Ecodesign of batch processes: optimal design strategies for economic and ecological bioprocesses

DIETZ Adrian\textsuperscript{a}, AZZARO-PANTEL Catherine\textsuperscript{b}, PIBOULEAU Luc\textsuperscript{b}, DOMENECH Serge\textsuperscript{b}
\textsuperscript{a}Laboratoire des Sciences du Génie Chimique, LSGC - Groupe ENSIC, UPR 6811
\textsuperscript{b}Laboratoire de Génie Chimique, LGC UMR CNRS 5503, ,ENSIACET INPT
5 rue Paulin Talabot - BP 1301
31106 Toulouse cedex 1 – France
Catherine.AzzaroPantel@ensiacet.fr

Abstract
This work deals with the multicriteria cost-environment design of multiproduct batch plants, where the design variables are the equipment item sizes as well as the operating conditions. The case study is a multiproduct batch plant for the production of four recombinant proteins. Given the important combinatorial aspect of the problem, the approach used consists in coupling a stochastic algorithm, indeed a Genetic Algorithm (GA) with a Discrete Event Simulator (DES). To take into account the conflicting situations that may be encountered at the earliest stage of batch plant design, i.e. compromise situations between cost and environmental consideration, a Multicriteria Genetic Algorithm (MUGA) was developed with a Pareto optimal ranking method. The results show how the methodology can be used to find a range of trade-off solutions for optimizing batch plant design.

Key-words: Multicriteria optimization - Genetic Algorithm - Batch plant design - Environmental impact

1. Introduction
Traditionally, system optimisation in chemical and process engineering applications has focused on maximising the economic objectives. Over the past 10 years, considerations for improving the environmental performance have started to be integrated into system optimisation alongside economic criteria. These included various waste minimisation approaches from the concept of mass pinch as a tool to derive cost-optimal Mass Exchange Networks with minimum emissions waste, through minimum waste water generation in process plants and waste treatments costs (Gundersen et al., 1988), to the concept of Zero Avoidable Pollution (Linninger et al., 1995). More recently, life cycle thinking (Azapagic and Clift, 1999) has started to be incorporated into the process design and optimisation procedures, thus establishing a link between the environmental impacts, operation and economics of the process. In this context, this paper proposes a methodology for the integration of environmental aspect into the development process, i.e. process ecodesign that is important from both environmental and the economic perspective.

Typically the approach involves three major steps.
(i) Process modelling and carrying out a Life Cycle Assessment study;
(ii) Formulation of the multi-objective optimisation problem in the LCA context;
(iii) Multi-objective optimisation and choice of the best compromise solution.

In this paper, biochemical plants are considered as an application of the proposed methodology, since they generate a wide range of liquid, solid, and gaseous waste streams that require treatment prior to discharge. The proposed work deals with the multi-criteria cost-environment design of multi-product batch plants, where the design variables are the plant configuration with equipment item sizes and parallel equipment number as well as the operating conditions. This article is organized as follows. Section 2 presents the general framework. Section 3 is devoted to the implementation of the proposed methodology. Typical results obtained on a biochemical plant are discussed in Section 4. Section 5 presents conclusions and suggests perspectives.
2. General framework

The framework for batch plant design proposed in this study (see Figure 1) integrates simple unit operation models into the batch plant wide model, which is then embedded in an outer optimization loop. The approach adopted in this work (Dietz et al., 2004) consists in coupling a stochastic algorithm, indeed a Multicriteria Genetic Algorithm (MUGA) with a Discrete Event Simulator (DES). The DES was developed using C++ object-oriented language, keeping the approach proposed in (Bérard et al., 1999). A four layer framework was proposed, resulting in the development of a standard library for the simulator classes that are general to any case, thus minimizing the task of treating different study cases or the variants of a given one (i.e. design or scheduling objectives). In this approach, at the lowest level, the common engine can be found. Most of the events at the next level are common to all batch plant simulations, but some case studies could need the definition of a particular event. Only few equipment items are common to all batch plants (i.e. storage vessels) whereas most of them are particular to each problem and must be defined. The upper layer is the supervisor, which must be generally adapted to each study case.

The objective of the master GA involved is to propose several good and even optimal solutions, whereas the DES checks the feasibility of the proposed configuration and evaluates different criteria with both economic and ecological targets.

![Figure 1. General methodology for optimal batch plant design](image)

Indeed, engineering design problems are usually characterized by the presence of many conflicting objectives that the design has to fulfil. Therefore, it is natural to look at the engineering design problem as a multiobjective optimization problem (MOOP) (Bhaskar et al., 2000; Coello, 2000; Ehrgott, 2000). As most optimization problems are multiobjective by nature, there are many methods available to tackle these kinds of problems. Lately there has been a large development of different types of multiobjective genetic algorithms, which is also reflected in the literature. The big advantage of genetic algorithms over other methods is that a GA manipulates a population of individuals. It is therefore tempting to develop a strategy in which the population captures the whole Pareto front in one single optimization run. This approach was adopted in this study.

The MUGA developed in this study involves different procedures:

1. A method for encoding solutions in strings of digits (or chromosomes); here, a chromosome represents a workshop configuration and the corresponding operating conditions. The encoding procedure will be briefly presented.
2. An initial population has to be randomly generated.
3. An evaluation function which takes a string as input and returns different fitness values which measure the quality of the solution that the chromosome represents relative to each criterion. Since this work is related to minimization cases (investment cost, limitation of pollution), the individual fitness $F_i$ is calculated by:

   $$F_i = C_{\text{max}} - C_i$$

   where $C_i$ is one of the objective function value for individual $i$, $C_{\text{max}}$ is the maximum objective function value computed on the current population for the corresponding criterion.
4. An adaptive plan involving evolution and mutation, based on string crossover and mutation operators.

The cycle (Evaluation, Selection, Crossover and Mutation) is repeated until a stop criterion is reached. After this cycle, the Pareto sort is applied. Concerning selection and multicriteria aspects involved, it
must be emphasized that for a given survival rate, the selection process is achieved via a classical Goldberg’s biased roulette wheel (Goldberg, 1994), relative to each criterion. For this purpose, the initial population is randomly partitioned into sub-populations (the number of sub-populations corresponds to the number of criteria considered simultaneously). A same number of individuals is chosen for each sub-population.

A binary system was chosen for encoding, as it simplifies the genetic operators, i.e., crossover and mutation. This encoding method presented was developed for the cases where the equipment items are identical at a given stage. The continuous variables were discretized and encoded in a binary way by a variable change (Figure 2), using the same bit number (i.e., eight bits). Figure 3 shows a code part used for operating stage encoding. For each stage, the equipment item number was encoded by a binary way (part A in Figure 3). The number of bits reserved to this variable set the maximum equipment item at the stage. For equipment size (L for large, M for Medium, S for Small), a number of bits equal to the available sizes for the equipment items was reserved (part B in Figure 3): the chosen size takes a positive value whereas zero is allocated to the other places. When equipment items are composed of several parts, the same approach is repeated for each component (part B and B' in Figure 3).

**Figure 2. Continuous variables encoding**

**Figure 3. Operating stage encoding method**

#### 3. Implementation

**3.1 Presentation of the illustrative example**

The previous methodology was applied to a batch plant for the production of proteins taken from the literature (Salomone et al., 1992; Montagna et al., 1994, Chiotti et al., 1996; Asenjo et al., 2000). This is a multiproduct batch plant, with four products to be manufactured by fermentation and eight treatment stages. This example is used as a test bench since short-cut models describing the unit operations involved in the process are available. The batch plant involves eight stages for producing four recombinant proteins, on one hand two therapeutic proteins, Human insulin (I) and Vaccine for Hepatitis B (V) and, on the other hand, a food grade protein, Chymosin (C) and a detergent enzyme, cryophilic protease (P). The methodology is generic for any plant producing recombinant proteins from yeast.

Figure 4 shows the flowsheet of the multiproduct batch plant considered in this study. All the proteins are produced as cells grow in the fermenter (Fer).

**Figure 4. Multiproduct batch plant for proteins**
Vaccine and protease are considered as being intracellular, hence, for these two products, the first microfilter (Mf1) is used to concentrate the cell suspension, which is then sent to the homogenizer (Hom) for cell disruption to liberate the intracellular proteins. The second microfilter (Mf2) is used to remove the cell debris from the solution proteins.

The ultrafiltration (Uf1), prior to extraction, is designed to concentrate the solution in order to minimize the extractor volume. In the liquid-liquid extractor (Ext), salt concentration (NaCl) is used to first drive the product to a poly-ethylene-glycol (PEG) phase and again into an aqueous saline solution in the back extraction.

Ultrafiltration (Uf2) is used again to concentrate the solution. The last stage is finally chromatography (Chr), during which selective binding is used to better separate the product of interest from the other proteins.

Insulin and chymosin are extracellular products. Proteins are separated from the cells in the first microfilter (Mf1), where cells and some of the supernatant liquid stay behind. To reduce the amount of valuable products lost in the retentate, extra water is added to the cell suspension.

The homogenizer (Hom) and microfilter (Mf2) for cell debris removal are not used when the product is extracellular. Nevertheless, the ultrafilter (Uf1) is necessary to concentrate the dilute solution prior to extraction. The final step of extraction (Ext), ultrafiltration (Uf2) and chromatography (Chr) are common to both the extracellular and intracellular products.

3.2. Environmental Impact Evaluation

Several methodologies for environmental impact (EI) consideration are available in the literature. The most important concept refers perhaps to the Life Cycle Assessment (LCA) (Burgess and Brennan, 1999) considering all the wastes generated in order to produce the different products in the upstream stages (i.e., raw material production, energy generation, etc.), in the study stage (i.e. solvents, non-valuable by-products, etc) and in the downstream steps (i.e. recycling, incineration, etc). The aim of LCA is to consider the wide chain in order to prevent pollution generation and to compare the different alternatives to produce a product. Another concept used the Pollution Balance (PB) principle to carry out a pollution balance (Cabezas et al., 1999) equivalent to the balance made for mass or energy. It means that a process can not only pollute but also consume a polluting product and will be a benign process.

Finally, the Pollution Vector (PV) methodology (Stefanis et al., 1995) consists in evaluating the environmental impact by means of an impact vector over different environments (i.e. water, air, etc) defined as the mass emitted on an environment divided by the standard limit value in this environment.

Given the production recipes for the different products and the general flow-sheet, the first step consists in applying the LCA methodology to determine all the products contributing to the environmental impact (fig. 4). For information availability reasons, the study was reduced to the process being studied, which is of course a limited application of LCA. Products (i.e. vaccine) and raw materials (glucose, NH3) were considered not having an environmental impact. After that, a PB is applied, using the PV to quantify the environmental impact. In this case, an adapted definition of the pollution vector was introduced, because the standard limit values for the polluting product were not found in the literature. This vector has two components; the former is the total biomass quantity released and the latter concerns the PEG volume used. Even if the solvent can be recycled, it cannot be carried out at 100%, so the environmental impact is considered to be proportional to this quantity. The pollution indexes were thus defined as the emitted quantities divided by the mass of the manufactured products. Let us remark that the environmental impact minimization can be viewed a multicriteria problem in itself.

The global index of each environmental impact criterion is defined as weighted sum respect to the production of each product index (eq. 1).

\[
I_k = \frac{I_k^{wm} \cdot P_{wm} + I_k^{pest} \cdot P_{pest} + I_k^{ity} \cdot P_{ity} + I_k^{pro} \cdot P_{pro}}{P_{wm} + P_{pest} + P_{ity} + P_{pro}}
\]

where \(I_k\) is the pollution global index, \(I_k^i\) is the \(k\) pollution index of \(i\) product defined as the kilograms of the pollutant \(k\) by kilograms of manufactured product \(i\) and \(P_i\) is the total production of the \(i\) product.

The cost criterion considered in this study is classically based on investment minimization because there was not enough information to evaluate the operational cost of the batch plant (raw material cost, utilities cost ...) and to embed it in a net present worth computation.
The optimization criterion involving investment cost for equipment and storage vessels $\text{ICost}$ is calculated using (1):

$$\text{ICost} = \sum_{i=1}^{N_{\text{OP}}} \sum_{j=1}^{N_{\text{EQ}_i}} \left( A_i + B_i V_{ij}^{\text{EQ}_i} \right) + \sum_{k=1}^{N_{\text{SV}}} \left( A_k + B_k V_{sk}^{\text{SV}_k} \right) \tag{2}$$

where $N_{\text{OP}}$ is the number of operations, $N_{\text{EQ}_i}$ is the number of equipment items (for operation $i$), $N_{\text{SV}}$ is the number of storage vessels, $A_i$, $B_i$ and $C_i$ are the cost coefficients for operation $i$, $A_s$, $B_s$ and $C_s$ are the cost coefficients for storage vessels, $V_{ij}$ is the volume of equipment $ij$ and $V_{sk}$ is the volume of storage vessel $k$.

The models representing the operation units involved in the global process are presented in detail in (Dietz et al., 2005) and will not be recalled here.

The optimization problem considered can be formulated as follows:

$$\begin{align*}
\text{min } f_1(y) \\
\text{min } f_2(x) \\
\text{s.t. } g(x,y) \leq H
\end{align*}$$

where, $f_1$ represents the investment cost and $f_2$ the environmental impact. The $x$ vector, $x = [x_1, x_2, \ldots, x_n]$ represents the operating conditions and $y = [y_1, y_2, \ldots, y_m]$ refers to batch plant configuration.

Without going into further detail, the optimization problem involves 44 variables, which may be either continuous (i.e. the operating conditions) or discrete (parallel equipment number, equipment size).

A set of data must be fixed by the user concerning the optimization problem definition before the implementation of the design methodology (see Dietz et al., 2005). For instance, the annual demand for each product is presented in Table 1.

### Table 1. Product demands

<table>
<thead>
<tr>
<th>Product</th>
<th>Production (kg/year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin</td>
<td>1500</td>
</tr>
<tr>
<td>Vaccine</td>
<td>1000</td>
</tr>
<tr>
<td>Chymosin</td>
<td>3000</td>
</tr>
<tr>
<td>Protease</td>
<td>6000</td>
</tr>
</tbody>
</table>

Three sizes are available for each equipment item: large (L), medium (M) and small (S). The classical expressions used for computing the investment cost of the equipment items follows a classical scaling law. Of course, a thorough economic study would also include the operating cost estimation and analysis of profitability. Since this kind of analysis only requires reliable economic data for a real process and does not induce additional difficulties in the chosen resolution strategy, a capital cost-based study was finally adopted for the preliminary economic evaluation of the project for manufacturing biological products.

Table 2 displays the parameters of the genetic algorithm used for multicriteria batch plant design. In this work, the generation number was fixed as twice the population size. The global survival rate is relatively low as compared to standard values for optimization of test mathematical functions (Dedieu et al., 2003). Moreover, a high mutation rate was set. Although a systematic study was not carried out to find these values, they were chosen from several preliminary tests and agree with previous works (Dedieu et al., 2003) where similar problems were treated. The elitism was used in order to avoid losing the best solution for each criterion. Considering the stochastic aspect of GAs, several optimization runs were carried out for each multicriteria optimization. Given that solutions obtained in one optimization run could be dominated by solution of another one, a Pareto sort procedure is applied to the set of solutions obtained at each optimization run, and the non-dominated solutions are those proposed by the methodology.

### Table 2. Genetic algorithm parameters

<table>
<thead>
<tr>
<th></th>
<th>Bicriteria Solvent-biomass</th>
<th>Bicriteria Cost-Solvent</th>
<th>Bicriteria Cost-Biomass</th>
<th>Tricriteria Cost-EI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>300</td>
<td>450</td>
<td>450</td>
<td>600</td>
</tr>
<tr>
<td>Generation number</td>
<td>600</td>
<td>900</td>
<td>900</td>
<td>1200</td>
</tr>
<tr>
<td>Survival rate</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Mutation rate</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Elitism by criterion</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
4. Results and Discussion

Only typical results concerning bicriteria cases are presented in what follows. Two strategies were tested either monoproduct or multiproduct campaigns with a batch of each product to be produced. In the latter case, the starting order of the different batches was fixed altering intracellular and extracellular products. The MUGA presented in this work was first used to demonstrate that the two EI criteria considered, that are respectively the total biomass quantity and the PEG volume, present antagonist goals (figure 5). Very similar results were obtained at each optimization run, so only the results after the final Pareto sort procedure are presented in figure 6. Moreover, it must be noted that slight differences are obtained between both production policies because the environmental impact depends only on the mass balance that is function of the continuous variables.

This antagonist behavior can be explained at the liquid – liquid extraction stage. The more solvent is used, the more efficient the stage becomes and, consequently, the fewer products are lost, reducing the environmental impact index computed as Kg of biomass released by Kg of final product.

The same approach was also applied to the cost – environment criteria. First, the amount of solvent used and the investment cost were considered.

For illustration purposes, figure 6 shows the results obtained at each optimization run for the monoproduct production policy, performed with an identical parameter set to guarantee the stochastic nature of the GA. In this case, the results are not superposed as it was the case for the bicriteria optimization biomass – solvent and show the need of carrying out several optimization runs for the same problem.

In figure 6, it can be seen that each optimization run is oriented to a section of the search region. The first optimization comes up with the better solution for the cost criterion, the second for the environmental criterion and the third is a compromise between both. The final Pareto sort procedure is carried out over these solutions. The final results for both production policies are presented in figure 7. Let us note that the Pareto zone is constituted of sparse points, since the adaptation function related to the investment cost takes discrete values.

Slight differences were found between both production policies. The antagonist behavior between these two criteria, investment cost – amount of solvent used, can be explained by a compromise between the solvent yield and the process global yield. When process yield is penalized, a bigger, and consequently more expensive, batch plant is required.

In order to evaluate the search performance of the GA, table 3 presents the best solution obtained at each optimization run for each criteria considered as well as the best solution obtained with a monocriterion approach. Even though the methodology was not able to find the best solution, the values are relatively near (around 5% more expensive for the investment cost criterion). It must be noted that in the monocriterion optimization (not presented here), the best value was obtained only once and, in the other cases, the solutions were around 2%-3% more expensive, which justifies the results when several criteria are taken into account simultaneously. The number of solutions obtained in each optimization run was around of 25. It is important to note that in each case, the solutions are uniformly distributed in the search space; this means that there is no preferential search region in the multicriteria search as shown in Figure 6.
It is also interesting to see where the results are placed with respect to the criterion not considered here, in this case the amount of biomass released. Table 4 presents the range of values for this criterion for both production policies.

<table>
<thead>
<tr>
<th>Biomass for Cost-Solvent</th>
<th>Solvent for Cost-Biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>Monoproduct</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiproduct</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

They have the same order of magnitude for both policies. Moreover, the minimal value of the range is close to the best value obtained in monocriterion optimization value which allows to predict less antagonism between investment cost and biomass released criteria.

The last bicriteria optimization considers the investment cost and biomass released. As for the previous case, three optimization runs were carried out for each production policy. The results obtained after the final Pareto sort procedure are presented in Figure 8 and are similar for both production policies as it was shown for the cost-solvent criteria.
5. Conclusions

A methodology was proposed for batch plant design, considering both investment cost and environmental impact minimization. An optimization scheme has been implemented using a multiobjective genetic algorithm with a Pareto optimal ranking method. This technique is ideally suited to this type of problem, where a number of conflicting considerations must be taken into account. The use of MUGA makes possible a robust optimization technique, across a non-linear search space (the objective functions are computed by the use of a discrete event simulator (DES) integrating shortcut unit operations models) linking multiple variables and objectives. The paper clearly shows that opportunities for process optimization and environmental impact minimization must be considered at the early stages of process development before the process is frozen due to regulatory reasons.

References


