A Performance Comparison Study with Information Criteria for MaxEnt Distributions

Ozer OZDEMIR and Aslı KAYA

Abstract— In statistical modeling, the beginning problem that has to be solved is the parameter estimation of the empirical distribution which fits the data appropriately, and then decision can be given by making goodness of fit test or using error measures to determine the accuracy of the predicted model. One of the frequently used measures for such purposes is Mean Square Error (MSE) which formula is defined as the difference between the actual observations and the response predicted by the model and is used to determine with how much error the model fit the data, Root Mean Square Error (RMSE), Mean Absolute Proportion Error (MAPE) etc. measures can be also used for the similar aims. According to this information; firstly, a simulation study is prepared. Then, maximum entropy (MaxEnt) distributions for the error distribution of data is demonstrated and the best fitted distribution is found based on the information criterion

Keywords—Information criteria, MaxEnt distributions, Monte Carlo simulation, Performance measures.

I. INTRODUCTION

In statistical modeling, the MSE is defined as the difference between the actual observations and the response predicted by the model and is used to determine whether the model does not fit the data or whether the model can be simplified by removing terms. MSE, RMSE and MAPE are defined as follow:

- mean absolute error : $\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right|$ (1)
- mean squared error : $\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$ (2)
- root mean squared error : $\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$ (3)

There are some important studies by using performance measures as MSE, RMSE and MAPE. Allen (2012) has proposed the MSE of prediction as a criterion for selecting variables [5]. Willmott and Matsuura (2005) have suggested that the RMSE is not a good indicator of average model performance and might be a misleading indicator of average error [6, 7]. Chai and Draxler (2014) have discussed some circumstances where using the RMSE will be more beneficial [7]. Goodwin and Lawton (1999) have proposed the use of a symmetric (or modified) MAPE [8]. Myttenaere et al. (2016) have studied the consequences of using the MAPE as a measure of quality for regression models [9].

Such criterions as MSE are important in order to compare the models to find the best one that fit the data. From this point of view, in this study information based statistical error criterion is proposed. Firstly Monte Carlo simulation study is designed and showed that the proposed criterion is satisfied. Then, MaxEnt distributions for the error distribution of gun shot angles is demonstrated and the best fitted distribution is found based on the information criterion. Finally it is indicated that it gives similar decision as with MSE. There are some important studies by using MaxEnt distributions.

Ishwar and Moulin (2005) have derived an analytical characterization of the MaxEnt distribution by using results from the theory of constrained optimization in infinite-dimensional normed linear spaces [10]. Philips et al. (2006) have studied maximum entropy modeling of species geographic distributions [11]. Sankaran and Zabaras (2006) have suggested a maximum entropy approach for property prediction of random microstructures [12].

II. MONTE CARLO SIMULATION

Since the simulation process involves generating chance variables and exhibits random behaviours, it has been called Monte Carlo simulation. Monte Carlo simulation is a type of simulation that relies on repeated random sampling and statistical analysis to compute the results. This method of simulation is very closely related to random experiments, experiments for which the specific result is not known in advance. In this context, Monte Carlo simulation can be considered as a methodical way of doing so-called what-if analysis [13].

Monte Carlo simulation, or probability simulation, is a powerful statistical analysis technique used to understand the impact of risk and uncertainty in financial, project management, cost, and other forecasting models. It was initially used to solve neutron diffusion problems in atomic bomb work at Alamos Scientific Laboratory in 1944. Different from a physical experiment, Monte Carlo simulation performs random sampling and conducts a large number of experiments on compute.

Ozer OZDEMIR is with the Department of Statistics, Anadolu University Eskisehir 26470 TURKEY (corresponding author to provide phone: 902223350580-4668; e-mail: ozerozdemir@anadolu.edu.tr).

Aslı KAYA is with the Department of Statistics, Anadolu University Eskisehir 26470 TURKEY

The outline of Monte Carlo simulation is depicted in Fig.1. Three steps are required in the simulation process: Step 1 – sampling on random input variables X, Step 2 – evaluating model output Y, and Step 3 – statistical analysis on model output.

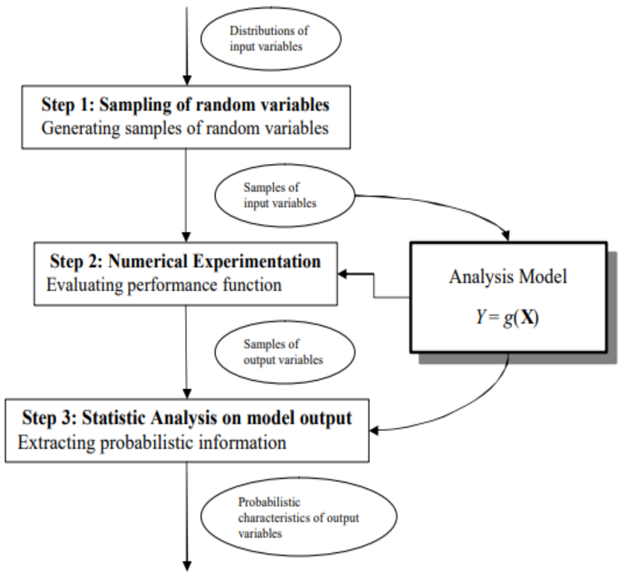

Fig.1. Monte Carlo simulation

In a Monte Carlo simulation, a random value is selected for each of the tasks, based on the range of estimates. The model is calculated based on this random value. The result of the model is recorded, and the process is repeated. A typical Monte Carlo simulation calculates the model hundreds or thousands of times, each time using different randomly-selected values. When the simulation is complete, we have a large number of results from the model, each based on random input values. These results are used to describe the likelihood, or probability, of reaching various results in the model [14].

III. Information Based Statistical Error Criterion

Information is defined as the variation of entropies between two systems. As much the variation of system's entropy is less so much is the loss or gain of information about the system. In other words, let choose a statistical sample as a system. You can compute the entropy of the actual sample and also the entropy of the empirical sample after modeling the actual sample. By observing the variation of entropy, one can obtain whether the information about the system is lost or gain. If one obtains two empirical models for the actual system, then by using the variation of entropy the models can be compared according in mean of information. From this point of view, information based statistical error criterion is calculated as follows:

$$I = H\left(y\right) - H\left(\hat{y}\right) \tag{4}$$

where H denotes Shannon entropy measure considered as:

$$H\left(y\right) = \sum_{i=1}^{n}\left(\frac{y_i}{\sum_{i=1}^{n} y_i}\right)\log\left(\frac{y_i}{\sum_{i=1}^{n} y_i}\right) \tag{5}$$

From central limit theorem it must be noted that $I \xrightarrow{P} 0,\ n \to \infty$, where $\xrightarrow{P}$ denotes the convergence in probability.

So, from Eq. (4), one can easily make interpretation such as that the estimate which has information near 0 best fits the observed data.

IV. Simulation Study

In application part of the study, random samples arbitrarily drawn from standard normal distribution are used to illustrate the practicability of the proposed criterion. For this aim, 100 random samples are generated respectively of n = 30, 50, 125, 500, 100.000, 1.000.000 number of sample units. It's well known that the entropy of normal distribution is given by $\log\sqrt{2\pi e\sigma^2}$. The experimental entropy is calculated for each generated sample as $\log\sqrt{2\pi e S^2}$, where S^2 is maximum likelihood estimator. [Fig. 2. - Fig. 7.] obviously points out that $I \xrightarrow{P} 0,\ n \to \infty$ is satisfied in experimental way.

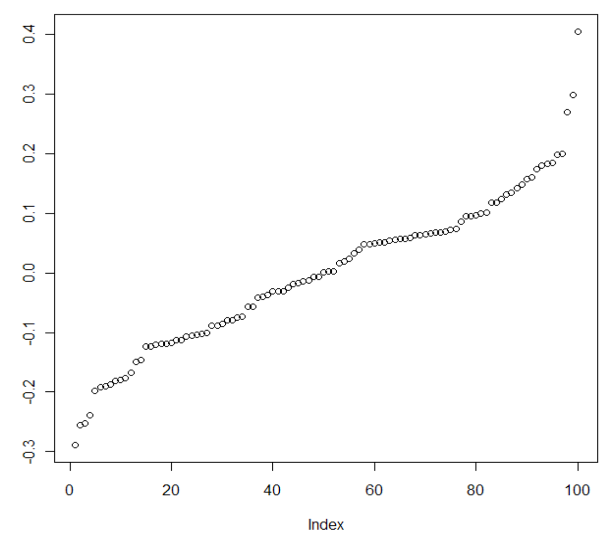

Fig. 2. The information measures results for $n = 30,\ -0.3 < I < 0.4$

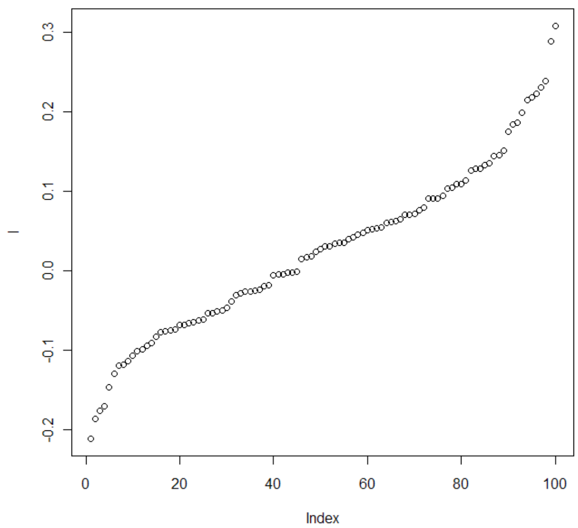

Fig. 3. The information measures results for $n = 50,\ -0.2 < I < 0.3$

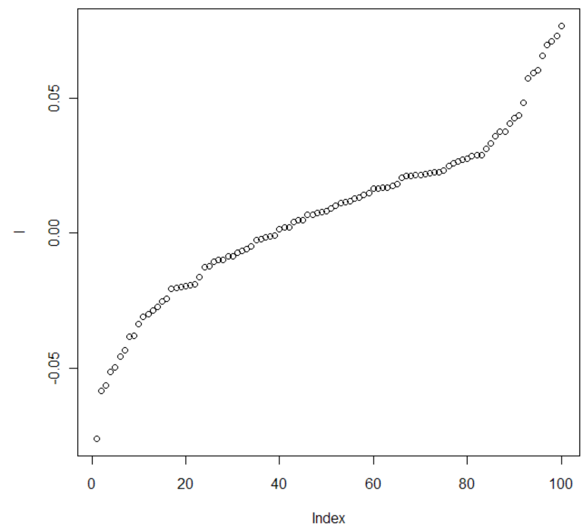

Fig. 5. The information measures results for $n = 500,\ -0.05 < I < 0.05$

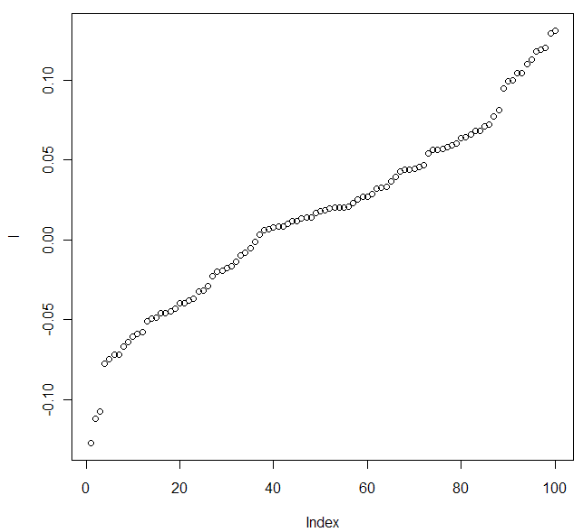

Fig. 4. The information measures results for $n = 125,\ -0.10 < I < 0.10$

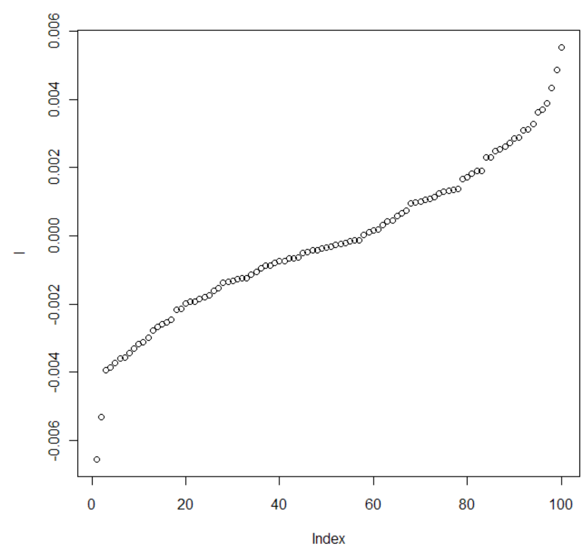

Fig. 6. The information measures results for $n = 100.000$,

$-0.006 < I < 0.006$

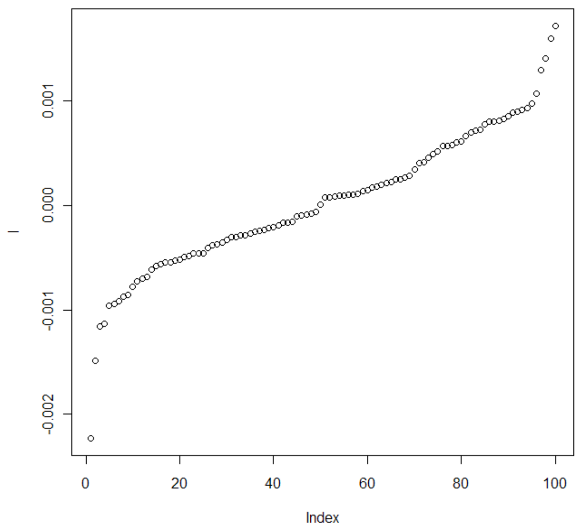

Fig. 7. The information measures results for $n = 1.000.000$, $-0.001 < I < 0.001$

From this point of view the error distribution of gunshot angles is used to illustrate the practicability of the proposed criterion in selection of appropriate model. This illustrative data presented in Table 1 is specially selected because it cannot be modeled by the known statistical distributions.

Table 1. The error distribution of gunshot angles.

Gun Shot Angle $_i$	-4;-3	-3;-2	-2;-1	-1;0
Frequency $_i$	6	25	72	133
Gun Shot Angle $_i$	0;1	1;2	2;3	3;4
Frequency $_i$	120	88	46	10

So, we obtained MaxEnt distribution subject to triple constraints in order to fit the data. The main advantage of modeling data by using these distributions is no requirements of parameter estimation. Hence, we can only be interested in the problem of goodness of fit.

Let remind the following entropy optimization problem ([3], [4]):

$$\max H = -\sum_{i=1}^{n} p_i \ln p_i \tag{6}$$

s. t. constraints:

$$\sum_{i=1}^{n} p_i g_j\left(x_i\right) = \mu_j, \quad j = 0,1,2,\ldots,m, \tag{7}$$

$$\sum_{i=1}^{n} p_i = 1, \ \ i = 1,2,\ldots,n \tag{8}$$

In the recent study,

$$K_0 = \left\{g_1(x),\ldots,g_r(x)\right\} = \left\{x, x^2, \log(x), \left(\log(x)\right)^2, \log\left(1+x^2\right)\right\}$$

is considered and triple constraints are performed as combination of the set K_0 such as

$$g^{(1)} = \left\{1, x, x^2\right\},\ g^{(2)} = \left\{1, x, \log(x)\right\}$$

$$g^{(3)} = \left\{1, x, \left(\log(x)\right)^2\right\},\ g^{(4)} = \left\{1, x, \log\left(1+x^2\right)\right\}$$

$$g^{(5)} = \left\{1, x^2, \log(x)\right\},\ g^{(6)} = \left\{1, x^2, \left(\log(x)\right)^2\right\}$$

$$g^{(7)} = \left\{1, x^2, \log\left(1+x^2\right)\right\},\ g^{(8)} = \left\{1, \log(x), \left(\log(x)\right)^2\right\}$$

$$g^{(9)} = \left\{1, \log(x), \log\left(1+x^2\right)\right\},$$

$$g^{(10)} = \left\{1, \left(\log(x)\right)^2, \log\left(1+x^2\right)\right\}$$

The solution of that mentioned entropy optimization problem subject to specific constraints takes on the following exponential form

$$p_i = e^{-\sum_{j=0}^{m} \lambda_j g_j(x_i)}, i = 1,2,\ldots,n \tag{9}$$

where $\lambda_i \ \left(i = 0,1,\ldots,m\right)$ are Lagrange multipliers.

So, several MaxEnt distributions can be obtained in order to fit the observed data. The calculated MaxEnt distributions for the error distribution of gunshot angles are illustrated in Fig. 8.

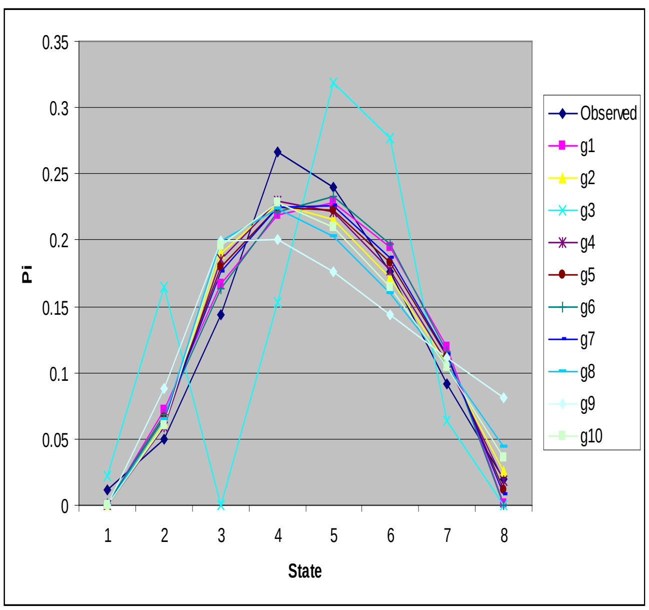

Fig. 8. MaxEnt distributions subject to triple constraints

From Fig. 8, it is realized that most of the MaxEnt distributions well fit to the data. Therefore, we require performance comparison criteria in order to determine the best fitted model. Therefore, estimates of mentioned MaxEnt Distributions are given in Table 2. and Table 3. Besides, calculated mean square error and information measures are presented.

Table 2. Estimates of MaxEnt Distributions subject to triple constraints and values of RMSE.

	Obs. Value	**P_** $g^{(1)}$	**P_** $g^{(2)}$	**P_** $g^{(3)}$	**P_** $g^{(4)}$	**P_** $g^{(5)}$
	0.012	0.0000	0.0000	0.0223	0.0000	0.0000
	0.05	0.0718	0.0622	0.1650	0.0594	0.0660
	0.144	0.1666	0.1902	0.0000	0.1854	0.1796
	0.266	0.2187	0.2276	0.1532	0.2300	0.2245
	0.24	0.2282	0.2155	0.3185	0.2219	0.2226
	0.176	0.1950	0.1719	0.2768	0.1776	0.1834
	0.092	0.1191	0.1068	0.0642	0.1076	0.1121
	0.02	0.0006	0.0257	0.0000	0.0181	0.0118
H	2.568	2.5017	2.5852	2.2588	2.5560	2.5482
RMSE		0.0249	0.0244	0.0896	0.0218	0.0229
I		0.0666	-0.0169	0.3095	0.0123	0.0201
	Obs. Value	**P_** $g^{(6)}$	**P_** $g^{(7)}$	**P_** $g^{(8)}$	**P_** $g^{(9)}$	**P_** $g^{(10)}$
	0.012	0.0006	0.0000	0.0000	0.0000	0.0000
	0.05	0.0687	0.0650	0.0649	0.0883	0.0600
	0.144	0.1637	0.1757	0.1989	0.1990	0.1963
	0.266	0.2212	0.2248	0.2238	0.2007	0.2281
	0.24	0.2325	0.2259	0.2029	0.1757	0.2097
	0.176	0.1971	0.1871	0.1599	0.1440	0.1647
	0.092	0.1162	0.1134	0.1052	0.1117	0.1048
	0.02	0.0000	0.0079	0.0444	0.0806	0.0365
H	2.568	2.4928	2.5301	2.6339	2.7272	2.6075
RMSE		0.0234	0.0224	0.0308	0.0476	0.0272
I		0.0755	0.0382	-0.0656	-0.1589	-0.0392

Table 3. Estimates of MaxEnt Distributions subject to triple constraints and values of MSE.

	Obs. Value	**P_** $g^{(1)}$	**P_** $g^{(2)}$	**P_** $g^{(3)}$	**P_** $g^{(4)}$	**P_** $g^{(5)}$
	0.012	0.0000	0.0000	0.0223	0.0000	0.0000
	0.05	0.0718	0.0622	0.1650	0.0594	0.0660
	0.144	0.1666	0.1902	0.0000	0.1854	0.1796
	0.266	0.2187	0.2276	0.1532	0.2300	0.2245
	0.24	0.2282	0.2155	0.3185	0.2219	0.2226
	0.176	0.1950	0.1719	0.2768	0.1776	0.1834
	0.092	0.1191	0.1068	0.0642	0.1076	0.1121
	0.02	0.0006	0.0257	0.0000	0.0181	0.0118
H	2.568	2.5017	2.5852	2.2588	2.5560	2.5482
MSE		0.0006	0.0006	0.0080	0.0005	0.0005
I		0.0666	-0.0169	0.3095	0.0123	0.0201
	Obs. Value	**P_** $g^{(6)}$	**P_** $g^{(7)}$	**P_** $g^{(8)}$	**P_** $g^{(9)}$	**P_** $g^{(10)}$
	0.012	0.0006	0.0000	0.0000	0.0000	0.0000
	0.05	0.0687	0.0650	0.0649	0.0883	0.0600
	0.144	0.1637	0.1757	0.1989	0.1990	0.1963
	0.266	0.2212	0.2248	0.2238	0.2007	0.2281
	0.24	0.2325	0.2259	0.2029	0.1757	0.2097
	0.176	0.1971	0.1871	0.1599	0.1440	0.1647
	0.092	0.1162	0.1134	0.1052	0.1117	0.1048
	0.02	0.0000	0.0079	0.0444	0.0806	0.0365
H	2.568	2.4928	2.5301	2.6339	2.7272	2.6075
MSE		0.0005	0.0005	0.0009	0.0023	0.0007
I		0.0755	0.0382	-0.0656	-0.1589	-0.0392

MaxEnt distribution with mean square error is found to be $P\_g^{(4)}$ which is also the best fitted distribution according to proposed information criterion. It is observed the criterion at a hand shows the similar results with RMSE/MSE.

V. Conclusion

By using the proposed information based statistical error criterion, the following results are determined:

- Simulation study is done by generating random samples from standard normal distribution. It's observed that in fact information based statistical error criterion converges to zero as sample size goes to infinity.
- MaxEnt distributions under triple moment vector function constraints are obtained for the error distribution of gunshot angles.
- Best fitted MaxEnt distribution is determined to be the one subject to $g^{(4)} = \left\{1, x, \log\left(1+x^2\right)\right\}$. It's examined that information error criterion agrees with MSE.

Since, the results given by proposed information based statistical error criterion doesn't contradict with the MSE. Since sometimes MSE gives contradictory decisions and there is no certain measure or test for the selection of an appropriate model because of the nature of the data. (See. [1], [2]), we recommend using that information based statistical error criterion in such cases. Hence, we are going to test the performance of the mentioned information error criterion in time series model selection.

References

[1] A. Aslanargun, M. Mammadov, B. Yazici & S. Yolacan, "Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting", *Journal of Statistical Computation and Simulation*, vol. 77(1), pp. 29–53, Feb. 2007.

[2] M. P. Clements and D. F. Nendry, "On the limitations of comparing mean square forecast errors", *Journal of Forecasting*, vol. **12**, pp. 617–637, December 1993.

[3] E. T. Jaynes, "Information theory and statistical mechanics", Physical Reviews, vol. 106, pp. 620—630, May 1957.

[4] J. N. Kapur and H. K. Kesavan, "Entropy Optimization Principles with Applications", Academic Press, Inc., USA, May 1992.

[5] D. M. Allen, “Mean Square Error of Prediction as a Criterion for Selecting Variables”, Technometrics, 13:3, 469-475, 2012.
[6] C. Willmott and K. Matsuura, “Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance”, Climate Research, Vol. 30, 79-82, 2005.
[7] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature”, Geoscientific Model Development, 7, 1247-1250, 2014.
[8] P. Goodwin and R. Lawton, “On the asymmetry of the symmetric MAPE”, International Journal of Forecasting, 15, 405-408, 1999.
[9] A. Myttenaere, B. Golden, B. Le Grand and F. Rossi, “Mean Absolute Percentage Error for Regression Models”, Neurocomputing, Vol. 192, 38-48, 2016.
[10] P. Ishwar and P. Moulin, “On the Existence and Characterization of the maxent distribution under general moment inequality constraints”, IEEE transactions on information theory, Vol. 51, Issue: 9, 3322-3333, 2005.
[11] S. J. Philips, R. P. Anderson and R. E. Schapire, “Maximum entropy modeling of species geographic distributions”, Vol. 190, Issue 3-4, 231-259, 2006.
[12] S. Sankaran and N. Zabaras, “A maximum entropy approach for property prediction of random microstructures”, Acta Materialia, Vol. 54, Issue 8, 2265-2276, 2006.
[13] G. S. Fishman, “ Monte Carlo: Concepts, algorithms, and applications. N.Y., USA: Springer-Verlag, 1995.
[14] A. M. Law,, and Kelton W. D., “Simulation modeling & analysis”, 3rd ed. N.Y., USA: McGraw-Hill, Inc, 2000.

Ozer Ozdemir was born in Turkey in 1982. He received his B.Sc., M.Sc. and Ph.D. degrees in statistics in the Department of Statistics at Anadolu University, Turkey, respectively in 2005, in 2008 and in 2013. He has worked as a Research Assistant from 2006-2008, as a Lecturer from 2008-2014 and as an Assistant Professor from 2014 in the Department of Statistics at Anadolu University, Turkey.
He has published over 50 international conference papers and journals in his research areas. His research interests include Applied Statistics, Simulation, Artificial Neural Networks, Fuzzy Logic, Fuzzy Modeling, Time Series, Computer Programming, Statistical Software and Computer Applications in Statistics.
Aslı Kaya was born in Turkey in 1991. She received her B.Sc. degree in Statistics in the Department of Statistics at Anadolu University, Turkey, in 2014. She received her B.Sc. degree in the Department of Business in Faculty of Management at Anadolu University, Turkey, in 2015. She still continues her education as a master student in the Department of Statistics at Anadolu University, Turkey from 2015.