

RESEARCH ARTICLE

## Using genetic algorithms for estimating Weibull parameters with application to wind speed

Melih Burak Koca , Muhammet Burak Kılıç\* , Yusuf Şahin 

*Department of Business Administration, Burdur Mehmet Akif Ersoy University, Turkey*  
*mbkoca@mehmetakif.edu.tr, mburak@mehmetakif.edu.tr, ysahin@mehmetakif.edu.tr*

### ARTICLE INFO

#### Article history:

*Received: 31 October 2018*

*Accepted: 26 June 2019*

*Available Online: 31 January 2020*

#### Keywords:

*Weibull distribution*

*Genetic algorithms*

*Wind speed modeling*

*Parameter estimation*

AMS Classification 2010:

*62P30, 65C60*

### ABSTRACT

Renewable energy has become a prominent subject for researchers since fossil fuel reserves have been decreasing and are not promising to meet the energy demand of the future. Wind takes an important place in renewable energy resources and there is extensive research on wind speed modeling. Herein, one of the most commonly used distributions for wind speed modeling is the Weibull distribution with its simplicity and flexibility. Maximum likelihood (ML) method is the most frequently used technique in Weibull parameter estimation. Iterative techniques such as Newton-Raphson (NR) use random initial values to obtain the ML estimators of the parameters of the Weibull distribution. Therefore, the success of the iterative techniques highly depends on the initial value selection. In order to deliver a solution to the initial value problem, genetic algorithm (GA) is considered to obtain the estimators of the model parameters. The ML estimators obtained using the GA and NR techniques are compared with the method of moments (MoM) estimators via Monte Carlo simulation and wind speed applications. The results show that the ML estimators obtained using GA present superiority over MoM and the ML estimators obtained using NR.



### 1. Introduction

The increase in population and the inadequacy of existing energy resources put the human being into the search of alternative energy resources over the course of human history. In the last decades, there is extensive research on renewable energy due to the decrease in fossil fuel reserves and the increase in environmental awareness. As a clean and never-ending resource, the wind has become an important energy resource and distinguished among the other renewable energy forms such as geothermal energy, hydro energy, solar energy, and biomass energy.

Converting the kinetic energy carried by wind to electrical energy is a clean and economical way to produce energy. Once the wind plant is set up, the maintenance cost is relatively low compared to other energy plants. However, the wind turbines and installation costs are high, therefore the wind energy potential of a region should be carefully estimated to determine the proper turbine type. Wind speed is the key factor in determining the wind energy potential of a region [1-4]. Statistical distributions are used to model wind speed and estimate energy potential. The

Weibull distribution is one of the most commonly used distributions in wind energy studies due to its simplicity and flexibility [1, 2, 4-10].

There are various techniques used in Weibull parameter estimation. Sohoni et al. [2] estimated the Weibull parameters using the method of moments (MoM). Seguro and Lambert [5] employed MoM, maximum likelihood (ML) method and modified maximum likelihood (MML) methods. They found that the ML method is more appropriate for the data sets in time series format. For the data sets in frequency distribution format, they recommended using MML method. Akgül et al. [6] compared the least square method, ML method and MML method. Although they found that ML is the most efficient method in overall, they mentioned that ML and MML has a similar efficiency for the large data sets, however, MML has less computational complexity. Arslan et al. [8] compared MoM, L-Moments (L-Mom) method and ML method, and showed that L-Mom method is more efficient for small data sets where ML method is more efficient for larger data sets. Kaplan [10] found that graphical method provides more efficiency than MoM in Weibull

\*Corresponding author

parameter estimation. Kollu et al. [11], Akpınar and Akpınar [12] used the ML method to estimate Weibull parameters in their studies. Teimouri et al. [13] compared their proposed L-moment estimator with several methods including the ML method, method of logarithmic moment, percentile method and MoM. They found that their proposed method and the ML method are the most efficient estimators. Akdağ and Dinler [14] proposed the power density method. They found it superior to commonly used methods including MoM and ML method. Saleh et al. [15] compared five different methods and recommended the mean wind speed method and the ML method for fitting Weibull distribution. Azad et al. [16] found MoM and ML method more efficient among several methods. Recently, Usta et al. [17] proposed a new estimation approach based on moments for estimating the Weibull parameters.

It is seen from the previous studies that the ML method is one of the most frequently used parameter estimation methods for the Weibull distribution. Due to the nonlinear nature of the log-likelihood function of the Weibull distribution, numerical methods such as Newton-Raphson (NR) should be employed. However, when the iterative techniques are employed, the success of the technique highly depends on the initial value selection. This study departs from the literature by delivering a solution to the initial value problem by using genetic algorithms (GA), which is a heuristic search algorithm and uses a set of solution (search space) instead of single points, for ML estimation of the Weibull parameters. GA is a useful approach in the solution of optimization problems and applied in various studies such as signal control optimization [18] or optimization of mixture parameters of high-performance concrete [19]. In parameter estimation, GA was previously used for negative binomial gamma mixture distribution [20], skew-normal distribution [21] and nonlinear regression [22]. Parameter estimation of Weibull distribution using GA was introduced by Thomas et al. [23] for breakdown times of insulating fluid dataset. GA presented a comparable good performance based on the maximization of the log-likelihood function. With this motivation, the applicability of GA is used in wind speed data modeling. To the best our knowledge, this is the first time GA is used to estimate the parameters of Weibull distribution in wind speed distribution modeling. Observations were obtained from an existing wind farm and different meteorological stations. The efficiency of ML method estimation using GA was compared with ML estimation using NR, and MoM. Mean absolute error (MAE), bias and Kolmogorov-Smirnov (K-S) test were used as decision criteria. The remainder of this paper is structured as follows: Section 2 gives basic information about the Weibull distribution, Section 3 gives detailed information about the parameter estimation methods, Section 4 presents the simulation experiments and wind speed data analysis. Section 5 includes the conclusion.

## 2. Weibull distribution

The probability density function (pdf) and cumulative distribution function (cdf) of Weibull distribution are respectively given by:

$$f(v; k, c) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^k\right], \quad v, k, c > 0 \quad (1)$$

and

$$F(v; k, c) = 1 - \exp\left[-\left(\frac{v}{c}\right)^k\right], \quad v, k, c > 0 \quad (2)$$

where  $v$  is the wind speed,  $k$  and  $c$  are the Weibull shape and scale (dimensionless) parameters respectively. Probability density plots for some different parameter values are given in Figure 1.

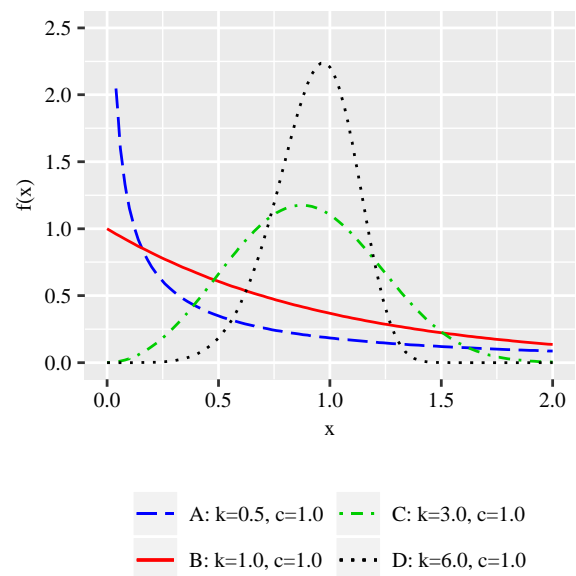


Figure 1. Probability density plots of the Weibull distribution for different parameters.

## 3. Parameter estimation methods

### 3.1. Method of moments estimation

MoM is based on equating sample moments with theoretical moments of respective distribution. To estimate the parameters of the Weibull distribution, coefficient of variation of the sample should be calculated and set equal to the theoretical coefficient of variation as follows [8]:

$$\widehat{CV}_{MoM} = \left[ \frac{(\sum_{i=1}^n v_i^2)n}{(\sum_{i=1}^n v_i)^2} - 1 \right] = \left[ \frac{\Gamma\left(\frac{2}{k} + 1\right)}{\left[\Gamma\left(\frac{1}{k} + 1\right)\right]^2} - 1 \right] \quad (3)$$

where  $n$  is the number of data points,  $\Gamma$  is the gamma function. When the shape parameter  $k$  is obtained from the Equation (3), scale parameter  $c$  can be calculated by:

$$\hat{c} = \left[ \frac{\frac{1}{n} (\sum_{i=1}^n v_i)}{\Gamma\left(\frac{1}{k} + 1\right)} \right]. \quad (4)$$

### 3.2. Maximum likelihood estimation

The ML method is based on the maximization of the log-likelihood function of the underlying distribution. The log-likelihood function of the Weibull distribution is given as follows:

$$\ln L(v; k, c) = n \ln k - nk \ln c + (k-1) \sum_{i=1}^n \ln v_i - c^{-k} \sum_{i=1}^n v_i^k. \quad (5)$$

By maximizing the log-likelihood function, taking derivative respect to each of the parameters and equating them to zero, the ML estimators of the shape and the scale parameters will be obtained as follows:

$$\hat{k} = \left[ \frac{\sum_{i=1}^n v_i^{\hat{k}} \ln v_i}{\sum_{i=1}^n v_i^{\hat{k}}} - \frac{\sum_{i=1}^n \ln v_i}{n} \right]^{-1} \quad (6)$$

and

$$\hat{c} = \left[ \frac{1}{n} \sum_{i=1}^n v_i^{\hat{k}} \right]^{\frac{1}{\hat{k}}}. \quad (7)$$

ML estimator of the shape parameter  $k$  includes nonlinear function, therefore, it can be solved by numerical techniques such as NR algorithm, Nelder-Mead algorithm, simulated annealing algorithm or GA. In this study, we used the NR algorithm and the GA in the maximization of the log-likelihood function given in Equation (5).

#### 3.2.1. Newton-Raphson algorithm

The steps of the NR algorithm are summarized in [21] as follows:

1. Determine the initial values  $k^{(0)}$  and  $c^{(0)}$  for  $k$  and  $c$ .
2. Compute the vector  $U(k^{(m)}, c^{(m)})$  and  $V(k^{(m)}, c^{(m)})$  for  $m = 0, 1, \dots$  where  $U$  and  $V$  are defined by:

$$U = \left( \frac{\partial \ln L}{\partial k}, \frac{\partial \ln L}{\partial c} \right)$$

and

$$V = \begin{bmatrix} \frac{\partial^2 \ln L}{\partial k^2} & \frac{\partial^2 \ln L}{\partial k \partial c} \\ \frac{\partial^2 \ln L}{\partial k \partial c} & \frac{\partial^2 \ln L}{\partial c^2} \end{bmatrix}.$$

3. Compute the values of  $k$  and  $c$  at  $(m+1)$ th iteration by using the following equation:

$$\begin{bmatrix} k^{(m+1)} \\ c^{(m+1)} \end{bmatrix} = \begin{bmatrix} k^{(m)} \\ c^{(m)} \end{bmatrix} - V^{-1}(k^{(m)}, c^{(m)}) U(k^{(m)}, c^{(m)})$$

4. Repeat the iterations until the convergence criterion is satisfied.

NR is a fast-converging powerful algorithm, however, it is dependent on the initial guess. Therefore, we considered the GA in the maximization of the log-likelihood function of the Weibull distribution.

#### 3.2.2. Genetic algorithm

GA is a heuristic search algorithm motivated by the principles of biological evolution of species, to obtain the estimators of the model parameters. Unlike the conventional optimization techniques, GA uses a set of initial solutions which are called as chromosome. A flowchart of GA is presented in Figure 2. The steps of the GA in this study are summarized as follows:

1. A range of possible solutions (search space) was defined as arbitrarily for both shape and scale parameters. A sensitivity analysis was carried out to determine the initial population size where it was taken 6, 10, 15, and 20 respectively. Most efficient outcomes were obtained when the initial population size was set to 6, therefore, initial population size was set to 6.
2. Each set of possible solutions is evaluated using the fitness function. The log-likelihood function of the Weibull distribution is the fitness function in this study.
3. The best solution in each iteration is kept as parent chromosome.
4. New offsprings are reproduced by crossover and mutation with the rate of 0.8 and 0.1 respectively. The size of the population including original parents, crossover and mutation offsprings is equal to the initial population size in step 1.
5. New population is evaluated as in step 2. Steps 3-5 are repeated.

The algorithm stops if the decision criterion is satisfied or the maximum number of iterations is achieved. A flowchart of the study is given in Figure 2.

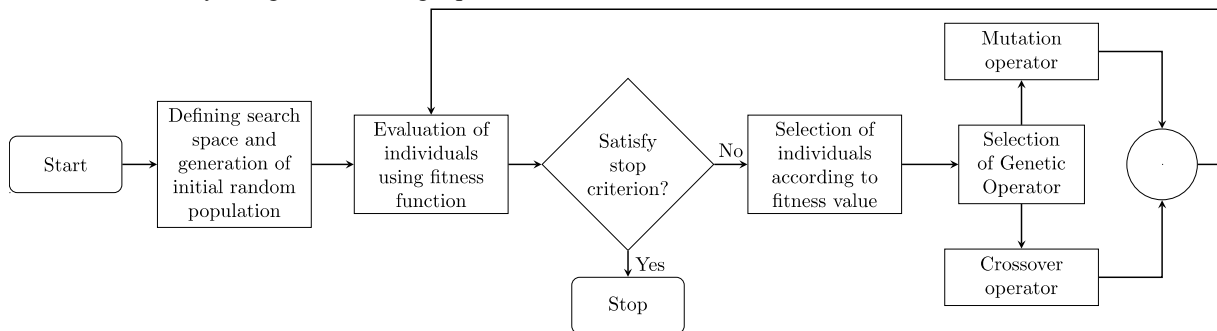


Figure 2. Flowchart of the GA used in this study.

**Table 1.** Parameter estimations, MAE and bias values for different simulation scenarios.

n	k	Method	$\hat{k}$			$\hat{c}$		
			Mean	MAE	Bias	Mean	MAE	Bias
20	0.5	MoM	0.6510	0.1637	0.1510	1.3549	0.5207	0.3549
		NR	0.5412	0.0840	0.0412	1.0823	0.3938	0.0823
		GA	0.5399	<b>0.0829</b>	<b>0.0399</b>	1.0570	<b>0.3685</b>	<b>0.0570</b>
	1	MoM	1.1172	0.1905	0.1172	1.0221	<b>0.1912</b>	0.0221
		NR	1.0823	0.1680	0.0823	1.0114	0.1915	0.0114
		GA	1.0822	<b>0.1679</b>	<b>0.0822</b>	1.0114	0.1915	<b>0.0114</b>
	3	MoM	3.2191	0.4729	0.2191	0.9975	0.0623	<b>-0.0025</b>
		NR	3.2470	0.5039	0.2470	0.9973	0.0640	-0.0027
		GA	3.1153	<b>0.3413</b>	<b>0.1153</b>	0.9949	<b>0.0622</b>	-0.0051
	6	MoM	6.4952	1.0341	0.4952	0.9977	0.0312	-0.0023
		NR	6.4939	1.0077	0.4939	0.9978	0.0321	<b>-0.0022</b>
		GA	6.2305	<b>0.6825</b>	<b>0.2305</b>	0.9967	<b>0.0312</b>	-0.0033
50	0.5	MoM	0.5818	0.1019	0.0818	1.2071	0.3412	0.2071
		NR	0.5172	0.0485	0.0172	1.0449	0.2514	0.0449
		GA	0.5171	<b>0.0484</b>	<b>0.0171</b>	1.0422	<b>0.2486</b>	<b>0.0422</b>
	1	MoM	1.0532	0.1144	0.0532	1.0105	<b>0.1221</b>	0.0105
		NR	1.0345	0.0971	0.0345	1.0101	0.1235	0.0101
		GA	1.0344	<b>0.0970</b>	<b>0.0344</b>	1.0097	0.1231	<b>0.0097</b>
	3	MoM	3.0866	0.2665	0.0866	0.9993	<b>0.0391</b>	-0.0007
		NR	3.1034	0.2912	0.1034	1.0007	0.0411	0.0007
		GA	3.0763	<b>0.2593</b>	<b>0.0763</b>	1.0000	0.0405	<b>0.0000</b>
	6	MoM	6.1854	0.5747	0.1854	0.9993	<b>0.0196</b>	-0.0007
		NR	6.2069	0.5824	0.2069	1.0000	0.0205	<b>0.0000</b>
		GA	6.1526	<b>0.5185</b>	<b>0.1526</b>	0.9997	0.0203	-0.0003
100	0.5	MoM	0.5495	0.0748	0.0495	1.1224	0.2469	0.1224
		NR	0.5086	0.0331	0.0086	1.0182	0.1675	0.0182
		GA	0.5085	<b>0.0330</b>	<b>0.0085</b>	1.0155	<b>0.1645</b>	<b>0.0155</b>
	1	MoM	1.0282	0.0839	0.0282	1.0034	0.0870	0.0034
		NR	1.0172	0.0662	0.0172	1.0034	0.0832	0.0034
		GA	1.0172	<b>0.0662</b>	<b>0.0172</b>	1.0034	<b>0.0832</b>	<b>0.0034</b>
	3	MoM	3.0470	0.1915	0.0470	0.9992	0.0280	-0.0008
		NR	3.0516	0.1985	0.0516	0.9999	0.0277	<b>-0.0001</b>
		GA	3.0112	<b>0.1808</b>	<b>0.0112</b>	0.9860	<b>0.0140</b>	-0.0140
	6	MoM	6.0963	0.4120	0.0963	0.9994	0.0140	-0.0006
		NR	6.1032	0.3971	0.1032	0.9998	0.0139	<b>-0.0002</b>
		GA	6.0939	<b>0.3868</b>	<b>0.0939</b>	0.9997	<b>0.0138</b>	-0.0003
500	0.5	MoM	0.5151	0.0369	0.0151	1.0339	0.1245	0.0339
		NR	0.5015	0.0140	0.0015	1.0029	0.0752	0.0029
		GA	0.5015	<b>0.0140</b>	<b>0.0015</b>	1.0029	<b>0.0752</b>	<b>0.0029</b>
	1	MoM	1.0063	0.0366	0.0063	0.9974	0.0387	-0.0026
		NR	1.0031	0.0280	0.0031	1.0003	0.0376	0.0003
		GA	1.0031	<b>0.0280</b>	<b>0.0031</b>	1.0003	<b>0.0376</b>	<b>0.0003</b>
	3	MoM	3.0090	0.0877	<b>0.0090</b>	0.9986	0.0126	-0.0014
		NR	3.0092	0.0840	0.0092	0.9999	0.0125	-0.0001
		GA	3.0092	<b>0.0840</b>	0.0092	0.9999	<b>0.0125</b>	<b>-0.0001</b>
	6	MoM	6.0193	0.1887	0.0193	0.9993	0.0063	-0.0007
		NR	6.0184	0.1680	0.0184	0.9999	0.0063	-0.0001
		GA	6.0176	<b>0.1666</b>	<b>0.0176</b>	0.9999	<b>0.0062</b>	<b>-0.0001</b>

**4. Application**

**4.1. Monte Carlo simulations**

In order to compare the parameter estimation methods for the Weibull distribution, a Monte Carlo simulation was conducted where the shape parameter is taken 0.5, 1, 3 and 6 and the scale parameter was fixed to 1. The

parameter sets used in the simulation can also be seen in Figure 1. The simulation was repeated 1000 times for each of the sample sizes of 20, 50, 100 and 500. MoM estimations were considered as the initial values for the NR. For the GA, the population size was chosen 6, mutation rate and crossover rate were fixed to 0.8 and 0.1 respectively. ML estimations using NR and GA were obtained via “maxLik” [24] and “GA” [25]

packages of R software. Mean absolute error (MAE) and bias are chosen as goodness-of-fit criteria for comparing the efficiencies of the parameter estimation methods. MAE and bias for the parameters  $k$  and  $c$  are given by:

$$MAE(\hat{k}) = \frac{1}{n} \sum_{i=1}^n |\hat{k}_i - k| \tag{8}$$

$$bias(\hat{k}) = \frac{1}{n} \sum_{i=1}^n (\hat{k}_i - k)$$

and

$$MAE(\hat{c}) = \frac{1}{n} \sum_{i=1}^n |\hat{c}_i - c| \tag{9}$$

$$bias(\hat{c}) = \frac{1}{n} \sum_{i=1}^n (\hat{c}_i - c).$$

Smaller values the absolute value of the bias and MAE indicate higher efficiency. Parameter estimations, absolute value of the bias and MAE for each parameter estimation method can be seen in Table 1. Accordingly,

best results are highlighted in bold.

It is seen from the simulation results that the GA approach was more efficient than NR and MoM in the estimation of the shape and scale parameters according to MAE and bias criteria. For the sample size of 20, 50 and 100, the GA approach provided the best efficiency for the shape parameter in each simulation scenario in terms of MAE and bias. For the sample size of 500, GA also provided the best efficiency for the shape parameter in each simulation scenario according to MAE.

In the estimation of scale parameter for the sample sizes of 20,50 and 100, GA provided the highest efficiency according to at least one of the decision criteria in almost each simulation scenario. For the sample size of 100, GA was the most efficient method in each simulation scenario according to MAE and bias. In overall, it can be said that GA is a very efficient method for small, moderate and large sample sizes. MAE and absolute values of the biases are also presented in Figures 3-6.

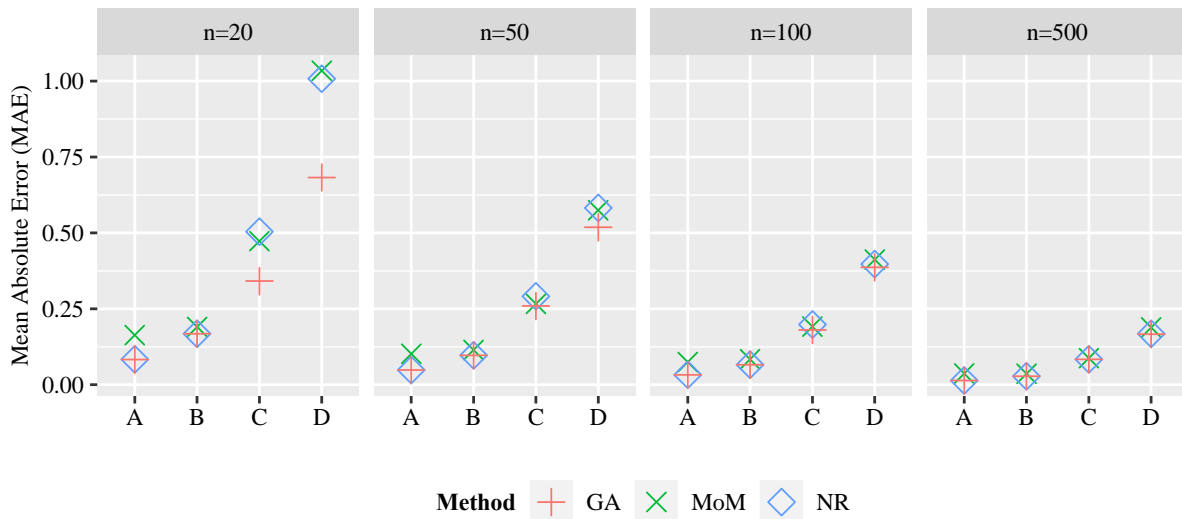


Figure 3. Comparison of the parameter estimation methods for  $k$  according to MAE criterion.

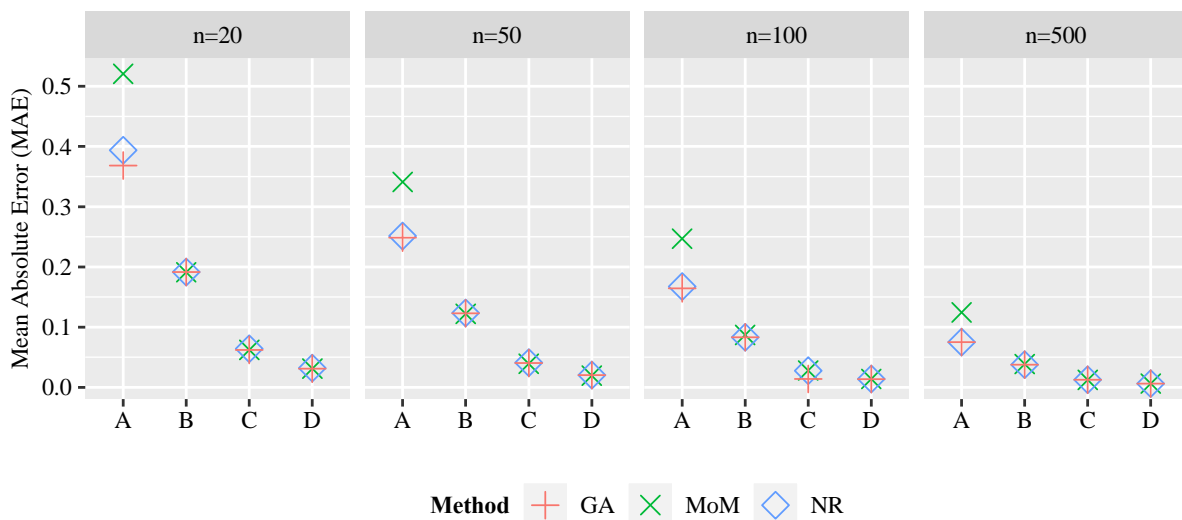


Figure 4. Comparison of the parameter estimation methods for  $c$  according to MAE criterion.

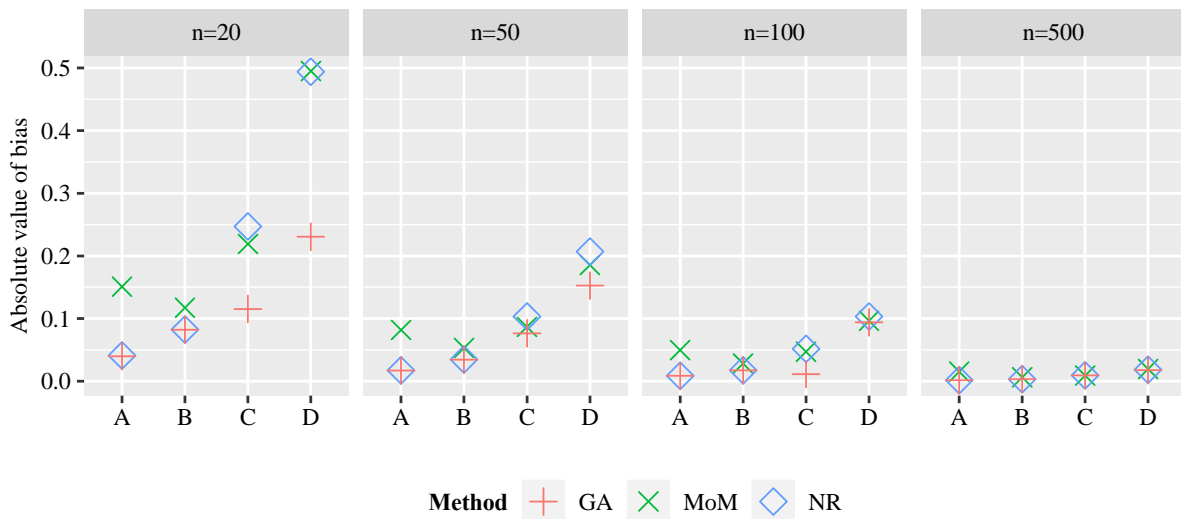


Figure 5. Comparison of the parameter estimation methods for  $k$  according to bias criterion

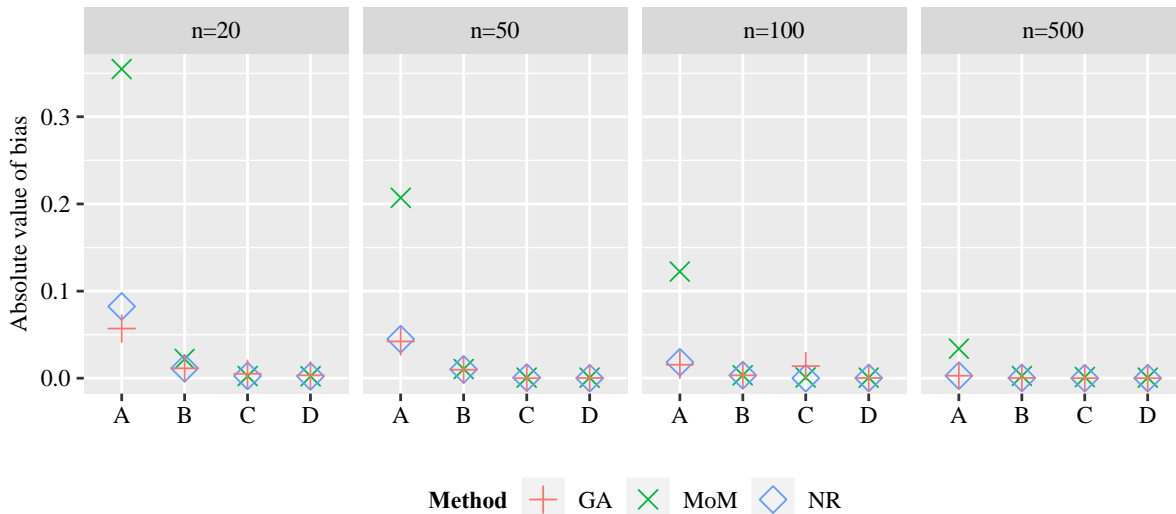


Figure 6. Comparison of the parameter estimation methods for  $c$  according to bias criterion.

Figure 3 presents the MAE values for the shape parameter  $k$ . GA presented more efficiency than NR and MoM in all simulation scenarios. NR was the second-best method. MAE values were decreased when the sample size was increased. However, when the value of the shape parameter was increased, MAE values were also increased.

Figure 4 shows the MAE values for the scale parameter  $c$ . GA was the most efficient method for the sample sizes of 20, 100 and 500. MoM was the most efficient for the sample size of 50. MAE values were decreased when the value of the shape parameter was increased. Similarly, MAE values were also decreased when the sample size was increased.

Figure 5 presents the absolute value of bias for the shape parameter  $k$ . GA presented the most efficient results. MoM presented better results than NR on some occasions. Similar to the MAE values, absolute values of the bias were decreased when the sample size was increased. However, when the value of the shape

parameter was increased, the absolute values of the bias were also increased

Figure 6 shows the absolute values of bias for the scale parameter  $c$ . GA was more efficient than other methods for most of the time. NR was the second-best method. With the increase in the value of shape parameter and sample size, absolute values of bias were decreased.

#### 4.2. Wind speed analysis

Wind speed observations obtained from three different locations, namely Belen Wind Farm (Belen), Gökçeada Meteorological Station (Gökçeada) and Datça/Deveboynu Feneri Meteorological Station (Datça) were used for the comparison of the parameter estimation methods. Belen data set was provided by Belen Electric Generation Co. Inc. Gökçeada and Datça data sets were provided by the Turkish State Meteorological Service. Information about the geographical coordinates of the stations, elevation, selected period of observations and collection process

are presented in Table 2. Accordingly, the wind speed data for the Belen Station were observed in 10-min basis. The wind speed data observed at other stations were collected on hourly basis. The descriptive statistics including mean, standard deviation, minimum

and maximum for the data sets used in this study are presented in Table 3. It can be seen from Table 3 that the average and the maximum wind speed were observed at Datça station.

**Table 2.** Geographical coordinates of the stations, selected period of observations and data collection process.

Station	Period of Observations	Collection Basis	Height	Latitude	Longitude	Elevation
Belen	01 Jan 2013 – 31 Dec 2014	10-min	80 m	36°28'42.2"N	36°12'45.0"E	744
Gökçeada	01 Jan 2010 – 31 Dec 2017	Hourly	10 m	40°11'27.6"N	25°54'27.0"E	79
Datça	01 Jan 2014 – 30 Apr 2017	Hourly	10 m	36°41'12.1"N	27°21'47.9"E	28

**Table 3.** Descriptive statistics for the data sets.

Station	Year	Mean	Std. Dev.	Min.	Max.
Belen	2013	7.2949	3.3345	0.4	24.9
	2014	7.3790	3.2631	0.4	24.9
Gökçeada	2010	4.3956	3.0779	0.1	16.9
	2011	4.5014	2.7362	0.1	16.6
	2012	4.1776	2.6393	0.2	14.9
	2013	3.6880	2.4716	0.1	16.9
	2014	3.7678	2.5150	0.1	14.9
	2015	4.3323	2.6961	0.1	18.0
	2016	4.2892	2.8234	0.1	16.7
Datça	2017	3.8563	2.6384	0.1	15.4
	2014	6.9937	4.5638	0.3	28.9
	2015	7.6496	4.9330	0.4	28.9
	2016	7.8554	5.1781	0.2	33.7
	2017	7.3715	4.8864	0.2	24.2

Weibull distribution is fitted at the monthly base for the Belen, Gökçeada and Datça data sets. To statistically test that monthly data sets come from Weibull distribution, the K-S test is separately applied to each data set.

K-S test is used for testing if a sample distribution belongs to a population with a specific distribution. K-S test statistic is the maximum difference between the empirical distribution  $F_0(x)$  and theoretical distribution  $S_N(x)$  [26].

$$d = \max |F_0(x) - S_N(x)| \quad (10)$$

After the K-S test process, monthly distributions that come from Weibull distribution are selected for further analysis ( $p\text{-value} > 0.05$ ). The parameter estimates and K-S test results for Belen, Gökçeada and Datça data sets are presented in Tables 4-6 respectively.

**Table 4.** Parameter estimations and K-S goodness-of-fit test results for Belen data set.

Date	Method	$\hat{k}$	$\hat{c}$	K-S	p-value
2014 - Feb	MoM	2.1104	6.7724	0.0182	0.1381
	NR	2.1191	6.7779	0.0197	0.0875
	GA	2.0984	6.7646	<b>0.0161</b>	<b>0.2438</b>

Table 4 shows that GA provides the best fit in terms of the K-S test for Belen data set.

**Table 5.** Parameter estimations and K-S goodness-of-fit test results for Gökçeada data set.

Date	Method	$\hat{k}$	$\hat{c}$	K-S	p-value
2010 - Oct	MoM	1.5527	4.8650	0.0443	0.1074
	NR	1.5571	4.8752	0.0453	0.0947
	GA	1.4954	4.8088	<b>0.0394</b>	<b>0.1990</b>
2011 - Apr	MoM	1.9337	5.7312	0.0517	0.0423
	NR	1.9002	5.7122	0.0501	0.0537
	GA	1.9023	5.7247	<b>0.0485</b>	<b>0.0673</b>
2011 - Nov	MoM	1.7368	5.6574	0.0363	0.2998
	NR	1.7307	5.6551	0.0363	0.3004
	GA	1.7299	5.6507	<b>0.0359</b>	<b>0.3144</b>
2012 - Dec	MoM	1.8973	5.4681	0.0440	0.1123
	NR	1.8750	5.4574	0.0414	0.1562
	GA	1.8737	5.4526	<b>0.0411</b>	<b>0.1615</b>
2013 - May	MoM	1.6149	4.4635	0.0463	0.0829
	NR	1.6212	4.4682	0.0472	0.0724
	GA	1.6148	4.4237	<b>0.0449</b>	<b>0.1001</b>
2015 - Jan	MoM	1.5037	5.6700	0.0611	0.0401
	NR	1.5681	5.7215	0.0591	0.0518
	GA	1.5673	5.7129	<b>0.0584</b>	<b>0.0562</b>
2015 - Feb	MoM	1.7919	6.6341	0.0480	0.0982
	NR	1.7502	6.6019	0.0484	0.0937
	GA	1.7509	6.6086	<b>0.0478</b>	<b>0.1016</b>
2015 - Apr	MoM	1.4922	5.0242	0.0370	0.3096
	NR	1.4876	5.0285	0.0365	0.3268
	GA	1.4863	5.0104	<b>0.0358</b>	<b>0.3493</b>
2015 - May	MoM	1.6151	4.0105	0.0524	0.0919
	NR	1.5916	4.0010	0.0488	0.1374
	GA	1.5905	3.9947	<b>0.0484</b>	<b>0.1435</b>
2015 - Oct	MoM	1.9727	4.9659	0.0514	0.0485
	NR	1.9312	4.9455	0.0487	0.0711
	GA	1.9301	4.9435	<b>0.0486</b>	<b>0.0721</b>
2015 - Nov	MoM	1.5010	5.2991	<b>0.0329</b>	<b>0.6387</b>
	NR	1.5066	5.3106	0.0343	0.5861
	GA	1.5034	5.3074	0.0338	0.6027
2016 - Jan	MoM	1.3632	5.1459	<b>0.0432</b>	<b>0.1321</b>
	NR	1.3710	5.1638	0.0453	0.1004
	GA	1.3683	5.1604	0.0447	0.1089
2016 - Mar	MoM	1.5731	5.4328	0.0466	0.1599
	NR	1.5608	5.4299	0.0448	0.1946
	GA	1.5582	5.4213	<b>0.0441</b>	<b>0.2078</b>
2016 - May	MoM	1.6266	4.0770	0.0457	0.3642
	NR	1.6111	4.0731	0.0430	0.4399
	GA	1.6109	4.0723	<b>0.0430</b>	<b>0.4422</b>

**Table 5.** (continued)

2016 - Jun	MoM	1.8951	3.0578	<b>0.0877</b>	<b>0.1657</b>
	NR	1.8507	3.0452	0.0947	0.1096
	GA	1.8535	3.0461	0.0942	0.1128
2016 - Nov	MoM	1.5976	5.3758	0.0532	0.0355
	NR	1.5571	5.3534	0.0474	0.0810
	GA	1.5569	5.3533	<b>0.0474</b>	<b>0.0814</b>
2017 - Dec	MoM	1.9966	6.8246	0.0516	0.1153
	NR	1.9906	6.8205	0.0513	0.1186
	GA	1.9909	6.8219	<b>0.0513</b>	<b>0.1197</b>

**Table 6.** Parameter estimations and K-S goodness-of-fit test results for Datça data set.

Date	Method	$\hat{k}$	$\hat{c}$	K-S	p-value
2014 - Jan	MoM	1.7162	6.0504	0.0570	0.0756
	NR	1.7787	6.0924	0.0566	0.0790
	GA	1.7672	5.9995	<b>0.0490</b>	<b>0.1783</b>
2014 - Aug	MoM	1.8009	8.4971	<b>0.0446</b>	<b>0.1031</b>
	NR	1.8379	8.5336	0.0461	0.0842
	GA	1.8373	8.5273	0.0457	0.0898
2015 - Jan	MoM	1.7307	10.8387	0.0431	0.1255
	NR	1.7483	10.8725	0.0437	0.1172
	GA	1.7501	10.9125	<b>0.0422</b>	<b>0.1408</b>
2015 - Mar	MoM	1.8631	9.6927	0.0409	0.1662
	NR	1.8842	9.7204	0.0418	0.1491
	GA	1.8683	9.7017	<b>0.0409</b>	<b>0.1660</b>
2015 - Jun	MoM	1.8422	7.4869	<b>0.0348</b>	<b>0.3482</b>
	NR	1.8781	7.5196	0.0418	0.1619
	GA	1.8051	7.5222	0.0363	0.2982
2015 - Oct	MoM	2.0635	8.5650	0.0463	0.0828
	NR	2.0843	8.5813	0.0493	0.0539
	GA	2.0452	8.5486	<b>0.0435</b>	<b>0.1198</b>
2015 - Nov	MoM	1.8870	9.2486	0.0440	0.1237
	NR	1.9013	9.2654	0.0467	0.0863
	GA	1.8814	9.2433	<b>0.0430</b>	<b>0.1398</b>
2015 - Dec	MoM	2.1032	10.2284	0.0383	0.2301
	NR	2.1241	10.2526	0.0407	0.1748
	GA	2.1181	10.2101	<b>0.0372</b>	<b>0.2585</b>
2016 - Jan	MoM	1.8773	10.8451	0.0279	0.6078
	NR	1.8728	10.8427	0.0278	0.6138
	GA	1.8451	10.8504	<b>0.0257</b>	<b>0.7112</b>
2016 - Feb	MoM	1.2831	9.1734	0.0370	0.3162
	NR	1.3318	9.2773	0.0442	0.1439
	GA	1.2793	9.1544	<b>0.0359</b>	<b>0.3519</b>
2016 - May	MoM	1.6948	8.4204	0.0314	0.4729
	NR	1.7220	8.4530	0.0365	0.2905
	GA	1.7033	8.3964	<b>0.0308</b>	<b>0.4997</b>
2016 - Sep	MoM	1.6449	8.8944	0.0478	0.0829
	NR	1.6563	8.9189	0.0505	0.0576
	GA	1.6342	8.7143	<b>0.0433</b>	<b>0.1469</b>
2016 - Oct	MoM	1.7423	9.4553	0.0403	0.1784
	NR	1.7605	9.4816	0.0442	0.1096
	GA	1.7254	9.4345	<b>0.0368</b>	<b>0.2652</b>
2017 - Mar	MoM	1.6872	7.9891	0.0400	0.1956
	NR	1.7152	8.0254	0.0425	0.1457
	GA	1.6913	8.0015	<b>0.0399</b>	<b>0.1990</b>

It can be seen from Table 5 that GA provides the highest efficiency in 14 of 17 months in terms of the K-S test results in Gökçeada data set. MoM provides the best fit in 3 months.

Table 6 shows that GA provides the best fit in 12 of 14 months. MoM is the second-best estimator and has the highest efficiency in 2 months for Datça dataset.

## 5. Conclusion

In this paper, we have obtained the ML estimators of the parameters of Weibull distribution using GA and NR techniques, and compared them with MoM. The efficiencies of the parameter estimation methods are evaluated based on MAE, bias and K-S test criteria. Results of the Monte Carlo simulation and real wind speed data analysis show that ML estimator using GA is more efficient than ML estimator using NR and MoM estimator in Weibull parameter estimation. Furthermore, it can be said that all data sets were observed in different geographical regions with different weather characteristics. GA showed superiority on these data sets including different types of weather conditions. Finally, arbitrary search spaces were used in this study which can be seen as a limitation. In the future works, we will focus on developing a data-based search space in GA for Weibull parameter estimation.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments. This study was based on an ongoing doctoral thesis by the first author. This work was supported by the Scientific Research Projects Coordination Unit of Burdur Mehmet Akif Ersoy University. Project number 0440-DR-17.

## References

- [1] Hou, Y., Peng, Y., Johnson, A.L. & Shi, J. (2012). Empirical analysis of wind power potential at multiple heights for North Dakota wind observation sites. *Energy Science and Technology*, 4(1), 1-9. DOI:10.3968/j.est.1923847920120401.289
- [2] Sohoni, V., Gupta, S. & Nema, R. (2016). A comparative analysis of wind speed probability distributions for wind power assessment of four sites. *Turkish Journal of Electrical Engineering & Computer Sciences*, 24(6), 4724-4735. DOI:10.3906/elk-1412-207
- [3] Turkan, Y.S., Aydogmus, H.Y. & Erdal, H. (2016). The prediction of the wind speed at different heights by machine learning methods. *An International Journal of Optimization and Control: Theories & Applications*, 6(2), 179-197. DOI:10.11121/ijocta.01.2016.00315
- [4] Koca, M.B., Kılıç, M.B. & Şahin, Y. Assessing wind energy potential using finite mixture distributions. *Turkish Journal of Electrical Engineering & Computer Sciences*, 27(3), 2276-2294. DOI:10.3906/elk-1802-109




- [5] Seguro, J.V., & Lambert, T.W. (2000). Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *Journal of Wind Engineering and Industrial Aerodynamics*, 85, 75-84. DOI:10.1016/S0167-6105(99)00122-1
- [6] Akgül, F.G., Şenoğlu, B., & Arslan, T. (2016). An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution. *Energy Conversion and Management*, 114, 234-240. DOI:10.1016/j.enconman.2016.02.026
- [7] Lun, I.Y.F., & Lam, J.C. (2000). A study of Weibull parameters using long-term wind observations. *Renewable Energy*, 20, 145-153. DOI:10.1016/S0960-1481(99)00103-2
- [8] Arslan, T., Bulut, Y.M., & Yavuz, A.A. (2014). Comparative study of numerical methods for determining Weibull parameters for wind speed modeling. *Renewable and Sustainable Energy Reviews*, 40, 820-825. DOI:10.1016/j.ser.2014.08.009
- [9] Safari, B. (2011). Modeling wind speed and wind power distributions in Rwanda. *Renewable and Sustainable Energy Reviews*, 15, 925-935. DOI:10.1016/j.ser.2010.11.001
- [10] Kaplan, Y.A. (2016). The evaluating of wind energy potential of Osmaniye region with using Weibull and Rayleigh distributions. *Süleyman Demirel University Journal of Natural and Applied Sciences*, 20(1), 62-71. DOI:10.19113/sdufbed.63806
- [11] Kollu, R., Rayapudi, S.R., Narasimham, S.V.L., & Pakkurthi, K.M. (2012). Mixture probability distribution functions to model wind speed distributions. *International Journal of Energy and Environmental Engineering*, 3(27). DOI:10.1186/2251-6832-3-27
- [12] Akpınar, E.K., & Akpınar, S. (2004). Determination of the wind energy potential for Maden-Elazığ, Turkey. *Energy Conversion and Management*, 45, 2901-2914. DOI:10.1016/j.enconman.2003.12.016
- [13] Teimouri, M., Hoseini, S.M., & Nadarajah, S. (2013). Comparison of estimation methods for the Weibull distribution. *Statistics*, 47(1), 93-109. DOI:10.1080/02331888.2011.559657
- [14] Akdağ, S.A. & Dinler, A. (2009). A new method to estimate Weibull parameters for wind energy applications. *Energy Conversion and Management*, 50, 1761-1766. DOI:10.1016/j.enconman.2009.03.020
- [15] Saleh, H., Abou El-Azm Aly, A. & Abdel-Hady, S. (2012). Assessment of different methods used to estimate Weibull distribution parameters for wind speed in Zafarana wind farm, Suez Gold, Egypt. *Energy*, 44, 710-719. DOI:10.1016/j.energy.2012.05.021
- [16] Azad, A.K., Rasul, M. G. & Yusaf, T. (2014). Statistical diagnosis of the best Weibull methods for wind power assessment for agricultural applications. *Energies*, 7, 3056-3085. DOI:10.3390/en7053056
- [17] Usta, I., Arik, I., Yenilmez, I. & Kantar, Y.M. (2018). A new estimation approach based on moments for estimating Weibull parameters in wind speed power applications. *Energy Conversion and Management*, 164, 570-578. DOI:10.1016/j.enconman.2018.03.033
- [18] Tu, T.V. & Sano, K. (2013). Genetic algorithm for optimization in adaptive bus signal priority control. *An International Journal of Optimization and Control: Theories & Applications*, 3(1), 35-43. DOI:10.11121/ijocta.01.2013.00138
- [19] Şimşek, B. & Şimşek, E.H. (2017). Assessment and optimization of thermal and fluidity properties of high strength concrete via genetic algorithm. *An International Journal of Optimization and Control: Theories & Applications*, 7(1), 90-97. DOI:10.11121/ijocta.01.2017.00345
- [20] Gençtürk, Y., & Yiğiter, A. (2016). Modelling claim number using a new mixture model: negative binomial gamma distribution. *Journal of Statistical Computation and Simulation*, 86, 1829-1839. DOI:10.1080/00949655.2015.1085987
- [21] Yalçınkaya, A., Şenoğlu, B., & Yolcu, U. (2018). Maximum likelihood estimation for the parameters of skew normal distribution using genetic algorithm. *Swarm and Evolutionary Computation*, 38, 127-138. DOI:10.1016/j.swevo.2017.07.007
- [22] Altunkaynak, B., & Esin, A. (2004). The genetic algorithm method for parameters estimation in nonlinear regression. *Gazi University Journal of Science*, 17(2), 43-51.
- [23] Thomas, G.M., Gerth, R., Velasco, T., & Rabelo, L.C. (1995). Using real-coded genetic algorithms for Weibull parameter estimation. *Computers & Industrial Engineering*, 29, 377-381. DOI:10.1016/0360-8352(95)00102-7
- [24] Henningsen, A., & Toomet, O. (2010). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26, 443-458. DOI: 10.1007/s00180-010-0217-1
- [25] Scrucca, L. (2013). GA: A package for genetic algorithms in R. *Journal of Statistical Software*, 53(4), 1-37. DOI: 10.18637/jss.v053.i04
- [26] Massey JR, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of American Statistical Association*, 46, 68-78. DOI:10.1080/01621459.1951.10500769

**Melih Burak Koca** received his M.S. degree in Applied Statistics from Purdue University in 2014. He is currently doing his Ph.D. in Quantitative Methods at Burdur Mehmet Akif Ersoy University.

 <http://orcid.org/0000-0003-2737-3643>

**Muhammet Burak Kılıç** received his M.S. degree in Statistics from Fırat University in 2011 and his Ph.D. in Statistics from Middle East Technical University in 2015. His research interests are directional statistics, Bayesian

models, and computational methods.

 <http://orcid.org/0000-0002-9597-1576>

**Yusuf Şahin** received the M.S. degree in industrial engineering from Pamukkale University in 2009 and a PhD degree in business administration from Suleyman Demirel University in 2014. He has been an assistant professor of business administration at Burdur Mehmet Akif Ersoy University since 2014. His field of study includes operations research, logistics, warehouse management, vehicle routing, meta-heuristics, and quantitative models.

 <http://orcid.org/0000-0002-3862-6485>

An International Journal of Optimization and Control: Theories & Applications (<http://ijocta.balikesir.edu.tr>)



This work is licensed under a Creative Commons Attribution 4.0 International License. The authors retain ownership of the copyright for their article, but they allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in IJOCTA, so long as the original authors and source are credited. To see the complete license contents, please visit <http://creativecommons.org/licenses/by/4.0/>.