Multiobjective strategies for New Product Development in the pharmaceutical industry

José L. Perez-Escobedo, Catherine Azzaro-Pantel*, Luc Pibouleau

Université de Toulouse, Laboratoire de Génie Chimique, UMR 5503 CNRS, ENSIACET INPT, 4, Allée Emile Monso, BP 84234-31432 Toulouse Cedex 4, France

A B S T R A C T

New Product Development (NPD) constitutes a challenging problem in the pharmaceutical industry, due to the characteristics of the development pipeline. Formally, the NPD problem can be stated as follows: select a set of R&D projects from a pool of candidate projects in order to satisfy several criteria (economic profitability, time to market) while coping with the uncertain nature of the projects. More precisely, the recurrent key issues are to determine the projects to develop once target molecules have been identified, their order and the level of resources to assign. In this context, the proposed approach combines discrete event stochastic simulation (Monte Carlo approach) with multiobjective genetic algorithms (NSGAII type, Non-Sorted Genetic Algorithm II) to optimize the highly combinatorial portfolio management problem. In that context, Genetic Algorithms (GAs) are particularly attractive for treating this kind of problem, due to their ability to directly lead to the so-called Pareto front and to account for the combinatorial aspect. This work is illustrated with a study case involving nine interdependent new product candidates targeting three diseases. An analysis is performed for this test bench on the different pairs of criteria both for the bi- and tricriteria optimization: large portfolios cause resource queues and delays time to launch and are eliminated by the bi- and tricriteria optimization strategy. The optimization strategy is thus interesting to detect the sequence candidates. Time is an important criterion to consider simultaneously with NPV and risk criteria. The order in which drugs are released in the pipeline is of great importance as with scheduling problems.

Keywords:
New Product Development
Portfolio management
Discrete event simulation
Optimization
Multicriteria genetic algorithm

1. Introduction

Traditionally, Process Systems Engineering (PSE) is concerned with the understanding and development of systematic procedures for the design and operation of chemical process systems, ranging from microsystems to industrial scale continuous and batch processes. This traditional definition of PSE has been broadened by the concept of the "chemical supply chain". Process Systems Engineering is now concerned with the improvement of decision making processes for the design and operation of the chemical supply chain. More precisely, it deals with the discovery, design, manufacture and distribution of chemical products in the context of many conflicting goals. The area of R&D and Process Operations has emerged among the major challenges in the PSE area: this topics, which has a shorter history than process design and control, expands upstream to R&D and downstream to logistics and product distribution activities.

In that context, optimal planning and scheduling for New Product Development (NPD) need increased attention to coordinate better product discovery, process development and plant design in the agrochemical and pharmaceutical industries. For downstream applications, areas that receive increased attention at the business level include planning of process networks, supply chain optimization, real time scheduling, and inventory control. Due to the pressure for reducing costs and inventories, in order to remain competitive, enterprise-wide optimization (EWO) that might be considered as an equivalent term for describing the chemical supply chain (Shapiro, 2001) has thus become a cornerstone in process industries.

Enterprise-wide optimization is an area that lies at the interface of Process Systems Engineering and Operations Research. As outlined in Grossmann (2005), a new generation of methods and tools that allow the full integration and large-scale solution of the optimization models, as well as the incorporation of accurate models for the manufacturing facilities is needed. Given the strong tradition that chemical engineers have in process systems engineering and in the optimization area (see Biegler & Grossmann, 2004 for a review), they are ideally positioned to make significant contributions in EWO.

The development of decision support strategies and systems for managing new product portfolios must be able to provide insights to managers on how to minimize risk while optimizing an objective or a set of objectives (e.g. maximization of expected net present

* Corresponding author.
E-mail address: Catherine.AzzaroPantel@ensiacet.fr (C. Azzaro-Pantel).

doi:10.1016/j.compchemeng.2011.10.004

value, minimization of time to market, etc.) in the presence of constraints. Moreover, the simultaneous consideration of all candidate projects is the key aspect in managing a NPD pipeline. This complexity has led to the common use of decomposition based in either strategic or operational strategies. Each of the two branches can be further subdivided according to the characteristics of the model used to support the decision making process.

An interesting contribution (Zapata, Varma, & Reklaitis, 2007) proposes a recent state-of-the art of the concerned problem. The main guidelines of their analysis are briefly recalled to position our work. First, a project can be analyzed in isolation (e.g. the net present value (NPV) of the project), or as performance assessment at the portfolio level (e.g. NPV of the portfolio), including all the interactions between projects. The time dimension distinguishes dynamic and static approaches. A dynamic model provides the specific state of the systems along each point of the time horizon (e.g. number of projects waiting for a given resource at a given time), while a static one uses average values to represent the system (e.g. average number of projects waiting for a given resource at any time). It is then possible to choose between deterministic and stochastic models. However, dynamic stochastic models can be viewed as either open loop or closed loop oriented. Open loop models only capture the response of the system to inputs from decision makers, while closed loop models also capture the response of the decision makers to the outcomes from the system.

Among the investigations dedicated to strategic decision support systems, the different techniques available depend on the type of data used, namely, qualitative and quantitative. On the one hand, the methodologies relative to static strategies are numerous in this area: it must be emphasized that a major drawback of such approaches is that they do not take into account project interactions. They include scoring methods (Coldrick, Longhurst, Ivey, & Hannis, 2005; Cooper, Edgett, & Kleinschmidt, 1999), analytical hierarchy approaches (Calantone, Benedetto, & Schmidt, 1999; Poh, Ang, & Bai, 2001) and fuzzy logic based approaches (Buyukozkan & Feyzioglu, 2004; Lin & Hsieh, 2004; Lin, Tan, & Hsieh, 2005). On the other hand, the methodologies that are based on quantitative information strive to provide a realistic simulation of the behaviour of each individual project along the time horizon considered, in order to determine what the possible outcomes are in terms of rewards and risk. This group includes dynamic deterministic strategies such as classical financial models (e.g. NPV, internal rate of return, etc.) (Cooper et al., 1999), as well as dynamic stochastic strategies, both closed loop such as real options (Copeland & Antikarov, 2001; Jacob & Kwak, 2003; Loch & Bode-Greuel, 2001; Newton, Paxson, & Widdicks, 2004; Santiago & Bifano, 2005), and open loop such as discrete event simulation (Chapman & Ward, 2002), and neural networks (Thieme, Song, & Calantone, 2000).

Most of the approaches that capture project interactions can be classified as dynamic stochastic open loop methodologies. An important contribution is the work of Blau, Pekny, Varma, and Bunch (2004) which proposes the use of stochastic optimization: the portfolio is modelled using a discrete event simulation and the optimization is implemented by a genetic algorithm; Rogers, Gupta, and Maranas (2002) formulate a real options decision tree that captures technical and market uncertainty as a stochastic MILP (Mixed Integer Linear Programming) that relates projects through a budget constraint. Rajapakse, Titchener-Hooker, and Farid (2005) present a decision support tool that uses sensitivity and scenario analysis on a discrete event model of the development pipeline. Finally, Ding and Eliashberg (2002) approach the problem of determining how many projects, that are assigned to develop the same product, have to be included in the pipeline to maximize the total expected profit. All of the techniques in this group are mainly focused on time independent decisions (excluding the work reported in Rogers et al., 2002) and therefore do not require closed loop models. Some work has been done to accommodate the higher level of complexity required by time dependent strategic decisions such as capacity expansion/contraction (Wan, Pekny, & Reklaitis, 2006). It must be yet pointed out that the non-Markovian nature of the associated decision problem has yet limited the size of the treated problem as expressed in project number.

At the operational level, decisions are time dependent and mostly Markovian in nature. This has motivated the development of operational decision support systems exclusively based on quantitative information and with a dynamic character (Honkomp, 1998; Jain & Grossmann, 1999; Subramanian, Pekny, Reklaitis, & Blau, 2001; Varma, 2005).

This literature review reveals that it is difficult to embed all the peculiarities of the problem in a generic formulation and to reconcile all levels at the involved scales. The complexity of the problem is attributed to several combined issues such as the stochastic behaviour of the system, the combinatorial aspect and consequently the size of industrial problems as well as the induced multilevel approach. Some recent works (Colvin & Maravelias, 2008, 2009) are trying to reconcile both strategic and operational levels, namely, the scheduling of clinical trials and the planning of the resources necessary to carry these trials out. A stochastic programming framework that addresses the two problems simultaneously is proposed in Colvin and Maravelias (2009). The underlying philosophy implies three levels: first, the structure of the problem is studied in order to reduce the number of pairs of scenarios; second, a finite-horizon approximation is developed so that problems can be formulated using fewer stages without compromising the quality of the solution; third, the sequential nature of the testing process is considered and modelled with a mixed-integer programming (MIP) formulation; a relaxation of this formulation is then used to obtain feasible and most often optimal solutions over the stages of interest. Finally, a rolling-horizon-based approach is implemented, where the decisions of the relaxed problem are used over few early periods and a new problem is formulated and solved as time evolves. This framework was recently improved including: (i) the selection and scheduling of R&D tasks with general precedence constraints under pass/fail uncertainty, and (ii) resource planning decisions (expansion/contraction and outsourcing). Furthermore, interdependencies between tasks in terms of probability of success, resource usage and market impact are considered with risk management approaches, taking into account conditional value at risk (Colvin & Maravelias, 2011), that was never considered in previous works. It must be also emphasized that all the reported approaches are based on a monoobjective optimization formulation even if the problem is multiobjective by nature.

This work is devoted to the development of a dynamic stochastic open loop methodology and involves a bi-and tricriteria optimization formulation of the NPD problem. It involves multi-stage decisions under uncertainty. The recurrent key issues are can be stated as follows: what are the projects to develop once target molecules have been identified? In what order? Which is the level of resources to assign? The proposed modelling approach is based on a discrete event simulator which is particularly useful for decision criteria evaluation, such as economic and risk metrics. This work can be viewed as an extension of the investigations previously dedicated to batch plant design and scheduling which are of major importance for such industries and which can be considered as part and parcel of the more general topics of NPD management. This kind of involves several criteria, the Net Present Value of a sequence, its associated risk (measured by an attractiveness ratio or by the so-called positivity probability) and the makespan that must be optimized simultaneously. Section 2 is first devoted to the key issues involved in New Product Development. Section 3 presents the principles of the discrete event simulation model developed and implemented for describing the pipeline behaviour.

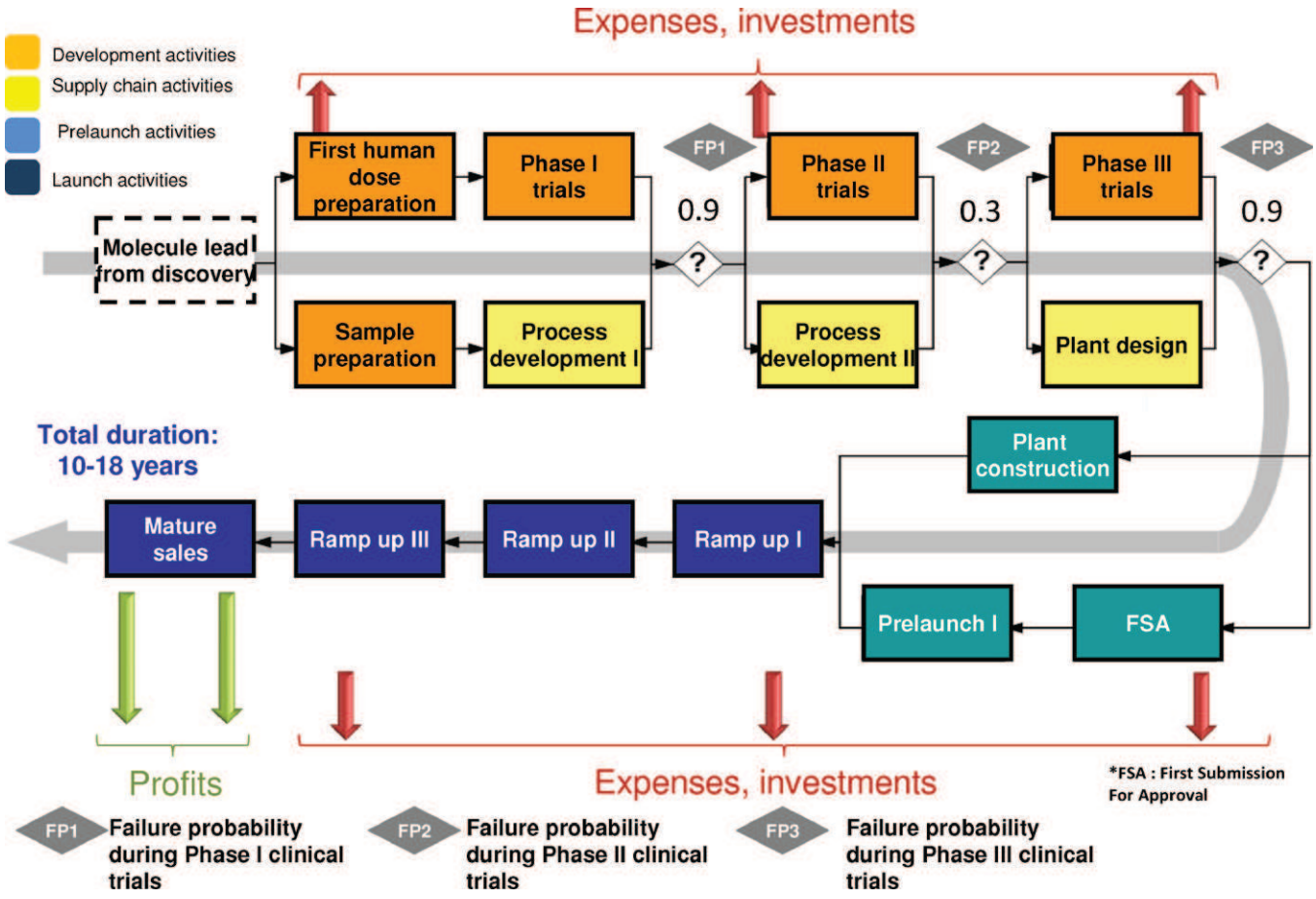

Fig. 1. Process for drug development.

Section 4 describes the formulation of the multiobjective optimization problem. The different optimization methods that may be used are briefly recalled with a special emphasis to Genetic Algorithms (GAs), that are particularly attractive for treating this kind of problem, due to their ability to directly lead to the so-called Pareto front. Among the various GAs, a discussion is then performed to select the most appropriate variant. The selected NSGAII algorithm is then applied to the treated case in Section 5. Sections 6 and 7 analyze and discuss the test bench examples and provide with some guidelines for the treatment of new cases from the bi and tricriteria viewpoints. Finally, Section 8 summarizes the major results of the paper.

2. Key issues in New Product Development

2.1. Life cycle of a pharmaceutical product

Basically, several stages are involved in the life cycle of a pharmaceutical product as it can be summarized in Fig. 1. In the Discovery stage, thousands of molecules are applied to targets developed to simulate various disease groups. Once an active molecule, i.e. a molecule that is identified to have a curative effect on the target, is discovered, various permutations of the structure of the molecule are tested to see if the activity can be enhanced. The most active molecule from these structure–activity relationships is tested for toxicological results on rats or mice. If no particular worrisome toxic endpoints are observed, the molecule is promoted to the status of "lead" molecule and becomes a candidate for development. In the Development stage, enormous sums of money and resources are committed to the lead molecule to first, observe its behaviour in healthy volunteers, secondly, in patients smitten with the disease and finally, in large scale clinical studies conducted in concert with the Food and Drug Administration (FDA). Coincident with these studies, process research and formulation work is conducted to both supply the drug for testing purposes as well as to design and construct a commercial plant if the product is launched. Other parallel studies involve extensive long-term (i.e. two years) chronic studies in animals to identify any indication of oncogenicity at different dosage levels. If the drug is effective in the clinical studies, has no unacceptable side effects and is blessed by the FDA, it moves to the Commercial Stage. Target markets are identified for a staged launch or "ramp-up" of the new compound. After a few years, a mature sales level is usually reached and maintained until patent coverage on the molecule expires and/or competition from generics is realized. Once generics are available, an attempt is usually made to get approval of the drug for alternative markets and perhaps in different dosage forms. Regardless, sales are diminished after expiration of the patent.

Some dependencies are considered for representing relationship between drugs for the same disease: financial dependency; technical dependency; manufacturing cost dependency; resources dependency. The system parameters are summarized in Fig. 2. It must be pointed out that a relationship between activity times and costs for specific drug candidates is typically considered as in Blau et al. (2004). This relationship is captured with a simple parameter called the degree of difficulty (DoD). Subjective estimates of DoD can be obtained from the various principal investigators, although the values may be different between work processes. However, since the focus is on project selection and sequencing rather than resource planning, the analysis can be simplified by using a single

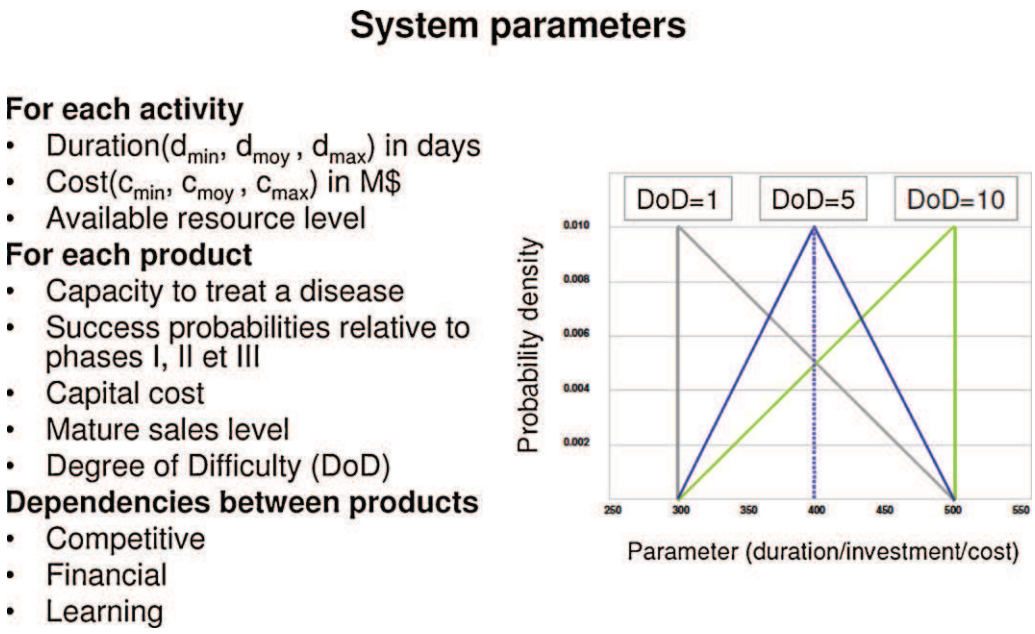

Fig. 2. System parameters.

value of DoD ranging from 1 (very easy) to 10 (very difficult). The case study used in this work is derived from the work of Blau et al. (2004). It must be highlighted that we have several fruitful discussions with a French pharmaceutical company to assess the validity of the example that is tackled here and that serves as a guideline of the methodological framework. Some typical problem data are displayed in Table 1. The interested reader can have all the complementary information in Perez-Escobedo (2010). By lack of place, all the data will not be extensively presented here. It must be said at that level that some slight differences exist between the example treated by Blau et al. (2004) and the one adopted here. They concern the type of dependencies between products. Actually, product development is generally influenced by the other products considered in the pipeline and by competitor products. It is assumed that all dependencies (except for technical dependency) will occur. When considered, this kind of dependency modifies the success probability. On the one hand, if the first drug in the sequence of drugs targeted for Disease I fails, the probability of technical success for all succeeding drugs decreases by a given percentage. On the other hand, if the first in the sequence for testing Disease I succeeds, the probability of technical success for all succeeding drugs for Disease I increases by a given percentage. It must emphasized that this technical dependency is not quite common in the pharmaceutical industry and is even a controversial issue from the fruitful discussions with French pharmaceutical managers. This explains why it has not been taken into account for modelling.

2.2. Classical approaches of resolution

A fundamental challenge in managing a pharmaceutical or biotechnology company is identifying the optimal allocation of finite resources across the infinite constellation of available investment opportunities. In that context, the optimal management of the new product pipeline has emerged at the forefront of all strategic planning initiatives of a company. This issue is traditionally identified as a complex one since it integrates various areas such as product development, manufacturing, accounting and marketing. The complexity of the problem is mainly attributed to the great

Table 1
Data for the 9-drug problem with resource limitations.

Activity	Duration (days)			Cost (M$)			Total available resources (M$)
	Min	ML	Max	Min	ML	Max	
FHDP	300	400	500	72	80	88	275
Sample prep	300	400	500	1.8	2	2.2	10
Phase I	225	300	375	70	80	90	350
Phase II	375	500	625	75	80	85	175
Phase III	575	775	975	150	200	250	250
Process develop I	600	800	1000	7	10	13	16
Process develop II	600	800	1000	7	10	13	16
Design plant	550	750	950	8	10	12	12
FSA	275	375	475	18	20	22	100
Prelaunch	75	100	125	45	50	55	550
Build plant	600	750	900	52	62	72	120
Ramp up I	250	350	450	9	12	15	25
Ramp up II	250	350	450	19	22	25	50
Ramp up III	250	350	450	35	40	45	100
Mature sales	250	350	450	46	53	60	150

variety of parameters and decision-making levels involved. A strategic investment plan should simultaneously address and evaluate in a proper manner the following four main issues: product management, clinical trials uncertainty, capacity management and trading structure. It is also generally viewed as a multistage stochastic portfolio optimization problem. The main challenge is to configure a product portfolio in order to obtain the highest possible profit, including any capacity investments, in a rapid and reliable way. These decisions have to be taken in the face of considerable uncertainty as demands, sales prices and outcomes of clinical tests that may not turn out as expected.

This kind of problem has recently received attention from the process systems engineering community utilizing previous works from the process planning and scheduling area. Various MILP optimization models are proposed in Schmidt and Grossmann (1996) for the scheduling of testing tasks with no resource constraints with a discretization scheme in order to induce linearity in the cost of testing. These models are extended in Jain and Grossmann (1999) to account for resource constraints. A simulation-optimization framework (Subramanian, Pekny, Reklaitis, & Blau, 2003) takes into account uncertainty in duration, cost and resource requirements as well as risk. An MILP model is proposed in Maravelias and Grossmann (2001) that integrates the scheduling of tests with the design and production planning decisions. A literature review of optimization approaches in the supply chain of pharmaceutical industries can be found in Shah (2004). The work of Blau et al. (2004) is based on a monoobjective Genetic Algorithm to optimize product sequence evaluated by a commercial discrete-event simulator.

This work lies in the perspective of implementing efficient optimization tools: the underlying idea is to use a multiobjective framework as already initiated by Aguilar-Lasserre, Azzaro-Pantel, Pibouleau, and Domenech (2007) to model both the conflicting nature of the criteria (i.e. risk minimization and profitability maximization) and the imprecise nature of some parameters (demand, operating times, . . .). In that context, this work aims at the development of an architecture that combines an optimization procedure and a simulation model to represent the dynamic behaviour of the pipeline with its inherent uncertainty and to help decision-making. The general objective is thus to propose a general methodology framework to support decisions and management of pharmaceutical products involved in their life cycle, from early stages to mature sales.

3. Discrete event simulation for NPD pipeline modelling

The main motivation of this work was to propose an optimization framework to select a set of R&D projects from a pool of candidate projects in order to maximize the expected benefits while coping with the uncertain nature of the projects. This is a challenging problem due to the characteristics of the development pipeline, namely, the presence of uncertainty, the interdependency between projects, the limited availability of resources, the overwhelming number of decisions due to the length of the time horizon and the combinatorial nature of a portfolio.

In that context, discrete event simulation is a common tool used to understand how a system works and how the different items interact each other. It must be said that discrete event simulation has been mostly confined to production systems (batch plant scheduling for production debottlenecking, batch plant design, etc.), but the trend in many industries of moving towards an integrated approach for supply chain management has expanded the areas in which this technology can be used. The analysis also highlights that all the processes involved in New Product Development are characterized by uncertainty at various levels of the pipeline: imprecise parameters for activity cost and durations as well as success probabilities at Phases I, II and III of the pipeline. In our research group, the development of discrete event simulators for batch plant design and scheduling has been a constant focus for the past decade (Baudet, Azzaro-Pantel, Domenech, & Pibouleau, 1999; Bérard et al., 1999; Dietz, Azzaro-Pantel, Pibouleau, & Domenech, 2005). Moreover, on the implementation side, it is not so easy to use commercial simulators capable of interacting with optimization packages or user written code. A major incentive to use discrete event simulation is that processes characterized by uncertainty and suitable for probabilistic modelling can be easily analyzed and synthesized using discrete event simulation by use of a Monte Carlo approach. An object-oriented model structure previously developed for batch plant scheduling and design was then extended to embed the case of product management, which is particularly adequate for reuse of both structure and logic. Its detailed presentation is not the purpose of this paper which has been presented in Perez-Escobedo (2010).

Table 2
Terminology in BPS and NPD project problems.

Batch plant scheduling (BPS)	NPD project
Product #*i*	Project related to a product #*i* (PRP #*i*)
Equipment item #*j*	Resource of a given step #*j*
Recipe #*k*	Succession of activities #*k* (also called recipe)
Unit operation of a recipe #*l*	Activity #*l*

A four layer framework was proposed in Bérard et al. (1999) based on the following items engine, event, object, supervisor, the aim being 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. Initially, the events in the next level are generic events common to all batch plant simulations: in this case, the definition must be adapted since we have to consider the whole life cycle of a project related to a product. In the same way, the objects taken into account present some similarities but differ in their appreciation: for instance, in batch plant scheduling problems (BPS), material resources are constituted by equipment whereas in NPD problems, resources may be viewed more globally. In fact, the main differences at this step occur from a terminology point of view and this can be easily transposed in the NPD formulation. The core of the simulator is the *Engine*, which has two functions: the former is to order the Events in its Calendar by their occurrence date whereas the latter is to activate them if the necessary resources are available; if not, it reports the Event to a next date. An Event represents a change of the real system at a given time. The class *Event* is a basis class from which the different events must be defined. If resources for this activity are available, the *Event* is activated; conversely, if resources are not available, the activity will be scheduled later. An *Event* is characterized by its occurrence date, its action over the system and a type that enables to give priorities when two or more *Events* have the same occurrence date. As a general rule, *Events* which release resources have priority over the others, and when *Events* have the same type, the classical FIFO rule (First In First Out) is applied. This will be useful when different projects compete for the same resources. The *Event* Class previously developed was generic enough to embed the NPD formulation (see Table 2).

We focused on developing a simulation decision support tool that uses probabilistic data in the form of durations of activities, resource requirements (modelled as capital and operating costs), clinical success probabilities and product sales, and computes a schedule. The resulting schedules and resource allocation levels can be used to infer efficient project prioritization and resource allocation policies under uncertainty.

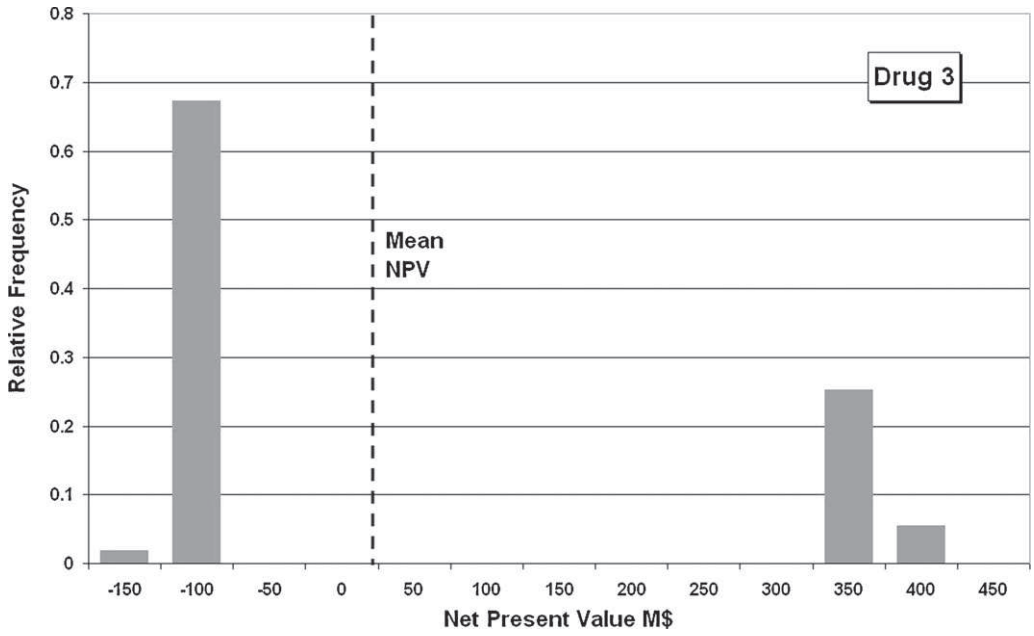

Fig. 3. Relative frequencies for product-project3.

Before considering a portfolio of products, it is interesting to examine the behaviour of each individual drug candidate. Using net present value (NPV) with an internal rate of return of 15% as the economic criterion, the behaviour of each drug can be simulated by using the discrete event simulator. A typical example is presented in Fig. 3. To take into account the imprecise nature of some parameters, NPV distribution was obtained from a sufficiently large number of Monte Carlo trials. A two-peaked distribution is observed which is typical of a new drug candidate in the pharmaceutical industry. The first peak corresponds to the loss of money in those instances when the drug fails to pass all the clinical trials. The second distribution corresponds to the returns following a successful product launch. Due to this bimodal distribution, the economic criterion must be clearly defined and necessarily associated with a risk criterion to evaluate the quality of sequence: it must be pointed out that the Expected Net Present Value (ENPV) that is commonly used for project evaluation must be considered carefully. It corresponds to a mean value between the positive and negative parts of the distribution. If considered at the optimization step, it represents a pessimistic value of the NPV but evolves in the same direction of the ENPV. The risk is appreciated by the computation of the positive values of NPV over the total number of samples that have been evaluated. Moreover, the makespan that is the time to market is also a criterion that needs to be considered at portfolio selection. This analysis allows to define the most important criteria that must be taken into account to define the best drug portfolio.

A special emphasis has then been devoted to uncertainty modelling in NPD. The uncertainty considered in this work is twofold: it is associated to project success, which is an endogenous uncertainty that can be represented by a discrete number of realizations (i.e., clinical trial failed or approved) but also with the uncertainty associated with time and cost parameters. Traditionally, two classes of methods of imprecision representation have become important: probability theory and non-probabilistic uncertainty modelling. The former class attempts to model uncertain parameters as random variables: imprecise parameters are associated with a probability distribution within a Monte Carlo framework. The concept of Degree of Difficulty was used to reflect the more or less difficulty to carry out a process task. The latter class includes interval computation and fuzzy set theory. Two approaches were implemented, a classical probability approach and an interval-based one. Both approaches have been illustrated by a numerical example which has shown that the tendencies obtained by the interval-based approach may be difficult to interpret for the decision maker, due to the growing uncertainty along the pipeline. Besides, the risk, which is taken into account via failure probability of some stages and which is strongly involved in the NPD process must be part and parcel of the modelling approach. At this level, it was difficult to model this parameter by an interval and the repetitive use of simulation with representative sampling was the adopted procedure to address this issue. All these reasons explain why there is no need to develop a proper interval-based framework for NPD problem with uncertainty. A more accurate analysis of an interval-based optimization method as an outer loop of the discrete-event simulation model for NPD has thus not been developed.

However, even if it is particularly useful for decision criteria evaluation, such as economic and risk metrics, the use of discrete event simulation as a stand-alone technology considerably limits the number of system configurations that can be considered. This has motivated the use of a hybrid simulation-optimization strategy that not only accurately captures the dynamics of the system but also provides a structured way to search for the optimal configurations according to several objective functions in a constrained space. The use of the discrete event simulator is particularly useful for criteria evaluation and will now be embedded in an outer optimization loop as summarized in Fig. 4.

4. Multiobjective optimization problem formulation

Real engineering design problems are generally characterized by the presence of many often conflicting objectives. This raises the issue about how different objectives should be combined to yield a final solution and to search for optimal solutions to the considered problem.

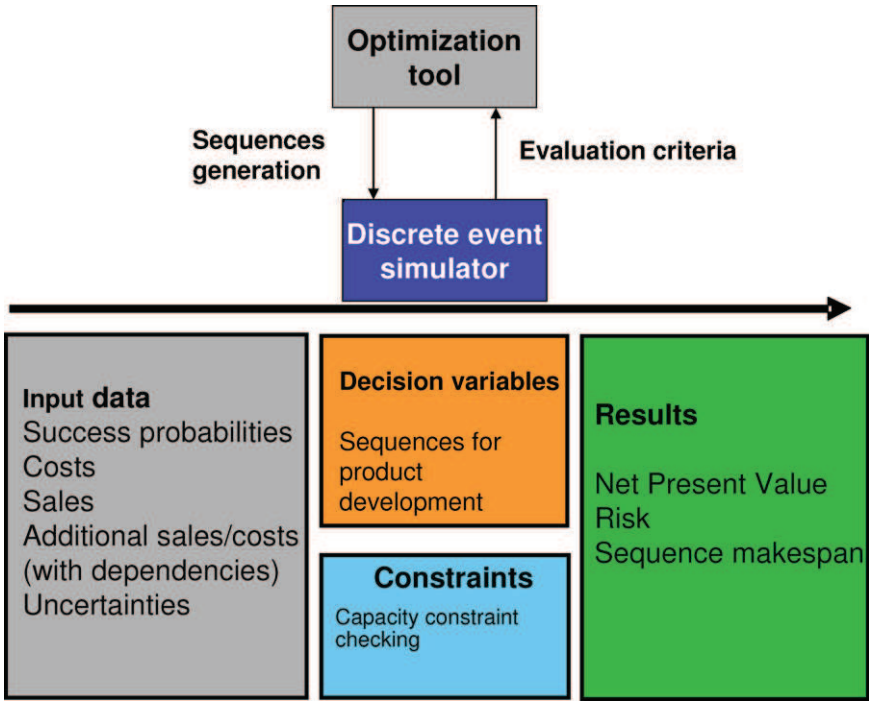

Fig. 4. SED-optimization framework.

4.1. General multiobjective optimization problem formulation

A general multiobjective design problem can be expressed as follows:

$$f(x) = (f_1(x), f_2(x), \ldots, f_k(x))^T$$
$$s.t.\ x \in S$$
$$x = (x_1, x_2, \ldots, x_n)^T$$

where $f_1(x), f_2(x), \ldots, f_k(x)$ are the k objective functions, $(x_1, x_2, \ldots, x_n)$ are the n optimization parameters and $S \subset R^q \times N^r : q + r = n$ is the solution or parameter space. The sub-space S might be defined by linear/nonlinear constraints linking both continuous and discrete variables.

One property commonly considered as necessary for any candidate solution to the multiobjective problem is that the solution is not dominated. The Pareto set consists of solutions that are not dominated by any other solutions. A solution **x** is said to dominate **y** if **x** is better or equal to **y** in all attributes, and strictly better in at least one attribute. Considering a minimization problem and two solution vectors $\mathbf{x}, \mathbf{y} \in \mathbf{S}$, **x** is said to dominate **y**, denoted $\mathbf{x} < \mathbf{y}$, if:

$$\forall i \in \{1, 2, \ldots, k\} : f_i(x) \leq f_i(y) \quad \text{and} \quad \exists j \in \{1, 2, \ldots, k\} : f_i(x) < f_i(y)$$

The space in $(R^v \times N^w : v + w = k)$ formed by the objective vectors of Pareto optimal solutions is known as the Pareto optimal frontier, **P**: any final design solution should preferably be a member of the Pareto optimal set. Pareto optimal solutions are also known efficient solutions when scalarization methods are used.

If the final solution is selected from the set of Pareto optimal solutions, there would not exist any solutions that are better in all attributes. In practice, the decision maker has to select a single solution by searching among the whole Pareto front, and it may be difficult to pick one "best" solution out of a large set of alternatives. Branke, Deb, Dierolf, and Osswald (2004) and Taboada and Coit (2006) suggest to pick the knees in the Pareto front, that is to say, solutions where a small improvement in one objective function would lead to a large deterioration in at least one of the other objectives.

4.2. General optimization methods

A great variety of applications, drawn from a wide range of investigation areas, can be formulated as complex optimization problems. This large number of optimization problems arises from models that have to enable, for industrial requirements, a truly realistic representation of the system they account for. Consequently, these models tend to show an increasing sophistication degree that derives into a higher complexity and, thus, solution difficulties. The complexity of the formulated models is basically due to the nature of the functions and of the variables involved in the optimization problem. The former ones may be not only nonlinear, but moreover, they also often prove to be nonconvex, which is a strongly penalizing characteristic in the typical minimization case. Then, for a constrained problem, determining the feasible space turns out to be a really difficult task. With regard to variable nature, most engineering problems consider both continuous and discrete variables, introducing discontinuities in the objective function and in the search space: those are called mixed-integer problems. Furthermore, the discrete variables induce an important combinatorial effect: this point is emphasized with NP-hard problems, for which no algorithm leading to polynomial solution times is known. In order to face these problems, a significant investigation effort has been carried out to develop efficient and robust optimization methods. At the beginning, this aim was pursued specially in the operational research and artificial intelligence areas. But, this trend was subsequently followed by the process system engineering community, since this one provides a wide number of applications formulated as complex optimization problems. A typical reference is constituted by design problems: heat or mass exchanger networks (Zamora & Grossmann, 1998), supply chain design (Guillen, Badell, Espuna, & Puigjaner, 2006), and multiproduct (Ravemark & Rippin, 1998) or multipurpose (Dedieu, Pibouleau, Azzaro-Pantel, & Domenech, 2003) batch plant design or retrofitting (Montagna & Vecchietti, 2003). As a consequence, a great diversity of optimization methods was implemented to meet the industrial stakes and provide competitive results. But, if they prove to be well-fitted to the particular case they pursue, the

performance of these techniques cannot be constant whatever the treated problem is. Actually, a method efficiency for a particular example is hardly predictable, and the only certainty we have is expressed by the No Free Lunch theory (Wolpert & Macready, 1997): there is no method that outdoes all the other ones for any considered problem. This feature generates a common lack of explanation concerning the use of a method for the solution of a particular example, and usually, no relevant justification for its choice is given a priori.

Optimization methods could be divided into derivative and non-derivative methods. The derivative or scalarization procedures aim at transforming the multiobjective optimization problem into a nonobjective one and solving it with classical NLP or MINLP tools. Non-derivative methods are particularly interesting for general engineering design problems. One reason is that non-derivative methods do not require any derivatives of the objective function in order to calculate the optimum. Therefore, they are also known as black box methods where numerical values of various objectives and/or constraints according to a given entrance vector x, are returned by computer codes. Another advantage of these methods is that they are more likely to find a global optimum, and not be trapped on local optima as gradient methods might do insofar as some degradations in objective functions can be admitted during the search.

For a general design problem, it is hard to express objective functions in terms of the design variables directly, which is particularly the case in our problem, since the performance functions are evaluated from a discrete event simulator. Therefore, there is no straightforward way of calculating the derivatives of the different objective functions. Another incentive to use non-derivative methods particularly Genetic Algorithms is that they are well-suited to tackle highly combinatorial problems.

4.3. Genetic algorithms

4.3.1. Genetic algorithm roadmap

Genetic algorithms (GAs) and the closely related evolutionary strategies are a class of non-gradient methods which has grown in popularity ever since Holland (1975) first published their in the early 1970s. The basic idea of GAs is the mechanics of natural selection. Each optimization parameter, (x_i), is coded into a gene as for example a real number or string of bits. The corresponding genes for all parameters, $x_1, \ldots, x_n$, form a chromosome, which describes each individual. A chromosome could be an array of real numbers, a binary string, a list of components in a database, all depending on the specific problem. Each individual represents a possible solution, and a set of individuals form a population. In a population, the fittest are selected for mating. Mating is performed by combining genes from different parents to produce children, called a crossover. Finally the children are inserted into the population where some mutations are randomly performed, and the procedure starts over again, thus representing an artificial Darwinian environment. The optimization continues until the population has converged (non evolution of statistical parameters like means, standard deviations, or domination ranks) or until a maximum number of generations predetermined has been reached.

The popularity of genetic algorithms has grown tremendously under recent years and they have been applied to a wide range of engineering problems (Altiparmak, Gen, Lin, & Karaoglan, 2009; Deb & Srinivasan, 2005; Dietz et al., 2005; Yoshikawa & Terai, 2005). There is also a large variety of genetic algorithms such as simple GA, steady state GA, GA with multiple populations, GA with crowding and sharing techniques (see Zitzler, Deb, & Thiele, 2000 for a complete set of references). The different GAs all have different features in order to solve various types of problems. There are also a number of multiobjective genetic algorithms which aim at converging the population on the Pareto optimal front instead of on just one single optimal point.

Multiobjective genetic algorithms are generally divided in non-Pareto and Pareto based approaches:

1. Non-Pareto based approaches: The first multiobjective genetic algorithm was VEGA (Vector Evaluating Genetic Algorithm) developed by Schaffer (1985). VEGA uses the selection mechanism of the GA to produce non-dominated individuals. Fourman (1985) presents a genetic algorithm using binary tournaments, randomly choosing one objective to decide each tournament. Kurasawe (1991) further developed this scheme by allowing the objective selection to be random, fixed by the user, or to evolve with the optimization process. All of these non-Pareto techniques tend to converge to a subset of the Pareto-optimal frontier, leaving a large part of the unexplored Pareto set.
2. Pareto based approaches: A non-dominated sorting to rank a search population according to Pareto optimality is introduced in Goldberg (1989). First, non-dominated individuals in the population are identified. They are given the rank 1 and are removed from the population. Then the non-dominated individuals in the reduced population are identified, given the rank 2, and then they are also removed from the population. This procedure of identifying non-dominated sets of individuals is repeated until the whole population has been ranked. Goldberg also discusses using niching methods and speciation to promote diversity so that the entire Pareto frontier is covered.

In the multiobjective GA (MOGA) presented by Fonseca and Fleming (1995, 1998) each individual is ranked according to a degree of dominance. The rankings are then scaled to score individuals in the population. In MOGA both sharing and mating restrictions are used in order to maintain population diversity.

The niched Pareto GA (NPGA) by Horn and Nafpliotis (1993) is Pareto-based but does not use ranking methods. Rather, Pareto domination binary tournaments are used to select individuals for the next generation. Zitzler and Thiele (1999) developed a multiobjective genetic algorithm called the strengthen Pareto evolutionary algorithm (SPEA)which uses two populations.

The non-dominated sorting GA (NSGA) of Srinivas and Deb (1995) implements Goldberg's concepts about the application of niching methods. In NSGA, non-dominated individuals in the population are identified, given a high initial individual score and are then removed from the population. These individuals are considered to be of the same rank. The score is then reduced using sharing techniques between individuals with the same ranking. Over the years, the main criticisms of the NSGA approach have been as follows: high computational complexity of non dominated sorting; lack of elitism; need for specifying the sharing parameter. All of these issues have been addressed in the improved version of NSGA, called NSGA-II. From the simulation results on a number of difficult test problems, it has been found that that NSGA-II outperforms two other contemporary MOEAs: Pareto-archived evolution strategy (PAES) (Connor & Tilley, 1998) and strength-Pareto EA (SPEA) (Goldberg, 1989) in terms of finding a diverse set of solutions and in converging near the true Pareto-optimal set. The way constraints are treated is briefly recalled in what follows.

4.3.2. Constraint handling

Constrained multiobjective optimization is the most common kind of problem in engineering applications. In general, three kinds

of constraints are considered: simple inequality ($\leq$), strict inequality ($<$), and equality:

$$\left.\begin{array}{l} g(x) \leq c1 \\ r(x) < c2 \\ h(x) = c3 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} constr1(x) = c1 - g(x) \geq 0 \\ constr2(x) = c2 - r(x) > 0 \\ constr3(x) = c3 - h(x) = 0 \end{array}\right.$$

where (g, r, h) are real-valued functions of a decision variable $x = (x_1, \ldots, x_n)$ on an n-dimension decisional search space U, and (c_1, c_2, c_3) are constant values. In the more general case, these constraints are written as vectors of the type:

$$\begin{array}{l} \overline{constr1}(x) = (constr1(x)_1, \ldots, constr1(x)_{n1}) \geq 0 \\ \overline{constr2}(x) = (constr2(x)_1, \ldots, constr2(x)_{n2}) > 0 \\ \overline{constr3}(x) = (constr3(x)_1, \ldots, constr3(x)_{n3}) = 0 \end{array}$$

where n_1, n_2, and n_3 are respectively, the number or inequality, strict inequality and equality constraints. This constraint formulation implies that each constraint value will be negative if and only if this constraint is violated. In practice, due to round-off error on real numbers, the equality constraint *constr3* is replaced by $\overline{contr3}(x) + \overline{\epsilon}$. In this expression, $\overline{\epsilon}$ is called a "precision vector" of the equality vector, and takes low values (less than 10^{-6} for example). This approximation is not necessary when equality constraint involves only integer or binary variables.

The constraint satisfaction implies the maximization of violated constraints in vectors *constr1*, *constr2*, and *constr3*. According to Fonseca and Fleming (1998), the satisfaction of a number of violated inequality constraints is a multiobjective maximization problem. A more simple solution consists in comparing the sum of values of violated constraints only, as in NSGA II algorithm of Deb, Pratap, Agarwal, and Meyarivan (2002), which implies there are no priority rules between constraints. This step is performed first, before the second one, which concerns the comparison of the objective function vectors. On four problems chosen from the literature (Deb et al., 2002), NSGA-II has been compared with another recently suggested constraint-handling strategy and proved to be more efficient. These results lead us to apply NSGA-II to the NPD problem.

4.4. Combinatorial aspects of the NPD problem and search space definitions

As above mentioned, evolutionary procedures, and particularly GAs, are well-suited for handling highly combinatorial problems. One of the objectives of the NPD optimization being the determination of the best sequence of products, this item introduces a very high combinatorial aspect in the problem. For example, as it is shown below, for a simple problem involving three diseases, two drugs for disease I, two for disease II and one for disease III, it exists 240 possible sequences, and this number grows up to 951,744 for the problem under consideration.

Given a problem involving M_D diseases. For each disease $di_i (i = 1, M_D)$, n_{d_i} therapeutic axis involving n_{d_i} drugs can be considered. A sequence is thus constituted by the union of sub-sequences of drugs, each devoted to a disease. A drug related to a disease $di_i (i = 1, M_D)$ is denoted with $p = 1, \ldots, n_{d_i}$. An integer value p is allocated to each drug in a partial sequence ranging from 1 to n_{d_i}. The drugs can be arranged as follows:

$$\underbrace{[1, \ldots, n_{d_1}]}_{M_1}\underbrace{[1, \ldots, n_{d_2}]}_{M_2}\underbrace{[1, \ldots, n_{d_{M_D}}]}_{M_D}$$

Let us consider the set S of all the possible sequences in which the number of products can vary between M_D (at least one drug per disease) and $n_{d_1} + n_{d_2} + n_{d_{M_D}}$ and in which all the permutations can be considered:

$$N_{TOT} = n_{d_1} + n_{d_2} + \cdots + n_{d_{M_D}}$$

$$Card(S) = N_{TOT}! + \sum_{p=1}^{N_{TOT}-M_D} (N_{TOT} - p)! \times \left(\sum_{i,j,\ldots,k} C^i_{n_{d_i}} C^j_{n_{d_j}} \ldots C^k_{n_{d_k}} \right),$$
$$i + j + \cdots + k = N_{TOT} - p; \quad i \leq n_{d_1}, j \leq n_{d_2}, \ldots, k \leq n_{d_{M_D}}$$

This can be applied to the example which serves as a test bench involving three diseases (four drugs for d_1; four drugs for d_2; one drug for d_3).

$$Card(S) = (9)! + (8)!(2C^3_4 C^4_4 C^1_1) + (7)!(2C^4_4 C^2_4 C^1_1 + 2C^3_4 C^3_4 C^1_1)$$
$$+ (6)!(2C^3_4 C^2_4 C^1_1) + (5)!(2C^2_4 C^2_4 C^1_1) + (4)!(2C^1_4 C^2_4 C^1_1) + (3)!(2C^1_4 C^1_4 C^1_1)$$

The total number of possibilities for this example is 951,744. This means that 951,744 possible portfolio drugs can be considered, taking into account that portfolios with less than 3 drugs are not possible due to the constraints defined for the model, at least one drug per disease.

5. Implementation of the NSGA II key procedures for NPD modelling

The methodology used for solving the NPD problems involves a two-step approach: at the lower level, the previously developed discrete event simulator is used to evaluate the product development sequences, according to different criteria: Net Present Value, risk metrics and makespan; at the upper level, a multiobjective procedure based on NSGAII principles is used to determine both the number of drug products in the sequence and the order in which the drugs are released in the pipeline.

5.1. Coding, crossover and mutation

A sequence is modelled by use of two types of chromosomes with an identical number of genes, equal to the number of products to consider in the global portfolio. To each product P_1 corresponds an index i which is the chromosome position i. The first chromosome $Chrom_1$ is related to the product order in a sequence. Genes are integer variables, ranging from 1 to the total number of products in a sequence. The value of each gene may occur only once in the chromosome. For a position i of a gene, its value corresponds to the product position i in the sequence. The second chromosome $Chrom_2$ is only constituted by binary variables, the unity value of a gene in position i (respectively 0) corresponding to the presence (respectively absence) of a product.

The chromosome corresponding to each sequence is then obtained by multiplying each gene of $Chrom_1$ with the corresponding one of $Chrom_2$, locus by locus. It must be highlighted that this coding is not unique which may introduce some bias in the search. Yet, a more attractive alternative would be to directly code the chromosomes representing a sequence with variable length in function of the product number in the sequence. Yet, this approach may lead to unfeasible individuals in the crossover phase, with a larger size than the one corresponding to the effective number of products in the sequence. The efficiency of the former procedure has been tested successfully through several examples and has thus been selected in this work. Fig. 5 illustrates the used coding representing a solution that is then evaluated by the simulator for 9 products. Crossover and mutation have been carried out by specific procedures for each type of chromosome.

Chromosomes dedicated to product order are haploid, yet, all the integer genes must be different, ranging from 1 to the total number of products in a sequence. For this purpose, a crossover operator

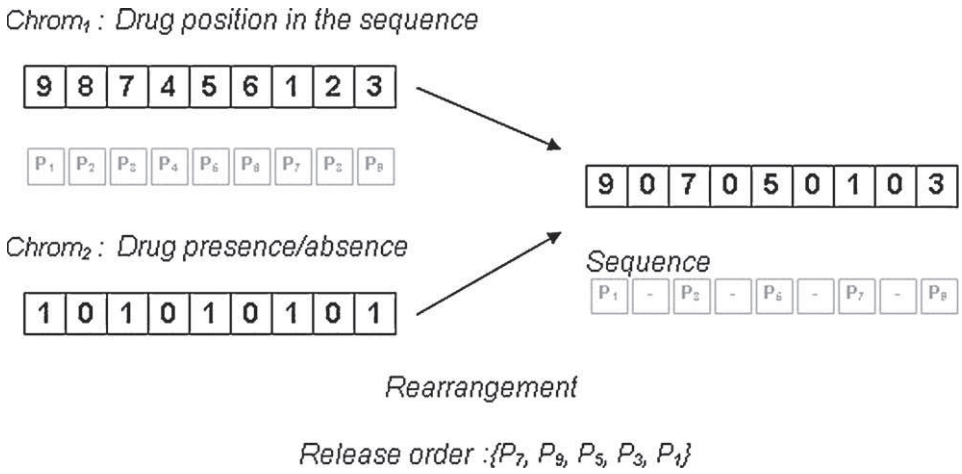

Fig. 5. Coding for generating a sequence.

with respect to genotype constraints without clone generation in the offspring genetic code has been carried out, the so-call MPX operator (Maximal Preservative X) (Andersson, 1999). This underlying idea is to insert a segment of a parent chromosome in the chromosome of the other parent so that the resulting crossover is closer to his parents. It is a two-point crossover and the two sons are obtained in a symmetrical manner. Concerning mutation, a classical mutation operator that randomly permutes two genes of a chromosome is used. This operator is applied to the individuals derived from crossover with an adapted rate (preferably 0.5). Then, the new offspring is placed in a new population.

5.2. Optimization parameters

The optimization parameters are presented in Table 3.

5.3. Optimization criteria

The optimization criteria are evaluated by use of the previously developed discrete event simulator. They involve the global Net Present Value of a sequence, classically computed from the average value of net present values of the samples. An actualization rate of 15% has been chosen:

$$f_1 : \frac{\sum_{j=1}^{n}\left[\sum_{i=1}^{W} NPV\right]_j}{n}$$

Another criterion is the risk, corresponding to the number of times a negative value of NPV is observed among the total number of samples. Note that risk f_2 is the complementary risk relative to the already mentioned positivity probability:

$$f_2 : \frac{\sum_{j=1}^{n}\left[\sum_{i=1}^{W} NPV\right]_j < 0}{n}$$

Table 3
Optimization parameters.

Optimization parameters	
Number of generations	200
Number of individuals per generation	80
Number of simulations per individual	300

Then, finally, the makespan of a sequence is computed from the average makespan of the samples:

$$f_3 : \frac{\sum_{j=1}^{n}(dur)_j}{n}$$

where $W \rightarrow$ number of drugs in the sequence; $n \rightarrow$ number of runs by sequence; $dur \rightarrow$ makespan for a sequence.

5.4. Constraints

It must be emphasized that the resource constraints have already been taken into account in the capacity requirements of each task in the pipeline within the discrete event simulator. The constraints that are considered here are related to the presence of at least one drug targeting a therapeutic axis. They can be formulated as follows:

$$\sum m_{d_i}^{p} \geq 1 \;\; with \;\; p = 1, \ldots, n_{d_i}$$

Let us recall that for the example, 3 diseases and 9 molecules have been considered: Disease 1: Products P2, P3, P6, P7; Disease 2: Products P4, P5, P8, P9; Disease M3: Product P1. These constraints involve at least one gene value equal to 1 for chromosome $Chrom_2$ in the loci corresponding to the genes of the products of the given disease g_i, that is:

$$g_2 + g_3 + g_6 + g_7 \geq 1$$
$$g_4 + g_5 + g_8 + g_9 \geq 1$$
$$g_1 > 1$$

6. Result presentation and discussion

The case study results are now discussed in the following sections, focusing on analyzing the Pareto front generated and identifying trends concerning portfolio composition. In all the optimization runs, unless explicitly mentioned, the initial population was generated randomly.

6.1. Introduction

Optimization runs were first carried out in a bicriteria way and then analyzed from a tricriteria viewpoint. To take into account the stochastic nature of the Genetic Algorithm, each optimization run is repeated 5 times (at least). The CPU time of each optimization run is

Table 4
Net Present Value and Risk for 4-drug sequences (9 drug-portfolio optimization).

Solutions	NPV (M$)	Risk (%)	Makespan by simulation (days)	Release order			
1,3,4	1546–1730	14–20	3721	P2	P7	P5	P1
2	1683	17	4055	P7	P5	P2	P1
5	1507	13	3604	P5	P7	P2	P1
6	1472	12	3721	P5	P2	P7	P1

difficult to evaluate, due to combined effects: first, it depends on the number of products in the sequence, second, the stochastic aspect of the Monte-Carlo approach used through simulation may lead to premature stop of the evaluation of a candidate. An optimization run takes around 36 h for this study with 9 drugs for 3 diseases.

6.2. Bicriteria optimization Net Present Value-Risk

6.2.1. Optimization NPV-Risk with random initialization

A first study concerns the Net Present Value and Risk to optimize simultaneously. As might be expected, random initialization of the population has resulted in a dispersion of the various solutions. The optimization procedure was considered to be converged when general progression of the Pareto front was insignificant. It can be observed that the risk variation lies between 10 and 40%. No solution exists for risk values greater than 40%. An interesting result concerns the number of drugs in the portfolio. For risk values comprised between 12 and 20%, the number of drugs in the sequences is equal to 4 and the drugs that can be systematically found are $[P_1 \wedge P_2 \wedge P_5 \wedge P_7]$ with an NPV value comprised between 1400 and 1700 M$. For risk values between 27% and 39% the number of drugs in the portfolio is equal to 6 and the products that can be systematically observed are $[P_1 \wedge P_2 \wedge P_3 \wedge P_5 \wedge P_6 \wedge P_7]$ with an NPV value comprised between 1800 and 2000 M$. The results are also presented in Tables 4 (4 drugs) and 5 (6 drugs). As it can be seen in Fig. 6, a first comment concerns the Pareto front solution. For all of them, the higher the risk, the higher the Net Present Value. Second, it can be highlighted that several solutions (solutions 1, 3, 4) are found several times $(P_2 \wedge P_7 \wedge P_5 \wedge P_1)$ and $(P_2 \wedge P_7 \wedge P_5 \wedge P_1 \wedge P_1)$. Furthermore in the coding of the GA, a sequence is not represented by a unique chromosome. Although some significant differences are observed above all for the risk criterion, these solutions can be considered as particularly attractive since the procedure has identified them several times as Pareto candidates.

The union of the Pareto fronts obtained from the optimization runs, can be visualized in Fig. 6 due to both the stochastic nature of the NPD model (a sequence is evaluated 300 times) and to the GA. This figure displays the non-dominated individuals obtained from 5 optimizations and are relative to sequences of 4 drugs (P1, P2, P5, P7) and (P1, P3, P5, P9).

For these two distinct behaviours, the solutions can be distinguished by the product order in the sequence. A closer look at these solutions indicate that the optimization strategy tends to eliminate long sequences, reproducing the so-called attrition phenomenon occurring in NPD problems. This is due to the complexity that is inherent in the model of the pharmaceutical drug development pathway and of the interactions between the various drugs. It is difficult to predict this behaviour without the use of a numerical tool. For the 12 sequences evaluated by simulation, the same behaviour as in the present section for sequences with four and six drugs, was already observed. The NPV varied from 80 to 685 M$ and the risk was in the range (0.77–0.38). However, the efficiency of the genetic algorithm improves considerably the performances. Another important observation is that strategies with differences in either drug selection, timing, can compete with similar reward versus risk profiles. Hence it is useful for the decision maker to identify and closely examine the different options that can yield the desired return and acceptable risk.

6.2.2. Optimization with 9-drug sequences in the first generation

An optimization run was then performed in order to study the influence of the number of products in the first generation. The idea is to initialize the AG with sequences containing exactly 9 drugs, in order to examine how the number of products evolves naturally along the generations. The results are presented in Table 6 for generations 1, 40, 80, 120, 160 and 200. It can be clearly observed that the sequences with a low number of products are favoured in the optimization process.

Fig. 7 displays the Pareto front in which sequences with 4, 6 and 7 products can be found. It must be emphasized that 7-drug sequences were not found in the previous optimization runs: this may be due to the fact that the number of generations needs to be increased. Table 7 presents the numerical values of NPV and risk as well as the release order. Finally, the main result here is that the natural evolution of the algorithm is towards the elimination of long sequences.

Table 5
Net Present Value and Risk for 6-drug sequences (9 drug-portfolio optimization).

Solutions	NPV (M$)	Risk (%)	Makespan by simulation (days)	Release order					
1	2206	39	5155	P6	P1	P3	P2	P7	P5
2	2203	37	4878	P6	P2	P1	P7	P5	P3
3	2097	36	4914	P2	P6	P1	P7	P5	P3
5	1900	28	4928	P7	P6	P3	P2	P5	P1
4,6	1887–2077	27–29	4869	P3	P6	P7	P2	P5	P1

Table 6
Evolution of the number of drugs by sequence.

Generation	Number of drugs in the sequence								
	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	80
40	0	0	0	0	0	11	69	0	0
80	0	0	0	0	0	30	50	0	0
120	0	0	0	0	0	45	35	0	0
160	0	0	0	24	0	31	25	0	0
200	0	0	0	21	0	35	24	0	0

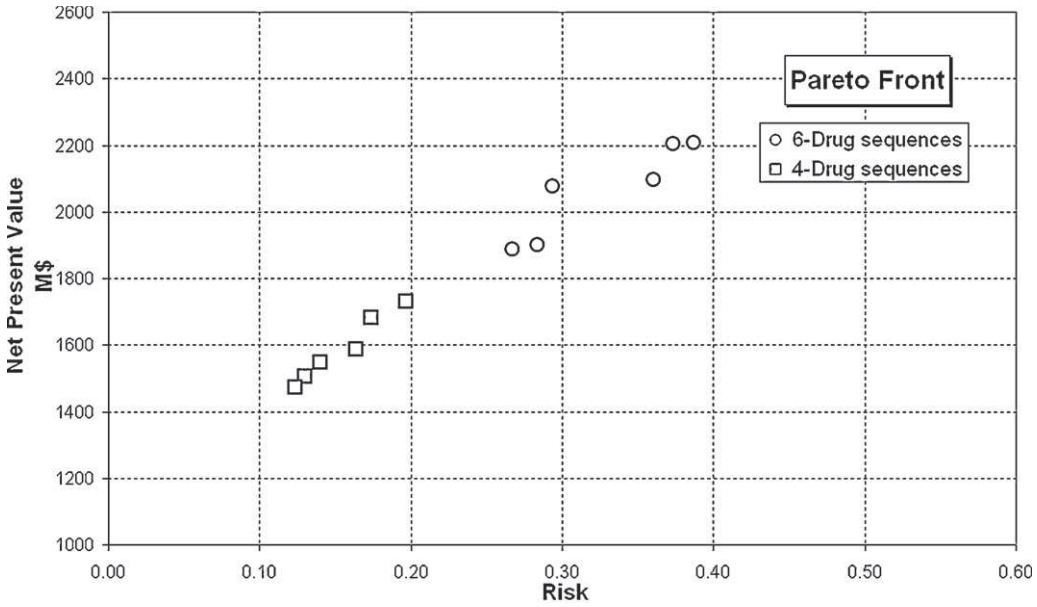

Fig. 6. Pareto Front. Net Present Value/Risk.

Table 7
Results of Net Present Value and Risk for 9-drug initialized portfolio (Pareto front solutions).

Solutions	NPV(M$)	Risk (%)	Release order						
1	798	22	P7	P1	P3	P5			
2	783	21	P7	P3	P5	P1			
3	690	18	P7	P1	P5	P3			
4	782	19.5	P7	P3	P5	P1			
5	700	19	P7	P3	P5	P1			
6	992	23	P7	P1	P3	P5			
7	1857	30	P7	P3	P6	P5	P2	P1	
8	1592	26.5	P7	P3	P6	P5	P2	P1	
9	1722	28.5	P1	P7	P6	P5	P2	P3	
10	1696	28	P7	P1	P6	P5	P2	P3	
11	1355	24.5	P6	P7	P1	P5	P9	P2	P3
12	1379	26	P7	P1	P6	P5	P9	P2	P3
13	1343	24	P7	P1	P6	P5	P9	P2	P3

6.2.3. Optimization with 9-drug sequences in the optimized portfolio

To confirm once more that long sequences are not interesting, an optimization is performed under the constraint of a 9-drug portfolio along the algorithm evolution. Here, the first population was generated randomly (without taking into account the constraint).

This is confirmed by the Pareto front (Fig. 8) only constituted of 3 sequences with an NPV between 1171 and 1313 M$ and a risk value between 35 and 44%. The numerical values of NPV and Risk for each non-dominated individual as well as the release order are presented in Table 8. As a conclusion, the results indicate that long sequences are not representative of attractive values both for NPV and risk. This explains why 9-drug individuals are eliminated from the Pareto front in the previous bicriteria optimization.

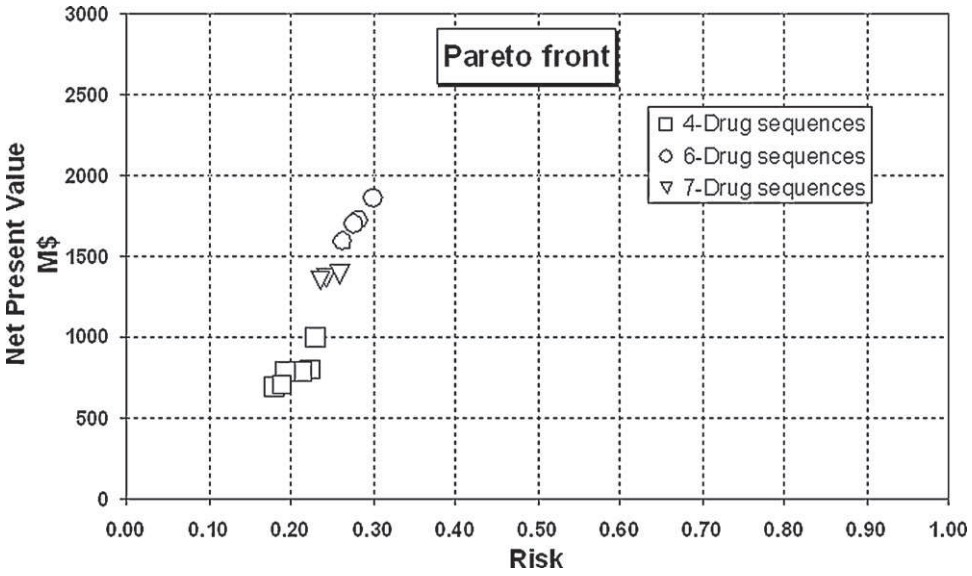

Fig. 7. Pareto front for the optimization Net Present Value/Risk for 9-drug initialized portfolio.

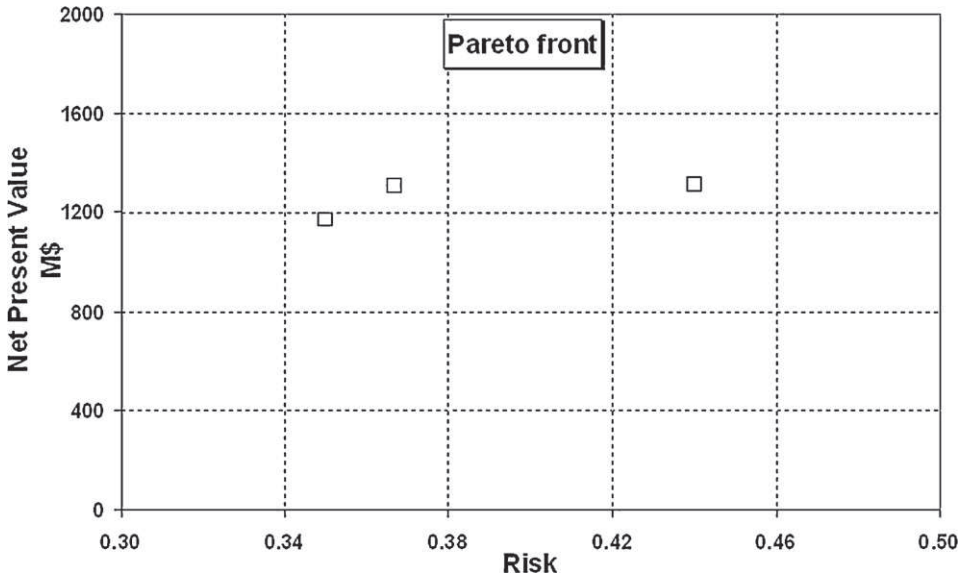

Fig. 8. Pareto front for the optimization Net Present Value/Risk for 9-drug optimized portfolio.

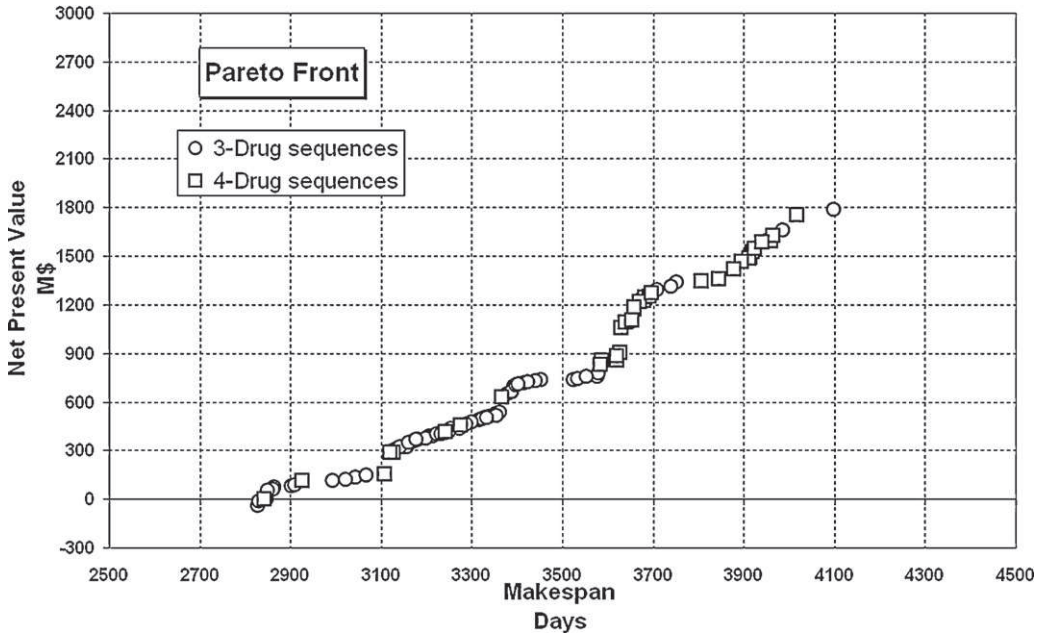

Fig. 9. Pareto Front. Net Present Value/Makespan.

6.3. Bicriteria optimization Net Present Value-Makespan

The second optimization study is based on Net Present Value-Makespan (expressed in days). The results relative to this pair of criteria are presented in Fig. 9. Solutions with negative values for net present values are found, which correspond to very low values of the time horizon, that obviously will not be considered by the decision maker. It can be highlighted that an important number of sequences with 3 or 4 products are found again. Among the 3-product sequence, the corresponding drugs are $[P1 \wedge (P2 \vee P6) \wedge (P5 \vee P8)]$. The 4-product portfolio involves the drugs $[P1 \wedge P2 \wedge P5 \wedge P7]$, that have been already identified as potential candidates for net present value-risk optimization (see Tables 9 and 10).

Table 8
Results of Net Present Value and Risk for 9-drug optimized portfolio (Pareto front solutions).

Solutions	NPV (M$)	Risk (%)	Release order								
1	1314	44	P9	P2	P3	P6	P4	P7	P5	P1	P8
2	1305	37	P9	P2	P3	P6	P4	P7	P5	P1	P8
3	1172	35	P9	P2	P3	P6	P4	P7	P5	P1	P8

6.4. Bicriteria optimization makespan-risk

The results of the bicriteria optimization makespan-risk are presented in Fig. 10. From these results, it can be seen that a decrease in risk has a strong impact in the pipeline duration, that can be quantified. Risk ranges from 13% to around 70% when the duration decreases from 11.5 to 7.7 years. The

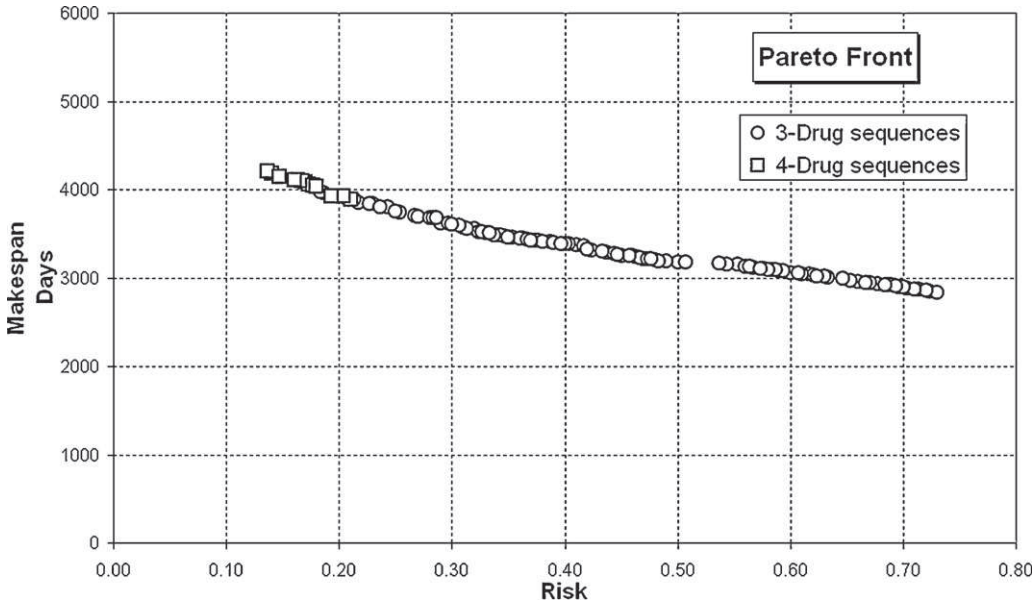

Fig. 10. Pareto Front Makespan/Risk.

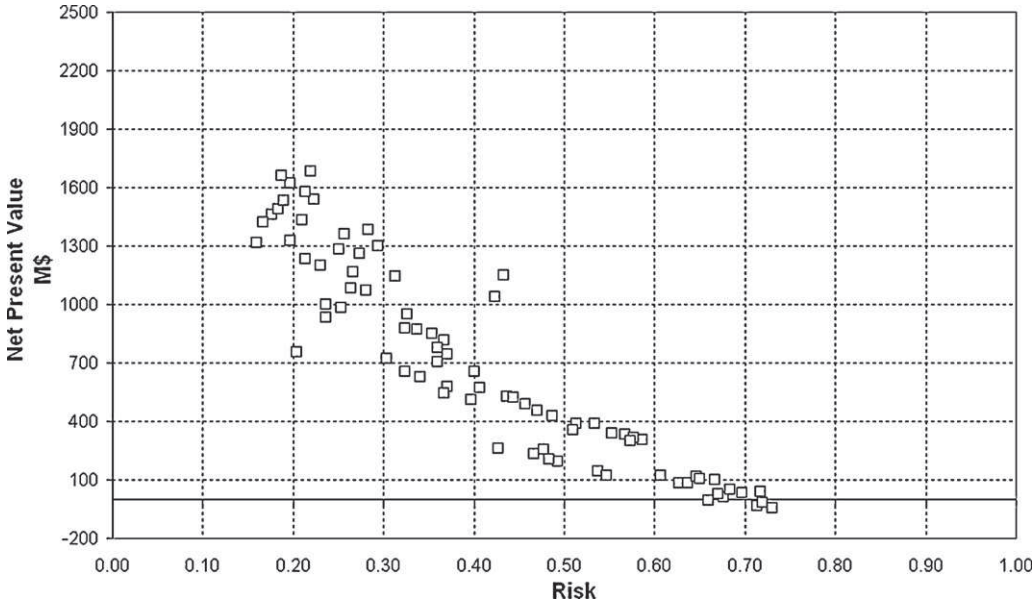

Fig. 11. Tricriteria solutions projection for *NPV Risk*.

sequences for which risk lies between 13 and 18% are exclusively composed of 3 products $[P1 \wedge (P2 \vee P7) \wedge (P5 \vee P8)]$. The sequences for which risk is comprised between 21 and 73% involve 4 products $[P1 \wedge P2 \wedge P5 \wedge P7]$. There is a small overlapping zone from 18 to 21% with mixed sequences of 3 and 4 products (the same ones as those previously found). Once more, the higher number of solutions presented relative to the combinatorial aspect of the problem is due to slight variations for clones, due to the stochastic aspects of the problems (see Tables 11 and 12).

Table 9
Net Present Value and makespan for 3-drug sequences.

Solutions	NPV (M$)	Makespan (days)	Risk by simulation	Release order		
1	1312	3739	54	P1	P5	P2
2,3,5	57–713	2850–3402	33	P5	P2	P1
4	136	3043	50	P8	P2	P1
6	11	2848	69	P1	P8	P2
6	2	2846	55	P5	P1	P2
7	-35	2827	73	P2	P1	P8

Table 10
Net Present Value and makespan for 4-drug sequences.

Solutions	NPV (M$)	Makespan (days)	Risk by simulation	Release order			
1	1752	4017	39	P5	P7	P2	P1
2	1357	3845	22	P2	P5	P7	P1
3,7	413–1253	3241–3691	39	P5	P1	P2	P7
4,5	856–1423	3587–3878	47	P7	P5	P2	P1
6	152	3107	45	P2	P1	P5	P7

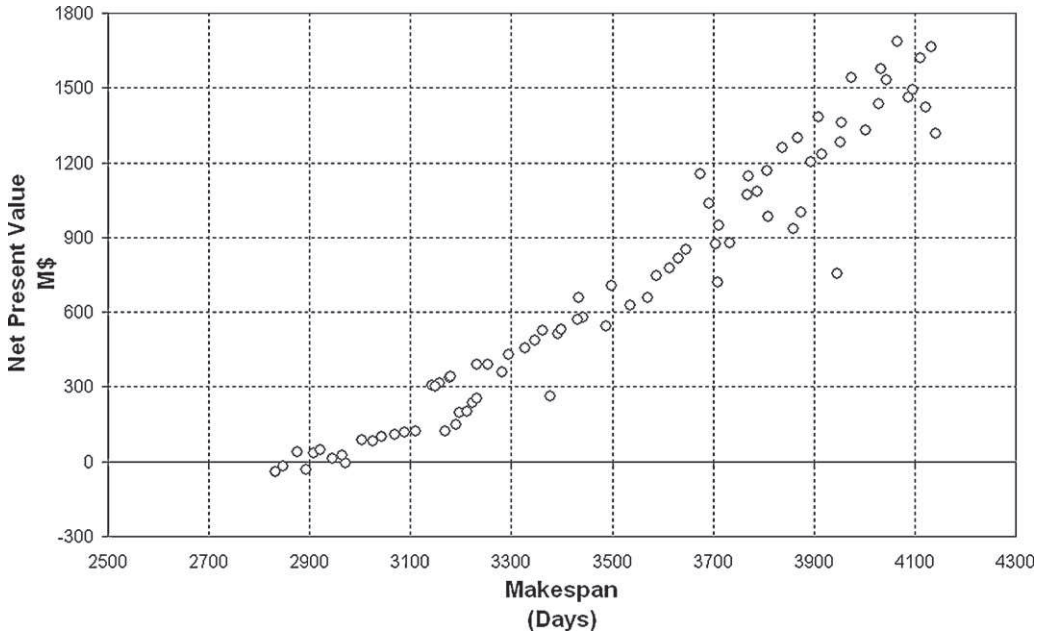

Fig. 12. Tricriteria solutions projection for *NPV Duration*.

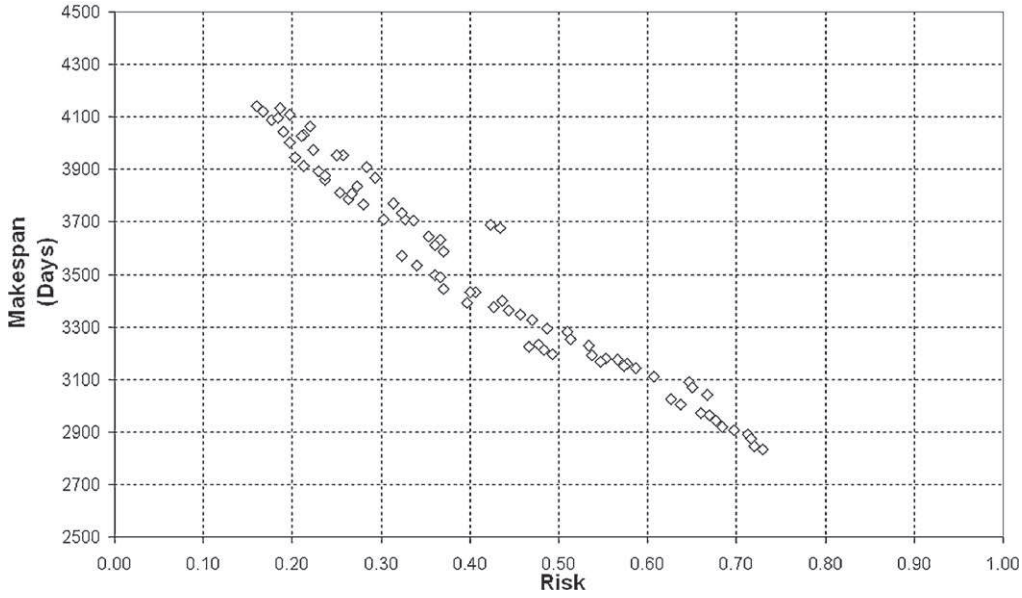

Fig. 13. Tricriteria solutions projection for *Risk Duration*.

As a partial conclusion of this bicriteria study involving time as a criterion, it can be said that decisions on timing are an important constituent of the portfolio development strategy as they are used to favourably organize cash flows. This is particularly important when having to consider the probability that a project will succeed, i.e. the financial impact of failed projects.

Table 11
Risk and duration for 3-drug sequences.

Solutions	Risk (%)	Makespan (days)	NPV by simulation	Release order		
1	40	3389	628	P7	P5	P1
2	35	3458	603	P7	P1	P5
3	30	3608	461	P1	P5	P2
4	25	3757	532	P2	P1	P5
5	21	3906	554	P5	P2	P1
6	18	3973	474	P1	P2	P5

7. Tricriteria optimization *NPV-Duration-Risk*

7.1. Study presentation

In this study, the tricriteria optimization is performed NPV-Duration-Risk. To make the interpretation easier, the results

Table 12
Risk and duration for 4-drug sequences.

Solutions	Risk (%)	Makespan (days)	NPV by simulation	Release order			
1,2,3	17–21	3889–4090	1375	P2	P5	P7	P1
4,5	16–17	4101–4105	1379	P7	P5	P2	P1
6,7,10,12,13	14–19	4038–4204	769	P5	P7	P2	P1
8	16–20	3922–4110	953	P5	P2	P7	P1
11	15	4150.48	584	P7	P2	P5	P1

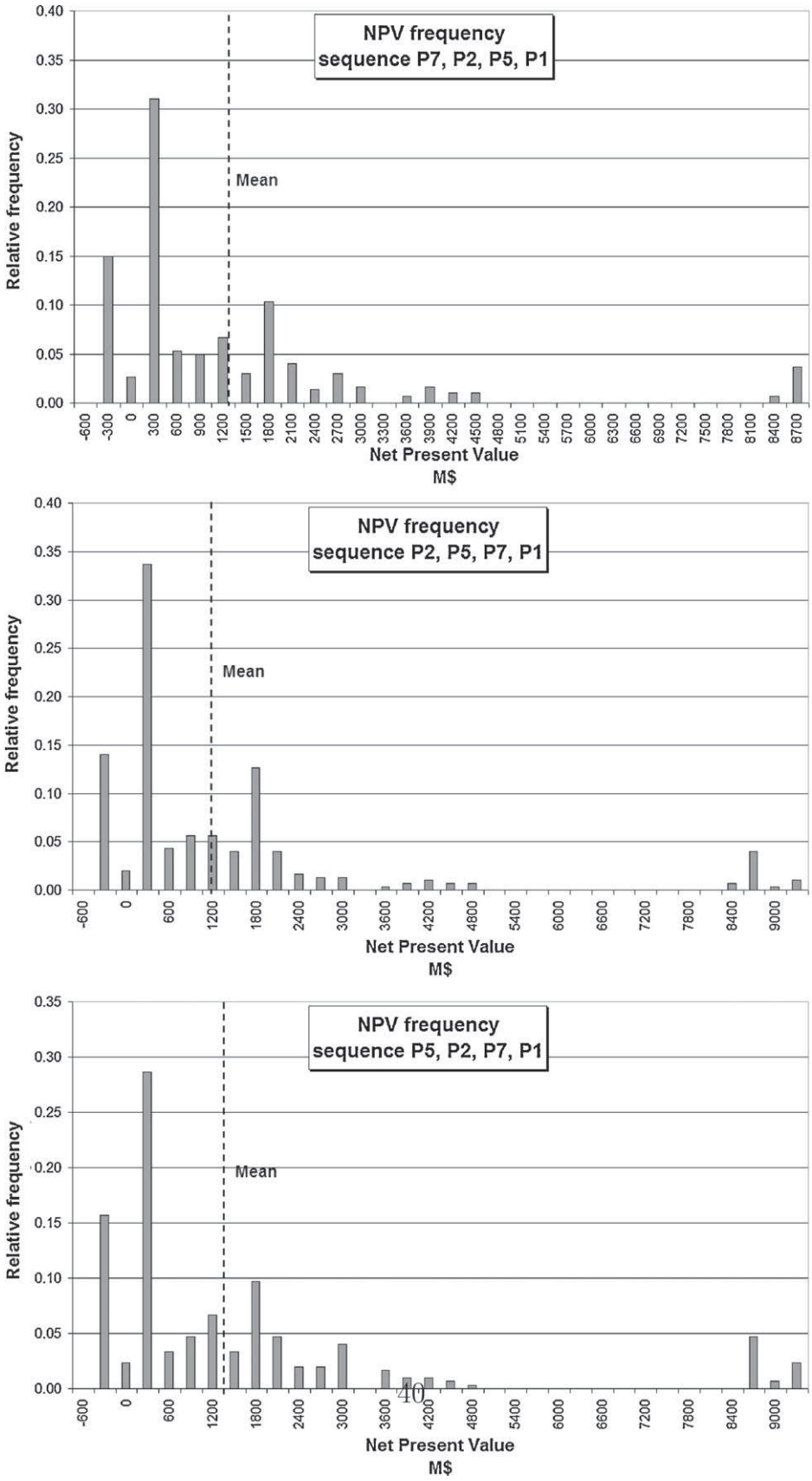

Fig. 14. Frequency and behaviour for non-dominated tricriteria optimization.

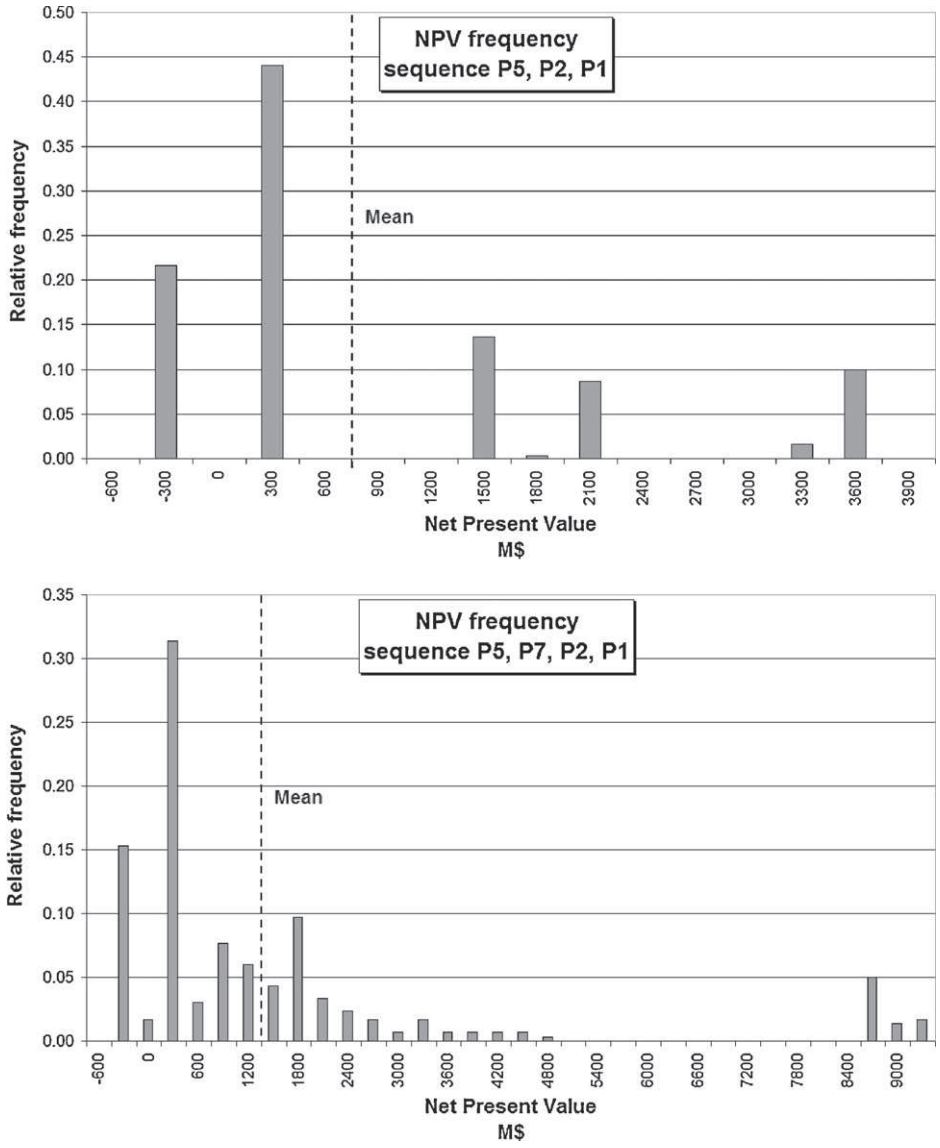

Fig. 15. Frequency and behaviour for non-dominated tricriteria optimization.

relative to a given pair of criteria are presented as a projection on a 2D-axis (see Figs. 11–13). Globally, it can be said that the same trends as the bicriteria approaches are observed, that is a small number of products in the portfolio favours the best compromise between the criteria. A closer examination at the evolution of NPV vs. risk needs some additional comments. Fig. 11 shows that for risk values corresponding to a 15–35% range, it seems that an increase in risk leads to an increase in NPV. This trend is no more observed when exploring risk values between 35 and 70%, where the higher the risk, the lower the NPV. This phenomenon can be now attributed to a strong decrease in makespan which is optimized simultaneously in this case. Since the bicriteria study has shown that the solution set is different according to the pair of criteria considered, the tricriteria analysis seems more consistent to find the most interesting solutions.

From the solutions obtained from the 3-criteria Pareto front, the decision maker can select a sequence, from a risk level that seems acceptable for him. The results presented in Table 13 all constitute potential candidates.

A closer examination of the solutions presented in Table 13 is presented in Figs. 14 and 15 where relative frequency is plotted vs. net present value. This kind of representation is more meaningful since the analysis of simulation results of some sequences has shown that they exhibit a bimodal behaviour. The mean net present value is interesting from an optimization viewpoint since the objective is to shift towards positive values for NPV: the interpretation is consistent here since mean NPV is combined with a risk criterion as measured by the ratio of the number of positive values for NPV to the total number of NPV evaluations. This two-peaked phenomenon is still observed for these 5 sequences, with more dispersed values for the 3-drug portfolio.

Solutions 1 and 2 are equivalent from the makespan criteria, but differ significantly from the risk and NPV criterion. Solution

Table 13
Some interesting solutions from tricriteria optimization.

Solutions	Risk (%)	Net Present Value (M$)	Duration (days)	Release order			
1	16	1316.29	4141.52	P7	P2	P5	P1
2	20	1620.7	4110.4	P2	P5	P7	P1
3	25	1280.99	3952.73	P5	P2	P7	P1
4	30	721.11	3709.74	P5	P2	P1	
5	35	850.35	3645.91	P5	P7	P2	P1

2 exhibits some peaks higher than 9000 M$. The decision maker has to decide if the higher risk induced by Solution 2 is justified. Solution 3 must be investigated if the time criterion is important to consider at that level, even if risk and NPV are lower than for solutions 1 and 2. Solutions 4 and 5 show the same order of magnitude for mean NPV and durations (risk is higher for solution 4). Yet, Solution 3 concerning the 3-drug sequence has poor performances (no peak higher than 3600 M$) and may be finally discarded by the decision maker. At this level of discussion, it is difficult to say more since the example has just a didactic value.

7.2. Conclusion of the bi- and tricriteria study

As a conclusion of this bicriteria analysis performed on the different pairs of criteria and of the tricriteria optimization, it can be highlighted that among the constellation of potential candidates, the optimization strategy seems efficient to detect the sequences which can be considered by the decision makers. Only a few sequences are detected. Among theses sequences, large portfolios cause resource queues and delays time to launch and are eliminated by the bicriteria optimization strategy. Small portfolio reduces queuing and time to launch appear as good candidates. The optimization strategy, based on NSGAII, that is particularly elitist, is interesting to detect the sequence candidates. Time is an important criterion to consider simultaneously with NPV and risk criteria. The order in which drugs are released in the pipeline is of great importance as with scheduling problems. The use of a decision analysis method (Pirdashti, Ghadi, Mohammadi, & Shojatalab, 2009) as TOPSIS will allow to select a sequences according to the decision maker preferences. The basic concept of this method is that the selected alternative should have the shortest distance from the negative ideal solution in geometrical sense (Pirdashti et al., 2009). A thorough analysis is proposed in Morales-Mendoza. et al. (2011).

8. Conclusions

The development of a multiobjective Genetic Algorithm optimization framework coupled with a discrete event simulator has been presented that addresses two key decisions simultaneously: portfolio management and scheduling of drug development and manufacturing. Two case studies were used to illustrate the capabilities of the framework and also highlighted that the scope of decisions that a drug developer may be confronted with can be vast and complex.

Our analysis on both case studies suggests that optimizing project priorities taking into account resource allocations yields a significantly improved portfolio performance, rather than a simple use of a bubble chart that cannot take into account the interdependencies between projects. Due to the complexity of this problem, a contribution of this work is in demonstrating a formulation based on techniques from evolutionary computation employed for an efficient search of the decision space and of the objective space. All the results tend to highlight that pharmaceutical product development strategies in the real world may be better analyzed when considering the impact of decisions holistically rather than only individually.

One important issue in this kind of problems is the decision flexibility that management can exercise during the course of the development path. Given the high failure rate and large potential investment loss, a perspective of this work could be to take into account for the valuation of product portfolios not only the uncertainties and risks but also the decision flexibility. The sequence of projects, as proposed in this work, is fixed during the whole horizon. It would be interesting to consider the possibility of a dynamic evolution of the sequence along the simulation path embedded in the optimization loop.

References

Aguilar-Lasserre, A. A., Azzaro-Pantel, C., Pibouleau, L., & Domenech, S. (2007). Enhanced genetic algorithm-based fuzzy multiobjective strategy to multiproduct batch plant design. *Analysis and Design of Intelligent Systems using Soft Computing Techniques*, *41*, 590–599.

Altiparmak, F., Gen, M., Lin, L., & Karaoglan, I. (2009). A steady-state genetic algorithm for multi-product supply chain network design. *Computer and Industrial Engineering*, *56*, 521–537.

Andersson, J. (1999). *On engineering system design – A simulation optimization approach* (Master's thesis Licentiate thesis). Department of Mechanical Engineering, Linkping University.

Baudet, P., Azzaro-Pantel, C., Domenech, S., & Pibouleau, L. (1999). Un couplage entre un algortihme gntique et un modle de simulation pour l'ordonnacement court terme d'un atelier discotinu de chimie fine. *RAIRO Operations Research*, *33*, 299–338.

Biegler, L., & Grossmann, I. E. (2004). Retrospective on optimization. *Computers and Chemical Engineering*, *28*, 1169–1192.

Blau, G. E., Pekny, J. F., Varma, V. A., & Bunch, P. R. (2004). Managing a portfolio of interdependent new product candidates in the pharmaceutical industry. *Journal of Product Innovation Management*, *21*, 227–245.

Branke, J., Deb, K., Dierolf, H., & Osswald, M. (2004). *Parallel problem solving from nature. Chapter viii: Finding knees in multi-objective optimization*. Berlin/Heidelberg: Springer. (pp. 722–731).

Bérard, F., Azzaro-Pantel, C., Pibouleau, L., Domenech, S., Navarre, D., & Pantel, M. (1999). Towards and incremental development of discrete-event simulators for batch plants: Use of object-oriented concepts. *Computers and Chemical Engineering*, *23*, 565–568.

Buyukozkan, G., & Feyzioglu, O. (2004). A fuzzy-logic-based decision-making approach for new product development. *International Journal of Production Economics*, *90*, 27–45.

Calantone, R., Benedetto, C. D., & Schmidt, J. (1999). Using the analytic hierarchy process in new product screening. *Journal of Product Innovation Management*, *16*, 65–76.

Chapman, C., & Ward, S. (2002). *Managing project risk and uncertainty: A constructively simple approach to decision making*. Wiley.

Coldrick, S., Longhurst, P., Ivey, P., & Hannis, J. (2005). An R&D options selection model for investment decisions. *Technovation*, *25*, 185–193.

Colvin, M., & Maravelias, C. (2009). Scheduling of testing tasks and resource planning in new product development using stochastic programming. *Computers and Chemical Engineering*, *33*(5), 964–976.

Colvin, M., & Maravelias, C. T. (2008). A stochastic programming approach for clinical trial planning in new drug development. *Computers and Chemical Engineering*, *32*(11), 2626–2642.

Colvin, M., Maravelias, C. T. (2011). R&D pipeline management: Task interdependencies and risk management. *European Journal of Operational Research, 215*, 616–628.

Connor, A., & Tilley, D. (1998). A tabu search method for the optimization of fluid power circuits. *Journal of Systems & Control Engineering*, *212*, 373–381.

Cooper, R., Edgett, S., & Kleinschmidt, E. (1999). New product portfolio management: Practices and performance. *Journal of Product Innovation Management*, *16*, 333–351.

Copeland, T., & Antikarov, V. (2001). *Real options: A practitioner's guide*. Texere.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, *6*, 182–197.

Deb, K., Srinivasan, A. (2005). *Innovization: Innovative design principles through optimization* (Technical Report). Kanpur: Department of Mechanical Engineering, Indian Institute of Technology.

Dedieu, S., Pibouleau, L., Azzaro-Pantel, C., & Domenech, S. (2003). Design and retrofit of multiobjective batch plants via a multicriteria genetic algorithm. *Computers and Chemical Engineering*, *27*, 1723–1740.

Dietz, A., Azzaro-Pantel, C., Pibouleau, L., & Domenech, S. (2005). A framework for multiproduct batch plant design with environmental consideration: Application to protein production. *Industrial and Engineering Chemistry Research*, *44*, 2006–2191.

Ding, M., & Eliashberg, J. (2002). Structuring the new product development pipeline. *Management Science*, *48*, 343–363.

Fonseca, C. M., & Fleming, P. J. (1995). Multiobjective genetic algorithms made easy: Selection, sharing and mating restriction. In *IEE/IEEE international conference on genetic algorithms in engineering systems*.

Fonseca, C. M., & Fleming, P. J. (1998). Multiobjective optimization and multiple constraint handling with evolutionary algorithms – Part I: A unified formulation. *IEEE Transactions on Systems, Man, & Cybernetics Part A: Systems & Humans*, *28*, 26–37.

Fourman, M. (1985). Compaction of symbolic layout using genetic algorithms. In *Conference on genetic algorithms*.

Goldberg, D. (1989). *Genetic algorithms in search and machine learning*. Addison Wesley.

Grossmann, I. (2005). Enterprise-wide optimization: A new frontier in process systems engineering. *AIChE Journal*, *51*, 1846–1857.

Guillen, G., Badell, M., Espuna, A., & Puigjaner, L. (2006). Simultaneous optimization of process operations and financial decisions to enhance the integrated planning/scheduling of chemical supply chains. *Computer and Chemical Engineering, 30*, 421.

Holland, J. H. (1975). *Adaptation in natural and artificial systems, an introductory analysis with application to biology, control and artificial intelligence*. Complex Adaptive Systems.

Honkomp, S. J. (1998). *Solving mathematical programming planning models subject to stochastic task success* (Ph.D. thesis). Purdue University, West Lafayette, IN.

Horn, J., Nafpliotis, N. (1993). *Multiobjective optimization using the niched Pareto Genetic Algorithm* (Technical Report). Urbana, USA: Illinois Genetic Algorithm Laboratory, Dept. of General Engineering, University of Illinois at Urbana-Champaign.

Jacob, W., & Kwak, Y. (2003). In search of innovative techniques to evaluate pharmaceutical R&D projects. *Technovation*, *23*, 291–296.

Jain, V., & Grossmann, I. (1999). Resource constrained scheduling of test in new product development. *Industrial and Engineering Chemistry Research*, *38*, 3013–3036.

Kurasawe, F. (1991). A variant of evolution strategies for vector optimization. *Parallel Problem Solving from Nature*, *496*, 193–197.

Lin, C., & Hsieh, P. (2004). A fuzzy decision support system for strategic portfolio management. *Decision Support Systems, 38*, 383–398.

Lin, C., Tan, B., & Hsieh, P. (2005). Application of the fuzzy weighted average in strategic portfolio management. *Decision Sciences*, *36*, 498–511.

Loch, C., & Bode-Greuel, K. (2001). Evaluating growth options as sources of value for pharmaceutical research projects. *R&D Management*, *31*, 231–248.

Maravelias, C., & Grossmann, I. (2001). Simultaneous planning for new product development and batch manufacturing facilities. *I & EC Research*, *40*, 6147–6164.

Montagna, J. M., & Vecchietti, A. R. (2003). Retrofit of multiproduct batch plants through generalized disjunctive programming. *Mathematical and Computer Modelling, 38*, 465.

Morales-Mendoza, L., Perez-Escobedo, J., Aguilar-Lasserre, A., Azzaro-Pantel, C., Pibouleau, L., & Domenech, S. (2011). Selecting the best alternative based on a hybrid multiobjective ga-mcdm approach for new product development in the pharmaceutical industry. In *Proceedings IEEE symposium on swarm intelligence* Paris, 11–15 May.

Newton, D., Paxson, D., & Widdicks, M. (2004). Real R&D options. *International Journal of Management Reviews*, *5–6*, 113–130.

Perez-Escobedo, J. (2010). *Multiobjective optimization of new product development in the pharmaceutical industry* (Ph.D. thesis). INPT.

Pirdashti, M., Ghadi, A., Mohammadi, M., & Shojatalab, G. (2009). Multi-criteria decision-making selection model with application to chemical engineering management decisions. *World Academy of Science, Engineering and Technology*, *49*, 54–59.

Poh, K., Ang, B., & Bai, F. (2001). A comparative analysis of R&D project evaluation methods. *R&D Management*, *31*, 63–75.

Rajapakse, A., Titchener-Hooker, N. J., & Farid, S. S. (2005). Modelling of the biopharmaceutical drug development pathway and portfolio management. *Computers and Chemical Engineering, 29*, 1357–1368.

Ravemark, D. E., & Rippin, D. W. T. (1998). Optimal design of a multiproduct batch plant. *Computer and Chemical Engineering*, *22*, 177.

Rogers, M., Gupta, A., & Maranas, C. (2002). Real options based analysis of optimal pharmaceutical research and development portfolios. *Industrial and Engineering Chemistry*, *41*, 6607–6620.

Santiago, L., & Bifano, T. (2005). Management of R&D projects under uncertainty: A multidimensional approach to managerial flexibility. *IEEE Transactions on Engineering Management*, *52*, 269–280.

Schaffer, J. (1985). Multiple objective optimization with vector evaluated genetic algorithms. In *First international conference on genetic algorithms*.

Schmidt, C., & Grossmann, I. (1996). Optimization models for the scheduling of testing tasks in new product development. *Industrial and Engineering Chemistry Research*, *35*, 3498–3510.

Shah, N. (2004). Pharmaceutical supply chains: Key issues and strategies for optimization. *Computer and Chemical Engineering, 28*, 929–941.

Shapiro, J. F. (2001). *Modeling the supply chain*. Duxbury.

Srinivas, N., & Deb, K. (1995). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, *2*, 221–248.

Subramanian, D., Pekny, J., Reklaitis, G., & Blau, G. (2001). A simulation-optimization framework for research and development pipeline management. *AIChE Journal*, *47*, 2226–2242.

Subramanian, D., Pekny, J., Reklaitis, G., & Blau, G. (2003). Simulation-optimization framework for stochastic optimization of R&D pipeline management. *AIChE Journal*, *49*, 96–112.

Taboada, H., & Coit, D. (2006). Data mining techniques to facilitate the analysis of the Pareto-optimal set for multiple objective problems. In *Industrial engineering research conference*.

Thieme, R., Song, M., & Calantone, R. (2000). Artificial neural network decision support systems for new product development project selection. *Journal of Marketing Research*, *37*, 499–507.

Varma, V. (2005). *Development of computational models for strategic and tactical management of pharmaceutical R&D pipelines* (Ph.D. thesis). Purdue University, West Lafayette, IN.

Wan, X., Pekny, J., & Reklaitis, G. (2006). Simulation based optimization for risk management in multi-stage capacity expansion. In *Computer-aided chemical engineering, 16th European symposium on computer aided process engineering and ninth international symposium on process systems engineering (Vol. 21)*.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*, 67.

Yoshikawa, M., & Terai, H. (2005). Hybrid genetic algorithm engine for high-speed floorplanning. In *Circuit theory and design, 2005. Proceedings of the 2005 European conference*.

Zamora, M. Z., & Grossmann, I. E. (1998). A global minlp optimization algorithm for the synthesis of heat exchanger networks with no stream splits. *Computer and Chemical Engineering*, *22*, 367.

Zapata, J., Varma, V., & Reklaitis, G. (2007). Impact of tactical and operational policities in the selection of a new product portfolio. *Computer and Chemical Engineering*, *32*, 307–319.

Zitzler, E., Deb, K., & Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, *8*, 173–195.

Zitzler, E., & Thiele, L. (1999). Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Transaction on Evolutionary Computation*, *3*, 257–271.