Delineation of Site-Specific Management Zones Using Estimation of Distribution Algorithms

Jonás Velasco

CONACYT-Centro de Investigación en Matemáticas (CIMAT), A.C.
Aguascalientes, Mexico

jvelasco@cimat.mx

Salvador Vicencio

Centro de Investigación en Matemáticas (CIMAT), A.C.
Aguascalientes, Mexico

salvador.vicencio@cimat.mx

Jose A. Lozano

Basque Center for Applied Mathematics (BCAM)
University of the Basque Country UPV/EHU

San Sebastian, Spain

ja.lozano@ehu.es

Nestor M. Cid-Garcia

Laboratorio Nacional de Geointeligencia
CONACYT-Centro de Investigación en Ciencias de Información Geoespacial

Aguascalientes, Mexico

nxtr.cd@gmail.com

Abstract

In this paper, we present a novel methodology to solve the problem of delineating homogeneous site-specific management zones (*SSMZ*) in agricultural fields. This problem consists of dividing the field into small regions for which a specific rate of inputs is required. The objective is to minimize the number of management zones, which must be homogeneous according to a specific soil property: physical or chemical. Furthermore, as opposed to oval zones, *SSMZ* with rectangular shapes are preferable since they are more practical for agricultural

technologies. The methodology we propose is based on evolutionary computation, specifically on a class of the estimation of distribution algorithms ($EDAs$). One of the strongest contributions of this study is the representation used to model the management zones, which generates zones with orthogonal shapes, e.g., L or T shapes, and minimizes the number of zones required to delineate the field. The experimental results show that our method is efficient to solve real-field and randomly generated instances. The average improvement of our method consists in reducing the number of management zones in the agricultural fields concerning other operations research methods presented in the literature. The improvement depends on the size of the field and the level of homogeneity established for the resulting management zones.

Keywords: Site-specific management zones; estimation of distribution algorithms; orthogonal shapes; evolutionary computation; combinatorial optimization

1 Introduction

The problem of delineating *site-specific management zones (SSMZ)* in agricultural fields consists of generating sub-regions within a plot for which a specific rate of inputs is appropriate (Doerge, 1999; Moral et al., 2010). According to Plant (2001); Roudier et al. (2008); Zhang et al. (2016), rectangular management zones are more practical for farmers by reducing the difficulties of adopting variable technologies and facilitating the use of agriculture machinery. Furthermore, the management zones with rectangular shapes are more applicable for farming in underdeveloped areas because farmers can easily apply these management zones to reduce fertilizer input, labor costs, and environmental waste without using advanced agriculture machinery.

The management zones must be homogeneous according to a specific soil property, physical or chemical, such as organic matter *(OM)*, nitrogen *(N)*, phosphorus *(P)*, potential of hydrogen *(pH)*, potassium *(K)*, sodium *(Na)*, and the sum of bases *(SB)*, which is a mix of several properties. The delineation of management zones is a critical decision problem in agriculture since the soil characteristics have a strong impact on the crop yield. The chemical properties determine the application of inputs, e.g., fertilizers and pesticides, while the water for irrigation depends on the physical properties.

The integration of some information technologies, called Precision Agriculture *(PA)*, such as the Global Positioning Systems *(GPS)*, Geographical Information Systems *(GIS)*, and remote sensors, helps to improve crop pro-

ductivity and makes farm management better. The idea of the *PA* is to face the variability of the soil properties and doing the right management practice at the right place and at the right time (Bongiovanni and Lowenberg-Deboer, 2004; Janrao and Palivela, 2015; Mulla, 2013). However, in some countries, farmers attitudes and perceptions suggest that they are resistant to adopting unfamiliar technologies to improve agricultural management practices (Anastasiadis and Chukova, 2019; Watoo and Mugera, 2019).

In this context, the generation of *SSMZ* arises from the need of *PA* to deal with several factors, which are variable in space and time, that affect productivity and crop quality. Some of these factors, such as the heterogeneity of the physical and chemical soil properties, directly affect the water balance, the dynamics of nutrients, and the response to the application of inputs (Ortega and Santibáñez, 2007). The *SSMZ* helps to minimize the impact of spatial variability, allowing the site-specific application of inputs and making more effective the agricultural planning (Betzek et al., 2018; Castrignanò et al., 2018).

An advantage of the *SSMZ* is the correct application of inputs in each region of the plot only where and when they are necessary according to the real requirements of the crop, its phenological stage, and the soil properties of the field, which allows a reduction of the environmental impact and a saving of resources and investment capital. These critical parameters and the crop prices must be considered to improve the decision-making process in agricultural fields (González et al., 2020; López et al., 2020). This contrasts with the conventional management agricultural practices, where the uniform applications of inputs are made throughout the whole production cycle considering, just in some cases, the phenological stage of the crop, which increases the production costs and the unnecessary waste of resources, especially water (Janrao and Palivela, 2015; Mulla, 2013). As in other water supply systems, the decision about the amount of water to be irrigated in each irrigation period directly impacts the total costs of farmers (Santos et al., 2020). The benefits of the site-specific management zones in vineyards and some crops have been demonstrated in Ortega and Santibáñez (2007).

Delineating efficiently site-specific management zones is a big challenge for farmers and decision-makers. The most typical methodology used to solve the problem is the clustering method, which uses soil samples in conjunction with procedures such as *fuzzy k-means*, *fuzzy c-means*, and *k-means* (Betzek et al., 2018; Monzon et al., 2018; Oldoni et al., 2019; Ohana-Levi et al., 2019). The information for these algorithms is obtained from different methods such as analysis of topographic maps, statistical information, remote sensing data, or semivariogram analysis (Albornoz et al., 2018; Fu et al., 2010; Gaviola et al., 2019; Georgi et al., 2018; Gili et al., 2017; Haghverdi et al., 2015;

Hornung et al., 2006; Li et al., 2007; Molin and de Castro, 2008; Ortuani et al., 2019; Tagarakis et al., 2013). Although these methods obtain homogeneous management zones, in some cases, the solution can be harder to apply due to the structure of the generated regions, which are disjoint or with circular or irregular shapes.

Concerning operations research methods, the work of Cid-Garcia et al. (2013) presents one of the first mono-objective mathematical formulations of integer linear programming (*MILP*) to delineate rectangular and homogeneous management zones minimizing the total variance of the field for a specific soil property, physical or chemical. The complexity of this mathematical formulation is demonstrated by using a reduction to the 2D-Bin Packing Problem, which is NP-Hard (Chung et al., 1982). Therefore, the computational time required to solve the problem can increase exponentially with the size of the instance.

In Albornoz et al. (2015), the previous work was improved, showing a bi-objective mathematical formulation of integer linear programming (*BILP*) where: *a)* the number of management zones is minimized, and *b)* the homogeneity level within these zones is maximized. Some decomposition approaches to approximate a solution for the *SSMZ* problem are developed in Albornoz and Ñanco (2016) and Albornoz et al. (2019). In Saez and Albornoz (2016) the authors propose an approach to delineate *SSMZ* under uncertainty conditions. Other works that integrate the delineation of rectangular management zones and the crop planning problems are presented in Cid-Garcia and Ibarra-Rojas (2019); Albornoz et al. (2020); Albornoz and Zamora (2020).

The operations research methods mentioned above are based on the work of Cid-Garcia et al. (2013) and only consider regions with rectangular or square shapes to generate site-specific management zones, avoiding zones with orthogonal shapes, e.g., T or L shape, which can be used to partitioning the field. Fig. 1 shows the delineation for an agricultural field using the organic matter as chemical soil property with 40 soil samples and around of 7.82 ha (256 m width and 305.6 m long) presented in the work of Cid-Garcia et al. (2013). Fig. 1a is the thematic map obtained with specialized software such as MapInfo, before the delineation, showing the variability of the field and the number of soil samples (black points). Green zones represent the ideal level of *OM*; red or blue zones denote upper or lower levels, respectively. Fig. 1b is the resulting delineation for the exact method proposed by Albornoz et al. (2015), and Fig. 1c is the resulting delineation using a clustering method with specialized software (*MapInfo*), which can be harder to adopt by farmers considering the resulting zones (disjoint and with irregular shapes).

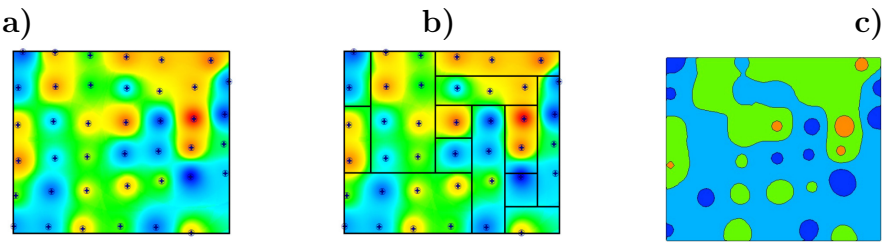

Figure 1: Delineation of *SSMZ* for an agricultural field close to Santiago, Chile, using organic matter as chemical soil property. Fig. 1a is the thematic map for the field before the delineation, Fig. 1b is the delineation using an operations research technique, and Fig. 1c is the delineation using clustering methods.

The objective of this article is to present a new methodology based on soil samples and an evolutionary algorithm, specifically on an estimation of distribution algorithm (*EDA*), to delineate agricultural fields, minimizing the number of management zones and satisfying a specific level of homogeneity for each zone (the details for the *EDA* are given in Section 2.2). We consider the computational complexity of the *SSMZ* problem and the possibility of obtaining management zones with orthogonal shapes, which can be harder to generate in the exact approaches mentioned above. Also, our algorithm generates the zones during its execution, instead of using a preprocessing stage to generate predefined management zones such as some previous approaches. Furthermore, the improvement of our method consists in reducing the number of management zones in the agricultural fields with respect to the operations research methods. This improvement depends on the size of the field (number of soil samples), and the level of homogeneity established for the resulting management zones.

The selection of an *EDA* approach is due to their applicability to solve other complex combinatorial optimization problems, which include multi-objective knapsack, routing, scheduling, forest management, portfolio management, environmental monitoring network design, and bioinformatics (Larrañaga and Lozano, 2002; Armañanzas et al., 2008; Hauschild and Pelikan, 2011; Ceberio et al., 2013; Wang et al., 2015). With the *EDA*, we generate zones with an orthogonal shape that can be used in the delineation of management zones on the field. To the best of our knowledge, this is one of the first approaches with these characteristics.

The rest of this paper is organized as follows. Section 2 presents the materials and methods to solve the *SSMZ* problem. Section 3 shows the experimental results for our *EDA*, and compares them with some exact ap-

proaches of the literature to validate its efficiency. Finally, Section 4 gives some conclusions and recommendations.

2 Materials and methods

In Fig. 2, we present an overview of the methodology used to solve the *SSMZ* problem. It is composed of two main steps: (I) collecting soil samples in the agricultural fields (see Section 2.1); and (II) designing and implementing a solution method based on an *EDA* algorithm (we call *EDA-SSMZ*) to obtain high-quality solutions in acceptable computational times (see Section 2.2).

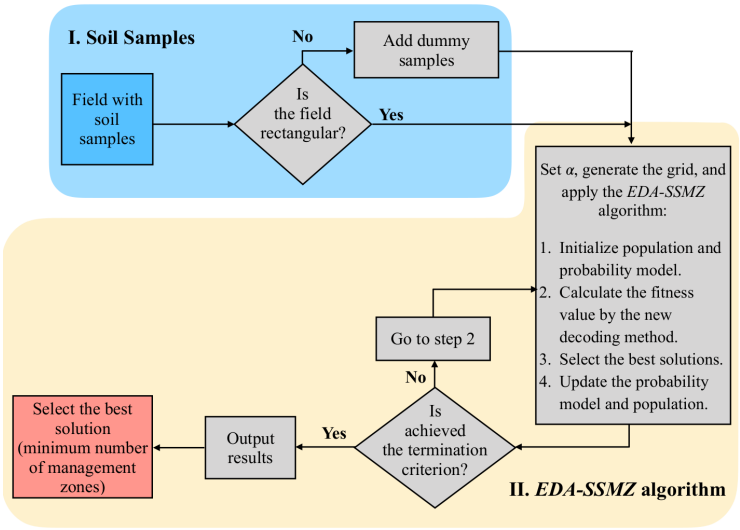

Figure 2: Summary of our methodology.

2.1 Soil samples

The first step in the *SSMZ* problem consists of collecting soil samples from the agricultural field that are represented by a grid where each soil sample is equidistant one from each other. The sampling is made to obtain information about the soil according to a specific soil property, chemical or physical. The chemical properties are used to determine the seeds, fertilizers, and pesticides to supply to the crop. The physical properties impact the amount of water needed in the irrigation process. The number of soil samples for each field

depends on the farmer's investment[1] and not necessarily on the field size, i.e., a plot with 10 ha can take the same number of soil samples as one with 30 ha. A way to visualize the soil variability of these properties is by generating thematic maps with specialized software, as in Fig. 3a, where it is possible to create a grid inside of the field with the resulting soil samples (see Fig. 3b). In this sense, each soil sample represents the center of each square in the grid.

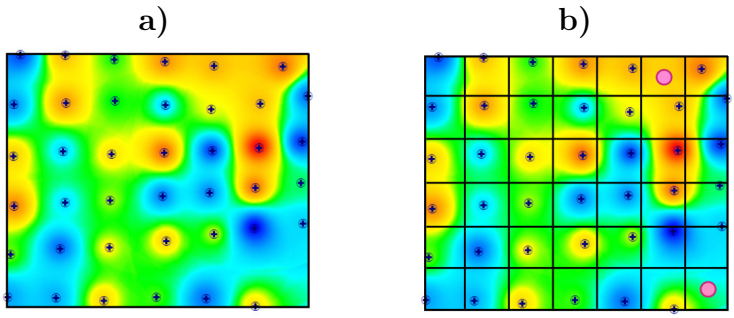

Figure 3: Generation of a grid for an agricultural field considering the organic matter (OM) as chemical soil property. Fig. 3a is the thematic map for the field and Fig. 3b is the resulting grid. The pink circles represent dummy samples.

When the fields are no initially with a rectangular or square shape, i.e., when the soil sampling does not generate a perfect grid, then the methodology adds dummy soil samples to complete the grid and enforces each soil sample as the center of each square. A perfect grid facilitates and improves the performance of the *EDA-SSMZ* algorithm. The number of dummy samples depends on the original soil sampling and the shape of the field. The *EDA-SSMZ* algorithm uses the variability in the soil samples to delineate the management zones considering a specific chemical or physical property. Therefore, to avoid using dummy soil samples as a new management zone in the final delineation, the value for each one is established by the decision-maker considering the same concentration as a specific neighbor (a spatial correlation among the soil samples can be considered). The procedures of the *EDA-SSMZ* ensure that a management zone cannot group only dummy soil samples. With this, the use of dummy samples is not a disadvantage for the method. Fig. 3b shows two dummy soil samples (pink circles) in the agricultural field (more details can be found in Cid-Garcia et al. (2013)).

[1]In Mexico, the costs per soil sample are around $50 dollars (INIFAP), which determines the number of samples in the fields of the farmer.

The relative variance *(RV)* constitutes an excellent criterion to prove the efficiency of a zoning method (Ortega and Santibáñez, 2007). Suppose a set M of management zones in the field, then the RV is defined as:

$$RV(M) = 1 - \frac{\sum_{m \in M}(n_m - 1)\sigma_m^2}{\sigma_T^2 \left[N - |M|\right]}, \tag{1}$$

where σ_T^2 is the total variance of the field, N is the total number of soil samples, $|M|$ is the number of management zones used to delineate the field, n_m the number of soil samples in zone m, and σ_m^2 is the variance within the management zone m. In this sense, σ_T^2 and σ_m^2 are calculated considering the formula of sample variance. According to the experts, to guarantee a homogeneous behavior of the zoning method, the relative variance must be greater than or equal to 0.5 (an alpha parameter (α)). The highest values for α mean the highest levels of homogeneity in the management zones. Notice that the range of values for α is $[0, 1]$.

$$1 - \frac{\sum_{m \in M}(n_m - 1)\sigma_m^2}{\sigma_T^2 \left[N - |M|\right]} \geq \alpha. \tag{2}$$

2.2 Estimation of Distribution Algorithms (EDA)

An estimation of distribution algorithm is a class of population-based optimization algorithm that extracts statistical information from the population of solutions, to generate new ones. The algorithm starts by generating a population of candidate solutions. These solutions are evaluated using an objective function. Based on this evaluation, a subset of solutions is selected using a selection method, and the population of the selected individuals is used to estimate a probability distribution. Finally, a new set of solutions is sampled from the estimated distribution, generating a new population of solutions, and the algorithm iterates again. The procedure ends when a stopping criterion, previously established, is reached. In our algorithm, the best solution represents the solution with the minimum number of management zones that satisfy the homogeneity level (α) established by the decision-maker, and the stopping criterion is fixed with a maximum number of iterations. For further reference about *EDAs*, the reader can consult (Larrañaga and Lozano, 2002).

There exist many different *EDAs* that differ one with each other by the probabilistic models used and their construction. In this work, we are going to consider the univariate marginal distribution algorithm (*UMDA*), which is the most basic *EDA* and was introduced by Mühlenbein and Paaß (1996) and Mühlenbein (1997) for binary optimization problems in the late 1990s.

Mathematically, a univariate model decomposes the probability of a candidate solution $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ into the product of probabilities of individual variables as

$$p(\boldsymbol{x}) = p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n)$$

where $p(x_i)$ is the probability of variable x_i, and $p(x_1, x_2, \ldots, x_n)$ is the probability of the candidate solution $(x_1, x_2, \ldots, x_n)$. In the case of binary problems, we can define the probability of x_i as $p(x_i = 1) = r_i$ and $p(x_i = 0) = 1 - r_i$, each x_i following a Bernouilli distribution with a parameter value equal to r_i. On the other hand, we use D, to represent the population of N individuals with n binary variables. In (3), we present a short example with a population size set of $N = 10$ and $n = 6$ binary variables per solution. Assuming that the initial population is obtained at random by sampling the following probability function $p(x_i = 1) = 0.5$ for $i = 1, \ldots, 6$, then a possible D is:

$$\begin{array}{llll}
\boldsymbol{x_1} = (0,1,1,1,1,0) & f(\boldsymbol{x_1}) = 2, & \boldsymbol{x_2} = (0,1,1,1,1,1) & f(\boldsymbol{x_2}) = 3, \\
\boldsymbol{x_3} = (1,0,0,1,1,0) & f(\boldsymbol{x_3}) = 1, & \boldsymbol{x_4} = (1,1,1,0,1,0) & f(\boldsymbol{x_4}) = 1, \\
\boldsymbol{x_5} = (0,1,0,0,0,1) & f(\boldsymbol{x_5}) = 2, & \boldsymbol{x_6} = (0,1,0,0,1,0) & f(\boldsymbol{x_6}) = 4, \\
\boldsymbol{x_7} = (0,0,1,1,1,0) & f(\boldsymbol{x_7}) = 4, & \boldsymbol{x_8} = (1,0,1,0,1,0) & f(\boldsymbol{x_8}) = 5, \\
\boldsymbol{x_9} = (0,1,0,0,0,0) & f(\boldsymbol{x_9}) = 5, & \boldsymbol{x_{10}} = (0,1,1,1,1,1) & f(\boldsymbol{x_{10}}) = 3,
\end{array} \tag{3}$$

where f is a fitness function, and $f(\boldsymbol{x})$ is the fitness value of each individual, $\boldsymbol{x}$. Similarly, we use D^{Se}, to represent the population of the selected Se individuals from D, where $Se < N$. This can be done using one of the standard selection methods that are common in evolutionary computation, and which use information from the fitness function. Hence, individuals with better fitness values have a bigger chance of being selected. Let us assume that our selection method is truncation and that we select half of the population, i.e., $Se = 5$. The population of selected individuals D^{Se} is represented by (4):

$$\begin{array}{l}
\boldsymbol{x_1} = (0,1,1,1,1,0), \\
\boldsymbol{x_2} = (0,1,1,1,1,1), \\
\boldsymbol{x_3} = (1,0,0,1,1,0), \\
\boldsymbol{x_4} = (1,1,1,0,1,0), \\
\boldsymbol{x_5} = (0,1,0,0,0,1).
\end{array} \tag{4}$$

In *UMDA*, the interest is to estimate $p(\boldsymbol{x} \mid D^{Se})$, that is, the joint probability distribution over one individual $\boldsymbol{x}$ being among the selected individuals

D^{Se}. Therefore, $p(x_i \mid D^{Se})$ with $i = 1, \ldots, 6$ is estimated from D^{Se} using its corresponding relative frequency, $p(x_i = 1 \mid D^{Se})$. Thus, using the information from (4), the univariate marginal frequencies are:

$$\begin{aligned} &p(x_1 = 1 \mid D^{Se}) = 2/5, \qquad p(x_2 = 1 \mid D^{Se}) = 4/5, \\ &p(x_3 = 1 \mid D^{Se}) = 3/5, \qquad p(x_4 = 1 \mid D^{Se}) = 3/5, \\ &p(x_5 = 1 \mid D^{Se}) = 4/5, \qquad p(x_6 = 1 \mid D^{Se}) = 2/5. \end{aligned} \tag{5}$$

Consequently, with the learned model and the values of (5), we can generate the next population D, where the first binary variable for each new candidate solution has a probability of 0.4 of being 1, and a 0.6 chance of being a 0. The second one has a 0.8 chance of being 1, and 0.2 of being a 0, and so on. Finally, we repeat the selection, estimation, and sampling steps, until a stopping criterion is reached, e.g., a maximum number of generations $Tmax$. The general form of the *UMDA* is as follows:

- **STEP 0 (Initialization):** Set $t \leftarrow 1$. $D_0 \leftarrow$ Generate $N > 0$ individuals randomly.
- **STEP 1 (Selection):** $D_{t-1}^{Se} \leftarrow$ Select $Se < N$ individuals for D_{t-1} according to a selection method.
- **STEP 2 (Estimation):** $p_t(x_i|D_{t-1}^{Se}) \leftarrow$ Compute the univariate marginal frequencies of the selected set.
- **STEP 3 (Sampling):** $D_t \leftarrow$ Generate N new individuals according to the distribution $p_t(\boldsymbol{x}) = \prod_{i=1}^{n} p_t(x_i|D_{t-1}^{Se})$. Set $t \leftarrow t + 1$.
- **STEP 4:** If termination criteria are not met, go to STEP 1.

2.2.1 EDA for the SSMZ problem

In this section, we define the methodology to solve the *SSMZ* problem using the population-based method of *UMDA* along with the objective function to evaluate the fitness for each candidate solution, which we call as *EDA-SSMZ* algorithm.

Individual Representation

Designing any iterative metaheuristic needs a representation of a solution. The individual representation plays a major role in the efficiency and effectiveness of any metaheuristic, and constitutes an essential step in designing a metaheuristic. A solution for the *EDA-SSMZ* problem represents a partition

of the field using a specific number of management zones. We implement an indirect encoding (representation) for the site-specific management zones problem. First, each possible edge inside of the grid is enumerated and labeled to create the representation of solutions for the *SSMZ* problem. Then, the candidate solution is generated with a vector of 1's or 0's represented by $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ where x_i indicates if the i^{th} edge inside of the grid (binary variable) appears in the solution or not. The size of the search space is determined by $2^{[n_c \cdot (n_r - 1) + n_r \cdot (n_c - 1)]}$ possible candidate solutions, where n_r and n_c represents the number of rows and columns inside of the grid, respectively.

Fig. 4 shows the representatiton (binary enconding) for a plot with 16 soil samples, 4×4, where Fig. 4a shows the sequence of the 16 soil samples numbered by $m_1, m_2, \ldots, m_{16}$, and the total edges inside of the grid (binary variables) numbered by $1, 2, \ldots, 24$. Fig. 4b illustrates a candidate solution for the *SSMZ* problem given by

$$\boldsymbol{x} = (0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0)$$

where it is possible to determine which edges inside of the grid are activated, and therefore, which soil samples correspond to which management zone. Fig. 4c illustrates a partition of the field with six management zones where q_1 contains m_1 and m_2; q_2 contains m_3, m_4, m_6, m_7, m_8, m_{10}, and m_{14}; q_3 contains m_5; q_4 contains m_9 and m_{13}; q_5 contains m_{11}; and q_6 contains m_{12}, m_{15}, and m_{16}. Note that, $q_i \subseteq S$, where $S = \{m_1, m_2 \ldots, m_M\}$ denotes the set of soil samples on the grid, and q_i represents a subset of the soil samples in the management zone i. We represent the set of management zones Q as a collection of subsets (partitions) of the soil samples $\{q_1, q_2, \ldots, q_k\}$, and $|Q|$ as the total number of management zones.

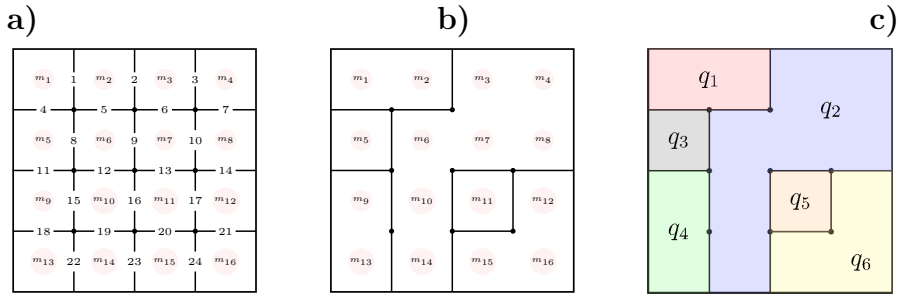

Figure 4: Individual representation for the SSMZ problem. Fig. 4a is the binary encoding, Fig. 4b is a candidate solution, and Fig. 4c is a partition of the field with six management zones.

Our indirect encoding can generate candidate solutions that represent the

same delineation of the field, e.g., Figs. 5a–5d show four candidate solutions that produce a similar result as Fig. 4b. In this particular case, these candidate solutions show a delineation of the field that corresponds to the same solution illustrated in Fig. 4c with six management zones. Notice that, in the candidate solutions of Figs. 5a–5d, not all activated edges separate two zones, i.e., some edges are isolated, which does not generate infeasibility. An isolated edge does not split a region of the plot into two different zones. For example, Figs. 5a–5d show the isolated edges for the four candidate solutions, which correspond to the edges labeled with the numbers 3, 6, 7, and 10, respectively. Moreover, the procedures of the *EDA-SSMZ* ignore these isolated edges to avoid them in the final delineation.

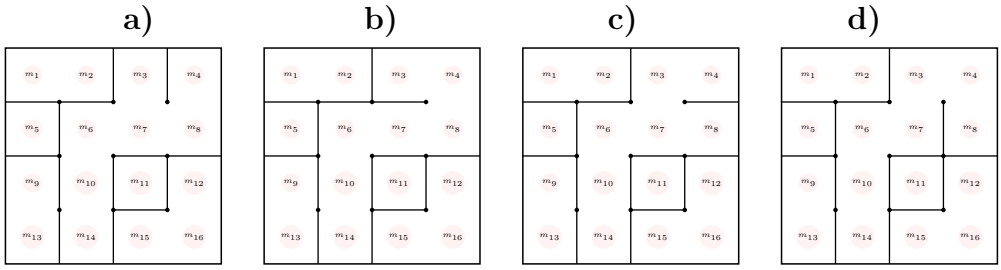

Figure 5: Four candidate solutions representing a similar delineation as Fig. 4b.

The *EDAs* build and maintain a probability distribution of the current population over the search space, from which the next generation of individuals is sampled. The fact that our algorithm obtains similar solutions tells us that it has converged, and therefore an edges pattern on the grid is learned. The above means that the probability of variable x_i could remain fixed at either zero or one, obstructing some search space regions in the next generation. On the other hand, the chance that some edges are active or not on the grid depends on whether their marginal frequencies are different from zero or one.

Fitness Function

The objective of the fitness function is to have the means to evaluate each one of the possible individuals so that the search algorithm can compare the different solutions and act in consequence to find the best solution. It is important to define it appropriately to assess the search for the *SSMZ* problem. For our *EDA-SSMZ*, the best solution represents the solution with the minimum number of management zones.

First, let us define our *SSMZ* problem as a combinatorial optimization problem, which can be described as finding a k-partition of the field by using a candidate solution, $\boldsymbol{x}$. The objective is minimizing the number of management zones (partitions) given by $f(\boldsymbol{x}) = |(q_1, \ldots, q_k)|$, subject to some constraints in the shape of these partitions and on the homogeneity level of the soil samples inside each partition $i \in K$, where $K \in \{1, ..., k\}$. The mathematical model for the *SSMZ* problem is expressed as follows:

$$\begin{aligned}
\text{Minimize} \quad & f(\boldsymbol{x}) = |(q_1, \ldots, q_k)| && && (6)\\
\text{subject to :} \quad & q_i \neq \emptyset, && i \in K && (7)\\
& \bigcup_{i \in K} q_i = S, && && (8)\\
& q_{i_1} \cap q_{i_2} = \emptyset, && i_1, i_2 \in K && (9)\\
& \delta(q_i) \text{ is connected}, && i \in K, && (10)\\
& h(Q, S) \geq \alpha. && && (11)
\end{aligned}$$

where (6) is the fitness function that minimizes the number of management zones. Constraints (7)–(9) define the k-partition: with no-empty zones, all soil samples are covered by the zones generated, and there is no intersection of soil samples between zones, respectively. Constraints (10) represent the contiguity of the soil samples, where $\delta(q_i)$ represents the set of soil samples that are adjacent to at least one soil sample of q_i. Constraint (11) guarantees the homogeneity of the management zones $Q = \{q_1, \ldots, q_k\}$. Note that constraints (7) to (10) can be easily handled by the representation proposed in the previous section. For constraint (11), it is necessary to introduce a special constraint handling strategy. In this work, we use a simple penalizing mechanism where infeasible solutions are considered during the search process.

The penalty function modifies the original fitness function $f(\boldsymbol{x})$ applied to a candidate solution $\boldsymbol{x}$ such that $f'(\boldsymbol{x}) = f(\boldsymbol{x}) + \bar{P}(\boldsymbol{x})$, where $\bar{P}(\boldsymbol{x})$ is a distance metric from the infeasible point to the feasible region $\mathcal{F}$ (this might be simply a count of the number of constraints violated). The penalty function $\bar{P}$ is zero for feasible solutions, and it increases with distance from the feasible region (for minimization problems). Equation (12) shows that one simple strategy is to calculate the homogeneity of the management zones (based on Equation (2))

$$h(Q, S) = \left(1 - \frac{\sum_{i \in K} \left[s(q_i) - 1\right] \sigma^2(q_i)}{\sigma_T^2(S) \left[|S| - |Q|\right]}\right), \tag{12}$$

and then we use the penalty function

$$\bar{P}(\boldsymbol{x}) = \begin{cases} 0, & \text{if} \quad h(Q,S) \geq \alpha, \\ \bar{M} - h(Q,S), & \text{if} \quad h(Q,S) < \alpha, \end{cases} \tag{13}$$

where α is a given homogeneity parameter, and the fixed number $\bar{M}$ is large enough that feasible solutions are preferred; $s(q_i)$ represents the number of soil samples in the management zone q_i; $\sigma_i^2(q_i)$ represents the variance in q_i, and finally, $\sigma_T^2(S)$ represents the total variance about the set of soil samples S in the field. The $\bar{M}$ value is setting as $\bar{M} = |S| \cdot 10$, where $|S|$ is the number of soil samples. With this, we ensure that $\bar{M}$ es large enough for this purpose. Therefore, in the case of some infeasible solution, $\bar{P}(\boldsymbol{x})$ is one order of magnitude more than $f(\boldsymbol{x})$. If the value of $\bar{M}$ is large enough, then infeasible points near the constraint boundary will be discarded, which may delay, or even prevent, the exploration of this region. On the other hand, if $\bar{M}$ is not large enough, then solutions in infeasible regions may dominate the feasible ones.

The fitness evaluation process requires a search process over an adjacency list to find connected soil samples for each management zone. In the worst case, the search process is $O(n \cdot |S|)$, where $|S|$ is the number of soil samples, and n is the number of edges inside of the grid. Finally, the algorithm determines which soil samples belong to which management zones and computes their homogeneity level.

EDA-SSMZ algorithm

The most representative steps for the proposed *EDA-SSMZ* are presented by the Algorithm 1 that takes as inputs the population size (N), the initial probability vector ($p_0(\boldsymbol{x})$), the selection size (S_e), the homogeneity parameter (α), and the maximum number of iterations ($Tmax$). With the procedure $InitializePopulation(N, p_0(\boldsymbol{x}))$, an initial population of N individuals is generated at random by sampling several Bernouilli distributions using the initial probability vector $p_0(\boldsymbol{x})$. Then, with $EvaluatePopulation(D, \alpha)$, a fitness function $f'(\boldsymbol{x})$ is evaluated, which weighs the infeasibilities using the objective (6) and the constraint (11). In this step, we evaluate each individual of the population D using the parameter of homogeneity α, and store the best individual in Q^{best} obtained with $GetBestSolution(D)$. Procedure $SelectBestSolutions(D_{t-1}, Se)$ selects the best Se individuals from population D_{t-1}, according to the fitness function. Then the joint probability distribution $p_t(\boldsymbol{x})$ is estimated with $CalculateMarginalFrequency(D_{t-1}^{Se})$, using the population of the selected individuals, D_{t-1}^{Se}. Procedure $GeneratePopulation(N, p_t(\boldsymbol{x}))$ generates the new population of solutions using the estimated probability

model $p_t(\boldsymbol{x})$. The algorithm evaluates the new individuals with $EvaluatePopulation(D, \alpha)$, gets the best individual of the population with $GetBestSolution(D)$, and stores it in $Best$. Finally, the algorithm updates the best solution, comparing the best current solution with the solution obtained in the previous iteration ($UpdateBestSolution(Q^{best}, Best)$). The above step preserves the best solution (or incumbent for short) to the current iteration. The *EDA-SSMZ* algorithm iterates until the maximum of iterations $Tmax$ has been reached and returns the best solution found, Q^{best}.

Algorithm 1 *EDA-SSMZ*

Input:

$\alpha :=$ homogeneity parameter
$S_e :=$ selection size
$N :=$ population size
$p_0(\boldsymbol{x}) :=$ initial probability vector
$Tmax :=$ maximum number of iterations

Output: Q^{best}: A feasible solution, k-partition of S (soil samples)

1: $t \leftarrow 1$
2: $D_0 \leftarrow InitializePopulation(N, p_0(\boldsymbol{x}))$
3: $EvaluatePopulation(D_0, \alpha)$
4: $Q^{best} \leftarrow GetBestSolution(D_0)$
5: **for** $t = 1, 2, ..., Tmax$ **do**
6: $\quad D_{t-1}^{Se} \leftarrow SelectBestSolutions(D_{t-1}, Se)$
7: $\quad p_t(\boldsymbol{x}) \leftarrow CalculateMarginalFrequency(D_{t-1}^{Se})$
8: $\quad D_t \leftarrow GeneratePopulation(N, p_t(\boldsymbol{x}))$
9: $\quad EvaluatePopulation(D_t, \alpha)$
10: $\quad Best \leftarrow GetBestSolution(D_t)$
11: $\quad Q^{best} \leftarrow UpdateBestSolution(Q^{best}, Best)$
12: $\quad t \leftarrow t + 1$
13: **end for**
14: **return** Q^{best}

3 Experimental Results

In this section, we present the experimental results to validate the performance of our *EDA-SSMZ* algorithm. In Section 3.1, we describe two exact approaches used to compare our methodology. In Section 3.2, we show the

set of instances used to test our algorithm. In Section 3.3, we present the calibration of the critical parameters for the *EDA*. Finally, Sections 3.4 and 3.5 show the experimentation and some graphical visualizations of the results, respectively.

3.1 Benchmark algorithms

We compare the *EDA-SSMZ* algorithm with the exact approaches of Cid-Garcia et al. (2013) and Albornoz et al. (2015) that proposed a mono-objective mathematical formulation of integer linear programming *(MILP)* and a bi-objective mathematical formulation of integer linear programming *(BILP)*, respectively. To the best of our knowledge, these approaches are the first ones in the literature to generate rectangular and homogeneous management zones by using operations research techniques.

3.2 Test problem instances

3.2.1 Real-field instances

To evaluate the performance of the algorithm, we used the real-field instances proposed by Cid-Garcia et al. (2013) and adapted in the work of Albornoz et al. (2015). These instances show an agricultural field with 40 soil samples, which extract information about the following soil properties: organic matter, potential of hydrogen, phosphorous, and sum of bases. For these instances, two dummy soil samples were considered to complete a perfect grid (the pink circles of Fig. 3b), and their values were fixed considering the neighbor of the left.

3.2.2 Randomly generated instances

Another set of instances was generated at random to evaluate the scalability of the algorithm. We use the data information for organic matter of the real-instances because this property showed more variability than the rest. To generate a random value for each soil sample, we consider a uniform distribution with the maximum and minimum value obtained from the *OM* property. These random values were generated using the *Mersenne Twister*, a strong pseudo-random number generator *(PRNG)*. In non-rigorous terms, a strong *PRNG* has a long period and statistically uniform distribution of values (Shema, 2012).

The instances were grouped into five classes according to the number of soil samples in the field, with a minimum of 42 and a maximum of 400 soil samples. Each class contains ten different instances considering alpha values

of 0.5, 0.7, and 0.9 for the homogeneity level (30 instances per class). Recall that alpha values (α) greater than or equal to 0.5 are desirable to guarantee a homogeneity management zone delineation. For this set of instances, we assume the plots have a rectangular or square shape. Therefore it is not necessary to add dummy soil samples.

Table 1 presents the characteristics for each class of the random instances. The first column represents the class of the instance. Columns 2 and 3 show the width and height of the plot regarding the number of soil samples, respectively. The fourth column represents the total number of soil samples. The last column is the total number of potential management zones computed for each plot according to the algorithm presented in Cid-Garcia et al. (2013), which is pseudo-polynomial. This set of potential management zones contains only zones with a rectangular or square shape, which depends on the size of the field (soil samples in the width and height of the plot), and the size of the zone with the minimum number of soil samples in the width and height. In real-life scenarios, the number of soil samples in the plots commonly corresponds to instances of class 1 or 2 (the set of random instances can be downloaded from *https://github.com/NxtrCd/Instances-EDA-SSMZ.git*).

Table 1: Characteristics for each instance group of the random instances.

Class	**Plot width**	**Plot height**	**Soil Samples**	**Management Zones**
1	6	7	42	588
2	10	10	100	3025
3	15	10	150	6600
4	15	15	225	14400
5	20	20	400	44100

3.3 Calibration of the EDA-SSMZ algorithm

Calibration of algorithms is one of the most important steps in order to obtain good results. To set the appropriate parameters for the *EDA-SSMZ* algorithm, we used the iterated racing procedure. This procedure focuses on the sampling parameter configurations according to a particular distribution, evaluating them using either the Friedman's test or the t-test, and refining the sampling distribution to bias the sampling towards the best configurations. To calibrate the parameters, we used Friedman's test and the *irace ver. 3.1.2112M*, a software package that implements the iterated racing procedure for metaheuristic parameter tuning (López-Ibáñez et al., 2016). For each class, the tuning procedure was performed using a budget of 5000 experiments

with the following ranges for the parameter sampling: $p_0(\boldsymbol{x}) \in [0.95, 0.99]$, $N \in [1000, 25000]$, $Se \in [15, 1500]$, and $Tmax \in [50, 100]$. For each iteration, the *irace* package determines elite parameter configurations and selects the best of them. The tuning procedure is conducted several times, and favorable parameter settings are selected. For the real-field instances the parameters obtained were $p_0(x) = (0.95, ..., 0.95)$, $N = 9902$, $Se = 669$, and $Tmax = 32$. For the generated instances the parameters were $p_0(x) = (0.99, ..., 0.99)$, $N = 9902$, $Se = 669$, and $Tmax = 32$, except for Class 5 where $Tmax = 60$, and when $\alpha = 0.9$ then $N = 24535$, and $Se = 200$.

3.4 Computational results

The computational experiments were carried out on a server with four Intel Xeon E5-2620 v2 Six-Core Processor @2.10 GHz, running the Linux operating system with Ubuntu Server release 18.04.2 LTS, and 128 GB of RAM. The *EDA-SSMZ* algorithm was implemented in C/C++ and replicated 50 times with the tuned parameter settings. The *MILP* and *BILP* approaches of Section 3.1 were executed in the same computer to compare the *EDA-SSMZ* algorithm.

3.4.1 Results for real-field instances

In this section, we evaluate the *EDA-SSMZ* algorithm by using the real-field instances described in Section 3.2. Table 2 shows the experimental results comparing the *BILP* and *MILP* approaches with the *EDA-SSMZ* algorithm. Numbers in bold are the best solutions. The first column presents the chemical soil property (*OM, pH, P,* and *SB*). The second column determines the homogeneity level of the management zones (α–parameter). Columns 3-6 present the results for the *EDA-SSMZ* showing the minimum, the average, the maximum, and the average time (in *seconds*) over 50 independent runs of the algorithm. Columns 7-10 are the results for the *BILP* and *MILP,* showing the number of zones (Z^*) and the execution time (in *seconds*) for each approach. The last column represents the improvement (in %) of the *EDA-SSMZ* (considering its best solution) in comparison with the other approaches. The percentage is computed as $|Min - Z^*|/Min * 100$. The last row shows the total number of management zones obtained by each approach.

The *EDA-SSMZ* algorithm obtains the best solutions in all the real-field instances. However, for ten instances, the *MILP* and *BILP* approaches reach similar solutions in comparison with our method. For *OM* and *SB*, the results of the *EDA-SSMZ* and the exact approaches, in almost all cases, are very different (high percentages of improvement). On the contrary, the results for

Table 2: Experimental results for the real-field instances.

Soil	Alpha	*EDA-SSMZ*				*BILP*		*MILP*		Improvement
Property	(α)	**Min**	**Avg**	**Max**	**Time (s)**	**Z***	**Time (s)**	**Z***	**Time (s)**	(%)
OM	1	**40**	40.0	40	23.28	**40**	0.01	**40**	0.01	0.00
	0.9	**17**	17.9	18	22.33	20	0.03	20	0.03	17.65
	0.8	**11**	11.9	12	22.73	17	0.04	17	0.12	54.55
	0.7	**9**	9.2	10	22.94	14	0.24	14	0.15	55.56
	0.6	**6**	6.0	7	23.17	11	0.17	11	0.08	83.33
	0.5	**5**	5.5	6	23.26	9	0.06	9	0.07	80.00
	0.4	**4**	4.0	4	23.33	7	0.06	7	0.07	75.00
	0.3	**2**	2.8	3	23.43	6	0.22	6	0.07	200.00
	0.2	**2**	2.2	3	23.47	5	0.07	5	0.14	150.00
	0.1	**2**	2.0	2	23.63	3	0.07	3	0.06	50.00
pH	1	**19**	19.0	19	22.30	24	0.01	24	0.01	26.32
	0.9	**13**	14.7	16	22.33	17	0.06	17	0.08	30.77
	0.8	**8**	8.8	9	23.15	10	0.05	10	0.10	25.00
	0.7	**6**	6.0	6	23.28	7	0.06	7	0.07	16.67
	0.6	**5**	5.0	5	23.39	**5**	0.06	**5**	0.08	0.00
	0.5	**4**	4.0	5	23.45	**4**	0.06	**4**	0.08	0.00
	0.4	**3**	3.8	4	23.52	4	0.06	4	0.08	33.33
	0.3	**3**	3.0	3	23.62	**3**	0.05	**3**	0.06	0.00
	0.2	**2**	2.2	3	23.60	3	0.07	3	0.14	50.00
	0.1	**2**	2.0	2	23.64	**2**	0.05	**2**	0.01	0.00
P	1	**32**	32.0	32	21.46	33	0.01	33	0.01	3.13
	0.9	**7**	7.94	8	23.28	9	0.05	9	0.13	28.57
	0.8	**4**	7.9	4	23.46	5	0.26	5	0.19	25.00
	0.7	**3**	3.0	3	23.54	**3**	0.05	**3**	0.01	0.00
	0.6	**2**	2.0	2	23.64	3	0.05	3	0.01	50.00
	0.5	**2**	2.0	2	23.67	3	0.04	3	0.01	50.00
	0.4	**2**	2.0	2	23.67	3	0.07	3	0.01	50.00
	0.3	**2**	2.0	2	23.66	3	0.05	3	0.04	50.00
	0.2	**2**	2.0	2	23.64	3	0.05	3	0.08	50.00
	0.1	**2**	2.0	2	23.61	**2**	0.05	**2**	0.01	0.00
SB	1	**40**	40.0	40	23.51	**40**	0.01	**40**	0.01	0.00
	0.9	**14**	14.0	14	22.60	20	0.02	20	0.02	42.86
	0.8	**9**	9.7	11	22.94	16	0.04	16	0.04	77.78
	0.7	**5**	5.1	6	23.16	12	0.05	12	0.20	140.00
	0.6	**3**	3.0	3	23.35	9	0.27	9	0.16	200.00
	0.5	**3**	3.0	3	23.43	7	0.08	7	0.17	133.33
	0.4	**2**	2.2	3	23.54	5	0.06	5	0.16	150.00
	0.3	**2**	2.0	2	23.50	4	0.07	4	0.07	100.00
	0.2	**2**	2.0	2	23.57	**2**	0.06	**2**	0.01	0.00
	0.1	**2**	2.0	2	23.51	**2**	0.06	**2**	0.01	0.00
Total		**303**	315.84	322		395		395		

pH and *P* are relatively similar (low percentages of improvement). That is because the data information for *OM* and *SB* exhibits more variability than the rest of the soil properties, and delineating the field with management zones that use orthogonal shapes allows using a minor number of management zones than that with regions with rectangular or square shapes.

It is also noteworthy that the percentage of improvement for the *EDA-SSMZ* algorithm is up to 200% higher compared with the other exact ap-

proaches, and the computational time for solving the instances is higher for the *EDA-SSMZ* than the exact methodologies, e.g., on average, around 23 seconds against less than one second. This behavior is expected because the solution space for the *EDA-SSMZ* is bigger. However, the execution time does not represent a disadvantage for the algorithm.

3.4.2 Results for random instances

In this section, we test the performance of the *EDA-SSMZ* by using the random instances explained in Section 3.2. The experimental results for each class of instances (Class 1–Class 5) are presented in Tables 3–7, which have similar format as Table 2, except for the first two columns that show the homogeneity level (α value) in the first column, and the number of instance in the second column.

Table 3: Experimental results for the random instances: Class 1.

Alpha α	**Instance**	***EDA-SSMZ***				***BILP***		***MILP***		**Improvement**
		Min	**Avg**	**Max**	**Time (s)**	**Z***	**Time (s)**	**Z***	**Time (s)**	**(%)**
0.5	1	**6**	6.00	6	23.06	11	0.23	11	0.18	83.33
	2	**5**	5.34	7	23.02	10	0.07	10	0.20	100.00
	3	**6**	6.62	7	22.87	8	0.06	8	0.18	33.33
	4	**4**	4.00	4	23.33	10	0.05	10	0.07	150.00
	5	**8**	8.00	8	23.09	11	0.27	11	0.10	37.50
	6	**3**	3.98	5	23.24	9	0.06	9	0.07	200.00
	7	**8**	9.38	10	23.08	13	0.07	13	0.15	62.50
	8	**4**	4.94	5	23.19	10	0.06	10	0.15	150.00
	9	**7**	7.78	9	22.96	13	0.06	13	0.12	85.71
	10	**5**	5.08	6	23.02	11	0.06	11	0.15	120.00
0.7	1	**10**	10.28	11	22.76	16	0.25	16	0.15	60.00
	2	**10**	10.94	12	22.55	18	0.05	18	0.07	80.00
	3	**9**	10.10	11	22.64	14	0.37	14	0.04	55.56
	4	**6**	7.16	9	23.00	13	0.04	13	0.05	116.67
	5	**10**	10.54	12	22.73	16	0.30	16	0.14	60.00
	6	**8**	9.48	11	23.00	14	0.05	14	0.19	75.00
	7	**13**	13.98	15	22.68	18	0.04	18	0.08	38.46
	8	**6**	7.20	8	22.88	14	0.06	14	0.16	133.33
	9	**12**	14.18	16	22.38	19	0.05	19	0.08	58.33
	10	**8**	8.20	9	22.66	15	0.04	15	0.12	87.50
0.9	1	**19**	19.00	19	22.03	24	0.02	24	0.03	26.33
	2	**21**	22.70	25	21.84	26	0.02	26	0.03	23.81
	3	**22**	23.34	25	21.81	25	0.02	25	0.13	13.64
	4	**22**	23.82	25	21.81	24	0.02	24	0.04	9.09
	5	**21**	23.48	25	21.87	25	0.02	25	0.02	19.05
	6	**16**	17.02	18	22.41	21	0.02	21	0.09	31.25
	7	**23**	24.62	25	21.77	26	0.17	26	0.02	13.04
	8	**16**	16.85	18	22.15	21	0.03	21	0.03	31.25
	9	**25**	26.06	27	21.55	29	0.02	29	0.02	16.00
	10	**19**	21.68	23	21.66	22	0.02	22	0.02	15.79
Total		**352**	381.75	411		506		506		

For all the random instances, the *EDA-SSMZ* can provide better solutions than the ones provided by the exact approaches. To highlight this behavior, we mark in boldface the best solution obtained for each instance. An important characteristic of the *EDA-SSMZ* is that even the worst solutions outperform the best ones obtained by the exact approaches. However, the computational time for the *EDA-SSMZ* increases considerably with the instance size and the homogeneity level in comparison with the exact approaches, e.g., for Class 5, with the alpha parameter equal to 0.9, in the worst case our algorithm takes around 3.5 hours in solving the instance, against the 45 seconds of the *MILP* approach. As in the real-field instances, this behavior is expected because the solution space for the *EDA-SSMZ* algorithm is bigger. Currently, it seems a long time for large instances, but, in

Table 4: Experimental results for the random instances: Class 2.

Alpha α	**Instance**	***EDA-SSMZ***				***BILP***		***MILP***		**Improvement**
		Min	**Avg**	**Max**	**Time (s)**	**Z***	**Time (s)**	**Z***	**Time (s)**	**(%)**
0.5	1	**8**	9.22	12	89.84	24	0.47	24	0.50	200.00
	2	**8**	8.82	10	90.12	22	0.45	22	0.52	175.00
	3	**10**	11.52	13	89.80	22	0.59	22	1.70	120.00
	4	**10**	12.22	14	89.86	24	0.42	24	0.53	140.00
	5	**5**	5.48	7	90.91	14	0.48	14	0.67	180.00
	6	**10**	10.80	12	90.01	25	0.42	25	1.46	150.00
	7	**9**	10.76	12	90.01	19	0.31	19	0.38	111.11
	8	**8**	9.94	13	90.52	17	0.47	17	0.58	112.50
	9	**8**	9.36	11	90.11	22	0.36	22	0.63	175.00
	10	**8**	8.54	9	90.17	14	0.59	14	0.72	75.00
0.7	1	**16**	17.04	19	88.23	36	0.19	36	0.35	125.00
	2	**18**	20.72	25	87.75	35	0.22	35	0.32	94.44
	3	**16**	16.56	18	88.22	34	0.26	34	0.34	112.50
	4	**20**	22.30	24	88.17	36	0.28	36	0.37	80.00
	5	**11**	11.94	13	89.38	25	0.30	25	0.43	127.27
	6	**18**	21.06	23	88.07	35	0.17	35	0.37	94.44
	7	**18**	20.26	23	88.69	32	0.30	32	0.62	77.78
	8	**15**	16.08	18	88.79	30	0.34	30	0.52	100.00
	9	**15**	16.68	19	88.72	31	0.22	31	0.32	106.67
	10	**13**	14.34	16	88.45	25	0.52	25	0.52	92.31
0.9	1	**41**	43.24	45	85.22	50	0.07	50	0.08	21.95
	2	**44**	46.90	49	84.77	57	0.22	57	0.15	29.55
	3	**39**	40.68	44	85.49	52	0.10	52	0.23	33.33
	4	**41**	43.58	47	85.08	55	0.08	55	0.09	34.15
	5	**32**	33.56	36	86.03	49	0.13	49	0.23	53.13
	6	**40**	41.92	44	85.23	51	0.07	51	0.18	27.50
	7	**39**	41.62	44	85.41	50	0.09	50	0.11	28.21
	8	**42**	44.22	46	84.71	51	0.13	51	0.16	21.43
	9	**42**	44.64	48	84.94	52	0.08	52	0.08	23.81
	10	**37**	39.30	41	85.30	44	0.15	44	0.23	18.92
Total		**641**	693.30	755		1033		1033		

practice, it is reasonable since the agricultural production cycle is every year or, in some cases, every six months. Furthermore, we highlight the number

of soil samples in real-fields corresponds, generally, to small instances (Class 1 or 2) that can be efficiently solved by our *EDA-SSMZ* in less than two minutes, as can be seen in Tables 3 and 4. Notice that, for large instances (Table 7), the homogeneity level (α–parameter) plays an important role in the execution of the algorithm, i.e., when the alpha value tends to 1, our algorithm requires more computational time. With this, we have detected a future research line where it is possible to apply the parallelization of the *EDA-SSMZ* to decrease the computational time for large instances.

Table 5: Experimental results for the random instances: Class 3.

Alpha	**Instance**	***EDA-SSMZ***				***BILP***		***MILP***		**Improvement**
α		**Min**	**Avg**	**Max**	**Time (s)**	**Z***	**Time (s)**	**Z***	**Time (s)**	**(%)**
0.5	1	**10**	12.56	14	184.62	30	1.31	30	1.63	200.00
	2	**12**	13.74	16	184.52	29	1.37	29	2.98	141.67
	3	**9**	11.84	14	184.12	32	1.36	32	1.95	255.56
	4	**9**	10.68	14	184.20	27	1.28	27	1.61	200.00
	5	**8**	8.74	10	184.11	24	1.35	24	3.59	200.00
	6	**8**	10.22	13	185.65	30	1.35	30	2.26	275.00
	7	**12**	13.70	16	185.01	32	1.32	32	1.96	166.67
	8	**10**	12.44	14	184.05	27	1.31	27	1.84	170.00
	9	**11**	11.94	14	184.10	28	1.52	28	3.23	154.55
	10	**9**	11.90	15	184.54	31	1.41	31	2.00	244.44
0.7	1	**23**	26.20	29	181.07	53	0.63	53	1.21	130.43
	2	**23**	25.26	28	181.87	46	1.18	46	1.36	100.00
	3	**20**	22.84	26	180.94	49	1.06	49	1.07	145.00
	4	**21**	22.64	24	181.66	42	0.83	42	1.14	100.00
	5	**16**	17.80	20	181.36	38	0.90	38	1.35	137.50
	6	**18**	20.44	24	182.69	46	1.11	46	1.48	155.56
	7	**25**	27.46	30	180.52	51	0.77	51	1.22	104.00
	8	**19**	20.24	22	181.85	43	1.05	43	1.29	126.32
	9	**19**	20.60	23	181.44	44	1.43	44	1.42	131.58
	10	**28**	31.20	34	179.76	51	0.60	51	0.78	82.14
0.9	1	**76**	78.06	82	173.82	84	0.21	84	0.38	10.53
	2	**58**	61.50	64	175.09	77	0.29	77	0.35	32.76
	3	**59**	61.86	65	175.10	82	0.29	82	0.34	38.98
	4	**61**	63.28	66	174.78	75	0.30	75	0.38	22.95
	5	**60**	62.72	65	174.90	73	0.24	73	0.43	21.67
	6	**60**	62.58	66	175.35	72	0.17	72	0.26	20.00
	7	**64**	66.94	70	174.88	77	0.30	77	0.40	20.31
	8	**53**	55.84	59	176.13	69	0.36	69	0.46	30.19
	9	**56**	59.18	62	175.21	73	0.28	73	0.36	30.36
	10	**71**	74.04	77	173.64	82	0.47	82	0.47	15.49
Total		**928**	998.44	1076		1517		1517		

Table 8 shows a summary of the experimental results for the random instances. The first and the second column present the class and the field size (total number of soil samples on the width and height). The third column is the homogeneity level (α–parameter). Columns 4-6 show the results for the *EDA-SSMZ* algorithm (the minimum, the average, and the maximum of management zones). Columns 7 and 8 are the results for the *BILP* and

Table 6: Experimental results for the random instances: Class 4.

Alpha α	Instance	*EDA-SSMZ*				*BILP*		*MILP*		Improvement
		Min	**Avg**	**Max**	**Time (s)**	**Z***	**Time (s)**	**Z***	**Time (s)**	**(%)**
0.5	1	**11**	13.50	16	549.33	39	4.09	39	4.78	254.55
	2	**18**	21.34	23	550.24	47	4.92	47	5.41	161.11
	3	**10**	12.42	15	553.42	41	3.69	41	5.58	310.00
	4	**14**	16.88	20	549.75	45	3.49	45	7.03	221.43
	5	**13**	19.82	21	550.75	43	4.46	43	5.69	230.77
	6	**7**	9.16	12	552.61	37	4.82	37	4.95	428.57
	7	**18**	21.36	25	551.79	47	4.11	47	8.36	161.11
	8	**10**	12.48	15	552.06	34	4.39	34	9.65	240.00
	9	**12**	15.92	19	550.08	42	3.55	42	9.08	250.00
	10	**11**	13.64	16	551.20	38	4.11	38	6.18	245.45
0.7	1	**24**	28.96	32	579.22	64	1.97	64	2.75	166.67
	2	**36**	39.50	43	574.92	69	2.95	69	4.38	91.67
	3	**25**	27.26	30	584.00	66	1.68	66	2.27	164.00
	4	**25**	28.26	30	574.06	71	1.75	71	2.46	184.00
	5	**28**	30.78	34	579.34	68	2.36	68	3.31	142.86
	6	**21**	24.80	29	583.24	63	3.35	63	4.49	200.00
	7	**30**	33.72	37	584.97	73	2.18	73	3.11	143.33
	8	**24**	28.46	32	583.46	62	2.89	62	3.13	158.33
	9	**29**	32.98	37	574.14	64	1.88	64	2.86	120.69
	10	**28**	30.12	33	577.07	63	3.26	63	5.27	125.00
0.9	1	**72**	77.30	84	565.32	108	0.57	108	0.62	50.00
	2	**84**	89.20	94	570.64	115	0.84	115	1.16	36.90
	3	**82**	85.58	89	569.74	112	0.90	112	1.21	36.59
	4	**76**	78.48	83	560.41	105	0.85	105	1.18	38.16
	5	**80**	83.64	88	565.92	114	0.48	114	0.74	42.50
	6	**74**	78.20	82	562.85	106	0.60	106	0.85	43.24
	7	**75**	78.28	81	568.91	111	0.54	111	0.73	48.00
	8	**82**	86.68	91	567.64	114	0.84	114	1.11	39.02
	9	**87**	94.74	104	557.68	116	0.57	116	0.73	33.33
	10	**79**	84.24	87	572.69	113	0.65	113	0.74	43.04
Total		**1185**	1297.70	1402		2190		2190		

MILP approaches, respectively. The last column shows the percentage of improvement obtained by the *EDA-SSMZ* algorithm. Finally, the last row presents the total of management zones obtained by each approach for all the classes.

For each class, we observe when the homogeneity level decreases (α tends to 0.5), then the average relative improvement of the *EDA-SSMZ* increases, and when the homogeneity level increases (α tends to 1), then the average relative improvement decreases. This behavior is not surprising since when a high level of homogeneity is imposed, and there is variability in the information of the soil sample, the *EDA-SSMZ* tends to assign each soil sample within an individual zone. Therefore, the number of possibilities for delineation is reduced, and the percentage of improvement decreases too. Furthermore, the percentage of improvement increases with the size of the instance, i.e., larger instances show better improvements than short ones. For exam-

Table 7: Experimental results for the random instances: Class 5.

Alpha α	Instance	*EDA-SSMZ*				*BILP*		*MILP*		Improvement
		Min	**Avg**	**Max**	**Time (s)**	**Z***	**Time (s)**	**Z***	**Time (s)**	**(%)**
0.5	1	**17**	20.54	23	5253.77	69	25.14	69	44.95	305.88
	2	**25**	28.96	33	4864.24	77	28.22	77	29.77	208.00
	3	**19**	23.14	27	4911.34	70	27.85	70	23.29	268.42
	4	**19**	22.88	26	5395.65	70	29.67	70	23.28	268.42
	5	**22**	26.88	33	4974.36	67	26.36	67	31.68	204.55
	6	**18**	22.48	26	4923.22	64	35.44	64	27.89	255.56
	7	**12**	16.54	20	4861.87	60	24.83	60	24.29	400.00
	8	**20**	23.28	27	5402.27	66	30.50	66	35.86	230.00
	9	**24**	28.10	33	4885.62	73	30.41	73	27.62	204.17
	10	**24**	28.52	33	4419.26	78	27.74	78	25.60	225.00
0.7	1	**38**	43.64	48	4201.84	114	15.75	114	18.51	200.00
	2	**50**	56.14	62	3991.92	130	16.06	130	19.61	160.00
	3	**43**	47.94	52	4041.12	113	15.03	113	12.01	162.79
	4	**42**	47.98	53	3957.44	113	16.48	113	12.03	169.05
	5	**52**	56.18	61	4047.37	116	20.06	116	17.34	123.08
	6	**38**	44.54	51	3927.64	104	20.17	104	22.18	173.68
	7	**34**	39.44	44	4011.76	103	13.20	103	19.59	202.94
	8	**43**	47.88	53	3857.57	106	18.99	106	18.29	146.51
	9	**49**	53.32	59	3800.13	118	17.86	118	14.67	140.82
	10	**50**	57.96	63	3950.53	124	18.25	124	16.40	148.00
0.9	1	**146**	164.48	174	10977.90	189	3.88	189	6.92	29.45
	2	**178**	189.10	207	10537.90	221	2.41	221	3.97	24.16
	3	**157**	167.54	178	12930.10	198	4.55	198	6.87	26.11
	4	**158**	166.54	183	10631.30	198	4.50	198	6.88	25.32
	5	**164**	173.14	187	10579.00	196	3.80	196	4.63	19.51
	6	**151**	161.10	172	10777.20	185	4.74	185	5.01	22.52
	7	**145**	160.64	174	10424.40	174	3.62	174	5.34	20.00
	8	**149**	164.48	174	10667.60	183	2.30	183	3.42	22.82
	9	**156**	167.90	180	10715.10	200	5.02	200	6.03	28.21
	10	**160**	171.30	183	10486.20	201	4.54	201	6.72	25.62
Total		**2203**	2422.56	2639		3780		3780		

ple, instances of Class 5 show an average percentage of improvement up to 277% for $\alpha = 0.5$, 159% for $\alpha = 0.7$, and 24% for $\alpha = 0.9$. In contrast with the 89% for $\alpha = 0.5$, 70% for $\alpha = 0.7$, and 19% for $\alpha = 0.9$ on the instances of Class 1.

Fig. 6 shows a summary for each class of the random instances presented in Tables 3–8. The *X-axis* represents the instance with its corresponding alpha-value, and the *Y-axis* the number of partitions used in the final delineation. Notice for all the instances, our *EDA-SSMZ* algorithm (black lines) improves the *BILP* and *MILP* approaches (red lines).

3.5 Visualization

In Figs. 7–9 we show some configurations obtained by the *EDA-SSMZ* algorithm, in comparison with the *BILP* approach, considering the organic mat-

Table 8: Experimental results for the random instances. A summary.

Class	Field size	Alpha	*EDA-SSMZ*			*BILP*	*MILP*	Improvement
		α	**Min**	**Avg**	**Max**	**Z***	**Z***	(%)
1	6x7	0.5	5.6	6.1	6.7	10.6	10.6	89.29
		0.7	9.2	10.2	11.4	15.7	15.7	70.65
		0.9	20.4	21.8	23.0	24.3	24.3	19.12
2	10x10	0.5	8.4	9.6	11.3	20.3	20.3	141.67
		0.7	16.0	17.6	19.8	31.9	31.9	99.37
		0.9	39.7	41.9	44.4	51.1	51.1	28.72
3	15x10	0.5	9.8	11.7	14.0	29.0	29.0	195.92
		0.7	21.2	23.4	26.0	46.3	46.3	118.40
		0.9	61.8	64.6	67.6	76.4	76.4	23.62
4	15x15	0.5	12.4	15.6	18.2	41.3	41.3	233.06
		0.7	27.0	30.4	33.7	66.3	66.3	145.56
		0.9	79.1	83.6	88.3	111.4	111.4	40.83
5	20x20	0.5	20.0	24.1	28.1	69.4	69.4	277.00
		0.7	43.9	49.5	54.6	114.1	114.1	159.91
		0.9	156.4	168.6	181.2	194.5	194.5	24.36
	Total		**530.9**	579.3	628.3	902.6	902.6	

ter as chemical soil property and fixing the homogeneity level (α–parameter) to 0.5, 0.7, and 0.9, respectively. Figs. 7a, 8a, and 9a represent the solution obtained by the *EDA-SSMZ* algorithm, and Figs. 7b, 8b, and 9b show the solution of the *BILP*. We can observe that our approach selects figures with orthogonal shapes to partitioning the field, which minimizes the number of management zones in the final delineation.

4 Conclusions

In this paper, we introduce a new methodology to solve the problem of delineating homogeneous site-specific management zones (*SSMZ*) in agricultural fields based on an estimation of distribution algorithm (*EDA*). This problem consists of partitioning the field in small regions considering a specific soil property, chemical or physical, such that the generated zones satisfy a determined level of homogeneity. To the best of our knowledge, this is the first approach that generates management zones with orthogonal shape, e.g., L or T, which minimizes the number of regions required in the final delineation of the field.

Our methodology was tested on a set of real-life instances, and it was compared with other operations research methodologies presented in the literature. Furthermore, a set of instances was generated at random to analyze

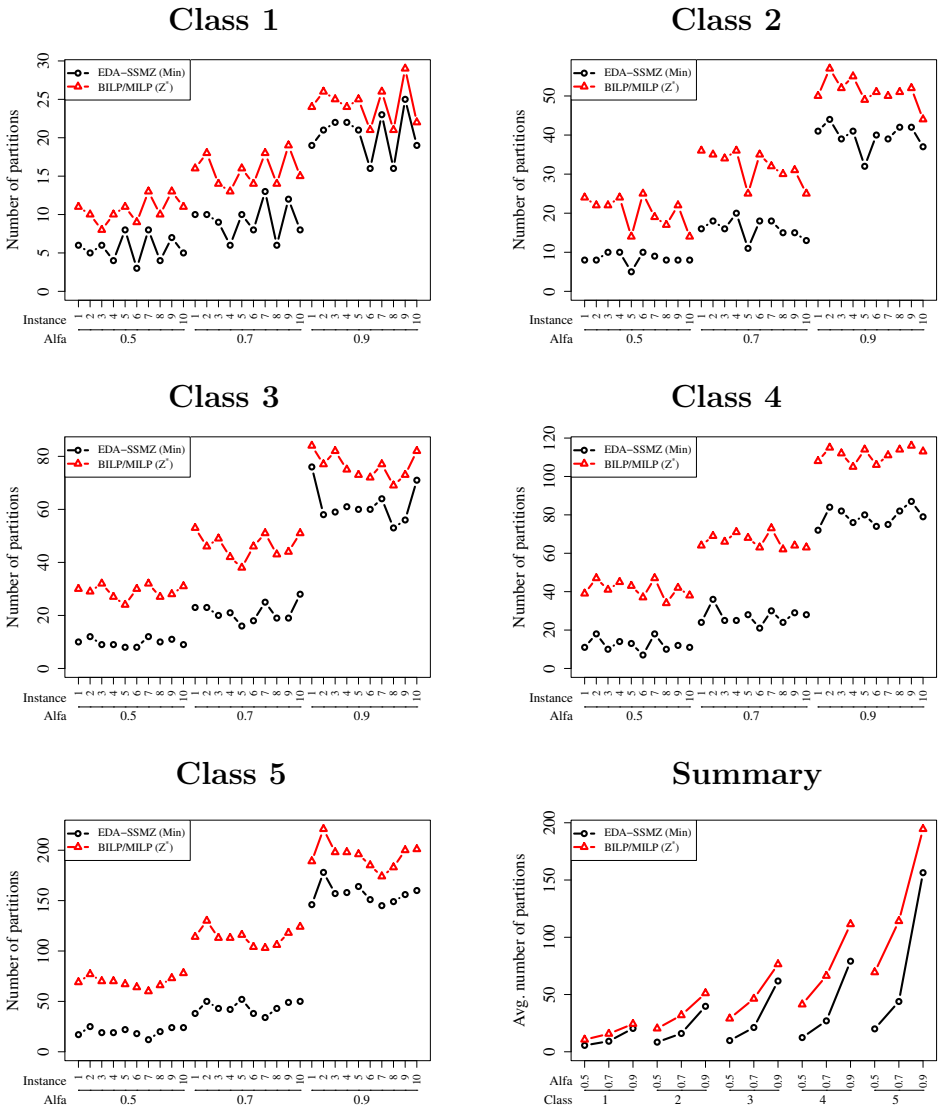

Figure 6: *EDA-SSMZ* vs. *BILP/MILP*: A summary for each class of randomly generated instances.

the scalability of the method. The experimental results show that our method is efficient and robust (the average/deviation behavior of the algorithm over different runs of the algorithm) to solve instances with different size for the *SSMZ* problem by improving the solutions presented by the other operations research approaches. According to the size of the instances, the *EDA-SSMZ* algorithm can find an average relative improvement is up to 277% when $\alpha = 0.5$, between 70% and 160% when $\alpha = 0.7$, and, at most, 40% when

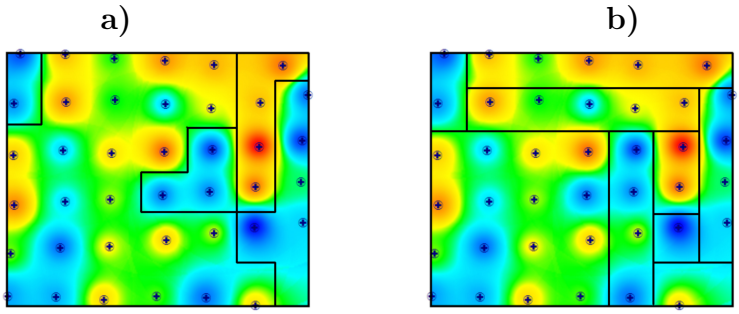

Figure 7: Management zones for organic matter when $\alpha = 0.5$. Fig. 7a shows the results for the *EDA-SSMZ* algorithm (five zones) and Figs. 7b shows the results for the *BILP* approach (nine zones).

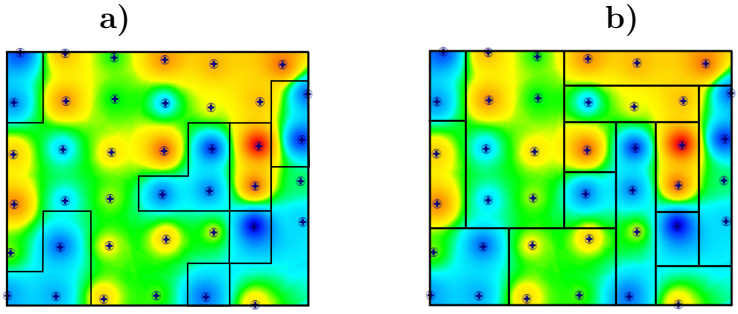

Figure 8: Management zones for organic matter when $\alpha = 0.7$. Fig. 8a shows the results for the *EDA-SSMZ* algorithm (nine zones) and Fig. 8b shows the results for the *BILP* approach (fourteen zones).

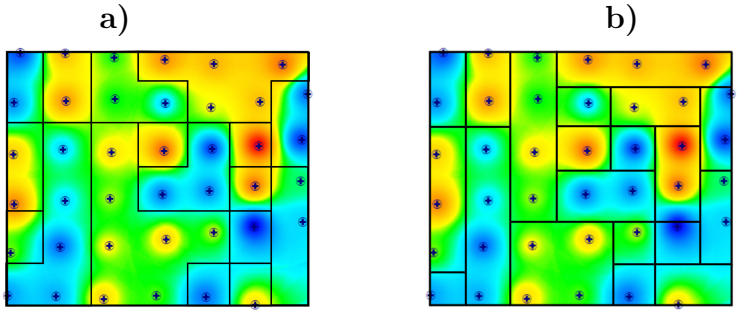

Figure 9: Management zones for organic matter when $\alpha = 0.9$. Fig. 9a shows the results for the *EDA-SSMZ* algorithm (seventeen zones) and Fig. 9b shows the results for the *BILP* approach (twenty zones).

$\alpha = 0.9$.

The *EDA-SSMZ* represents the agricultural field with a grid composed of edges and soil samples, from which and adjacency list is created. The execution time for the *EDA-SSMZ* increases considerably for large instances because our fitness evaluation process requires a search process (like depth-first search) over the adjacency list to find which soil samples belong to which management zone. Moreover, the fitness evaluation process is applied in a sequential way in each generation and for each individual of the population. Therefore, this is an important issue to improve. In particular, a parallel evolutionary approach can effectively reduce the computational time of the *EDA-SSMZ* and lead to an increased exploration and better diversity, compared to sequential one. In future research, our attention will concentrate on two main parallel approaches: the evolution of parallel populations and the parallelization of the fitness evaluation process. Additionally, we consider formulating our combinatorial optimization problem as a mixed-integer linear program *(MILP)* to obtain optimal solutions for the *SSMZ* problem or develop new methods for generating acceptable lower bounds. The main difficulty for both strategies is the procedure of finding connected components to calculate the evaluation of the management zone.

Acknowledgments

This study was partially supported by the Chairs Program of the National Council of Science and Technology (CONACYT) projects 843 and 2193. Salvador V. wishes to acknowledge graduate scholarship from CONACYT. Jose A. Lozano is partially supported by the Basque Government through the BERC 2018-2021 program, IT1244-19 and ELKARTEK program (3KIA KK-2020/00049) by the Spanish Ministry of Science, Innovation and Universities: BCAM Severo Ochoa accreditation SEV-2017-0718, TIN2016-78365-R and PID2019-104966GB-I00.

References

Albornoz, E.M., Kemerer, A.C., Galarza, R., Mastaglia, N., Melchiori, R., Martínez, C.E., 2018. Development and evaluation of an automatic software for management zone delineation. *Precision Agriculture* 19, 3, 463–476.

Albornoz, V.M., Cid-García, N.M., Ortega, R., Ríos-Solís, Y.A., 2015. A Hierarchical Planning Scheme Based on Precision Agriculture. In Plà-

Aragonés, L.M. (ed) *Handbook of Operations Research in Agriculture and the Agri-Food Industry*, Springer, New York, pp. 129–162.

Albornoz, V.M. and Ñanco, L.J., 2016. An Empirical Design of a Column Generation Algorithm Applied to a Management Zone Delineation Problem. In Fonseca, R.J., Weber, G.W., and Telhada, J. (eds) *Computational Management Science*, Springer, Cham, pp. 201–208.

Albornoz, V.M., Ñanco, L.J., Sáez, J.L., 2019. Delineating robust rectangular management zones based on column generation algorithm *Computers and Electronics in Agriculture* 161, 194–201 pp. 201–208.

Albornoz, V.M., Véliz, M.I., Ortega, R., Ortíz-Araya, V., 2020. Integrated versus hierarchical approach for zone delineation and crop planning under uncertainty. *Annals of Operations Research* 286, 1-2, 617–634.

Albornoz, V.M., Zamora, G.E., 2020. Decomposition-based heuristic for the zoning and crop planning problem with adjacency constraints. *TOP* 28, 3, 1–18.

Anastasiadis, S., Chukova, S., 2019. An inertia model for the adoption of new farming practices. *International Transactions in Operational Research* 26, 2, 667-685.

Armañanzas, R., Inza, I., Santana, R., Saeys, Y., Flores, J.L., Lozano, J.A., Van de Peer, Y., Blanco, R., Robles, V., Bielza, C., et al., 2008. A review of estimation of distribution algorithms in bioinformatics. *BioData Mining* 1, 1, 1 12.

Betzek, N.M., de Souza, E.G., Bazzi, C.L., Schenatto, K., Gavioli, A., 2018. Rectification methods for optimization of management zones. *Computers and Electronics in Agriculture* 146, 1–11.

Bongiovanni, R., Lowenberg-Deboer, J., 2004. Precision Agriculture and Sustainability. *Precision Agriculture* 5, 4, 359–387.

Castrignanò, A., Buttafuoco, G., Quarto, R., Parisi, D., Rossel, R.V., Terribile, F., Langella, G., Venezia, A., 2018. A geostatistical sensor data fusion approach for delineating homogeneous management zones in Precision Agriculture. *Catena* 167, 293–304.

Ceberio, J., Irurozki, E., Mendiburu, A., Lozano, J.A., 2013. A distance-based ranking model estimation of distribution algorithm for the flowshop scheduling problem. *IEEE Transactions on Evolutionary Computation* 18, 2, 286–300.

Chung, F.R., Garey, M.R., Johnson, D.S., 1982. On packing two-dimensional bins. *SIAM Journal on Algebraic Discrete Methods* 3, 1, 66–76.

Cid-Garcia, N.M., Albornoz, V., Rios-Solis, Y.A., Ortega, R., 2013. Rectangular shape management zone delineation using integer linear programming. *Computers and Electronics in Agriculture* 93, 1–9.

Cid-Garcia, N.M., Ibarra-Rojas, O.J., 2019. An integrated approach for the rectangular delineation of management zones and the crop planning problems. *Computers and Electronics in Agriculture* 164, 104925.

Doerge, T., 1999. *Management zone concepts. The Site-Specific Management Guidelines*. Potash and Phosphate Institute, South Dakota State University.

Fu, Q., Wang, Z., Jiang, Q., 2010. Delineating soil nutrient management zones based on fuzzy clustering optimized by PSO. *Mathematical and Computer Modelling* 51, 11-12, 1299–1305.

Gavioli, A., de Souza, E.G., Bazzi, C.L., Schenatto, K., Betzek, N.M., 2019. Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosystems engineering* 181, 86–102.

Georgi, C., Spengler, D., Itzerott, S., Kleinschmit, B., 2018. Automatic delineation algorithm for site-specific management zones based on satellite remote sensing data. *Precision Agriculture* 19, 4, 684–707.

Gili, A., Álvarez, C., Bagnato, R., Noellemeyer, E., 2017. Comparison of three methods for delineating management zones for site-specific crop management. *Computers and Electronics in Agriculture* 139, 213–223.

González, X.I., Bert, F., Podestá, G., 2020. Many objective robust decision-making model for agriculture decisions (MORDMAgro). *International Transactions in Operational Research*.

Haghverdi, A., Leib, B.G., Washington-Allen, R.A., Ayers, P.D., Buschermohle, M.J., 2015. Perspectives on delineating management zones for variable rate irrigation. *Computers and Electronics in Agriculture* 117, 154–167.

Hauschild, M., Pelikan, M., 2011. An introduction and survey of estimation of distribution algorithms. *Swarm and evolutionary computation* 1, 3, 111–128.

Hornung, A., Khosla, R., Reich, R., Inman, D., Westfall, D.G., 2006. Comparison of site-specific management zones: Soil-color-based and yield-based. *Agronomy Journal* 98, 2, 407–415.

Janrao, P., Palivela, H., 2015. *Management zone delineation in Precision agriculture using data mining: A review.* In 2015 International Conference on Innovations in Information, Embedded and Communication Systems, IEEE, pp. 1–7.

Larrañaga, P., Lozano, J.A., 2002. Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Kluwer Academic Publishers, Norwell, MA, USA.

Li, Y., Shi, Z., Li, F., Li, H.Y., 2007. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture* 56, 2, 174–186.

López-Ibáñez, M., Dubois-Lacoste, J., Cáceres, L.P., Birattari, M., Stützle, T., 2016. The irace package: Iterated racing for automatic algorithm configuration. *Operations Research Perspectives* 3, 43–58.

López, I.D., Grass, J.F., Figueroa, A., Corrales, J.C., 2020. A proposal for a multi-domain data fusion strategy in a climate-smart agriculture context. *International Transactions in Operational Research.*

Molin, J.P., de Castro, C.N., 2008. Establishing Management Zones Using Soil Electrical Conductivity And Other Soil Properties By The Fuzzy Clustering Technique. *Sci. Agric. (Piracicaba, Braz.)* 65, 567–573.

Monzon, J.P., Calviño, P., Sadras, V.O., Zubiaurre, J., Andrade, F.H., 2018. Precision agriculture based on crop physiological principles improves whole-farm yield and profit: A case study. *European Journal of Agronomy* 99, 62–71.

Moral, F., Terrón, J., Marques Da Silva, J.M., 2010. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. *Soil and Tillage Research* 106, 2, 335–343.

Mühlenbein, H., 1997. The equation for response to selection and its use for prediction. *Evolutionary Computation* 5, 3, 303–346.

Mühlenbein, H., Paaß, G., 1996. From recombination of genes to the estimation of distributions I. Binary parameters. In Voigt, H.M., Ebeling,

W., Rechenberg, I., and Schwefel, H.P. (eds) *International conference on parallel problem solving from nature*, Springer, pp. 178–187.

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering* 114, 4, 358–371.

Ohana-Levi, N., Bahat, I., Peeters, A., Shtein, A., Netzer, Y., Cohen, Y., Ben-Gal, A., 2019. A weighted multivariate spatial clustering model to determine irrigation management zones. *Computers and Electronics in Agriculture* 162, 719–731.

Oldoni, H., Terra, V.S.S., Timm, L.C., Júnior, C.R., Monteiro, A.B., 2019. Delineation of management zones in a peach orchard using multivariate and geostatistical analyses. *Soil and Tillage Research* 191, 1–10.

Ortega, R.A., Santibáñez, O.A., 2007. Determination of management zones in corn (Zea mays L.) based on soil fertility. *Computers and Electronics in Agriculture* 58, 1, 49–59.

Ortuani, B., Sona, G., Ronchetti, G., Mayer, A., Facchi, A., 2019. Integrating Geophysical and Multispectral Data to Delineate Homogeneous Management Zones within a Vineyard in Northern Italy. *Sensors* 19, 18, 3974.

Plant, R.E., 2001. Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture* 30, 1-3, 9–29.

Roudier, P., Tisseyre, B., Poilvé, H., Roger, J.M., 2008. Management zone delineation using a modified watershed algorithm. *Precision Agriculture* 9, 5, 233–250.

Saez, J.L., Albornoz, V.M., 2016. *Delineation of Rectangular Management Zones Under Uncertainty Conditions.* In ICORES 2016 - Proceedings of the 5th International Conference on Operations Research and Enterprise Systems, pp. 271–278.

Santos, M.O., Soler, E.M., Furlan, M.M., Vieira, J.C.M., 2020. A mixed integer programming model and solution method for the operation of an integrated water supply system. *International Transactions in Operational Research*.

Shema, M., 2012. Hacking web apps: detecting and preventing web application security problems. Newnes, Waltham, MA, USA.

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., Gemtos, T.A., 2013. Management zones delineation using fuzzy clustering techniques in grapevines. *Precision Agriculture* 14, 1, 18–39.

Wang, J., Tang, K., Lozano, J.A., Yao, X., 2015. Estimation of the distribution algorithm with a stochastic local search for uncertain capacitated arc routing problems. *IEEE Transactions on Evolutionary Computation* 20, 1, 96–109.

Watto, M., Mugera, A., 2019. Wheat farming system performance and irrigation efficiency in Pakistan: a bootstrapped metafrontier approach. *International Transactions in Operational Research* 26, 2, 686–706.

Zhang, X., Jiang, L., Qiu, X., Qiu, J., Wang, J., Zhu, Y., 2016. An improved method of delineating rectangular management zones using a semivariogram-based technique. *Computers and Electronics in Agriculture* 121, 74–83.