Comparing the performances of two techniques for the optimization under parametric uncertainty of the simultaneous design and planning of a multiproduct batch plant

Guillermo A. Durand[1], M. Susana Moreno[1], Fernando D. Mele[2], J. Marcelo Montagna[3], Alberto Bandoni[1,*]

[1] PLAPIQUI, Camino La Carrindanga km7, Bahía Blanca, Argentina
{gdurand,smoreno, abandoni}@plapiqui.edu.ar
[2] Universidad Nacional de Tucumán, Av. Independencia 1800, S.M. de Tucumán, Argentinafmele@herrera.unt.edu.ar
[3] INGAR (CONICET-UTN), Avellaneda 3657, Santa Fe, Argentina
mmontagna@santafe-conicet.gov.ar

Abstract. This paper addresses the comparison between two techniques for the optimization under parametric uncertainty of multiproduct batch plants integrating design and production planning decisions. This problem has been conceived as a two-stage stochastic mixed integer linear programming (MILP) in which the first-stage decisions consist of design variables that allow determining the batch plant structure, and the second-stage decisions consist of production planning continuous variables in a multiperiod context. The objective function maximizes the expected net present value. In the first solving approach, the problem has been tackled through mathematical programming considering a discrete set of scenarios. In the second solving approach, the multi-scenario MILP problem has been reformulated by adopting a simulation-based optimization scheme to accommodate the variables belonging to different management levels. Advantages and disadvantages of both approaches are demonstrated through a case study. Results allow concluding that a simulation-based optimization strategy may be a suitable technique to afford two-stage stochastic programming problems.

Keywords:uncertainty, two-stage stochastic programming, simulation-based optimization

1 Introduction

The concepts involved in decision-making under uncertainty are closely linked to those of optimization under uncertainty. Literature on optimization under uncertainty very often divides the approaches into two categories: "wait and see" and "here and now". In the "wait and see" approaches, one has to wait until an observation on the random elements is made, and then solve the deterministic problem. Conversely, a

"here and now" problem involves optimization over some probabilistic measure of the system performance –usually the expected value. It should be noted that many realistic problems have both "here and now", and "wait and see" approaches embedded in them. The trick to overcome this situation is to divide the decisions into these two categories and use a coupled approach [1].

In this regard, many advances have been observed in the supporting theory, including algorithmic developments and computational capabilities for solving this class of problems, most of which fall into one of these two areas: multistage stochastic programming and stochastic optimal control.

In multistage stochastic programming, decisions are based on past observations and decisions before the future events occur [2]. A finite set of scenarios is often generated to represent the space, therefore, the stochastic program becomes a deterministic equivalent program, whose size can easily grow out of hand for a large number of scenarios, making the direct solution approaches numerically intractable, thus requiring methods of decomposition or aggregation [3].

Stochastic optimal control describes a sequential decision problem in which the decision-maker chooses an action in the state involved at any decision stage according to a decision rule or policy. Dynamic programming provides the framework for designing algorithms to compute an optimal control policy. However, for large problems, dynamic programming also suffers numerically from dimensionality. Both approaches -stochastic programming and optimal control- are essentially equivalent, but they exhibit differences in formulation and solution, with the consequent advantages and disadvantages for specific problems (see [4], [5], [6], and [7]).

Efficient numerical solution proposals can be achieved by combining several techniques that belong to each approach. The resulting strategy needs to be adapted to solve the specific problem, defining some approximations or heuristic-based methods. The works [8] and [9] are relevant examples in this regard.

In the literature, it is not easy to find comparisons regarding the two-stage stochastic scheme solved by mathematical programming and by the heuristic simulation-based optimization approach. Therefore, this work presents such comparison so as to demonstrate advantages and disadvantages of both approaches.

2   Description of the techniques used in the comparison

2.1   Two-stage stochastic programming

The two-stage stochastic linear programming (2SSLP) problem can be stated as in [2]:

$$\min_{x} c^{T} x + E_{\omega} Q(x;\omega)$$
$$Ax = b \tag{1}$$
$$x \geq 0$$

where:

$$Q(x;\omega) = \min_{y} d_{\omega}^{T} y$$

$$T_{\omega} x + W_{\omega} y = h_{\omega}$$

$$y \geq 0$$

Here E_{ω} is the expectation, and ω denotes a scenario or possible outcome with respect to the probability space $(\Omega; P)$. The variables x are the first-stage variables, as they have to be decided upon before the stochastic variable ω realizes. The variables y are second-stage variables: they can be assessed after the outcome of ω is known.

In this work discrete distributions P are only considered, specifically through a Monte Carlo sampling technique, so it can be written:

$$E_{\omega} Q(x;\omega) = \sum_{\omega \in \Omega} p(\omega) Q(x;\omega) \tag{2}$$

Therefore, a large LP can be formulated. This LP is the deterministic equivalent of the problem:

$$\min_{x} c^{T} x + \sum_{\omega \in \Omega} p(\omega) d_{\omega}^{T} y_{\omega}$$

$$Ax = b$$

$$T_{\omega} x + W_{\omega} y_{\omega} = h_{\omega} \quad \forall \omega \tag{3}$$

$$x \geq 0, y_{\omega} \geq 0$$

The sequence of events in (3) is as follows: first the decision maker implements the first stage decisions x. Then the system will be subjected to the random process described by $(\Omega; P)$, which results in an outcome $\omega \in \Omega$. Finally, the decision maker will execute the second stage decisions y_{ω} accordingly.

2.2 Simulation-based optimization strategy

The simulation-based optimization (SbO) algorithm proposed for the comparison is a variation of the one presented in [10] and [11] and can be followed from Fig. 1. It involves an outer loop which corresponds to a genetic algorithm (GA) strategy and an inner loop which is a Monte Carlo sampling of the uncertain parameters over an LP deterministic planning model. As in the stochastic programming technique, variables must be divided into first-and second-stage variables.

At the beginning of the algorithm initial estimations for the first stage variables are provided to the GA outer loop, which in turn runs the inner loop taking samples of the uncertain parameters. For each sampling, an LP deterministic planning model obtains the values of the second-stage decisions variables and gives a value for the objective function. Enough samplings of the LP model are done to obtain a representative value of the population of objective function values. Usually, the expected value is utilized and then is returned to the GA outer loop, which uses it to search for the optimum combination of first-stage variables' values. A filter is utilized to avoid the use of the inner loop when an already-tried combination of first-stage variables is chosen again (the filter returns the same expected value of the objective function).
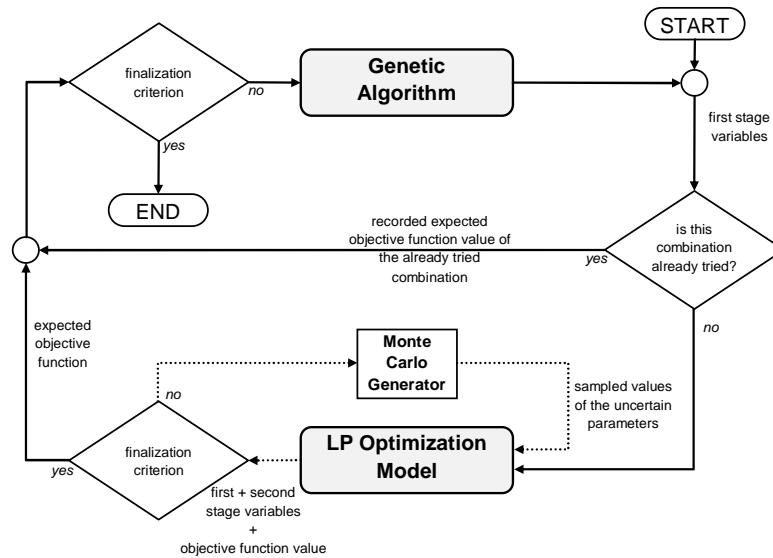
Fig. 1. Simulation-based optimization strategy.

The variation introduced in this work with respect to the SbO strategy described in [10] and [11] is that the simulation step in the inner loop is replaced with an LP optimization for each sample of the uncertain parameters. The reason for this variation is to make this technique and the stochastic programming more comparable.

3 Design and planning of a multiproduct plant in a mutiperiod environment. Overall description

In a multiperiod environment, consider a multiproduct plant that processes a set I of products i over a time horizon H, which is divided into $t = 1, 2,...,NT$ specified time periods H_t, not necessarily of the same length. Every product i follows the same production sequence throughout a set J of batch processing stages and a set K of semi-continuous stages that form NL semi-continuous subtrains. The production of product i at every time period t requires a given processing time pt_{ijt} in batch stage j and the size/duty factors SF_{ijt}/D_{ikt} for each batch/semi-continuous stage.

In order to reduce idle times in the plant, out of phase duplication of batch units and the introduction of intermediate storage tanks between batch stages are allowed. Additionally, in-phase duplication is admitted at semi-continuous stages, so each stage k can consists of one or more units of the same size. Let M_j denote the set {1, 2,..., M_j^U} of possible number of equal units that can be allocated in parallel in each batch stage j. And let G_k be the set {1, 2,..., G_k^U} of units that can be duplicated in

parallel in semi-continuous stage k. Thus, $m \in M_j$ identical parallel units can operate out of phase in stage j and $g \in G_k$ parallel units operate in phase in stage k. When an intermediate tank is allocated, the original process is decoupled in subprocesses, upstream and downstream of the tank, each one having different batch size and limiting cycle time.

Also, the design problem involves the selection of equipment sizes for both batch and semi-continuous units, and intermediate storage tanks from a number of available discrete sizes. Thus, the batch unit size of stage j, V_j, and the size of semi-continuous unit at stage k, R_k, are restricted to values from the sets $SV_j = \{ v_{j1}, v_{j2}, ..., v_{j,nj} \}$ and $SR_k = \{ \omega_{k1}, \omega_{k2}, ..., \omega_{k,mk} \}$, respectively. In the same way, the size of storage tanks VT_j is restricted to values from the set $ST_j = \{ vt_{j1}, vt_{j2}, ..., vt_{j,gj} \}$

The plant operates in a single product campaign (SPC) mode in every time period, and when storage tank are not allocated, a Zero Wait (ZW) policy is employed.

Product demands are not known to the decision maker with certainty, but it is assumed that the uncertainty can be represented by a set of scenarios S. Each scenario $s \in S$ has a known probability p_s that reflects the likelihood of each scenario to take place with $\sum_{s \in S} p_s = 1$. Moreover, these scenarios are described through lower and upper bounds on product demand levels in each time period t, DE_{its}^L and DE_{its}^U. The amounts of raw materials consumed are determined by mass balances with a given parameter F_{it} that accounts for the process conversion of raw material to product i during period t. Costs and availability of raw materials vary from period to period and are assumed to be known. Also, prices of final products in each time period and maximum available storage capacities are problem data.

In every scenario s, production planning decisions allow to determine at each period t and for each product i, the amount to be produced q_{its}, the number of batches n_{its}, and the total time T_{its} to produce product i. Furthermore, at the end of every period t, the levels of both final product IP_{its} and raw material inventories IM_{its} are obtained. The total sales QS_{its}, the amount of raw material purchased C_{its}, and the raw material to be used for the production RM_{its} of product i in each time period t are determined with this formulation. In this model, it is assumed that each product requires a unique raw material that it is not shared by other products. However, more sophisticated transformation processes can be easily incorporated.

In summary, a two-stage stochastic MILP model was developed. First-stage decisions consist of design variables that allow determining the batch plant structure. Second-stage decisions consist of planning variables (continuous variables) to determine the production, purchases, and inventories of raw materials and products for each period throughout the time horizon under each scenario, given the plant structure decided at the first-stage.

The objective function maximizes the expected net present value (*ENPV*) over a set of scenarios S.

$$\max ENPV = \sum_{s \in S} \sum_{t \in T} \sum_{i \in I} p_s \, np_{it} \, QS_{its} - \sum_{s \in S} \sum_{t \in T} \sum_{i \in I} p_s \, \kappa_{it} \, C_{its}$$

$$- \sum_{s \in S} \sum_{t \in T} \sum_{i \in I} p_s \left[\varepsilon_{it} \left(\frac{IM_{i,t-1,s} + IM_{its}}{2} \right) H_t + \sigma_{it} \left(\frac{IP_{i,t-1,s} + IP_{its}}{2} \right) H_t \right] \qquad (4)$$

$$- \sum_{j \in J} CB_j - \sum_{k \in K} CS_k - \sum_{j \in J}^{J^U - 1} CT_j$$

The economic criterion in Eq. (4) is calculated by the probabilistic average of the difference between the revenue due to product sales and the overall costs in each scenario s, with the latter consisting of the cost of raw materials, inventory costs, and the capital investment cost corresponding to batch units, semi-continuous units, and storage tanks. Parameters np_{it}, κ_{it}, ε_{it}, and σ_{it} are the corresponding cost coefficients for each term associated with production planning decisions. Note that all the above cost coefficients take into account the time value of money, in other words, they are discounted prices for each time period with a specified interest rate.

Due to space reasons, the detailed formulation is not provided in this manuscript, but readers can see [12] for more details.

4   Implementation and results

In order to create a manageable problem, the example from [12] has been modified to account for only 4 periods, but still with an operation horizon of one year. Therefore some parameters have been changed as shown in Tables 1 and 2. All other parameters have been kept unmodified.

Table 1. Data for the model

t	Costs of raw materials, κ_{it} ($/kg)					Prices of products, np_{it} ($/kg)				
	A	B	C	D	E	A	B	C	D	E
1	2.200	0.500	1.200	0.600	1.300	55.00	47.00	40.00	42.00	48.00
2	1.733	1.500	2.067	1.400	1.600	53.00	48.00	42.67	44.00	50.67
3	1.733	1.500	2.500	1.400	1.600	53.00	48.00	44.00	44.00	50.67
4	2.200	1.500	1.633	0.600	1.000	55.00	46.00	41.33	42.00	48.00
	$\varepsilon_{it} = 0.0002$ $/kg/hr					$\sigma_{it} = 0.0015$ $/kg/hr				

As explained before, the demands are the uncertain parameters in the formulation, which are specified by a range of lower-upper bounds (DE^L_{it}-DE^U_{it}). For each scenario, the nominal upper bound on product demands in every time period is affected with a factor generated following a normal probability distribution $N(1, 0.20)$. The lower bounds on demands for each product are set to 50% of the upper demands. Therefore, the number of uncertain parameters for this work is 20.

Table 2. Nominal upper bounds on demands

t	DE^U_{it} (x 10^3 kg) [†]				
	A	B	C	D	E
1	6.40	6.90	8.50	7.50	7.30
2	5.00	6.40	7.90	7.30	10.00
3	6.00	7.20	6.90	7.30	8.90
4	4.50	8.00	8.40	7.50	6.90

† Lower bounds on demands, DE^L_{it}, are calculated as $0.5 \cdot DE^U_{it}$.

The problem was solved with the techniques described in Section 2 in a AMD A6-3620 APU system (4 cores)at 2.20GHz with 8Gb of RAM memory, 64bit operating OS, using in both cases the GAMS 24.0.2/CPLEX 12.5.0 solver for MILP/LP resolutions (2SSLPtechnique and inner loop of the SbO method) and the MATLAB R2008a package for the outer loop, with the GA toolbox used with default options. For the SbO technique it was determined that 50 samplings in the inner loop per solution were sufficient to obtain a representative *ENPV* value (i.e., in consecutive tests at or near the optimal solution, the *ENPV* did not vary enough as to change the combination of outer loop variables at the optimal solution).

4.1 Determination of the optimal solution

The characteristics of the problem described in the previous section allow determining its optimal solution before applying the techniques proposed in this work.

In Eq. (4) the expected net present value is equal to the incomes from sales (1st term) minus the cost of raw materials (2nd term), minus the cost of inventories (3rd term) minus the capital cost for all the equipment installed in the plant (4th term is the cost for batch units, 5th term is the cost for semi-continuous units, and 6th term is the cost for storage tanks).

In both techniques, the planning decisions (handled by the second-stage variables in the 2SSLP method and the inner loop LP in the SbO approach), give flexibility to the plant in order to fulfill at least minimum demands. However, this flexibility is limited by the decisions taken in the first-stage or outer loop, i.e., the plant structure. In some scenarios, for a given plant structure, it can happen that no combination of planning variables allow fulfilling demands, leading to infeasibilities. If a given plant structure does not fall in infeasibilities for all the generated scenarios, the *ENPV* becomes the cost of the installed plant plus the aggregated effect of all the scenarios. This aggregated effect is the mean of the applied perturbation, which (for the normal distribution perturbations utilized in this work) is the nominal value for the demand. Therefore, the *ENPV*, for a given plant structure for which no infeasibilities are found, is equal to the case where no parametric uncertainty is considered. On the other hand, if the installed plant cannot comply with minimum demands in one or more of the generated scenarios, the 1st term of Eq. (4) becomes infeasible or null (depending on the utilized solving technique) for that scenario and the aggregated effect decreases with respect to the nominal value.

For the present problem, values of non-uncertain parameters and perturbations applied to demands have been chosen to produce the situation described in the previous paragraph for first-stage variables combinations at the optimum solution and near it. Therefore, solving the problem without considering parametric uncertainty gives a solution whose corresponding first-stage variables are the same as those obtained in the optimal solution for the problem considering uncertainties.

The optimal plant structure for the non-uncertainty case is (V_2, V_3, V_6) = (2000 L, 1500 L, 100 L) with 2 units for the first batch stage and 1 and 2 units for the next two batch stages, and (R_1, R_4, R_5, R_7) = (30 HP, 2.5 m^2, 3 m^2, 30 HP) for semi-continuous stages, with 2, 2, 3 and 1 units operating in parallel respectively, and a storage tank after the second batch stage with size (VT_2) = (4000). The objective function value for this solution is $2,766,549.07.

4.2 Results

Table 3 shows the solution performance of both techniques and the size of the problem that each solved.

Table 3. Solution performance of the 2SSLP and SbO techniques

	2SSLP	SbO
Objective function value	$2740130.32	$2784888.82
CPU time	23306.908 secs	25206.702 secs[†]
Problem size	50 scenarios	Population: 200
	131924 eqs.	Generations: 81
	137109 continuous vars.	Unique solutions: 3100
	70 binary vars.	Best solution found at generation 16
Valid solutions[††]	$2^3*3^4*6^3*6^4*5^2$ = 4534963200	
Found optimal solution?	Yes	Yes

[†] Algorithm stopped by time limit (25000 seconds) violation at generation 81.

[††] Total quantity of valid combinations of the integer variables.

Fig. 2 shows the convergence of the objective function value with the number of scenarios utilized. The dotted line represents the average *ENPV* value. As expected, the computing time to solve the example in each instance increases with the number of scenarios. When the *ENPV* value starts to stabilize the computing time is in the range of 20000-25000 seconds. Therefore, the instance of 50 scenarios has been chosen as the optimal number and its statistics presented in Table 3.
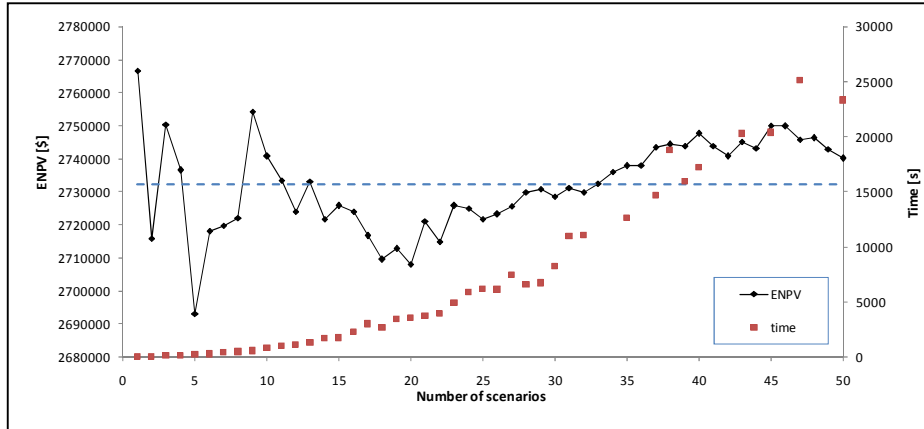
Fig. 2. Objective function convergence for the 2SSLP technique.

Fig. 3 presents the solving performance of the SbO technique. For a better comparison with the 2SSLP method, the algorithm was forced to run until a computing time limit of 25000 seconds. But, it can be seen that the optimal solution is found very early in the process, in generation 16. In the following generations this solution is held as the optimal, but the algorithm continues looking for other solutions that could improve the result, therefore maintaining a great gap between the best and the worst individual in each generation. The objective function value in the worst individual of each generation varies greatly due to the presence or not of unfeasible scenarios, but the average number per individual stabilizes around generation 10.
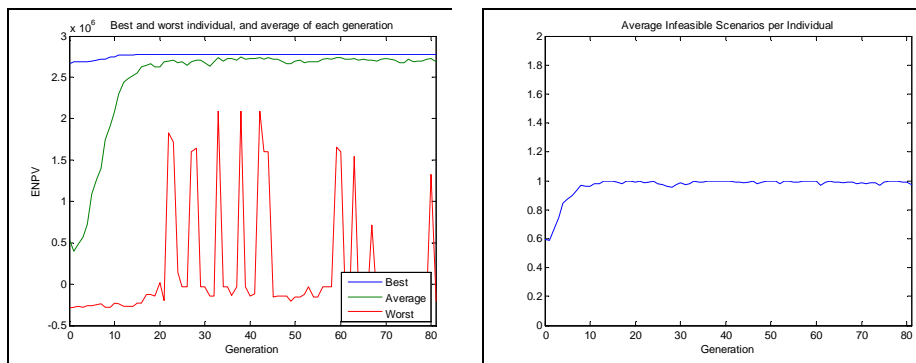

Fig. 3. Objective function convergence for the 2SSLP technique.

As can be seen from Table 3 and Figs. 2 and 3, while both techniques find the optimal solution, the SbO algorithm takes less than half of the computing time. The *ENPV* of both techniques differ from the non-uncertainty considered case. The reason for this difference is that the discretization of the uncertain parameters does not allow for a totally smooth representation, but in both cases the difference is less than 1%.

We can conclude that the SbO technique is better in terms of computing time, but, since it utilizes a metaheuristic algorithm, it cannot determine if it has reached the optimal solution.

5 Conclusions

This paper compares the performance of two techniques for optimization under parametric uncertainty in solving the simultaneous design and planning of a multiproduct plant. The problem presented in this work is very interesting and difficult to solve, even more if uncertainty is considered. However, its characteristics can be modified to allow for its use as a benchmark for different solving approaches, since an optimal solution for the problem can be found before these techniques are applied.

The techniques compared in this work are a rigorous two-stage stochastic programming and a hybrid simulation-based optimization algorithm. The characteristics of both methods where manipulated for a more fitted comparison. The SbO approach had a better performance in terms of computing time, while both of them reached the optimal solution. However, since SbO utilizes a metaheuristic algorithm it cannot ensure having found the best solution, thus leading to a potentially waste of computation power and time.

As a future work, the two-stage stochastic programming with relaxation techniques should be included in the comparison, since their characteristics can lead to a better performance.

References

1. Diwekar, U. Optimization under Uncertainty: An Overview. SIAG/OPT Views-and-News, 13 (1), 1-8 (2002).
2. Birge, Z., Louveaux, S. Principles on Stochastic Programming. New York (USA): Springer-Verlag (1997).
3. Balasubramanian, J., Grossmann, I. E. Approximation to multistage stochastic optimization in multiperiod batch plant scheduling under demand uncertainty. Industrial and Engineering Chemistry Research, 43, 3695-3713 (2004).
4. Cheng, L., Subrahmanian, E., Westerberg, A. W. Design and planning under uncertainty: Issues on problem formulations and solutions. Computers & Chemical Engineering, 27, 781-801 (2003).
5. Cheng, L., Subrahmanian, E., Westerberg, A. W. A comparison of optimal control and stochastic programming from a formulation and computation perspective. Computers & Chemical Engineering, 29, 149-164 (2004).
6. Duenas, A., Petrovic, D. An approach to predictive-reactive scheduling of parallel machines subject to disruptions. Annals of Operations Research, 159 (1), 65-82 (2000).

7. Kuster, J., Jannach, D., Friedrich, G. Applying Local Rescheduling in response to schedule disruptions. Annals of Operations Research, 180 (1), 265-282 (2010).
8. Cheng, L., Subrahmanian, E., Westerberg, A. W. Multiobjective decisions on capacity planning and inventory control. Industrial and Engineering Chemistry Research, 43, 2192-2208 (2004).
9. Jung, J. Y., Blau, G., Pekny, J. F., Reklaitis, G. V. & Eversdyk, D. A simulation based optimization approach to supply chain management under demand uncertainty. Computers & Chemical Engineering, 28, 2087-2106 (2004).
10. Durand, G.A., Mele, F.D., Bandoni, J.A. Determination of storage tanks location for optimal short-term scheduling in multipurpose/multiproduct batch-continuous plants under uncertainties. Ann. Oper. Res. DOI: 10.1007/s10479-011-0970-8 (2011).
11. Durand, G.A, Mele F.D., Guillén-Gosálbez G., Bandoni J.A. Design and planning of the bioethanol supply chain via simulation-based optimization: The case of Argentina, 1° Simposio Argentino de Informática Industrial, 41° JAIIO. Presentación oral. Edición en CD. Páginas 72-81. La Plata (Argentina) del 24–29/8/2012 (2012).
12. Moreno, M.S., Montagna, J.M., Iribarren, O.A. Multiperiod optimization for the design and planning of multiproduct batch plants, Comput. Chem. Eng., 31, 1159-1173 (2007).