How Preferences Affect Productivity in the Sugarcane Harvest Problem – a Comparative Study of a Two-Steps MOEA

Diogo F. Pacheco, Tarcísio D. P. Lucas, Fernando B. de Lima Neto
Department of Computing and Systems - Polytechnic School of Pernambuco
University of Pernambuco, Recife-PE, Brazil.
E-mails: {dfp, tdpl, fbln}@dsc.upe.br

Abstract

In this paper we propose a two-level MOEA to help on the sugarcane harvest decision support. This problem is multi-objective in nature, as it contains agronomical and logistic objectives considered simultaneously. Two different sets of heuristics were used during harvest decisions, namely crisp and fuzzy prioritization schema. They are both tested and compared here with regards to effective help to decision makers – via traditional metrics and attainment to decision scenarios. Simulations show that the productivity was increased in all hypothetical scenarios investigated because of the two-level MOEA.

1. Introduction

Food and Agriculture Organization of the United Nations (FAO) considers sugarcane as one of the most important commodity in the world [1]. Its end products are sugar and ethanol. These two products are highly relevant to tackle social problems such as starvation and renewable fuel. This paper aims at improving production levels of sugarcane by offering intelligent computer decision support applications to the harvest of this commodity.

The management of sugarcane crops is very demanding. The harvest is deemed to be critical to the success of the entire process. Harvesting encompasses complex decisions that influence, directly, in the final production figures. During harvest, the manager must provide the sequence of sugarcane lots to be cropped. This decision is bounded to some logistic factors such as manual labor, transportation and mill capacity. This means than several intrinsic agricultural objectives have to be maximized [2] and at the same time, these logistic costs have to be minimized. A complicating aspect is that these two groups of factors are usually conflicting.

In many areas, the harvest decision making is carried out in an empirical or in a heuristic manner. In such cases, the managers' expertise is solely and heavily used to overcome the difficulties associated to the task. Some production data such as lot's planting date and species agronomical behavior are often used.

Lima Neto has proposed back in 1997-8 that agronomical production indicators could be forecasted using a Multilayer Perceptron (MLP) Artificial Neural Networks (ANN) [3]. Therefore, several papers have been produced outperforming the forecasting module and aggregating a search module that uses the forecasted indicators in order to suggest harvest decisions [4][5][6]. Other researches created models only to estimate the sugarcane crop production, with statistics [7] or using ANN with Geographic Information System (GIS) [8].

Originally, the sugarcane harvest problem (SHP) was dealt as a single-objective problem. Only recently, Pacheco *et al.* [6] has incorporated multi-objective optimization techniques into the solution formulation. In this referred work, for the first time performance comparisons considered the whole crop year. So far, concerned papers have considered only agronomic objectives, leaving out logistics, which are critical in decision-making process.

In this work we investigate how some decision maker preferences, applied on harvest decisions, can influence in the final crop production. To optimize each harvest day (*i.e.* step) we successively apply the multi-objective hybrid intelligent decision suite model [9]. In each interaction an approximation set (Pareto front) is constructed even though only one decision is required. A heuristic module is then utilized [9].

However, unlike previous works where the same heuristic was applied in all steps – across the entire harvest [6], we use another EMOO application to help choosing a better combination of heuristics to optimize production levels. We also develop fuzzy-heuristics to facilitate the interaction *men-optimizer* and to soften the crisp decisions. Logistic criteria were also handled and incorporated in the optimizations.

The final experiments were performed with two state-of-arts multi-objective evolutionary algorithm (MOEA): NSGA-II and PESA-II, with different heuristic sets; and revealed how the productivity can be increased by performing different heuristics sequences.

This paper is organized as follows. In section 2 there is a brief review on the intelligent techniques utilized, including an overview on MO-HIDS. In section 3 we detail our contribution, namely, (i) incorporating a meta-application of decision preferences, (ii) creating a heuristic set and (iii) incorporating logistic data into the decision problem formulation. The experimental methods and results are described in section 4. Finally, in section 5, we present the conclusion and some future lines of investigation.

2. Background

2.1 Intelligent Techniques Used

Artificial Neural Networks (ANN) - are popular intelligent techniques inspired in the human brain [10]. They try to emulate how brains process information to perform intelligent operations such as pattern recognition, classification, regression and forecast. ANN have been successfully applied in many domains and problems such as agriculture, medicine, chemistry and economy [11].

Fuzzy Logic (FL) - the concept of fuzzy membership functions enables the construction of computer applications that deal well with the uncertainty present in most real-world problems [12]. Thus, by using FL, one avoids linguistic mistakes common in man-machine interfaces [13].

Evolutionary Multi-Objective Optimization (EMOO) - is one direct application of Darwin's evolutionary concepts into multi-objective problems. It is the combination of evolutionary techniques, such as genetic algorithms and evolutionary strategies with MO theories. Several multi-objective evolutionary algorithms (MOEA) developed and have been used to solve complex problems comprising large decision space [14].

2.2 Encapsulated Framework

Multi-Objective Hybrid Intelligent Suite for Decision Support (MO-HIDS) – combined different intelligent techniques such as ANN, FL and EMOO to support managers in the decision making process [9]. It was first developed to help on tackling the *inverse problem.* Such ability suits perfectly for dealing with the sugarcane harvest problem (SHP) [3][4][5]. In SHP, an application of MO-HIDS includes a predictive module to forecast production indicators (*e.g.* PCC, FIBER, ATR and TCH – see details in 3.1.1). Next, these values afford the EMOO module to perform an optimization, retrieving the good candidates for the harvest decision (close to the Pareto front). Finally, the heuristic module (*e.g.* using fuzzy-logic) implements actual help to the decision. Fig. 1 shows an overview of MO-HIDS [9].

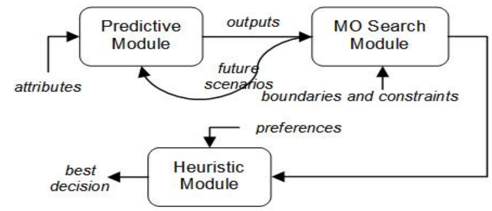

Fig. 1. MO-HIDS overview, extracted from Pacheco *et al.* [9].

3. Modeling the Problem

3.1 System Overview

In this work, we use a meta-application of EMOO. That is, a first level MOEA (of two levels) defines which heuristic should be applied in the harvest sequence. Then in the second level, another MOEA is repeatedly applied to solve single harvest days. Fig. 2 illustrates this schema.

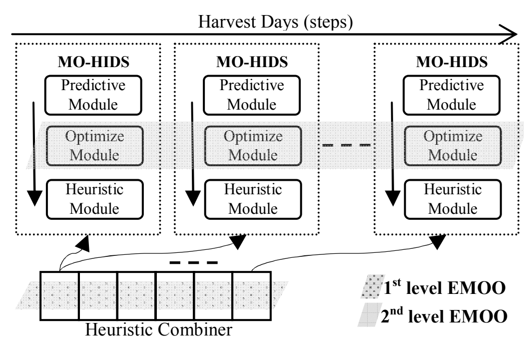

Fig. 2. Harvest system that uses an EMOO meta-application defining the heuristic for a second use of EMOO.

3.1.1 First Level Problem. The second level's output is an approximation set, thus, a decision making process is necessary. To help with this selection, a MOEA is applied to provide a heuristic harvest sequence from a set of heuristics previously defined; the creation of this set is detailed next. Because of that, this first level, determines how harvest decisions should be made.

3.1.2 Second Level Problem. This level is located inside the MO-HIDS [9], where the SHP is actually solved. The optimization task can be defined as to find sets of lots (*i.e.* solutions) to be harvested, that maximize the agronomical objectives, namely, *PCC* (6), *FIBER* (7) and *ATR* (8). This, at the same time, minimizes the one logistic objective considered, *i.e.* overall distance between lot and mill (9) – the used rule is explained later. The formulation also includes two constraints that inform the desired tonnage that guarantees continuous fulfillment of sugarcane to the factory (*i.e.* mill) demands during harvest (1). After this optimization, the heuristic set by the first level is applied, retrieving a unique solution among all candidates. In all equations, $x_i = 1$ indicates that the lot *i* was selected.

$$crop_{prod} = \sum_{i=1}^{n} x_i * tonnage_i \tag{1}$$

$$tonnage_i = TCH_i * AREA_i \tag{2}$$

$$prod^i{}_{pcc} = PCC_i * tonnage_i \tag{3}$$

$$prod^i{}_{fiber} = FIBER_i * tonnage_i \tag{4}$$

$$prod^i{}_{atr} = ATR_i * tonnage_i \tag{5}$$

$$obj_1^i = -\sum_{i=1}^{n} x_i * prod^i{}_{pcc} \tag{6}$$

$$obj_2^i = -\sum_{i=1}^{n} x_i * prod^i{}_{fiber} \tag{7}$$

$$obj_3^i = -\sum_{i=1}^{n} x_i * prod^i{}_{atr} \tag{8}$$

$$obj_4^i = calculateOverallDis\tan ce() \tag{9}$$

$$\rho_i = \left(obj_1^i + obj_2^i + obj_3^i\right) / obj_4^i \tag{10}$$

Every lot is defined by its attributes indicating quality, productivity and logistics. The first group – quality indicators – is given by: *PCC* (apparent percentage of sugar in the cane juice), *FIBER* (calorific power in the dry residue remaining after the extraction of juice) and *ATR* (recoverable total sugar). The productivity indicator considered is *TCH* (sugarcane tonnage per hectare). Moreover, the logistic indicators are *AREA* (lot surface area), *DISTANCE* (from lot to factory) and *ID* (a unique identification) [6].

3.2 The Heuristic Set

The heuristic set contains eleven heuristics: five crisp and six fuzzy.

Four crisp heuristics hardly represent the objectives, as they return the maximum value for each respective target. They are: *REDUCE_COST*, *INCREASE_PCC*, *INCREASE_ATR* and *INCREASE_FIBER*. The fifth balances production and costs, *i.e.* It returns the solution which has the maximum productivity ρ (10).

The fuzzy heuristics are defined by setting objectives priorities or equivalences. For each objective in each solution it is attributed a value according to its relative variation in the approximation set, *i.e.* the solution that has the worst value for an objective is set to 0 (zero) and the best, to 1 (one). Membership functions classify the solution in three different classes: *BAD, REGULAR* and *GOOD* (see Fig. 3). The maximum priority includes only *GOOD* solutions; the next, *REGULAR* and *GOOD*; otherwise, any classification is accepted. To be eligible by a fuzzy heuristic the solution has to fit in the criteria defined to all objectives. If there is more than one eligible solution, the highest in the preferable objective is chosen.

According to this, the six fuzzy heuristics are:

- Prioritizing one agronomical indicator (AI), then the costs and the others AI: *PCC > COST > FIBER=ATR*, *FIBER > COST > PCC=ATR* and *ATR > COST > PCC=FIBER*;
- Prioritizing AI or costs: *PCC=FIBER=ATR > COST* and *COST > PCC=FIBER=ATR*;
- Any prioritization: *PCC=FIBER=ATR=COST*.

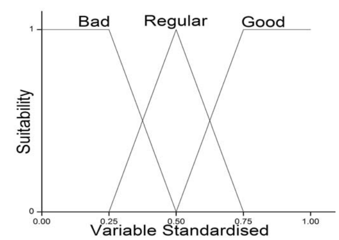

Fig. 3. Membership functions used.

3.3 Incorporating Logistic Data into SHP

This paper also incorporates a logistic objective in the optimization. This increases the problem complexity [6] at the same time that makes the model more realistic.

Although it would be ideal to have the actual configuration of lots in the field (*i.e.* distance between them, their neighborhood etc), only the distance of each lot to the factory was available (in the data set used in all simulations).

To overcome this incomplete data set, we have assumed a hypothetical spiral physical distribution of lots surrounding the factory. This is based on their distances (to the factory), see Fig. 4. In the figure, each square represents a lot (of different areas) and various types of neighborhoods. Black points mean the selected lot, while the grey ones are their neighbors. The concept of neighborhood of a lot is important as it directly reflects on decisions such as workforce, and machinery and transport assignment. None of these others logistic issues are dealt by the current model.

The heuristic to calculate costs was to consider the overall distance to neighborhood of lots, the steps are:

1. Group all selected lots in neighborhoods;

2. For each group, the considered distance is the mean of all lots;
3. The overall distance is the summarization of all group distances.

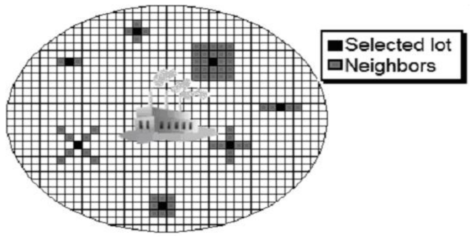

Fig. 4. Hypothetical physical distributions of lots and neighbors.

4. Simulations

In this section we describe how experiments were performed, including parameter setup, performance metrics and used approaches. The two EMOO applications are detailed.

The SHP, as defined previously, is formulated with four objectives. As a multi-objective problem it can be evaluated using traditional metrics, such as: generalized spread [15], hypervolume [16], among others. As a real problem it can be evaluated by attainment to scenarios considering the preferences of a decision maker. In this work, SHP was evaluated in both manners.

The experiments were performed with a real data set from a Brazilian sugar mill composed by 590 lots. All these lots must be harvest in a 6 months period. Hence, this task was divided in 12 fortnights (second level replications).

4.1 Parameter Setting

Since the problem has two levels, two settings of parameters were defined and utilized. In both cases, the fine-tuning regarding parameterization was found experimentally.

4.1.1 First Level Setting. The algorithm NSGA-II [14] was used to set six different heuristics, one per month (the whole harvest period), for each replication of the second level problem. Note that the same heuristic is applied in two fortnights due to similarities of agronomical indicators. The algorithm used real representation, binary tournament, SBX crossover and polynomial mutation operators [14]. Table 1 shows the setup for the 1st level algorithm.

4.1.2 Second Level Setting. This level is responsible for the main optimization. Initial experiments confronting different representations (binary, binary with decoders and symbolic), operators (single-point, CX and PMX crossovers; bit-flip and swap mutation) and algorithms (NSGA-II, PESA-II and SPEA2 – only used in initial simulations) suggested that an order-based representation, with swap mutation and a pairwise, *ALGORITHM & CROSSOVER*, such as *NSGA-II & CX* and *PESA-II & PMX* [14] might be more efficient. Due to space limitation, some comparative details were omitted. Fig. 5 shows runtime comparisons. Table 1 shows the setup for the 2nd level algorithm.

Table 1. First and second level EMOO settings.

Parameters	1st level	2nd level
population size	30	200
# of evaluations	1000	2000
crossover rate	0.9	0.9
mutation rate	1 / 6	1 / 590
desired tonnage	-	50000
archive size[1]	-	200
bi-section[1]	-	5
distribution index[2]	20	-

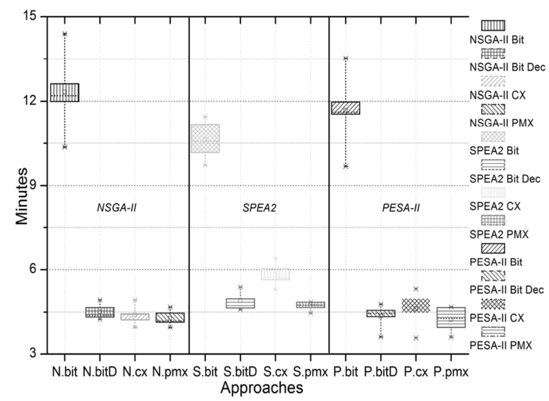

Fig. 5. Different approaches run-time comparisons.

4.2 Results

To perform the experiments, we created fifteen different approaches related to the heuristic set (see section 3.2). Table 2 describes all these approaches. Note that to be equivalent and comparable with a benchmark previous work [6], approaches *1..5* do not use the first level EMOO as they apply the same heuristic in all steps.

4.2.1 Multi-Objective Evaluation. This evaluation can only be done in approaches that use the first level EMOO, as their outputs are approximation solution sets. The applied metrics define a unique measure that classifies the whole sets. Thus, we confronted approaches *6..15* considering the following metrics:

- Generalized Spread (GS) - This is an extension of the Spread metric [17] to be applied to problems with more than two objectives. The distance is calculated from every point to its nearest neighbor. As the approximation set becomes more suitably (spread-

[1] This is applicable to PESA-II.
[2] This is applicable to SBX and polynomial operators.

out) distributed, including extreme solutions, GS output values tend to zero, $\lim GS = 0$ [15];

- Hypervolume (HV) - This metric was originally proposed by Zitzler and Thiele [16]. It calculates the volume covered by members of an approximation set. For each solution in the set a hypercube is constructed with reference point (a vector of worst objective function values) and the solution as the diagonal corners of the objective [14]. In this paper, we have used the modified version of Veldhuinzen [18] that measures the HV ratio between the solution set found and a utopian vector as the true Pareto front [14]. As the obtained set is deemed to be near the Pareto-optimal set, $\lim HV = 1$.

Table 2. Approaches used in the experiments.

Id	Approach	Description
1	*DISTANCE--*	Uses of *REDUCE_COST* in all steps
2	*PCC++*	Uses of *INCREASE_PCC* in all steps
3	*ATR++*	Uses of *INCREASE_ATR* in all steps
4	*FIBER++*	Uses of *INCREASE_FIBER* in all steps
5	*balance*	Uses of *p* in all steps
6	*3cpPESA*	Combines 1..3 crisp preferences with PESA
7	*3cpNSGA*	Combines 1..3 crisp preferences with NSGA
8	*4cpPESA*	Combines 1..4 crisp preferences with PESA
9	*4cpNSGA*	Combines 1..4 crisp preferences with NSGA
10	*5cpPESA*	Combines 1..5 crisp preferences with PESA
11	*5cpNSGA*	Combines 1..5 crisp preferences with NSGA
12	*fpPESA*	Combines the 6 fuzzy preferences with PESA
13	*fpNSGA*	Combines the 6 fuzzy preferences with PESA
14	*allPESA*	Combines all heuristics with PESA
15	*allNSGA*	Combines all heuristics with NSGA

In Fig. 6, approaches are plotted confronting both metrics, GS and HV. Note that every point means the whole attained set for each approach. The ideal Pareto is the region formed by trade-off solution (GSxHV). The Pareto reached by them is composed only by PESA-II handlings without fuzzy heuristics. In this way, by the multi-objective evaluation, the approaches that are closer to the *Ideal Pareto* and more diversified (represent more fairly the trade-offs between objectives) are *3cpPESA*, *4cpPESA* and *5cpPESA*.

4.2.2 Real Problem Evaluation. In a previous work [6], this evaluation was carried out by choosing one specific objective, e.g. PCC production, and considering the sum of all harvest days. Avoiding tendentious affirmatives, we have created five hypothetical scenarios to compare different approaches. Four of these scenarios consider specific objectives and the last one balances the gains and losses of all objectives together.

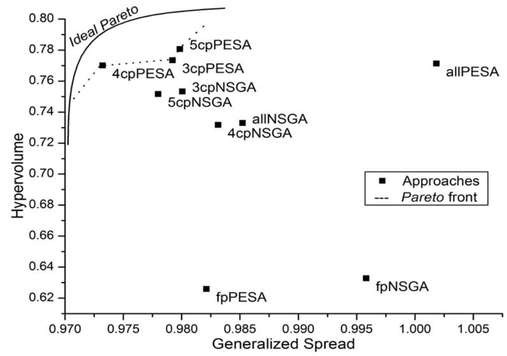

Fig. 6. MO plot HV x GS of approaches.

Despite of MO evaluation, where the entire solution set is considered, now just one solution is selected, i.e. the one that more enhance each scenario. The best five solutions in each hypothetical scenario are described in Table 3 as their approaches. Note that except to the COST scenario, the variation between approaches only presented a small percentage. Because of the big scale, this difference should not be ignored.

Table 3. Top 5 among 15 approaches for each hypothetical scenario.

Obj.	Rank 1º	2º	3º	4º	5º
PCC	*3cpNSGA* 100%	*5cpNSGA* 99.92%	*fpNSGA* 99.87%	*5cpPESA* 99.86%	*fpPESA* 99.85%
FIBER	*5cpNSGA* 100%	*fpNSGA* 99.97%	*3cpNSGA* 99.96%	*4cpNSGA* 99.94%	*allPESA* 99.93%
ATR	*3cpNSGA* 100%	*4cpPESA* 99.98%	*allNSGA* 99.98%	*4cpNSGA* 99.96%	*3cpPESA* 99.93%
COST	*5cpPESA* 0.00%	*3cpPESA* 2.10%	*allPESA* 2.58%	*4cpPESA* 2.83%	*3cpNSGA* 8.17%
Balance	*5cpPESA* 2.981	*allPESA* 2.958	*3cpPESA* 2.954	*4cpPESA* 2.949	*5cpNSGA* 2.896

Table 4. Rank of approaches according to the punctuation criterion in hypothetical scenarios.

Rank	Approach	Points
1	*3cpNSGA*	14
2	*5cpPESA*	12
3	*5cpNSGA*	10
4	*4cpPESA*	8
4	*3cpPESA*	8
4	*allPESA*	8
5	*fpNSGA*	7
6	*4cpNSGA*	4
7	*allNSGA*	3
8	*fpPESA*	1

To perform a more clear interpretation of these comparisons, we created a punctuation rank where the best approach in a scenario receives 5 points and the worst gets 1 point. According to this, an order can be established as shown in Table 4. Thus, considering the real problem possible situations, *3cpNSGA* is the approach that returns better solutions in different scenarios, closely followed by *5cpPESA* and *5cpNSGA*.

5. Conclusions

This paper investigated how preferences affect the productivity in the sugarcane harvest problem. By using a meta-application of a multi-objective evolutionary algorithm to combine different heuristics during the harvest cycle, the productivity was increased in all hypothetical scenarios created in comparison to situations were the same heuristic is applied in the whole harvest.

This work has incorporated logistic data into optimization bringing more realism to the results, as the experiments were performed with real data.

According to Fig. 7-A, we are tempted to consider non-conflicting objectives, but a more careful analysis in Fig. 7-B exposes the real existing trade-off. This means that despite of the minimal increases found, especially on agronomical objectives, they are greatly welcome. The reason is that the sugarcane industry deals with large sums of money.

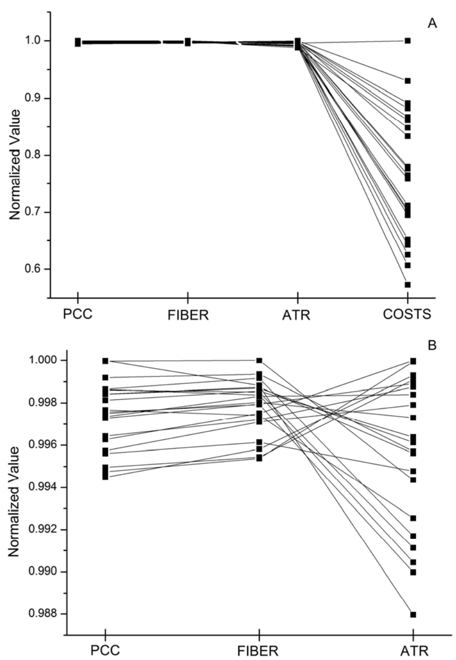

Fig. 7. Trade-off graph from the *5cpPESA* approach: all objectives together (A) and only agronomical objectives (B).

Finally, it is not clear whether PESA-II performs better than NSGA-II as both achieve good results. Moreover, the use of fuzzy logic in the heuristics definition does not increase the final results.

As future works, we suggest investigate the progressive articulation of preferences in second level algorithms as they can possibly reduce the time consuming of this meta-application.

6. References

[1] Food and Agricultural Organization of United Nations, “Major Food and Agricultural Commodities and Producers”, The Statistics Division, Economic and Social Department, available on http://www.fao.org/es/ess/top/commodity.html?lang=en&item=156 &year=2005, accessed in March 2008.

[2] G. Alexander, “Sugarcane physiology: a comprehensive study of the Saccharum source-to-sink system”. Amsterdam: Elsevier, 1973.

[3] F.B.L. Neto, “Managerial Decision Support, based on Artificial Neural Networks” (in Portuguese), Master Dissertation. Department of Informatics, Federal University of Pernambuco, Brazil, 1998.

[4] D.F. Pacheco, “An Intelligent Decision Support System for Agriculture Harvest” (in Portuguese), Technical Report presented as Graduation Monograph, Department of Computing and Systems, University of Pernambuco, Brazil, 77p. 2006.

[5] F.R.S. Oliveira, D.F. Pacheco and F.B.L. Neto, “Intelligent Support Decision in Sugarcane Harvest”, Proceedings 4th World Congress of Computers in Agriculture, Orlando, pp. 456-462, 2006.

[6] D.F. Pacheco, T.D.P. Lucas e F.B.L. Neto, "How to obtain fair managerial decisions in sugarcane harvest using NSGA-II", Proceedings of 7th International Conference on Hybrid Intelligent Systems (HIS’07), Kaiserslautern, Germany, 2007.

[7] B.A. Keating, M.J. Robertson, R.C. Muchow, N.I. Huth, “Modelling sugarcane production systems I. Development and performance of the sugarcane module”, Field Crops Research, Vol. 61, Issue 3, pp. 253-271, 1999.

[8] M.C.A. Picoli, "Estimate sugarcane yield using combining artificial neural networks: case study Catanduva plant", Master's Dissertation (in Portuguese), INPE, S. José dos Campos, 90p, 2006.

[9] D.F. Pacheco, F.R.S. Oliveira, F.B.L. Neto, "Including Multi-objective Abilities in the Hybrid Intelligent Suite for Decision Support". In: International Joint Conference on Neural Networks-IJCNN (WCCI2008), Hong Kong, China, 2008.

[10] S. Haykin, “Neural Networks – A Comprehensive Foundation”. Prentice-Hall International Editions. New Jersey, USA, 1994.

[11] D.W. Patterson, “Artificial neural networks, theory and applications”. Prentice Hall, Singapore, 1996.

[12] S.Russel and P.Norvig, "Artificial intelligence: a modern approach", 2nd Edition, Prentice-Hall Editions, NJ, USA, 1995.

[13] H.-J. Zimmermann, “Fuzzy set theory and its applications (3rd ed.)”, Kluwer Academic Publishers, Norwell, MA, 1996.

[14] K. Deb. “Multi-Objective Optimization using Evolutionary Algorithms”, John Wiley & Sons, UK, 2001.

[15] A. Zhou, Y. Jin, Q. Zhang, B. Sendhoff, E. Tsang, "Combining model-based and genetics-based offspring generation for multi-objective optimization using a convergence criterion". In IEEE Congress on Evolutionary Computation CEC’06, pp. 892-899, 2006.

[16] E. Zitzler and L. Thiele. “Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study”. Parallel Problem Solving from Nature - PPSN-V, Sept. 1998, pages 292-301.

[17] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II”, IEEE Transactions on Evolutionary Computation, vol.6, no.2, pp.182-197, 2002.

[18] D. V. Veldhuizen, “Multiobjective evolutionary algorithms: classifications, analyses, and new innovations”, PhD Thesis, Dayton, OH: Air Force Institute of Technology, 1999.