A flexible neural network-fuzzy data envelopment analysis approach for location optimization of solar plants with uncertainty and complexity

A. Azadeh, M. Sheikhalishahi, S.M. Asadzadeh
Department of Industrial Engineering
Center of Excellence for Intelligent Based Experimental Mechanics
College of Engineering, University of Tehran, Iran

Abstract-**This study presents a flexible approach for optimization the location of solar plants. It is composed of artificial neural network (ANN) and fuzzy data envelopment analysis (FDEA). The intelligent approach of this study is applied to an actual location optimization of solar plants in Iran. First, FDEA is validated by DEA, and then it is used for ranking of solar plant units (SPUs) and the best α -cut is selected based on test of Normality. Also, several ANNs are developed through multi layer perceptron (MLP) for ranking of solar plants and the best one with minimum MAPE is selected for further considerations. Finally, the preferred model (FDEA or ANN) is selected based on test of Normality. The implementation of the flexible approach for solar plants in Iran identifies FDEA at α = 0.3. This indicates that the data are collected from the uncertain and fuzzy environment. This is the first study that presents a flexible approach for identification of optimum location of solar plants with possible noise, non-linearity, vague and fuzzy environment.**

Keywords*:* Fuzzy Data Envelopment Analysis (FDEA); Artificial Neural Network (ANN); Location Optimization; Solar Plant Unit (SPU); Uncertainty; Complexity

I. Introduction

Solar energy is the most ancient source, and it is origin of almost all fossil and renewable types, but due to environmental and technical consideration, recently solar energy attracted a lot of attention in all over the world Solar plants are one of the most used applications of solar energy which have great potential for supplying energy, especially in the remote and shiny regions [1].

The bibliography by [2] lists more than 1500 references dealing with location and layout problems, and many more contributions have appeared since then. There are 4 components that characterize location problems; these are: (1) customers, who are presumed to be already located at points or on routes, (2) facilities that will be located, (3) a space in which customers and facilities are located, and (4) a metric that indicates distances or times between customers and facilities [3]. A survey of many distinct applications of location models is provided by [4], ranging from traditional applications involving newspaper transfer points [5], solid waste transfer points [6 and 7], bank branches [8], and motels [9] to the more unusual location problems such as the location of a church camp [10], the determination of apparel sizes [11], ingot sizes [12], and the location of rain gauges [13]. For a list of location applications, readers are referred to [14].

In this paper a flexible approach consists of fuzzy data envelopment analysis (FDEA) and artificial neural network (ANN) is applied for ranking and assessment of potential place for locating solar plants. DEA has been used as an optimization method for indicating the most efficient location [1]. Since DEA is focused on frontiers or boundaries, any noise or error from data can cause a variation in the obtained solutions by DEA. Therefore, the input or output data should be accurate in order to successful application of DEA [15]. In some real world problems, the data for evaluation of DMUs are often not precisely defined and may be cannot accurately measured. This means that the inputs and outputs are uncertain and fuzzy. Therefore FDEA is used to overcome this weakness. Furthermore in some cases the data are corrupted and are associated with complexity and non-linearity. Thus ANN is applied to deal with these kinds of problems. The preferred model is selected based on test of Normality according to central limit theorem, because the data are collected from various sources and are associated with accumulated error.

II. Method: The Flexible ANN-FDEA Approach

We proposed a new approach namely, ANN-FDEA to alleviate these problems. According to the proposed approach, first the standard inputs and outputs are determined, collected and preprocessed. Then, the best model between FDEA with different α values and ANN is selected via test of Normality. Before selecting the preferred model, FDEA must be validated by DEA, such that there is high correlation between DEA and FDEA at α =1. In the following sections the FDEA and ANN models are described.

A. Fuzzy Data Envelopment Analysis

There are many works which propose FDEA models for performance assessment [15-18]. In this paper we use the FDEA model developed by Jahanshahloo et al. [17]. Considering *n* DMUs, *m* fuzzy inputs and *r* fuzzy outputs, the Fuzzy DEA model for a sample DMU *p* is written as the following fuzzy LP-model.

$$\tilde{Z} = \max \sum_{r=1}^{r} u_r \tilde{y}_{rp} - \sum_{i=1}^{m} v_i \tilde{x}_{ip} \quad (1)$$

$$\begin{aligned} s.t. \quad & \sum_{r=1}^{r} u_r \tilde{y}_{rj} - \sum_{i=1}^{m} v_r \tilde{x}_{ij} \le \tilde{0}; \ j = 1,2,...n \\ & (m \times \tilde{x}_{ip}) v_i \ge \tilde{1}; \quad i = 1,2,...,m \\ & \sum_{r=1}^{r} u_r \tilde{y}_{rp} - \sum_{i=1}^{m} v_r \tilde{x}_{ip} + \tilde{1} \le r \times \tilde{y}_{rp}; \quad r = 1,2,...r \\ & u_1, u_2, .. u_r \ge 0 \\ & v_1, v_2, ..., v_m \ge 0 \end{aligned}$$

$\tilde{Z}; \tilde{y}_{rj}; \tilde{x}_{ij}; \tilde{1}$ and $\tilde{0}$ are triangular fuzzy numbers and $\tilde{1} = (1,0,0)$ and $\tilde{0} = (0,0,0)$. The objective of Model 1 is to increase the value to zero. For more details the interested reader could see [17]. A non- symmetrical triangular number is defined for each variable with respect to the average, the lower limit as a minimum, and the upper limit as a maximum score of each indicator for the human resources of each DMU. After running the FDEA with different α -cuts the results of FDEA with crisp inputs (i.e. α = 1) have been compared with the DEA ones. If it is shown that the results of DEA and FDEA with this α -cut have a high degree of correlation, it can be concluded that the proposed FDEA is superior to the formerly used DEA approaches in literature, since it can be used for evaluation and ranking of the DMUs with both crisp and fuzzy input parameters. Note that if the Spearman correlation tests for comparing DEA against FDEA with α = 1 result in a low degree of correlation between DEA and FDEA, the results of FDEA are still valid and it can be concluded that the system under study deals with high degrees of complexity and uncertainty.

B. Artificial Neural Network

ANNs consists of an inter-connection of a number of neurons. There are many varieties of connections under study, however here we will discuss only one type of network which is called Multi Layer Perceptron (MLP). In this network the data flows forward to the output continuously without any feedback. A typical two layer feed forward model is used for forecasting demand. The input nodes are the previous lagged observations while the output provides the forecast for the future value. Hidden nodes with appropriate nonlinear transfer functions are used to process the information received by the input nodes. The model can be written as follows:

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-i} + \beta_{0j} \right) + \varepsilon_t \quad (2)$$

Where m is the number of input nodes, n is the number of hidden nodes, f is a sigmoid transfer function such as the logistic, $f(x) = \dfrac{1}{1+\exp(-x)}$, α_j $(j = 0, 1, ..., n)$ is a vector of weights from the hidden to output nodes and $\{\beta_{ij},\ i = 1, 2, \ldots, m;\ j = 0, 1, \ldots, n\}$ are weights from the input to hidden nodes. α_0 and β_{oj} are weights of arcs leading from the bias terms which have values always equal to 1. Note that Equation (2) indicates a linear transfer function is employed in the output node as desired for forecasting problems. The MLP's most popular learning rule is the error back propagation algorithm. Back Propagation learning is a kind of supervised learning introduced by [19] and later developed by [20]. The attraction of MLP has been explained by the ability of the network to learn complex relationships between input and output patterns, which would be difficult to model with conventional algorithmic methods. The disadvantage of ANN is that because the network finds out how to solve the problem by itself, its operation can be unpredictable. The following general architecture is designed for ANN: The training algorithm is set to trainlm (Lavenberg-Marqurdet back propagation) with one hidden layer. The transfer function in hidden layer is set to be hyperbolic tangent sigmoid transfer function (tansig) and linear transfer function (purelin) in output layer.

III. Experiment: The Case Study

A set of technical, geographical and social factors for location optimization of solar plant units (SPUs) are considered in this paper. The required data indicators, which were used in the proposed model, were collected from the National Statistics Bureau of Iran. These parameters were defined as indicators (inputs and outputs) as follows:

- *Population and human labor:* Undoubtedly, establishment of industrial facilities must be performed in the locations that have adequate production consumers. So in this viewpoint the vicinity of solar plants to the center of the city that has a higher population is an advantage (we will discuss this matter later from other points of view). Therefore, we have used the population of selected regions in each city as an output parameter.
- *Distance of power distribution networks:* As noted for establishing power plants, selection of regions with high degree of population aggregation is a plus. But for solar plants another aspect must be considered. In general, construction of solar plants is not still economical and they are not comparable with conventional fossil fuel plants. Therefore, establishment of solar plants in the rural and remote locations is a better alternative. For this purpose, solar plants for the shiny locations are excellent solution. With regard to this viewpoint and unlike prior view, construction of solar plants in the high distances from center of the cities is better. With this assumption this parameter has an input structure.
- *Land cost:* The land is original infrastructure for construction of each plant. This is more important for solar than other plants because they considerably need

more land than other methods of energy generation. It can be seen that the required land for solar energy is approximately equal to 30 times as conventional fossil fuel plants [21]. This means that land cost is an important parameter for indicating the location of power plants. Hence, land cost in different regions of each city is calculated and used as one of the indicators for the FDEA model. This parameter has an input structure.

- *Solar global radiation:* We can suggest that solar total radiation is the most important parameter for selecting a region as a candidate for establishing solar plants. The effective parameters on the quantity of solar global radiation are geographical and climatological parameters such as: temperature, wind speed, humidity, vapor pressure, precipitation, etc. This parameter has an increasing trend and therefore it is considered as an output indicator.
- *Intensity of natural disasters occurrence:* Establishment of the plants in a location with a high degree of disaster incidence could be hazardous and increasing related maintenance costs. Therefore, the foundation of plants must be performed in as safe and secure place that have a low statistics of related natural disasters. The value of this parameter must has decreasing trend and therefore we consider it as an input indicator.
- *Quantity of proper geological areas:* In this viewpoint material, grade and other conditions of substruction of the plant are important. For this parameter, the available area of suitable grounds in proximity of each city are collected and used as one of the effective factors. This parameter is used as an output indicator to its increasing trend.
- *Quantity of availability of the water:* Supplying the consumable water for operation of the plants is a vital issue. Therefore, the construction of solar plants must be initiated at the locations with capability of water supply. For this purpose the quantity of suitable grounds with capability of water accumulation are calculated and used as one of the output parameters of related FDEA model.
- *Quantity of proper topographical areas:* For establishing each structure, existence of a set of topographical conditions is necessary. Determining the suitable places from topographical aspect is also important. For this purpose, the quantity of proper grounds in proximity of each city is determined. This parameter has an output structure because of its increasing trend.

The required data for population, human labor and land cost were gathered from the National Statistics Bureau of Iran. For the values of distance of power distribution networks, there were no exact available data. Thus, their impact was determined based on the given scores by the experts. Therefore, this parameter has an increasing structure. Therefore, in spite of input structure of original parameter, in the final model, we consider this indicator as one of the output parameters of related FDEA model. This shows the uncertainty and ambiguity in data set. For indicators such as quantity of proper geological and topographical areas and availability of the water, a geographical information system (GIS) technique has been used, which has been performed by Iran Energy Efficiency Organization. This process was accomplished by numerating related maps of each city and using existing numeral maps in the scale of 1:500,000. In accordance to this method, the required values of each indicator were calculated. Also the quantities of solar global radiation were provided through National Meteorological Organization of Iran. As mentioned earlier, we have considered 2 parameters for calculation of the indicator intensity of natural disasters occurrence: earthquake and torrent. These 2 parameters have been recognized as more general effective parameters in Iran as well as their quantities were more available. These parameters have considered with equal coefficients for calculating the final values of the indicator. For the other locations in the other countries different kind of parameters can be used for this indicator such as sand tornados or intensive local winds. In the following section the results of ranking solar plant units by preferred model are described.

IV. Results and Analysis

The DEA and FDEA are used for ranking 150 SPUs. The input variables are land cost and intensity of natural disasters occurrence. The output variables are population and human labor, solar global radiation, distance of power distribution networks, quantity of proper geological areas, quantity of availability of the water, quantity of proper topographical areas. The results of ranking by DEA and FDEA at α =1 are shown in Figure 1.

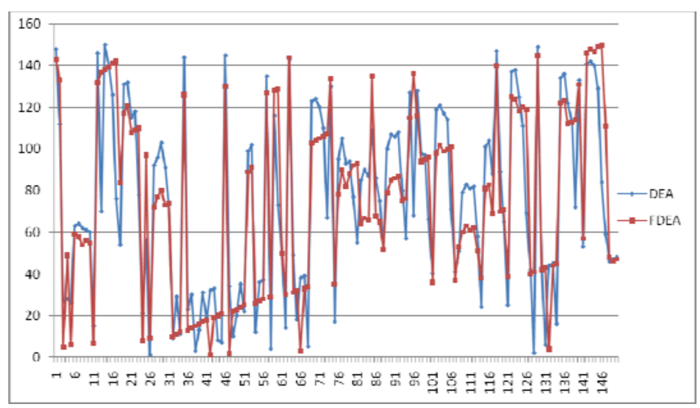

Figure 1: Results of DEA and FDEA at α =1

Next, the results of FDEA is verified and validated by DEA at α =1. To compare the results of DEA and FDEA, a non-parametric method has been utilized. One of the methods used for testing the independence of the ranking data is the Spearman nonparametric test. Using the $\sqrt{n-1}\gamma_{sp}$ criteria, H_o is tested for identifying the independence of the two stated methods on a pair-wise comparison basis. Table 1 indicates the Spearman correlation measures for several α values.

Table 1: The results of Spearman correlation test between DEA and FDEA for different α values

α values	**Spearman correlation index**
0.1	0.46
0.3	0.45
0.5	0.48
0.7	0.63
0.9	0.78
1	0.87

The spearman correlation scores indicate an acceptable high correlation among the ranks obtained by DEA and FDEA. Hence, for α =1 H_o (i.e. independence of the obtained ranks) is rejected and H_1 is accepted which shows the correlation among the ranks obtained by two methods (pair-wise comparisons). Therefore, the ranking results obtained by DEA and FDEA at α =1 are verified with relatively high degrees of confidence. According to the proposed approach, the best α value for FDEA is selected based on test of Normality. Table 2 shows Normality test for different α values using Kolmogorov-Smirnov Test.

Table 2: Test of Normality according to p-value

***α* values**	*p*-value
0.1	0.029
0.3	0.080
0.5	0.020
0.7	0.007
0.9	0.000
1	0.000

The results of Kolmogorov-Smirnov Test indicate that the FDEA at α = 0.3 is the preferred model for ranking SPUs because it has higher p-value in comparison to other α values. It also addresses the high ambiguity and uncertainty associated with the environment under study. Thus, the FDEA at α = 0.3 is selected to compare with ANN. The best SPU's which are selected by ANN and FDEA at α =0.3 are shown in Table 3.

Table 3: The best locations selected by FDEA (α =0.3) and ANN

RANK	FDEA (α =0.3)	ANN
1	SPU 145	SPU 81
2	SPU 147	SPU 97
3	SPU 150	SPU 19
4	SPU 149	SPU 82
5	SPU 148	SPU 143
6	SPU 146	SPU 137
7	SPU 129	SPU 83
8	SPU 127	SPU 21
9	SPU 128	SPU 99
10	SPU 131	SPU 22

As mentioned before, the preferred model is selected based on test of Normality. In the proposed case study, the *p*-values for FDEA at α = 0.3 and ANN are 0.08 and 0.00, respectively. Thus, the best model for location optimization of solar plants is FDEA at α = 0.3. This proves the existence of tremendous uncertainty and noise in the data set. Therefore, if DEA method was used for ranking of SPUs, the results of ranking could be misleading. Hence, the preferred locations by DEA would be drastically different from the ones selected by FDEA at α = 0.3. This is shown in Table 4. For instance, if DEA is used instead of FDEA, the preferred location would be SPU121, whereas by using FDEA at α = 0.3, the best location is selected as SPU145. Thus, the results of DEA would be incorrect for decision making process. This is due to data uncertainty and environmental noise.

Table 4: The best locations selected by FDEA (α =0.3) and DEA

RANK	FDEA (α =0.3)	DEA
1	SPU 145	SPU 121
2	SPU 147	SPU 79
3	SPU 150	SPU 132
4	SPU 149	SPU 150
5	SPU 148	SPU 24
6	SPU 146	SPU 81
7	SPU 129	SPU 139
8	SPU 127	SPU 138
9	SPU 128	SPU 127
10	SPU 131	SPU 141

V. Conclusion

DEA is one of the mostly used methods that can be employed for evaluating the relative efficiency of SPUs of location optimization. Since its required data cannot be precisely measured in some cases, the theory of uncertainty plays a significant role in DEA. Moreover, in some cases data are associated with complexity and non-linearity. Therefore, in this paper a flexible FDEA-ANN approach was proposed for location optimization of solar plants. First, the input and output data are specified and collected. DEA is applied for ranking the SPUs and is used for validation and verification of the results of FDEA at α =1 by Spearman correlation test. It was concluded that in the case of having low level of uncertainty the results of DEA and FDEA are relatively equal. Also, the best structure of ANN was selected based on MAPE. Finally, the best model between FDEA (with different α - cuts) and ANN is selected based on test of normality. This is because the required data are collected from various sources and are associated with accumulated error. In the case of location optimization of solar plants in Iran, the best model was selected as FDEA at α =0.3. However, in different cases, other models may be selected as the preferred model. Also, it was shown that because of existence of uncertainty and noise in the data set, using DEA instead of FDEA leads to the misleading decisions. The suggested approach of this study can be utilized for various solar location optimization problems.

VI. References

[1] Azadeh, A., Ghaderi, S.F. and Maghsoudi, A (2008) Location optimization of solar plants by an integrated hierarchical DEA PCA approach, Energy Policy, Vol.36, pp. 3993–4004.

[2] Domschke, W. and Drexl, A (1985) Location and layout planning: An international bibliography. Lecture Notes in Economics and Mathematical Systems, Vol. 238. Springer, Berlin.

[3] Bhatnagar, R. and Sohal, A.S (2005) Supply chain competitiveness: measuring the impact of location factors, uncertainty and manufacturing practices. Technovation, Vol. 25, pp. 443–456.

[4] Eiselt, H.A (1992) Location modeling in practice. American Journal of Mathematical and Management Sciences, Vol. 12, pp. 3–18.

[5] Jacobsen, S.K. and Madsen, O.B.G (1980) A comparative study of heuristics for a two-level routing—location problem. European Journal of Operational Research, Vol. 5, pp. 278–287.

[6] Marks, D. and Liebman, J (1971) Location models: solid wastes collection examples, Journal of the Urban Planning and Development Division. In: Proceedings of the American Society of Civil Engineers, Vol. 97, No. UP1, April, pp. 15–30.

[7] Wirasinghe, S.C., Waters, N.M (1983) An approximate procedure for determining the number, capacities, and locations of solid waste transfer stations in an urban region. European Journal of Operational Research, Vol. 12, pp. 105–111.

[8] Hopmans, A.C.M (1986) A spsatial interaction model for branch bank accounts. European Journal of Operational Research, Vol. 27, pp. 242–250.

[9] Kimes, S.E. and Fitzsimmons, J.A (1990) Selecting profitable hotel sites at La Quinta Inns. Interfaces, Vol. 20, pp. 12–20.

[10] Huxley, S.J (1982) Finding the right spot for a church camp in Spain. Interfaces, Vol. 12, pp. 108–114.

[11] Tryfos, P (1986) An integer programming approach to the apparel sizing problem, Journal of the Operational Research Society, Vol. 37, pp. 1001–1006.

[12] Vasko, F., Wolf, F. and Stott, K (1987) Optimal selection of ingot sizes via set covering, Operations Research, Vol. 35, pp. 346–353.

[13] Hogan, K (1990) Reducing errors in rainfall estimates through rain gauge location, Geographical Analysis, Vol. 22, pp. 33–49.

[14] Current, J., Daskin, M. and Schilling, D (2002) Discrete network location models. In: Drezner, Z., Hamacher, H. (Eds.), Facility Location: Applications and Theory. Springer, Berlin, pp. 81–118.

[15] Lertworasirikul, S., Fang, S. C., Joines, J. A. and Nuttle, H. L. W (2003) Fuzzy data envelopment analysis (DEA): a possibility approach, Fuzzy Sets and Systems, Vol. 139, pp. 379-394.

[16] Guo, P (2010) Fuzzy data envelopment analysis and its application to location problems, Information Science, Vol. 179, pp. 820-829.

[17] Jahanshahloo, G. R., Soleimani-damaneh, M. and Nasrabadi, E (2004) Measure of efficiency in DEA with fuzzy input–output levels: a methodology for assessing, ranking and imposing of weights restrictions, Applied Mathematics and Computation, Vol. 156, No. 1, pp. 175-187.

[18] Wen, M. and Li, H (2009) Fuzzy data envelopment analysis (DEA): Model and ranking method, Journal of Computational and Applied Mathematics, Vol. 223, No. 2, pp. 872-878.

[19] Werbos, P (1974) Beyond regression: new tools for prediction and analysis in the behavioral sciences, Doctoral dissertation, Harvard, Cambridge, MA.

[20] Rumelhart, D.E. and McClelland, J.J (1986) Rumelhart PDP Research Group and McClelland, Parallel distributed processing, MIT Press, Cambridge, MA.

[21] Ramanathan, R (2001) Comparative risk assessment of energy supply technologies: a data envelopment analysis approach, Energy, Vol. 26, No. 2, pp. 197–203.