Productive performance of small peri-urban farms using self-organizing maps and data envelopment analysis

A. P. S. Rubem[1], A. L. de Moura[1], E. de Oliveira[1],
J. C. C. B. Soares de Mello[1], L. A. Alves[2] & R. S. Tavares[1]
[1]*Federal Fluminense University, Brazil*
[2]*Petrobras, Brazil*

Abstract

In this paper, we aim to analyze the productive performance of plots cultivated by family farmers. We use an alternative Data Envelopment Analysis (DEA) approach to calculate the relative efficiency of such plots. Notwithstanding, DEA's basic assumption includes the homogeneity of the production units under analysis. Herein, as the chemical composition of the soil varies considerably among plots, directly influencing their fertility, and, thus, their productivity levels, the plots shall be grouped into homogeneous clusters first. For that, we use Self-Organizing Maps, based on their comparative advantages to other methods. Then, the relative efficiencies of plots within each cluster will be assessed using separate DEA models. Nonetheless, a direct comparison among the scores of plots from different clusters is not feasible, because the relative efficiency of a production unit can solely be compared to those inserted in the same set of analysis. Hence, to overcome such inconvenience, we further apply a technique that allows compensating for the non-homogeneity of plots. The results indicate that, when countervailing the effects of the chemical composition of the soil, the plots with favorable conditions do not necessarily present better productive outcomes.
Keywords: sustainable development, data envelopment analysis, efficiency analysis, peri-urban spaces, self-organizing maps.

1 Introduction

Since 2006, the Family Agriculture Project in Strip of Ducts (or PAF Ducts – the acronym in Portuguese) is an agricultural occupation strategy for the peri-urban areas that overly Petrobras refinery pipelines. The area, located in the municipality of Nova Iguaçu, Rio de Janeiro, Brazil, was divided in plots of 1,000 m² each. Each one assigned to a particular family.

The activity developed meets some ecological concerns and focuses primarily at the cultivation of fruits, tubers and vegetables. The farmers cultivate about 50% of the total area of each plot, while the other part is alternately intended for fallow. The average labor force employed is two workers per year. All plots make use of electrical energy and have a catchment basin for manual irrigation. Besides, technical assistance and inputs are offered at equitable basis.

The objective of this study is to evaluate the productive performance of such plots, in accordance to some criteria, which aim at great stability (i.e. less seasonal variation) in production, large variety of items produced and high volume of products offered for sale. For that, we use an alternative approach that combines Data Envelopment Analysis (DEA [1]), Self-Organizing Maps (SOMs [2]) and a compensating algorithm proposed by [3].

The option for DEA was grounded on its ability of dealing with multidimensional problems and various units of measurement. Nevertheless, DEA models assume the homogeneity of the production units under analysis (the plots). However, in a preliminary analysis, it was identified that the chemical composition of the soil varies notably among the plots, influencing their fertility, and, thus, their productive levels. Therefore, the plots needed to be grouped in homogeneous clusters firstly. For such purpose, we use SOMs based on their comparative advantage to other clustering methods [4]. Next, the relative efficiencies of the plots within each cluster will be calculated using separate DEA models.

Due to the disjointedness of the clusters formed, a direct comparison among plots from different clusters is not feasible, once the relative efficiency of a production unit can solely be compared to those units inserted in the same set of analysis. To surpass such inconvenience, we will further apply the algorithm of [3] to compensate for the non-homogeneity of the plots.

Section 2 brings a review of DEA´s use in agricultural activities. In section 3, we present the methodology used herein. Section 4 describes the application to the problem under analysis. In section 5, we present the results derived. Section 6 discusses the results in terms of the chemical composition of the soil. Finally, in section 7, we present some conclusions and a few suggestions for future work.

2 DEA's application in agriculture

Since the early 1990s, DEA has become widely used in the evaluation of relative efficiency in different areas [5]. The agricultural activity represents a field where the application of DEA models is quite fruitful. Classic DEA models [1, 6] were used to analyse the relative efficiency of energy consumption in the agriculture by [7]. In [8], the production relative efficiencies and productivity changes of agricultural cooperatives were measured using DEA and the Malmquist index. In [9], the beef cattle production was examined using a unitary input DEA model combined to relative efficiency measures generated by the inverted DEA frontier. In [3], DEA models were used to compare the performance of farms cultivated with different technologies, and an algorithm for calculating the relative efficiency scores taking into account such non-homogeneity was proposed. In [10], the performance of producers were analysed, using SOMs and cross-evaluation DEA models. In [11], a unitary input model was used for the evaluation of intercropping.

Here, as [9, 11], we apply a unitary input DEA model, but unlikely we use SOMs to cluster the production units and allow the homogeneity in the subset to be evaluated by each DEA model. In this sense, our proposal differs from [10], which used the DEA cross-efficiencies as inputs to the SOMs procedure for *ex-post* clustering. Additionally, we apply the algorithm developed in [3] to enable the direct comparison of plots from different clusters.

3 Methodology

The methodology applied herein comprises three sequential steps. First, we use SOMs [2] to segregate the plots and assure their homogeneity in terms of the chemical properties of the soil. Second, we use a separate DEA model for each cluster, and evaluate the relative efficiency of each plot within the cluster. Finally, we use an algorithm [3] to countervail for non-homogeneity among plots.

3.1 Self-organizing maps: fundamental aspects

Although there exist several methods that evaluate the similarity between a set of units to create homogeneous subsets (clusters), we opted for SOMs. In the literature, we find some works [10, 12] that discuss the combined use of SOMs and DEA models.

SOMs represent a type of neural network, i.e. computer models of artificial intelligence that incorporate certain capabilities of the human brain, where sensory inputs are represented by topologically organized maps [12]. Particularly, SOMs emulate the unsupervised learning by considering neuron neighborhood, whose structures are arranged in grid. The most used topology is the interconnected two-dimensional one, where the neurons are represented by rectangular, hexagonal or random grid knots of neighbor neurons [12].

SOMs procedure comprises three processes. In the competitive process, each neuron is initialized with a vector of inputs, and then the neurons compete to become active. The choice of the winner neuron is usually based on the Euclidean distance, as performed here. Next, the winner has its weights adjusted to respond to the stimulus (synapse), and a cooperative process between the winner and its topological neighbors is simulated, so that the neighbors receive adjustments as well. The topological neighborhood is normally defined by a Gaussian function, as herein. Finally, the adaptive process takes place by the adjustment of the synaptic weights, considering that the learning rate decreases over time to avoid that new information compromise the knowledge accumulated.

In synthesis, SOMs yield a topological mapping, segregating the input data based on their similarities. Comparatively to other neural networks, SOMs present the advantage of transforming patterns of high-dimensionality in discrete maps, usually single or two-dimensional. For further insights, see e.g. [12].

3.2 Data envelopment analysis

3.2.1 Basic concepts

DEA is a non-parametric method based on mathematical programming that calculates the relative efficiency of a set of production units using multiple inputs (resources) and producing multiple outputs (products). These units are known as decision-making units (DMUs). In DEA, the basic premise is homogeneity of the DMUs. Here, as the DMUs (plots) in the set of analysis do not operate in similar environments (due to differences in the chemical conditions of the soil), we use an alternative DEA approach to overcome the lack of homogeneity.

Each DMU's relative efficiency score is optimized, by comparing the resources used and the quantities produced to the levels of the others. The result is an efficient frontier. The DMUs lying on the frontier are efficient (score of 100%); the others are inefficient (score of less than 100%).

The most used DEA models are: CCR [1] and BCC [8]. The first assumes constant returns-to-scale, while the latter works under variable returns-to-scale, replacing the axiom of proportionality by convexity. These models present equivalent formulations (envelope and multipliers), which provide the same efficiency scores for each DMU, as they are dual problems. Traditionally, there are two possible radial orientations for such models: input orientation, which seeks to minimize the resources while the production levels remain fixed; and the output orientation, which implies the increase in quantities produced while the resource levels remain unchanged.

3.2.2 Unitary input DEA model

In this study, as every agricultural plot in the set of analysis operate with quite similar inputs, we use the DEA model with a unitary, constant and single input, wherein the input denotes the very existence of the DMU. The unitary input avoids the mathematical inconsistencies that arise in a model without inputs. We apply the approach of [13], in which the CCR and BCC models are equivalent, and use an output orientation. The linearized unitary input DEA model, output-oriented, in the envelope formulation, herein applied, is given by the following linear programming:

$$\text{Max } h_o \tag{1}$$

$$\text{s. t. } \sum_{k=1}^{n} \lambda_k y_{jk} \geq h_o y_{j0}, \forall j \tag{2}$$

$$\sum_{k=1}^{n} \lambda_k \leq 1 \tag{3}$$

$$\lambda_k \geq 0, \forall k. \tag{4}$$

In eqns. (1)–(4), h_o denotes the inverse of the relative efficiency of the DMU under analysis (DMU$_o$); y_{jk} is the jth output ($j = 1, ..., s$) of DMU$_k$ ($k = 1, ..., n$); and $\{\lambda_k\}$ and is the individual contribution of each DMU in the formation of DMU$_o$'s target. In fact, this model resembles a multi-criteria additive model, where the alternatives (DMUs) assign weights to each criterion (outputs), ignoring any judgment of an eventual decision-maker.

3.2.3 Homogenization technique

Although the literature comprises several proposals to overcome the lack of homogeneity [14, 15], we follow the line of action (as in [3, 12, 16]) that applies handicap factors to compensate for the DMUs non-homogeneity.

The assumption is that, after grouping the DMUs into homogeneous clusters, the efficient DMUs of each cluster equally share good management practices. Nonetheless, these DMUs do not present scores of 100%, when compared to DMUs from other clusters, due to different exogenous conditions that characterize each particular cluster. Thus, comparisons among clusters shall be done by taking into account solely the efficient DMUs of each cluster. This comparison allows identifying the cluster that benefits from exogenous variables to compensate the disadvantaged clusters, giving a prior benefit to those DMUs under disadvantage.

Differently from [12, 16] that apply the handicap factor to the inputs and outputs, respectively; we follow [3] that corrects the relative efficiency measures directly. The algorithm herein applied obeys the following steps.

(1) Cluster the DMUs in homogeneous groups.
(2) Run a specific CCR model for each cluster, and select the efficient DMUs.
(3) Run a CCR model with the efficient DMUs from step 2.
(4) Calculate the average relative efficiencies of the DMUs from step 3, separated in their original clusters.
(5) Run a CCR model with all DMUs in the set of analysis.
(6) Use the average relative efficiencies of step 4 as a compensating factor to the relative efficiency measures of the disadvantaged clusters, by dividing each DMU's relative efficiency score found in step 5 by the average relative efficiency of step 4 assigned to its original cluster. If any value obtained is greater than one, we need to perform the corresponding normalization.

The compensated relative efficiency scores are those derived in step 6.

4 Application and results

In the following, we apply the methodology described in section 3 to the evaluation of the agricultural plots of PAF Ducts. The data refer to year 2012.

4.1 Step 1: clusters definition

We start choosing the environmental variables to be used as input vectors in the SOMs procedure. Based on the chemical composition of the soil of the plots in the set of analysis, we selected the variables: potential of hydrogen (pH); potassium content (K); percentage of organic matter (OM); base saturation (BS); and boron content (B). Such variables are widely recognized as relevant to express the fertility of the soil. Table 1 exhibits the input data used to initialize the SOMs procedure, which come from a research conducted by the Center of Analysis of the Federal Rural University of Rio de Janeiro, in May 2012.

Table 1: Input data used for SOMs procedure.

Plot	pH	K (ppm)	OM (%)	BS (%)	B (ppm)
P3	7	107	26.4	90	0.12
P6	7.4	26	19.1	100	0.25
P7	7.6	36	18.3	100	0.21
P9	7	43	17.4	91	0.12
P10	7.2	29	15.9	100	0.21
P11	7.1	31	19.1	96	0.17
P12	7.3	31	16.6	96	0.12
P15	7.3	60	20.0	98	0.17
P26	7.1	95	20.9	95	0.12

We used the software MATLAB® version 7.10.0, and opted for a hexagonal topology to the grid, as it showed good results and represents a usual practice among experts. Then, we tested different grid dimensions to determine the best arrangement of clusters. We restricted the tests to the (2x1), (3x1) and (2x2) grids, to assure a minimum number of plots in each cluster for the DEA analysis. Table 2 shows the cluster configurations, according to the grid dimensions tested.

Table 2: Clusters formation, according to the grid dimension tested.

Plot	Dimension		
	(2x1)	(3x1)	(2x2)
P3	2	3	4
P6	1	1	1
P7	1	1	1
P9	1	2	2
P10	1	1	1
P11	1	2	2
P12	1	2	2
P15	1	2	3
P26	2	3	4

The (2x2) grid yields four clusters, one formed by a single plot (P15). As we will apply a separate DEA model to each cluster, a comparison of this plot (which originates eight DMUs, as later explained) with others would be impaired, without the application of the homogenization algorithm used herein. Once this is not a limitation to our approach, we opted for the (2x2) grid dimension, because it better distinguished the lack of homogeneity among the plots (i.e. generated a larger number of clusters).

4.2 DEA modeling

To evaluate the productive performance of the plots, primary data were gathered on the availability of items put for sale at weekly fairs. In 2012, the production activity was concentrated in the period from February to October. This occurred mainly due to the climate conditions in the region.

The dataset comprises the agricultural items offered, listed in accordance to the amount put for sale by item, and the average prices (in Brazilian Real – R$) practiced by each of the nine plots comprised in the analysis. Thereby, our unitary input DEA modeling was designed with two outputs: the variety of products available for sale (y_1); and the estimated revenue based on the average price of products (y_2). The "variety of products" is the amount of different items available for sale for each DMU (combination plot-month). A large variety of items denotes a better performance of the plot in an effort to meet the market's needs, as well as a greater ability to deal with seasonality. The "estimated revenue" is the amount of product available for sale multiplied by the average selling price of each product. This value was used as the actual revenue data were not available, and it standardizes the production of distinct items in monetary units as well, thus making possible the sum of the production from different crops.

As the output data were collected monthly, we apply the model to the period from February to October 2012, and pool all observations together in the analysis, through a longitudinal data approach, as done in [17]. This is one of the ways to increase the number of DMUs [18], since we regard the same plot as a distinct DMU in different months. Thence, the DMUs are each plot-month combination (i.e. "P6-Feb" is a different DMU from "P6-Mar"). The assumption is that the technology and the environmental conditions remain stable over the period of time concerned, what seems fairly acceptable to our case study. Otherwise, we would need to use, e.g. the Malmquist index [19] instead.

In the analysis, we solely take into account the plot-month combinations with non-nil outputs, totalizing 68 DMUs, which are distributed among the four clusters defined by the SOMs procedure. As a deterministic method, DEA does not depend on a large number of observations for the validity of its application, unlike statistical approaches. In such sense, the number of DMUs in each cluster (varying from 25 to 8) meets the minimum advised by [20].

4.2.1 Step 2: evaluation of DMUs within each cluster

The linear program used to compute the relative efficiency of each DMU through our proposed unitary input DEA-CCR model is obtained by replacing the values of the outputs above defined in the general formulation of subsection 3.2.2, eqns. (1)–(4), provided that a separate model is run for each one of the four clusters previously defined.

For that, we applied the software SIAD [21] version 3.0 (available at http://www.uff.br/decisao/Siadv3.zip), and calculated the relative efficiency scores of each DMU, in relation to the others belonging to the same cluster. These results are shown in table 3. We may observe that, in the whole set of analysis, seven DMUs were 100% efficient within their own cluster (two in C1, two in C2, two in C3, and one in C4, as marked in grey). Notably, five out of the seven cluster-efficient DMUs refer to plots operating in August.

Table 3: Relative efficiency scores for each DMU in relation to the others in the same cluster.

C1		C2		C3		C4	
DMU	Eff	DMU	Eff	DMU	Eff	DMU	Eff
P6-Feb	13.3%	P11-Feb	15.4%	P15-Feb	30.9%	P3-Feb	42.3%
P7-Feb	46.7%	P12-Feb	15.4%	P15-Mar	42.5%	P3-Mar	73.1%
P6-Mar	56.7%	P9-Mar	30.8%	P15-May	100.0%	P3-Apr	80.8%
P7-Mar	76.7%	P11-Mar	65.4%	P15-Jun	44.8%	P3-May	73.1%
P6-Apr	33.3%	P12-Mar	34.6%	P15-Jul	100.0%	P3-Jun	61.5%
P7-Apr	63.3%	P9-Apr	23.1%	P15-Aug	70.8%	P3-Jul	76.9%
P10-Apr	53.3%	P11-Apr	51.3%	P15-Sep	74.6%	P3-Aug	100.0%
P6-May	93.3%	P9-May	69.2%	P15-Oct	45.2%	P3-Sep	65.4%
P7-May	73.3%	P11-May	88.5%			P26-Sep	7.7%
P10-May	56.7%	P12-May	53.8%			P3-Oct	88.5%
P6-Jun	73.3%	P9-Jun	65.4%				
P7-Jun	70.0%	P11-Jun	57.7%				
P10-Jun	36.7%	P12-Jun	34.6%				
P6-Jul	93.3%	P9-Jul	84.6%				
P7-Jul	90.0%	P11-Jul	88.5%				
P10-Jul	36.7%	P12-Jul	69.2%				
P6-Aug	100.0%	P9-Aug	100.0%				
P7-Aug	100.0%	P11-Aug	98.8%				
P10-Aug	43.3%	P12-Aug	100.0%				
P6-Sep	83.6%	P9-Sep	53.8%				
P7-Sep	84.2%	P11-Sep	67.5%				
P10-Sep	26.7%	P12-Sep	61.5%				
P6-Oct	80.1%	P9-Oct	61.5%				
P7-Oct	76.7%	P11-Oct	50.0%				
P10-Oct	40.7%	P12-Oct	50.0%				

4.2.2 Steps 3 and 4: cluster of efficient DMUs

Next, we separate the 100% efficient DMUs of each cluster in a cluster of efficient units, and apply the same unitary input DEA-CCR model previously used to these seven DMUs (step 3). From the scores obtained, we calculate the average scores of the DMUs in the cluster of efficient units, taking into account the other efficient units from their original clusters (step 4). These results are shown in table 4.

The fact that only the DMUs from cluster C1 got a score of 100% in the cluster of efficient units suggests this is the only cluster operating in optimal environmental condition, while the others show soil disadvantages that negatively affect their productive outcomes, despite any other managerial inefficiencies. DMUs "P6-Aug" and "P7-Aug" were deemed as 100% efficient because they individually exhibit the best ratio at each one of the partial productivity measures, i.e. the largest revenue and the widest variety of items produced, respectively. This is a well-known and widely reported feature of DEA models [22].

Table 4: Relative efficiency scores for the DMUs in the cluster of efficient units and the average scores by the original cluster.

Original cluster	DMU	Score in this cluster	Average by original cluster
C1	P6-Aug	100.0%	100.0%
	P7-Aug	100.0%	
C2	P9-Aug	86.7%	80.6%
	P12-Aug	74.5%	
C3	P15-May	60.0%	56.7%
	P15-Jul	53.3%	
C4	P3-Aug	87.5%	87.5%

4.2.3 Steps 5 and 6: overall evaluation of DMUs by countervailing the lack of homogeneity among clusters

In the following, we apply the same unitary input DEA model to all the DMUs comprised in set of analysis (step 5), and refer to this as the "all-units" model. Then, starting step 6, we use the reciprocal of each average relative efficiency score from step 4 (cluster of efficient units) as a compensating factor for each disadvantaged cluster (C2, C3 and C4). For that, we multiply the compensating factor assigned to each cluster by the relative efficiency scores found to each DMU in step 5 (all-units model) taking into account its corresponding original cluster.

As the compensating procedure resulted in two relative efficiency scores greater than one, we had to perform the corresponding normalization, dividing the scores so far obtained by their maximum value. Therefore, all DMUs in cluster C1 had their relative efficiency scores reduced in relation to those calculated using the all-units model, while the relative efficiency scores of the DMUs from the other clusters have all increased.

Table 6 displays the DMU's relative efficiency scores calculated using the all-units model, as well as their compensated (after the normalization) scores. After the compensation and subsequent normalization, "P9-Aug" was the single 100% efficient DMU, which was originally allocated to cluster C2, where it was 100% efficient as well (see table 3).

From the results in table 6, we may deduce that the plots P10 and P26 were those that faced the worst managerial practices. The data analyzed suggest that fertilization intended to complement and elevated levels of K and OM in P10, and B in P26, may contribute to the increase of productive outcomes. Another relevant aspect is that most farmers shall make efforts to maintain good levels of production along the year, not solely during May, July and August.

5 Conclusions

This study provided an evaluation of the productive performance of family farms of PAF Ducts project. In the analysis, we used a unitary input DEA model combined to the SOMs procedure, to set homogeneous clusters, in accordance to

Table 5: Relative efficiency scores calculated by the all-units DEA-CCR model, as well as the compensated normalized scores.

Cluster	DMU	All-units model	Compensating algorithm	Cluster	DMU	All-units model	Compensating algorithm
		Eff	Eff			Eff	Eff
C4	P3-Feb	36.7%	39.0%	C3	P15-Jun	26.7%	43.8%
C1	P6-Feb	13.3%	12.4%	C4	P3-Jul	66.7%	70.8%
C1	P7-Feb	46.7%	43.4%	C1	P6-Jul	93.3%	86.8%
C2	P11-Feb	13.3%	15.4%	C1	P7-Jul	90.0%	83.7%
C2	P12-Feb	13.3%	15.4%	C2	P9-Jul	73.3%	84.6%
C3	P15-Feb	16.7%	27.4%	C1	P10-Jul	36.7%	34.1%
C4	P3-Mar	63.3%	67.3%	C2	P11-Jul	76.7%	88.5%
C1	P6-Mar	56.7%	52.7%	C2	P12-Jul	60.0%	69.2%
C1	P7-Mar	76.7%	71.3%	C3	P15-Jul	53.3%	87.5%
C2	P9-Mar	26.7%	30.8%	C4	P3-Aug	87.5%	93.0%
C2	P11-Mar	56.7%	65.4%	C1	P6-Aug	100.0%	93.0%
C2	P12-Mar	30.0%	34.6%	C1	P7-Aug	100.0%	93.0%
C3	P15-Mar	23.3%	38.3%	C2	P9-Aug	86.7%	100.0%
C4	P3-Apr	70.0%	74.4%	C1	P10-Aug	43.3%	40.3%
C1	P6-Apr	33.3%	31.0%	C2	P11-Aug	74.9%	86.4%
C1	P7-Apr	63.3%	58.9%	C2	P12-Aug	74.5%	86.0%
C2	P9-Apr	20.0%	23.1%	C3	P15-Aug	40.0%	65.6%
C1	P10-Apr	53.3%	49.6%	C4	P3-Sep	56.7%	60.2%
C2	P11-Apr	43.3%	50.0%	C1	P6-Sep	83.6%	77.8%
C4	P3-May	63.3%	67.3%	C1	P7-Sep	84.2%	78.3%
C1	P6-May	93.3%	86.8%	C2	P9-Sep	46.7%	53.8%
C1	P7-May	73.3%	68.2%	C1	P10-Sep	26.7%	24.8%
C2	P9-May	60.0%	69.2%	C2	P11-Sep	56.7%	65.4%
C1	P10-May	56.7%	52.7%	C2	P12-Sep	53.3%	61.5%
C2	P11-May	76.7%	88.5%	C3	P15-Sep	20.6%	33.8%
C2	P12-May	46.7%	53.8%	C4	P26-Sep	6.7%	7.1%
C3	P15-May	60.0%	98.5%	C4	P3-Oct	76.7%	81.5%
C4	P3-Jun	53.3%	56.7%	C1	P6-Oct	80.1%	74.5%
C1	P6-Jun	73.3%	68.2%	C1	P7-Oct	76.7%	71.3%
C1	P7-Jun	70.0%	65.1%	C2	P9-Oct	53.3%	61.5%
C2	P9-Jun	56.7%	65.4%	C1	P10-Oct	40.7%	37.9%
C1	P10-Jun	36.7%	34.1%	C2	P11-Oct	43.3%	50.0%
C2	P11-Jun	50.0%	57.7%	C2	P12-Oct	43.3%	50.0%
C2	P12-Jun	30.0%	34.6%	C3	P15-Oct	26.7%	43.8%

criteria related to soil fertility. Furthermore, we applied an algorithm that enables to compensate for the non-homogeneity of the plots. So that, a direct comparison among the relative efficiency scores from different clusters became possible.

In addition, it was found that both the relative efficiency and the maintenance of the soil fertility in the plot result from the interaction of several variables related to the chemical composition of the soil. Among the variables analyzed, it was found that the levels of pH, K, B, OM and BS were those that greatly contributed to the promotion of relative efficiency. This suggests that the proper soil management helps the sustainability of the agricultural activity, fomenting the preservation or even increasing soil fertility.

Table 6: Average scores using the within-cluster model and the compensating (normalized) algorithm per plot, month and cluster.

Plot	Average scores		Month	Average scores	Cluster	Average scores	
	Within cluster	Compensating algorithm		Compensating algorithm		Within cluster	Compensating algorithm
P3	63.8%	67.8%	Feb	25.5%	C1	64.1%	59.6%
P6	69.7%	63.2%	Mar	51.5%	C2	59.6%	58.4%
P7	75.7%	70.3%	Apr	47.8%	C3	63.6%	61.7%
P9	54.3%	61.0%	May	73.1%	C4	66.9%	54.8%
P10	42.0%	39.1%	Jun	53.2%			
P11	64.8%	63.0%	Jul	75.7%			
P12	52.4%	50.7%	Aug	82.2%			
P15	63.6%	54.8%	Sep	51.4%			
P26	6.7%	7.1%	Oct	58.8%			

Remarkably, the definition of clusters, through the SOMs, combined to the use of the DEA model, proved very promising. Besides, it corroborated the connection between the levels of the chemical elements present in the soil composition and related to its fertility with the relative efficiency in agricultural activity. We believe the methodological integration proposed herein may contribute to the improvement of the management of family agriculture with ecological concerns, as it may effectively be employed to assist small farmers in the decision-making process (e.g. what to plant, how many varieties, when to start etc.).

A possible extension for this work consists of using the so-called dynamic clustering [23] combined to the DEA model, replacing the (static) clustering method applied herein. By doing this, although indirectly, an overall comparison among all DMUs can be made, even in the clustered model, provided that no cluster is disjoint in relation to all the others.

References

[1] Charnes, A., Cooper & W., Rhodes, E., Measuring efficiency of decision making units. *European Journal of Operational Research*, 2(6), pp. 429–444, 1978.

[2] Kohonen, T., Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), pp. 59–69, 1982.

[3] Gomes, E.G., Soares de Mello, J.C.C.B. & Freitas, A.C.R., Efficiency measures for a non-homogeneous group of family farmers. *Pesquisa Operacional*, 32(3), pp. 561–574, 2012.

[4] Mangiameli P., Chen, S.K. & West, D., A comparison of SOM neural networks and hierarchical clustering methods. *European Journal of Operations Research*, 93(2), pp. 402–417, 1996.

[5] Liu, J.S., Lu, L.Y., Lu, W.M. & Lin, B.J., Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, 41(1), pp. 3–15, 2013.

[6] Banker, R.D., Charnes, A. & Cooper, W.W., Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), pp. 1078–1092, 1984.

[7] Mousavi-Avval, S.H., Rafiee, S. & Mohammadi, A., Optimization of energy consumption and input costs for apple production in Iran using data envelopment analysis. *Energy*, 36(2), pp. 909–916, 2011.

[8] Candemir, M., Özcan, M., Günes, M. & Deliktas, E., Technical efficiency and total factor productivity growth in the hazelnut agricultural sales cooperatives unions in Turkey. *Mathematical and Computational Applications*, 16(1), pp. 66–76, 2011.

[9] Gomes, E.G., Abreu, U.G.P., Soares de Mello, J.C.C.B., Carvalho, T.B. & Zen, S., Unitary input DEA model to identify beef cattle production systems typologies. *Pesquisa Operacional*, 32(2), pp. 389–406, 2012.

[10] Soares de Mello, J.C.C B., Gomes, E.G., Angulo-Meza, L., Biondi Neto, L., Abreu, U.G.P., Carvalho, T.B. & Zen, S., Ex-post clustering of Brazilian beef cattle farms using SOMs and cross-evaluation DEA models (Chapter 4). *Applications of self-organizing maps*, ed. M. Johnsson, Intech: Croatia, pp. 67–88, 2012.

[11] Lima, J.S., Bezerra Neto, F., Gomes, E.G., Negreiros, M.Z., Pontes, F.S., Medeiros, M.A. & Barros Júnior, A.P., Agroeconomic evaluation of intercropping rocket and carrot by uni- and multivariate analyses in a semi-arid region of Brazil. *Ecological Indicators*, 41, pp. 109–114, 2014.

[12] Angulo-Meza, L., Biondi Neto, L., Brandão, L.C., Andrade, F.V.S., Soares de Mello, J.C.C.B. & Coelho, P.H.G., Modelling with self-organizing maps and data envelopment analysis: A case study in educational evaluation (Chapter 4). *Self-organizing maps - applications and novel algorithm design*, ed. J.I. Mwasiagi, Intech: Crotia, pp. 71–88, 2011.

[13] Lovell, C. & Pastor, J.T., Radial DEA models without inputs or without outputs. *European Journal of Operational Research*, 118(1), pp. 46–51, 1999.

[14] Haas, D.A. & Murphy, F.H., Compensating for non-homogeneity in decision-making units in data envelopment analysis. *European Journal of Operational Research*, 144(3), pp. 530–544, 2009.

[15] Cook, W.D., Harrison, J., Imanirad, R., Rouse, P. & Zhu, J., Data Envelopment Analysis with Nonhomogeneous DMUs. *Operations Research*, 61(3), pp. 666–676, 2013.

[16] Bertoloto, R.F. & Soares de Mello, J.C.C.B., Eficiência de portos e terminais privativos brasileiros com características distintas. *Journal of Transport Literature*, 5(2), pp. 4–21, 2011.

[17] Gomes, E.G., Soares de Mello, J.C.C.B., Souza, G.S., Angulo-Meza, L. & Mangabeira, J.A.C., Efficiency and sustainability assessment for a group of farmers in the Brazilian Amazon. *Annals of Operations Research*, 169(1), pp. 167–181, 2009.

[18] Podinovski, V.V. & Thanassoulis, E., Improving discrimination in data envelopment analysis: Some practical suggestions. *Journal of Productivity Analysis*, 28(1–2), pp. 117–126, 2007.

[19] Malmquist, S., Index numbers and indifference surfaces. *Trabajos de Estadistica y de Investigacion Operativa*, 4(2), pp. 209–242, 1953.
[20] Golany, B. & Roll, Y., An application procedure for DEA. *Omega*, 17(3), pp. 237–250, 1989.
[21] Angulo-Meza, L., Biondi Neto, L., Soares de Mello, J.C.C.B. & Gomes, E.G., ISYDS - Integrated System for Decision Support (SIAD - Sistema Integrado de Apoio à Decisão): a software package for data envelopment analysis model. *Pesquisa Operacional*, 25(3), pp. 493–503, 2005.
[22] Ali, A.I, Computational aspects of Data Envelopment Analysis (Chapter 4). *DEA: Theory, Methodology and Applications*, eds. A. Charnes, W.W Cooper, A.Y. Lewin, L.M. Seiford, Springer: New York, pp. 63–88, 1994.
[23] Golany, B. & Thore, S., Restricted best practice selection in DEA: An overview with a case study evaluating the socio-economic performance of nations. *Annals of Operations Research*, 73, pp. 117–140, 1997.