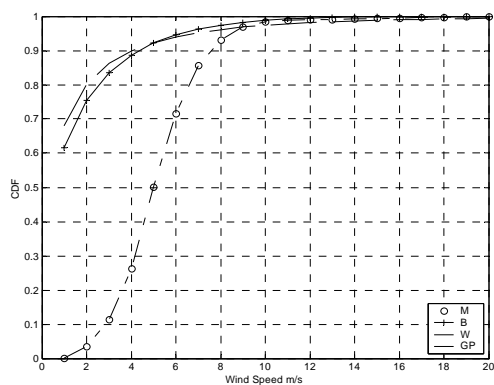
Table 2: GP Results



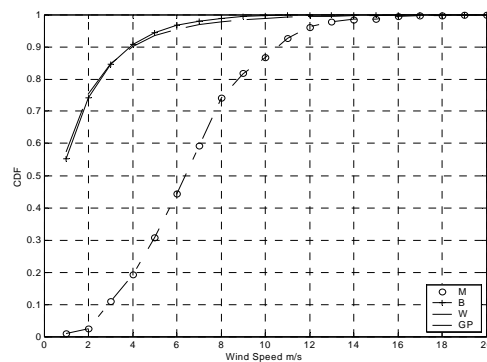
(a) June, 12 noon



(b) March, 12 noon



(c) September, 12 noon



(d) December, 12 noon

Fig. 3. Cumulative distribution Functions.

REFERENCES

- [1] R. Walpole and R. Myers, "Probability and statistics for engineers and scientists", Macmillan Publisher Co., 1993.
- [2] A. R. Abdelaziz and Z. M. Salameh, "A new statistical distribution function sensitive to renewable energy systems", International Journal of Electric Machines & Power Systems, vol.26, no. 6, July 1998, pp.659-667.
- [3] L. L. Freris, "Wind energy conversion systems", Prentice-Hall, 1990.
- [4] B. Bugdan, "Design and performance analysis of a stand-alone wind/photovoltaic hybrid system", doctoral dissertation, University of Massachusetts-Lowell, USA, 1996.
- [5] R. Billinton, H. Chen, and R. Ghajar, "A sequential simulation technique for adequacy evaluation of generating systems including wind energy", IEEE Transactions on Energy Conversion, pp. 728-734, 1996.
- [6] J.Koza, "Genetic programming: on the programming of computers by means of natural selection", Cambridge, MA, The MIT Press, 1992.