How to Obtain Fair Managerial Decisions in Sugar Cane Harvest Using NSGA-II

Diogo Ferreira Pacheco   Tarcísio Daniel P. Lucas   Fernando B. de Lima Neto

Department of Computing Systems - Polytechnic School of Engineering
Pernambuco State University, Recife-PE, Brazil.
E-mails: {dfp, tdpl, fbln}@dsc.upe.br

Abstract

The world's demand for sugar and particularly for renewable fuels such as ethanol requires an increase in production in sugar mills. The use of artificial neural networks (ANN) posed as a predictive core associated with the algorithm NSGA-II aims at helping decision makers to optimize the multi-objective harvest problem. This paper presents two approaches and the good results achieved as compared with other classical techniques.

1. Introduction

Global warming is producing some crucial changes in the world, not only in the climate but in the prospects for our future lives. These changes are bringing about behavioral, social and economics transformations.

Renewable fuels are one step in the direction of decreasing global warming. The resources that remain in the Earth are preserved, but still produce the fuel our economy needs. Recently, the USA has approved the increased use of ethanol that must be mixed with gasoline. Ethanol, a renewable fuel, reduces pollution and greenhouse gas emissions, does not pollute ground water, is cheaper than gasoline and supports local farmers because it is made principally from corn and sugar cane [1]. In this scenario, with the worldwide increasing demand, production and productivity need to be improved.

Brazil, the largest producer of sugarcane in the world, produced 420,121,000 tons in 2005 at sugar cane costs of approximately 20.77 Int$[1]/ton [2]. Therefore, a minimal increase in production represents millions of Int$ profit (1% increase, means Int$ 87M).

[1] Int$ have been calculated based on 1999-2001 international prices.

There are a multitude of factors and activities that affect the production of a sugar cane factory. The harvest, for instance, is a factor that greatly influences productivity at the same time it is very difficult to manage. Every day, the harvest manager has to choose a subset between hundreds of lots to be harvested following the constraint of milling capacity, during the harvest period – usually six months. He often takes his decisions empirically because the lack, scarceness or uncorrelated existing data [3].

A previous work [4] applied an artificial neural network (ANN) correlating the available data to predict three important productivity indicators:

- TCH – measure the sugarcane tonnage per hectare – hence, it is a production indicator;
- PCC – measure the apparent percentage of sugar in the cane juice. Thus, it indicates quality (of sugar and ethanol);
- Fiber – measure the calorific power in the dry, fibrous residue remaining after the extraction of juice from the crushed sugarcane stalks. Good fiber indicator makes possible a sugar mill be more self-sufficient regarding its steam and burned electricity production. Fiber is also a source of secondary products such as: animal feed and raw material for paper manufacture.

By having a prior knowledge of future values of these three indicators, the decision maker could be much more precise in his decisions. However, as mentioned above, this particular kind of forecasting is already tackled and reasonably addressed in the literature [3][4][6][7].

On the other hand the decision maker faces a multi-objective problem, namely discovering the combination of lots that maximizes the production of PCC and Fiber constrained to a minimal tonnage that guarantees energy power for the sugar mill to operate without interruption. This is a combinatorial hard problem [5].

Pacheco *et al.* [6] refined the indicator predictions and then applied linear decision models to support managers, suggesting which lots should be harvested [3]. Oliveira *et al.* [7] used genetic algorithm to achieve better decisions in this problem.

The present work used ANN to predict the three production indicators TCH, PCC and Fiber. It also applied evolutionary multi-objective optimization (EMOO), particularly, the fast non-dominated sorting genetic algorithm (NSGA-II) to obtain a fair decision, regarding the harvest. Experimental results demonstrate the efficiency of the applied method as compared to classical ones.

2. Background

2.1. Artificial Neural Networks

Artificial neural networks are an intelligent technique inspired by the human brain [8]. They try to emulate how brains process information to perform intelligent operation such as pattern recognition, classifications and regressions. ANNs have been successfully applied in diverse areas like agriculture, medicine, chemistry and economy, among others.

2.2. Evolutionary Computation

Evolutionary computation is a set of techniques inspired by Darwin's theory of evolution. Genetic Algorithms (GA), Evolution Strategies (ES), Evolutionary Programming (EP) and Genetic Programming (GP) are the main techniques. The concepts of natural selection, *i.e.* preservation of the fittest individual and inheritance of characteristics through generations are easily transposed to the resolution of search and optimization problems [9]. Amongst these algorithms GA is the most widely used and has been constantly tested in several kinds of application with success; this work also uses GA.

2.3. Multi-Objective Optimization

Real (decision) problems usually deal with multiple objectives (criteria). These purposes are often in disagreement with each other.

A multi-objective optimization problem (MOOP) itself ideally has two goals: to find solutions as close as possible to the Pareto-optimal front and to obtain these solutions well spread over this front. Both goals aim at representing, as clearly as possible, the trade-off between decision objectives [10].

It is a long time since the first technique was developed to solve MOOP. However, these initial conventional methods, referred to here as classical, functioned by converting a MOOP into a single-objective optimization. This has many implications that will not be discussed here [10].

2.3.1. Classical Techniques. There are two main families in MOOP methods. The elementary approach is simple and naive, but, depending on the problem, can obtain satisfactory results. The best known methods are *Pros-Cons Analysis*, *Maximin & Maximax Methods*, *Conjunctive and Disjunctive Methods*, and *Lexicographic Method* [11].

The first family is based on Multi-Attribute Utility Theory (MAUT) and aims to aggregate different objectives into a single function that has to be maximized. The main methods in this family are *MAUT Additive Linear, SMART, Generalized Means* and *AHP* [12].

The second family represents the Outranking methods which follow the concept of outranking proposed by Roy (1968). The basic idea is to create subsets according to this outranking as small as possible. The most representative methods are in two families *ELECTRE* and *PROMETHEE* [13].

2.3.2. Multi-Objective Evolutionary Algorithms. In general we can classify multi-objective evolutionary algorithms (MOEA) in two classes: the algorithms that include ranking based on Pareto dominance and the algorithms that do not. Historically, this classification is preceded by two different generations, where the first does not introduce the concept of elitism. As members of first generation we have *VEGA*, *MOGA*, *NSGA* and *NPGA* [14], where the second one is represented by *SPEA*, *PAES* and *NSGA-II* [15].

3. Modeling the Problem

3.1. System overview

In this paper, we applied ANN, followed by the NSGA-II algorithm to the sugarcane harvest problem. Figure 1 shows how the system is conceived. The three modules that comprise the system are detailed below:

- Predictive module: it consists of a Multi-Layer Perceptron (MLP) [8] trained with real data from a sugar mill in Brazil. Its function is to predict the three productivity indicators (TCH, PCC, Fiber);
- MO module: this module is responsible for achieving a multi-objective optimal solution set

that balances trade-offs between maximization of PCC production and Fiber production; in it we have implemented the algorithm NSGA-II;

- Decisor module: this is where decisions actually happen by application of a heuristic. The one implemented suggests that (i) PCC production should be preferred in detriment of Fiber production; (ii) the tonnage production must not be less than 0.5% or greater than 5% of the desired minimal tonnage. This difference in boundaries is because it is very expensive if the factory stops without energy that could be originated from sugarcane bagasse.

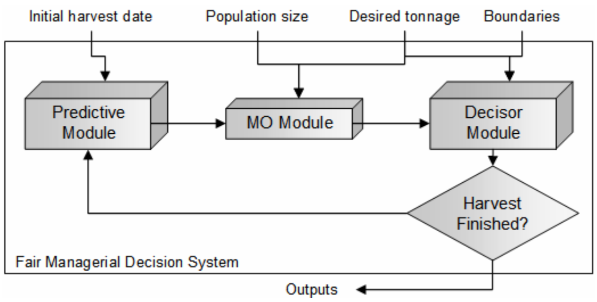

Figure 1. Schematic view of a fair managerial decision system for sugar cane harvest

The system starts by receiving input such as: initial harvest date, desired minimal tonnage and GA population size. The date is conveyed to the predictive module that generates the predicted indicators for the multi-objective module.

Next, the NSGA-II algorithm of the MO module utilizes the minimal tonnage in fitness calculation. It produces a set of optimal solutions of same size as the population size. This set is then forward to the decisor module, which carries out the heuristic that selects only one solution for the harvest.

If there are any remaining lots, the process continues – through new harvest dates generation and until all lots be yielded. Finally, when the harvest is finished, the manager has predicted values which correspond to PCC and Fiber production for all harvest year.

3.2. Representation of Individuals

The individuals used in NSGA-II represent the available lots (*i.e.* not harvested) in the fields. It uses a bit stream representation, where *0* indicates not harvest, while *1* means the opposite. Figure 2 illustrates an individual when there are 40 available lots and the system suggested 15 to be harvested.

Genes of an individual (*i.e.* available lots)

100**1**00**1**000 **111**00000**1**0 **1**000**11**000**1** 00**1**0**1**0**1**00**1**

Figure 2. Individual representation of a suggestion of 15 lots in 40 available ones

At each interaction a new mapping is implemented between the indices of individuals and the identifiers of the lots. This allows all lots to be harvest at one time. It is important to notice that until the end of the gathering of the production is not reached and new harvest dates are provided, the selected lots will be harvested. Consequently, the number of lots available decrease so does the individual size (of the GA).

3.3. GA Initial Population

Individuals are randomly generated according to the quantity of available lots in the field. Each bit in the genotype has a 50% likelihood of being activated.

3.4. Multi-Objective Definition

As previously mentioned a NSGA-II individual represents the whole harvest. It is evaluated by its PCC and Fiber indicators. Each candidate harvest is defined by a combination of which lots are to be cropped. A lot has many attributes; some of them were also used to configure the MLP to predict the indicators that are used in MOO such as PCC, TCH and Fiber. Other useful attributes in this phase are: area (in hectare) and the lot identification code. The derived lot variables are calculated as demonstrated in (1), (2) and (3).

$$tonnage = TCH * area \tag{1}$$

$$pcc\mathrm{Re}leased = PCC * tonnage \tag{2}$$

$$fiber\mathrm{Re}leased = Fiber * tonnage \tag{3}$$

Two approaches (A and B) were conceived using NSGA-II; the main difference between them is how the constraints are incorporated into the algorithm. Both use harvest variables of equations (6), (4), (5), (7) and (8). In all of them x is the bit relative to the index i of the individual (*ind*). Table 1 has formulas descriptions.

Table 1. Formula definitions

Formula	Description
α	Difference between produced and desired tonnage.
β	Number of lots chosen to be harvested.
χ	Total tonnage production.
δ	Total PCC production.
ε	Total Fiber production.

ϕ	Penalty level.
φ	Penalty function.
ω	First fitness function (regarding PCC production).
ξ	Second fitness function (regarding Fiber production).
ψ	Third fitness function (actually, penalty function converted into an objective).

$$\beta(ind) = \sum_{i=1}^{n} x_i \quad (4)$$

$$\chi(ind) = \sum_{i=1}^{n} x_i * tonnage_i \quad (5)$$

$$\alpha(ind) = \chi - desired \quad (6)$$

$$\delta(ind) = \sum_{i=1}^{n} x_i * pcc\mathrm{Re}\,leased_i \quad (7)$$

$$\varepsilon(ind) = \sum_{i=1}^{n} x_i * fiber\,\mathrm{Re}\,leased_i \quad (8)$$

- *Approach A* – considers only two objectives: maximize PCC production and maximize Fiber production. The constraint is absorbed by using a penalty function in fitness calculation. The penalty index is demonstrated in (9) and (10). The two fitness functions are calculated in (11) for PCC production and in (12) for Fiber production; these fitnesses used this dual concept, therefore NSGA-II here carries out two minimizations.

$$\phi(ind) = \varphi\left(\frac{\alpha}{desired}\right) \quad (9)$$

$$\varphi(diff) = \begin{Bmatrix} 20 \to diff < -0.02 \\ 10 \to diff > 0.05 \\ 0 \to otherwise. \end{Bmatrix} \quad (10)$$

$$\omega(ind) = -\left(\frac{\delta}{\beta} - \phi * \left(\frac{\delta}{\beta * \chi}\right) * |\alpha|\right) \quad (11)$$

$$\xi(ind) = -\left(\frac{\varepsilon}{\beta} - \phi * \left(\frac{\varepsilon}{\beta * \chi}\right) * |\alpha|\right) \quad (12)$$

- *Approach B* – considers the constraint as another objective (in addition to PCC and Fiber). The optimization here should be performed with three objectives. That is maximization of PCC production and Fiber production, and minimization of the difference between desired and harvested tonnage. For implementation purposes the dual concept was applied again and the optimization is in the direction of three minimizations. Equations (13), (14) and (15) formalize fitness calculations for PCC and Fiber production, and difference between produced and desired tonnage, respectively.

$$\omega(ind) = -\left(\frac{\delta}{\beta}\right) \quad (13)$$

$$\xi(ind) = -\left(\frac{\varepsilon}{\beta}\right) \quad (14)$$

$$\psi(ind) = |\alpha| \quad (15)$$

4. Experimental Setup

Experiments were performed with a real data set obtained from a sugar factory in Brazil. The data correspond to information on 418 lots. The following hypothetical scenario was assumed:

- All lots have to be harvested. The previous considered papers [3], [4], [6] and [7] evaluated only a single harvest day. On this work we are evaluating the entire production of a crop year;
- A minimal desired tonnage, randomly selected, was fixated at 4000 tons for each prediction. This means that the factory needs this tonnage to continue working and cannot mill much more than that in the selected period. Therefore, when the heuristic is applied, the acceptable boundaries are from 3980 to 4200 tons. These values, correspond to less 0.5% and more 5% adopted by the utilized heuristic;
- Only two predictions are performed per month, *i.e.* one for each fortnight;
- The harvest must be finished (*i.e.* all lots yielded) at most in 12 interactions. This corresponds to a 6 months harvest;
- If, at the 12th interaction the limit of 4200 ton is achieved and there are still remaining lots to be cropped, they will all be harvested.

4.1. Parameter Setting

The parameters used in NSGA-II were experimentally chosen. In particular, a new stop criterion was used in parallel to the number of generations. It represents how many times the entire population is in the first Pareto-front. When this attribute assumed high values the experiments always finished with the maximum generation number. However, when this attribute assumes low values, as its value was decreasing, the simulations were stopped earlier and produced better results.

Table 2 illustrates the best parameter setting found, according to the experiments. We also tested experimentally which crossover operator was most adaptable to this problem. After the one point and two point crossover had been tested, it was the uniform crossover that achieved the best results.

Table 2. Best parameter settings experimentally found

Parameter	Value
Crossover rate	90%
Mutation rate	5%
Population size	100
Number of generations	90
Times all solutions in first Pareto-front	1

4.2. Results of Experiments

NSGA-II was successfully applied in both approaches. In both situations the population reaches the Pareto-front. However, approach B obtains a better diversity than approach A. In addition, approach B has the best output values. Figure 3 and Figure 4 show the population convergences for approach A and B, respectively. Notice the two and three objectives used.

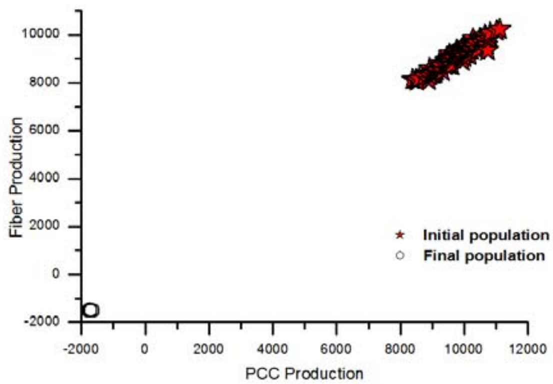

Figure 3. Population convergence - *Approach A*

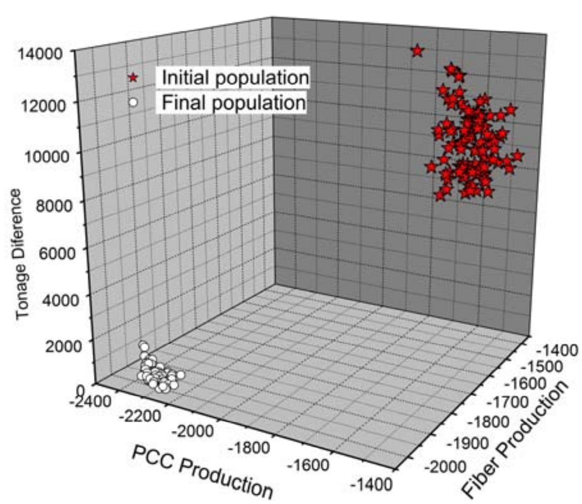

Figure 4. Population convergence - *Approach B*

The difference between the initial population and the final population is huge in both approaches. This indicates that the decision space is vast. It is clear that number of possible combinations is very high which adds to the complexity of the problem.

Figure 5 and Figure 6 illustrates the surface created by the last generation in the Pareto-front for approaches, A and B, respectively.

The results obtained are illustrated in next sub-section on Table 3; they are the average of 10 interactions. The standard deviation for all three outputs was less than 1%.

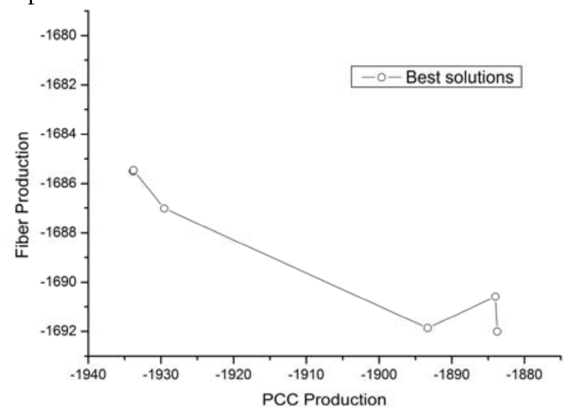

Figure 5. Pareto-front surface for approach A

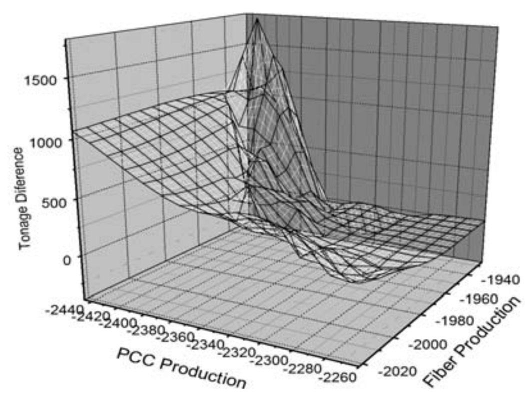

Figure 6. Pareto-front surface for approach B

4.3. Comparisons with Other Techniques

In order to assess the efficiency of the proposed method and application, we compared the results obtained here with other methods using the same data and with the same hypothetical scenario. Table 3 shows the production values obtained for tonnage, PCC and fiber for each of the distinct methods.

As was expected, the evolutionary approaches A and B (*i.e.* 2 and 3 aims, respectively) performed better than the classical ones. Using the Lexicographic method the production was about 46 thousand tonnages while the approach *B* (NSGA-II) achieved 52 thousand. It means a difference of more than Int$124,000.

Table 3. Outputs production values produced by different methods

Technique	Production		
	Tonnage	PCC	Fiber
NSGA-II (3 aims)	52644.65	778419.15	740252.40
NSGA-II (2 aims)	51337.62	756915.75	727215.22
AHP	47520.80	701325.18	701280.50
MAUT	47395.93	700453.01	700079.16
Lexicographic	46642.91	692220.13	690656.33

Figure 7 shows how much better is the best technique – NSGA-II using 3 aims (*i.e.* targets) – compared to the others tested. This figure shows the best method as 1.00 representing 100% satisfaction of constraint and other methods proportionally displayed.

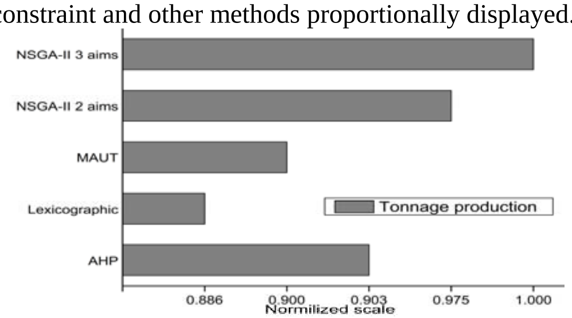

Figure 7. A relative comparison between tonnage production using different methods

5. Conclusions

Nowadays decisions on the harvest are mostly made empirically. The adoption of intelligent techniques tailored to help decision makers on these tasks can increase crop production. This work has shown that the application of evolutionary algorithms within a multi-objective formulation can be even more beneficial to the decision process.

NSGA-II as was modeled can aggregate much property and fairness to the manager decision. The evaluation considering all the harvest of the entire set of lots represented an actual contribution to this problem while previously this potential was stood up.

Approach B, which considers the problem constraint as another goal was found to be a good avenue of thinking. Since the attribution of a penalty function may cause the same problems faced by MO classical method. That is, the use of high level information prematurely (by defining the penalty function) does not allow measure the impacts of these decisions.

As for future works we suggest modeling the sugarcane harvest considering not only agronomical factors but also, logistic ones. This would make this application even more realistic and more necessary as the decision dimensions will go up. We propose to use Fuzzy Logic to soften the thresholds used. Finally, we suggest further investigations on transforming penalty functions into objectives. New stop criteria may also be valuable.

6. References

[1] Biodiesel and Ethanol Investing, "Ethanol Fuel Benefits", available on http://www.biodieselinvesting.com/ethanol-fuel-benefits/, accessed in May 2007.

[2] 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.ht ml?lang=en&item=156&year=2005, accessed in May 2007.

[3] D.F. Pacheco, "An Intelligent Decision Support System for Agriculture Harvest" (in Portuguese), Technical Report presented as Graduation Monograph, Department of Computing Systems, Polytechnic School of Engineering, Pernambuco State University, Recife - Brazil, 2006.

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

[5] C.H. Papadimitriou and K. Steiglitz, "Combinatorial Optimization: Algorithms and Complexity". Prentice-Hall, Inc., 1982.

[6] D.F. Pacheco, F.S. Regueira and F.B.L. Neto, "Using Artificial Neural Networks in Sugar Cane Harvest to Predict PCC, TCH and Fiber" (in Portuguese), Alcoolbrás Magazine, S. Paulo - Brazil, v. 90, 2005, pp. 60-63.

[7] 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, Florida, 2006, pp. 456-462.

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

[9] A.E. Eiben and J.E. Smith, "Introduction to Evolutionary Computing", Springer, New York, 2003.

[10] K. Deb. "Multi-Objective Optimization using Evolutionary Algorithms", John Wiley & Sons, UK, 2001.

[11] I. Linkov, et. al., "Multi-criteria decision analysis: A framework for structuring remedial decisions at the contaminated sites", Springer, New York, 2004, pp. 15-54.

[12] J. Fülöp, "Introduction to decision-making methods", BDEI-3 Workshop, Washington, 2005.

[13] J. Figueira et. al., "Multiple Criteria Decision Analysis: State of the Art Surveys", Springer, New York, 2004.

[14] C.A.C. Coello, "Metaheuristics for Multiobjective Optimization", Tutorial on IEEE Symposium Series on Computational Intelligence, 2007.

[15] K. Deb et. al., "A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II", KanGAL report 200001, Indian Institute of Technology, Kanpur, India, 2000.