Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Génie des procédés et de l'Environnement

Présentée et soutenue par :
José Luis Pérez-Escobedo

le : 03 Juin 2010

Titre :

Optimisation du développement de nouveaux produits dans l'industrie pharmaceutique par algorithme génétique multicritère

---

JURY

M J. M. Reneaume, ENSGTI , M. G. Defaye ENSCP Rapporteurs
Mme C. Azzaro-Pantel, INPT/ENSIACET, Membre (Directeur de thèse)
M L. Pibouleau, INPT/ENSIACET, Membre
M X. G. Yuan, Chemical Engineering Research Center Tianjin, Membre
Mme P. Zarate. ENSIACET/IRIT. Membre

Ecole doctorale :
Mécanique, Energétique, Génie civil et Procédés (MEGeP)

Unité de recherche :
Laboratoire de Génie Chimique

Directeur(s) de Thèse :

Mme Catherine Azzaro-Pantel

Rapporteurs :

M Jean-Michel Reneaume et M Guy Defay

Title

Multiobjective optimization of New Product Development in the pharmaceutical industry

Abstract

New Product Development (NPD) constitutes a challenging problem in the pharmaceutical industry, due to the characteristics of the development pipeline, namely, the presence of uncertainty, the high level of the involved capital costs, the interdependency between projects, the limited availability of resources, the overwhelming number of decisions due to the length of the time horizon (about 10 years) and the combinatorial nature of a portfolio. 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 copying 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 (NSGA II type, Non-Sorted Genetic Algorithm II) to optimize the highly combinatorial portfolio management problem. An object-oriented model previously developed for batch plant scheduling and design is then extended to embed the case of new product management, which is particularly adequate for reuse of both structure and logic. Two case studies illustrate and validate the approach. From this simulation study, three performance evaluation criteria must be considered for decision making: the Net Present Value (NPV) of a sequence, its associated risk defined as the number of positive occurrences of NPV among the samples and the time to market. They have been used in the multiobjective optimization formulation of the 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. NSGA II has been adapted to the treated case for taking into account both the number of products in a sequence and the drug release order. From an analysis performed for a representative case study on the different pairs of criteria both for the bi- and tricriteria optimization, the optimization strategy turns out to be efficient and particularly elitist 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 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.

Key words

Portfolio management, New Product Development, Discrete Event Simulation, Multiobjective optimization, Multicriteria Genetic Algorithms

Titre

Optimisation multiobjectif du Développement de Nouveaux Produits dans l'industrie pharmaceutique

Résumé

Le développement de nouveaux produits constitue une priorité stratégique de l'industrie pharmaceutique, en raison de la présence d'incertitudes, de la lourdeur des investissements mis en jeu, de l'interdépendance entre projets, de la disponibilité limitée des ressources, du nombre très élevé de décisions impliquées dû à la longueur des processus (de l'ordre d'une dizaine d'années) et de la nature combinatoire du problème. Formellement, le problème se pose ainsi : sélectionner des projets de R&D parmi des projets candidats pour satisfaire plusieurs critères (rentabilité économique, temps de mise sur le marché) tout en considérant leur nature incertaine. Plus précisément, les points clés récurrents sont relatifs à la détermination des projets à développer une fois que les molécules cibles sont identifiées, leur ordre de traitement et le niveau de ressources à affecter. Dans ce contexte, une approche basée sur le couplage entre un simulateur à événements discrets stochastique (approche Monte Carlo) pour représenter la dynamique du système et un algorithme d'optimisation multicritère (de type NSGA II) pour choisir les produits est proposée. Un modèle par objets développé précédemment pour la conception et l'ordonnancement d'ateliers discontinus, de réutilisation aisée tant par les aspects de structure que de logique de fonctionnement, a été étendu pour intégrer le cas de la gestion de nouveaux produits. Deux cas d'étude illustrent et valident l'approche. Les résultats de simulation ont mis en évidence l'intérêt de trois critères d'évaluation de performance pour l'aide à la décision : le bénéfice actualisé d'une séquence, le risque associé et le temps de mise sur le marché. Ils ont été utilisés dans la formulation multiobjectif du problème d'optimisation. Dans ce contexte, des algorithmes génétiques sont particulièrement intéressants en raison de leur capacité à conduire directement au front de Pareto et à traiter l'aspect combinatoire. La variante NSGA II a été adaptée au problème pour prendre en compte à la fois le nombre et l'ordre de lancement des produits dans une séquence. A partir d'une analyse bicritère réalisée pour un cas d'étude représentatif sur différentes paires de critères pour l'optimisation bi- et tri-critère, la stratégie d'optimisation s'avère efficace et particulièrement élitiste pour détecter les séquences à considérer par le décideur. Seules quelques séquences sont détectées. Parmi elles, les portefeuilles à nombre élevé de produits provoquent des attentes et des retards au lancement ; ils sont éliminés par la stratégie d'optimistaion bicritère. Les petits portefeuilles qui réduisent les files d'attente et le temps de lancement sont ainsi préférés. Le temps se révèle un critère important à optimiser simultanément, mettant en évidence tout l'intérêt d'une optimisation tricritère. Enfin, l'ordre de lancement des produits est une variable majeure comme pour les problèmes d'ordonnancement d'atelier.

Mots-Clés

Gestion du portefeuille de produits, Développement de nouveaux produits, Simulation par événements discrets, Optimisation, Algorithmes génétiques multicritères.

Contents

Chapter 1

Introduction, Aims and Scope

1.1 General context

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 making use of the concept of the "chemical supply chain" as shown in Figure 1.1. Process Systems Engineering is now concerned with the improvement of decision making processes for the creation 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.

To support the expansion to R&D, 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 increasing pressure for reducing costs and inventories, in order to remain competitive in the global marketplace, enterprise-wide optimization (EWO) that might be considered as an equivalent term for describing the chemical supply chain (see Shapiro [2001]) has thus become the "holy grail" in process industries.

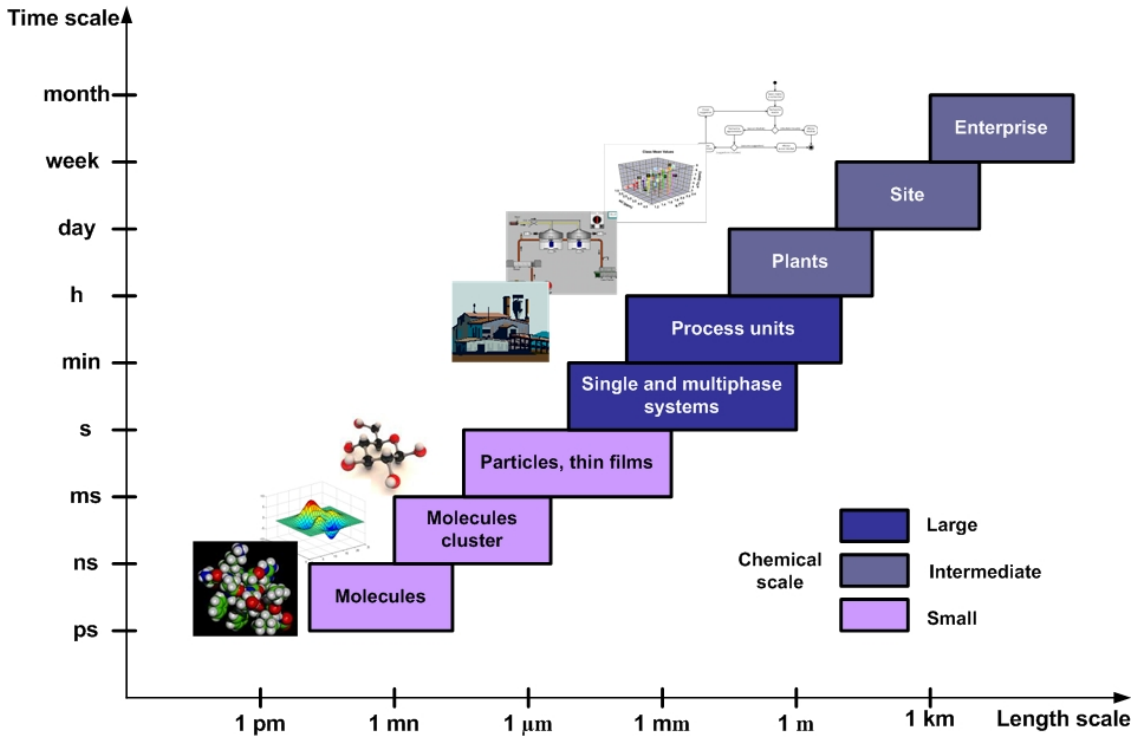


Figure 1.1: Chemical Supply Chain [Grossmann and Westerberg, 2000]

Enterprise-wide optimization is an area that lies at the interface of chemical engineering (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 and Grossmann [2004] for a recent review), they are ideally positioned to make significant contributions in EWO. This motivates the research challenges of the thesis work, which is devoted to the management of the so-called New Product Development management for pharmaceutical/biotechnology industry. This work is 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. Even if this thesis work was not supported by an industrial partnership, it must be highlighted that we have several fruitful discussions with a French pharmaceutical company to assess the validity of the examples that will be tackled here and that will serve as a guideline of the methodological framework.

1.2 Key issues in New Product Development

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 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. Schmidt and Grossmann [1996] proposed various MILP optimization models for the scheduling of testing tasks with no resource constraints with a discretization scheme in order to induce linearity in the cost of testing. Jain and Grossmann [1999] extended these models to account for resource constraints. Subramanian et al. [2003] proposed a simulation-optimization framework that takes into account uncertainty in duration, cost and resource requirements and extended this model to account for risk. Maravelias and Grossmann [2001] proposed an MILP model 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 mono-objective Genetic Algorithm to optimize product sequence evaluated by a commercial discrete-event simulator.

This work lies in this perspective: the underlying idea is to use a multiobjective framework as already initiated by [Aguilar-Lasserre et al., 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.

1.3 Portfolio selection

1.3.1 Portfolio management definition

Formally, portfolio management is a dynamic decision process, whereby a business's list of active new product and R&D projects is updated and revised. In this process, new projects are evaluated, selected and prioritized; existing projects may be accelerated, killed or de-prioritized; and resources are allocated or re-allocated to the active project. The portfolio decision process is characterized by uncertain and changing information, dynamic opportunities, multiple goals and strategic considerations, interdependencies among projects and multiple decision-makers and locations. Even if portfolio management is viewed as a very important task in industry, there is no consensus about the best strategy, perhaps because there are too many different business practices, much confusion about which strategy is the best, and no easy answers as reported in Cooper et al. [1999].

The work presented here has not the ambition to treat all the issues involved but to give a solution to the most critical ones.

1.3.2 Risk assessment

A balanced whole portfolio provides the investor with protections and opportunities with respect to a wide range of contingencies. The investor should build toward an integrated portfolio which best suits his needs [Markowitz, 1959].

A portfolio analysis starts with information concerning individual securities. It ends with conclusions concerning as a whole. The purpose of the analysis is to find portfolios which best meet the objectives of the investor [Markowitz, 1959].

Various types of information concerning securities can be used as the raw material of a portfolio analysis. One source of information is the past performance of individual securities. A second source of information is the belief of one or more security analysts concerning future performances. When past performances of securities are used as inputs, the outputs of the analysis are portfolios which performed particularly well in past. When beliefs of security analysts are used as inputs, the outputs of the analysis are the implications of these beliefs for better and worse portfolios [Markowitz, 1959].

Uncertainty is a salient feature of security investment. Economic forces are not understood well enough for predictions to be beyond doubt or error. Even if the consequences of economic conditions were understood perfectly, non-economic influences can change the course of general prosperity, the level of the market, or the success of a particular security [Markowitz, 1959].

The subject of risk [Kaplan and Garrick, 1981] has become very popular and is involved in various fields, far beyond the subject of this thesis: business risk, social risk, economic risk, safety risk, investment risk, military risk, political risk, etc...

Distinction between Risk and Uncertainty The notion of risk involves both uncertainty and some kind of loss or damage that might be received [Kaplan and Garrick, 1981]. Symbolically, it could be written this as:

$$risk = uncertainty + damage$$

This equation expresses the first distinction. As a second, it is great to differentiate between the notions of "Risk" and "Hazard".

Distinction between Risk and Hazard In the dictionary hazard is defined as "a source of danger". Risk is the "possibility of loss injury" and the "degree of probability of such loss". Hazard, therefore, simply exists as a source. Risk includes the likehood of conversion of that source into actual delivery of loss, injury, or some form of damage [Kaplan and Garrick, 1981]. This idea is symbolically in the form of an equation:

$$Risk = \frac{hazard}{safeguards}$$

This equation also brings out the thought that we may make risk as small as we like by increasing the safeguards [Kaplan and Garrick, 1981] but may never, as a matter of principle, bring it to zero. Risk is never zero, but it can be small.

1.3.3 Objectives of a portfolio analysis

A portfolio analysis must be based on criteria which serve as a guide for the decision maker.

The proper choice of a criteria depends on the nature of the investor. For some investors, taxes are a prime consideration; for others, such as non-profit corporations, they are irrelevant. Institutional considerations, legal restrictions, relationships between portfolio returns and the cost of living may be important to one investor and not to another. For each type of investor, the details of the portfolio analysis must be suitably selected [Markowitz, 1959].

Two objectives, however, are common to all investors:

1. They want "return" to be high. The appropriate definition of "return" may differ from investor to investor. But, in whatever sense is appropriate, they prefer more of it to less of it.
2. They want this return to be dependable, stable, not subject to uncertainty.

The portfolio with highest "likely return" is not necessarily the one with least "uncertainty of return". The most reliable portfolio with an extremely high likely return may be subject to an unacceptably high degree of uncertainty. The portfolio with the least uncertainty may have an undesirable small "likely return". Between these extremes would lie portfolios with varying degrees of likely return and uncertainty [Markowitz, 1959]. It must be said at that level that the proper choice among efficient portfolios depends on the willingness and ability of investor to assume risk.

For this purpose, this work must develop a strategy that separates efficient from inefficient portfolios, helps the investor and investment manager to carefully select the combination of likely return and uncertainty that best suit his needs, and finally determines the portfolio which provides this most suitable combination of risk and return.

1.3.4 Portfolio selection techniques in industrial practice

No method has a monopoly in the field of portfolio management. It is quite common that a company uses multiple methods or techniques for portfolio management. These techniques, in rank order of popularity, are as follows [Cooper et al., 1999]:

- *Financial methods*, where profitability, return, payback, or economic value of the project is determined, and projects are judged and rank ordered on this criterion: 77% of businesses use this approach.
- *Business strategy methods*, where the business's strategy is the basis for allocating money for different types of projects. For example, having decided strategy, different buckets or envelopes of money for different project types are established and projects are rank ordered within buckets: 64.8% of businesses use a strategic approach.
- *Bubble diagrams*, where projects are plotted on on X-Y portfolio map (the X-Y axes are various dimensions of interest, such as reward versus probability of success): 40.6% of businesses employ bubble diagrams.
- *Scoring models*, where projects are rated or scored on a number of criteria on scales, then the ratings are added to yield a project score (this score then becomes the basis as a ranking/prioritizing tool: 37.9% of businesses employ scoring models for portfolio management.
- *Checklists*, where projects are evaluated via a list of yes/no questions (and each project must achieve all or a certain percentage of "yes" answers): only 20.9% of businesses use checklists for project selection and porfolio management.

The percentage cited add up to well over 100% (241.5%), suggesting that, on average, the typical business relies on about 2.4 times different portfolio management methods. Using multiple methods -the notion of hybrid approach of portfolio management- appears to be the right answer, however [Cooper et al., 1999].

Another kind of classification for studies on R&D portfolio management is considered in Wang and Hwang [2007]. Studies on R&D portfolio management can be divided into three categories: strategic management tools, benefit measurement methods, and mathematical programming approaches. The strategic management tools, such as bubble diagram, portfolio map, and strategic bucket method, are used to emphasize the connection of innovation projects to strategy or illuminate issues of risk or strategic balances of the portfolio. Benefits measurement methods determine the preferability figure of each project. A number of approaches, such as the merit-cost value index, the analytical hierarchy process, net present value, and option pricing theory, have been developed in the literature to estimate the benefit of an R&D project. The projects with the highest score may be selected sequentially. The major drawback of most benefits measurements approaches is that neither uncertainty nor resources interactions among projects can be captured. In recent years, some studies used the criterion of conditional stochastic dominance or the mean-Gini analysis to make the decisions to handle R&D uncertainties for risk-averse decision makers.

Mathematical programming models optimize some objective functions subject to constraints related to resources, project logics, technology, and strategies.

The NPD problem is clearly based on an optimization formulation.

The main contributions in this field are presented below.

1.4 Related optimization works

1.4.1 General classification

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 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. The complexity of the problem has led to the common use of decomposition based strategies, resulting in two completely independent bodies of decision support literature: **strategic/tactical** and **tactical/operational**. Each of the two branches can be further subdivided according to the characteristics of the model used to support the decision making process. A taxonomy is proposed in Zapata et al. [2007]. Figure 1.2 shows the taxonomy of the main criteria to be considered when characterizing the level of detail of the model used in the decision support strategy for NPD portfolio management.

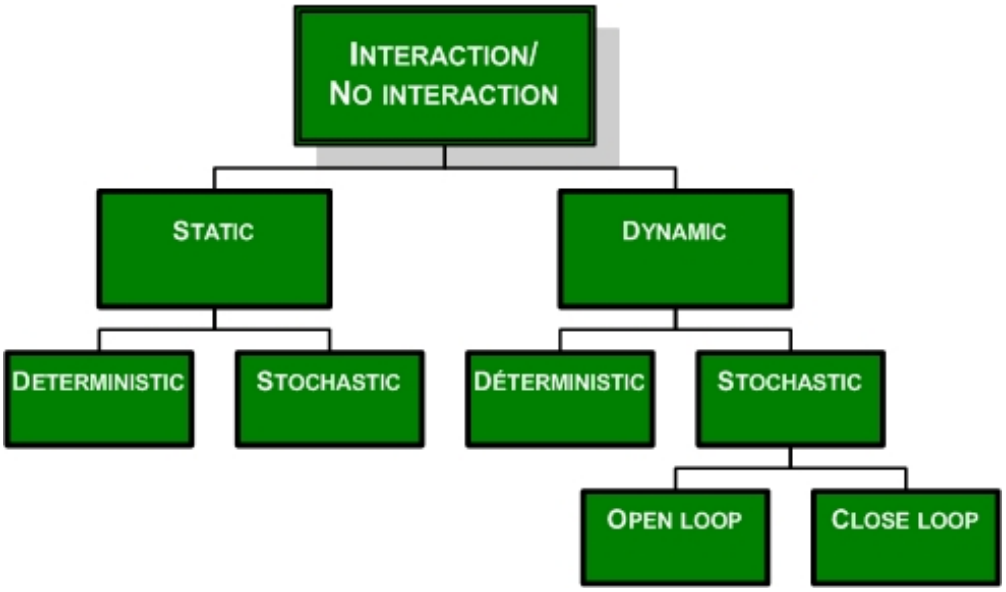


Figure 1.2: Taxonomy of the level of detail of decision support strategies from Zapata et al. [2007]

The first sublevel reflects the fact that a project can be analyzed in isolation based on certain company standards (e.g. the net present value (NPV) of the project), or as part of the bigger picture where the performance is assessed at the portfolio level (e.g. NPV of the portfolio), including all the interactions between projects. The time dimension is found one level down in the classification. A **dynamic model** provides the specific state of the systems along each point of the time horizon (e.g. number of projects waiting for resource x at time t), while **a static one** uses average values to represent the system (e.g. average number of projects waiting for resource x at any time). Within **static and dynamic** classes it is possible to choose between **deterministic and stochastic** models. However, **dynamic stochastic models** have an additional partition: **open loop versus closed loop**. **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.

In the strategic decision support systems literature, the different techniques available are shaped by the type of data used, namely, qualitative and quantitative. Strategies that are based on qualitative data and do not take into account project interactions are static. They have the main objective of translating the vagueness of adjectives used in classifications (e.g. excellent, poor, good, etc.) into structured forms that allow a quantitative comparison of the projects in the portfolio. The methodologies in this area can be grouped into scoring methods [Cooper et al., 1999, Coldrick et al., 2005], analytical hierarchy approaches [Calantone et al., 1999, Poh et al., 2001] and fuzzy logic based approaches [Buyukozkan and Feyzioglu, 2004, Lin and Hsieh, 2004, Lin et al., 2005]. On the other hand, the methodologies that are based on quantitative information strive to provide a realistic simulation of the behavior 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 (IRR), etc.) ([Cooper et al., 1999], as well as dynamic stochastic strategies, both closed loop such as real options [Copeland and Antikarov, 2001, Loch and Bode-Greuel, 2001, Jacob and Kwak, 2003, Newton et al., 2004, Santiago and Bifano, 2005], and open loop such as discrete event simulation [Chapman and Ward, 2002], and neural networks [Thieme et al., 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 et al. [2004] which proposes the use of stochastic optimization: the portfolio is modeled using a discrete event simulation and the optimization is implemented by a genetic algorithm; Rogers et al. [2002] formulates a real options decision tree that captures technical and market uncertainty as a stochastic MILP that relates projects through a budget constraint. Rajapakse et al. [2005] presents a decision support tool that uses sensitivity and scenario analysis on a discrete event model of the development pipeline. Finally, Ding and Eliashberg [2002] approaches 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. For that purpose, they use a static strategy based on a statistical model in which the outcome of each project follows a binomial distribution. All of the techniques in this group are mainly focused on time independent decisions (excluding the work by 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 et al., 2006], but the non-Markovian nature of the associated decision problem has limited such strategies to portfolios with a modest number of projects.

At the operational level, decisions are time dependent. Although their number is substantially larger and their interactions much more complex than those required at the strategic level, they are mostly Markovian in nature (i.e. information about the current state of the system is sufficient to characterize the system and be able to make decisions). This has motivated the development of operational decision support systems exclusively based on quantitative information and with a dynamic character. These techniques have mainly focused on scheduling and resource allocation. They can be divided in two main subgroups according to the type of solution strategy, namely conventional optimization and simulation optimization. In the first subgroup the problem is formulated as a resource constrained project scheduling problem (RCPSP) MILP in which the model is deterministic and the stochastic nature of the system is reflected in the use of expectation in the objective function and constraints. Along these lines, Honkomp [1998] proposed a discrete time MILP formulation that maximizes the total expected net present value of the projects in the pipeline and constrains the allocation of resources based on an overbooking strategy; Jain and Grossmann [1999] presented a continuous time formulation that minimizes the total expected cost of the portfolio and allows the inclusion of outsourcing as an additional degree of freedom in the optimization. The second subgroup determines allocation and scheduling policies by learning from the responses of a discrete event model of the system to changes implemented by a RCPSP MILP [Subramanian et al., 2001, Varma, 2005], thus making this technique stochastic.

This work will be devoted to the development of a dynamic stochastic open loop methodology.

1.4.2 Presentation of classical approaches for dynamic stochastic open loop methodologies for NPD

Among the classical approaches, three contributions must be mentioned:

- A formal statement of the portfolio optimization problem is as follows [Blau et al., 2004]: select a set of new drug candidates,and sequence them for the development process in such a way that the economic return expressed as the expected positive new present value (EPNPV) is maximized for a given level of risk measured as the probability of losing money. Let us note at that level that EPNPV is defined as the expected value over the positive axis of the NPV distribution. The information about the negative part of the distribution will be conveyed by using a risk measure called probability of losing money the area under the negative axis of the NPV distribution.

  Stated as a mathematical program the portfolio optimization problem is to

  Maximize EPNPV (over all available drug selections and sequences)

  Subject to

  $P(NPV < 0) < \beta$ (Risk constraint)

  where β is the risk of losing money.

  The methodology proposed by Blau et al. [2004] involves a commercial discrete event simulation tool (CSIM) embedded in a mono-objective optimization tool based on Genetic Algorithms. It can be summarized as follows:

  1. An initial list of 10 sequences of drug candidates is generated, some from the bubble chart using individual drug analysis and other at random.
  2. For every sequence, the probability distributions associated with the activities for each of its selected drug candidates are modified or are "preprocessed" to account for dependencies between products.
  3. The behavior of each sequence is simulated by using a discrete event simulator.
  4. The results from these simulations are used by a genetic algorithm to search for improved drug sequences.

  The optimization criterion measures how closely the sequence not only maximizes economic performance but also minimizes the probability of losing money. The so-called "fitness" function, Z_k, is calculated for each of the candidate sequences in the current population by normalizing the EPNPV and risk as follows:

$$Z_k = \alpha\left(\frac{EPNPV_k - EPNPV_{min}}{EPNPV_{max} - EPNPV_{min} + \gamma}\right) + (1-\alpha)\left(\frac{Risk_{max} - Risk_k}{Risk_{max} - Risk_{min} + \gamma}\right)$$

  where $k = 1, 2, \ldots, n$ candidate sequences; $EPNVP_{min}$ and $EPNVP_{max}$ are the minimum and maximum expected positive NPV, respectively, in the current population; $Risk_{max}$ and $Risk_{min}$ are the maximum and the minimum risk probabilities in the current population; and γ is the small positive value that prevents division by zero. The nonnegative number α (between zero and one inclusive) is inversely proportional to the cost per unit violation of the risk constraint, written at a risk tolerance level of β.

- Varma et al. [2008] proposed a framework called SIM-OPT as an integrated resource management tool with the goals of maximizing the portfolio's expected net present value (ENPV), controlling risk and reducing drug development cycle times. The framework includes three key components: (1) a stochastic simulation of the pharmaceutical work flow process modeled as a discrete event system, (2) a "resource manager" based on a mixed integer linear programming formulation that schedules and allocates resources as a function of demands from the simulated work process and (3) a "strategy learner" that evaluates the impact of various resource strategies on the financial and cycle time performance of the simulated pipeline and draws key learnings. The output is a recommended set of resource management strategies and their impacts on expected return, risk and cycle time metrics. Stated as a mathematical program the portfolio optimization problem is to

  Maximize ENPV (π)
  $\pi \in \prod$
  Subject to

  Portfolio risk$\leq \beta_{Risk}$

  ATM$\leq \beta_{ATM}(i),\ \forall i \in \{1, \ldots, N\}$

  where π is the optimal policy over the set of all control policies $\prod$. The "portfolio risk" can be defined either in terms of P (portfolio NPV $<$ M), where M can be an acceptable loss value or in terms of the standard deviation of the NPV distribution. β_{Risk} is a risk tolerance factor. The ATM_i is the average time to market for the *ith* drug and $\beta_{ATM}(i)$ is the upper bound on the launch of drug i in the event of clinical success.

- In Rogers et al. [2002], a stochastic optimization model (OptFolio) of pharmaceutical research and development portfolio management is presented using a real options valuation approach for making optimal project selection decisions. In this work, only main phases of pharmaceutical R&D are considered: three clinical trial phases, FDA (Food and Drug Administration) approval and product commercialization. According to these researches, one obvious shortcoming of the NPV approach is that it assumes that all future cash flows are static, neglecting the real-world choices to stop investing in the project or change course because of market circumstances. Yet Blau et al. [2004] consider that the Real Options Valuation method has been used effectively only to evaluate single projects.

  The problem solved by Rogers et al. can be stated as follows:

  Given a set of candidate drugs in various stages of development, estimates of the probability of clinical success, duration, and investment required for the remaining stages and forecasts for the future market values, determine the optimal drug developmental portfolio that maximizes ROV.

  Some significant works for dynamic stochastic open loop methodologies for NPD are summarized in Table 1.1 but concern exclusively monocriterion approaches.

This work will be devoted to a combined approach of simulation of the NPD pipeline and strategy optimization. It must be highlighted that the multicriterion feature of the NPD problem must be taken into account and that the various criteria must be thoroughly studied.

Reference	*Optimization method*	*Criteria*
Blau et al. 2004	Genetic algorithm	Maximization of Net Present Value
Varma et al. 2008	MILP Sim-Opt	Maximization of Net Present Value Risk Minimization Minimization of average time to market
Rogers et al. 2002	MILP Real Options Valuation (ROV)	Real Options Valuation Maximization

Table 1.1: Summary of classical approaches for dynamic stochastic open loop methodologies for NPD

1.5 Dissertation outline

This introduction (Chapter 1) has presented the key features of the New Product Development problem, the aims and scope of this PhD dissertation.

Chapter 2 describes the activities involved in the NPD problem and the life cycle of a pharmaceutical product. A typical pharmaceutical R&D pipeline serves as a guideline and will be used in the following chapters.

Chapter 3 is devoted to the presentation of the discrete event simulator used to model the various paths and the precedence relations between NPD activities. The simulator extends the previous works carried out in our research group for batch plant design and scheduling. Two case studies are used to validate and illustrate the proposed approach. The uncertainty associated to cost and durations are modeled by probability approaches and Monte-Carlo simulations. The use of a discrete event simulator is particularly useful for decision criteria evaluation, such as economic and risk metrics.

Chapter 4 deals with imprecision modeling involved in the NPD problem. The objective is to investigate alternative approaches to represent imprecision in order to determine the final strategy that could be then selected at the optimization step. An interval-based method is used and the results are compared with those obtained with the probability approach.

Chapter 5 is the core of the methodology: the discrete event simulator is embedded in an outer multiobjective optimization loop. 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. The test bench examples are analyzed and some guidelines for the treatment of new cases are provided.

Chapter 6 concludes this work by summarizing the main development and results. Possible directions for further research and indications for potential applications are given as well.

Chapter 2

Analysis of New Product Development process

2.1 Introduction

Within the scope of Research and Development Pipeline management problem, several New Product Development (NPD) projects compete for a limited pool of various resource types. The discovery and development of new drugs is a very lengthy and costly process. Each project product usually involves a series of testing tasks prior to product commercialization. If the project fails any of these tasks, then all the remaining work on that product is stopped and the investment in the previous testing tasks is wasted. In this Chapter, we are attempting to present the different steps involved in the process in a generic manner. A flow diagram of the activities involved in the development of a new pharmaceutical product is proposed in Figure 2.1. Although some differences may exist referring to various industrial practices, we consider it as generic enough to embed various formulations. Our focus is on providing the key parameters (cost, duration ...) associated with the drug development process. A typical example will serve as a guideline to illustrate the presentation.

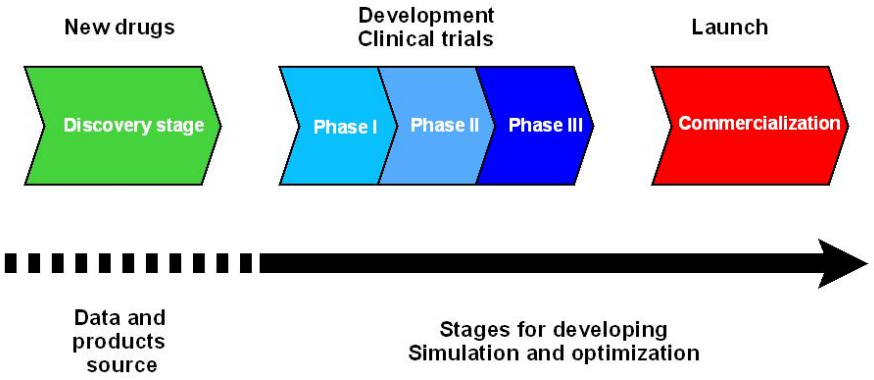


Figure 2.1: Process for drugs development.

2.2 Life Cycle of a Pharmaceutical Product

Basically, three stages are involved in the life cycle of a pharmaceutical product: discovery, development and commercialization as it can be seen in Figure 2.1. 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 behavior 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 formulations 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.

2.2.1 A typical pharmaceutical R&D pipeline

In what follows, each stage is described in more details. It must be pointed out at that level that the complexity, creativity and iterative nature of the discovery process make it difficult to describe this step in a high degree of details. Figure 2.2 is a simplified network flow diagram of the classical activities involved in the development and commercialization of a new drug candidate.

First human dose preparation (FHDP). This planning activity is relative to the preparation of administration to healthy volunteers. More practically, it includes pharmaco-kinetic studies involving adsorption, distribution, metabolism and excretion from the body as well as determining suitable dose levels.

Phase 1. In this stage, first clinical trials are carried out and drugs are administered to healthy volunteers. At the same time, acute/chronic and reproductive studies are also conducted in animals (mice/rats). Positive results will allow to the drug to go on the process where an unacceptable behaviour in human and animal studies can terminate the study.

Phase 2. Drug is administered to unhealthy human patients with the disease by using the results of dosing studies from Phase I. Coincident with these studies are long-term oncogenic toxicological studies in animals and market research to obtain sales estimates. If the compound fails to treat the disease or is inferior to competitive products, it is de-staged or returned to the discovery phase for modification.

Phase 3. Large-scale clinical studies are carried out on unhealthy human patients. The FDA is involved and indicates benchmarks for giving their approval. In addition to confirming the efficiency, these studies identify drug-drug interactions, human demographics, etc. This most expensive phase of the development process requires extensive global coordination and cooperation. The results should confirm what was learned in Phase II but on a much larger scale, otherwise the compound may be terminated.

First submission for approval. All information (efficacy, toxicology, process, drug-drug interactions, side effects, etc.) obtained is gathered and submitted it to the FDA. Simultaneously, the marketing strategy is evolving, price negotiations are being conducted with suppliers/distributors, and promotional materials are being developed. The building of a commercial plant is in progress. Approval for selling the new drug is the anticipated outcome.

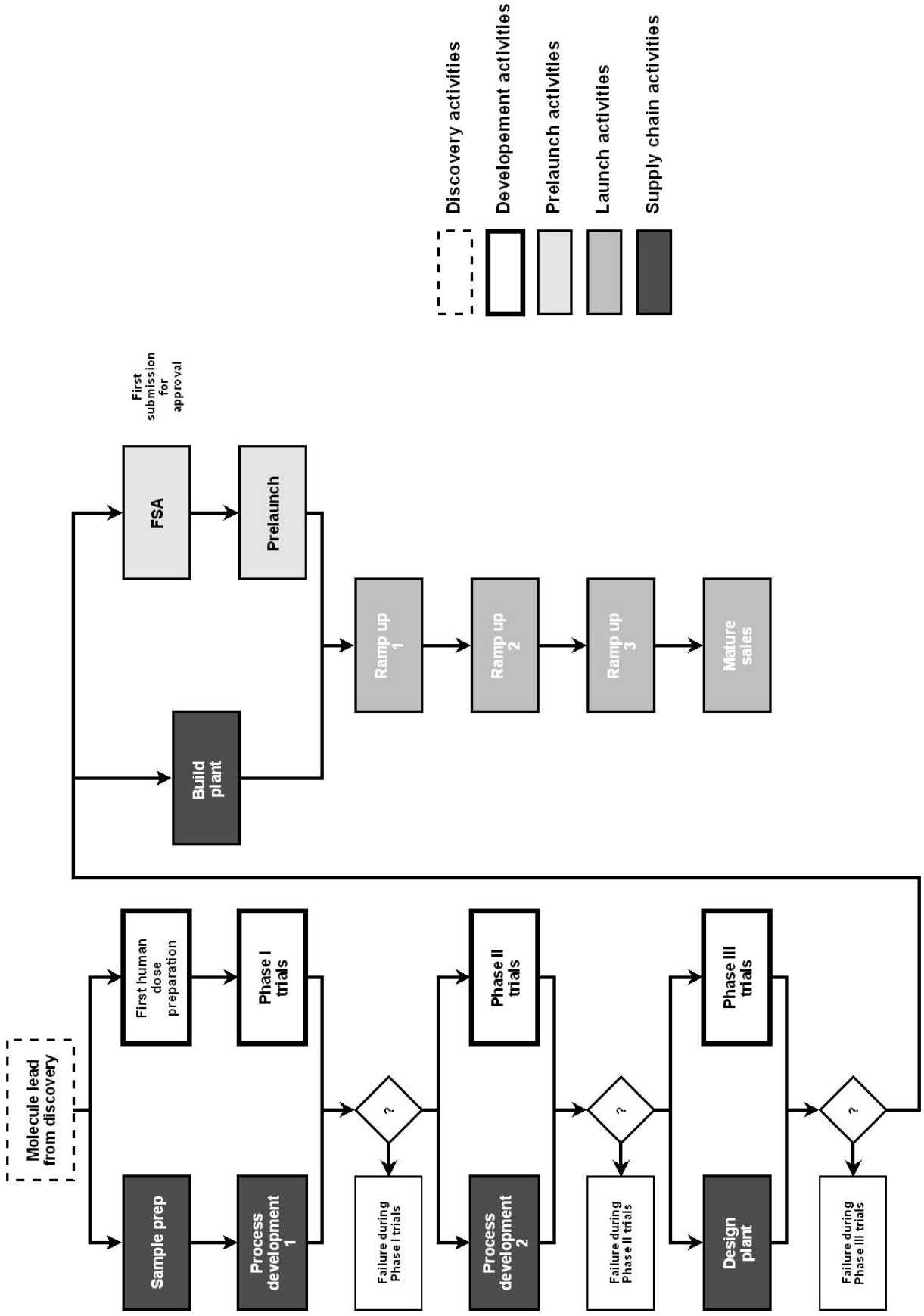


Figure 2.2: A typical pharmaceutical R&D pipeline.

Pre-launch activities. This is the final stage before launch-approval has been received from FDA; a global penetration strategy has been completed; the commercial plant has been built and started-up; the promotional campaign launched. This phase ends when the new drug is distributed.

Launch activities. The product is launched over a period of years in various global markets until mature sales levels are reached the ramp-up period. Mature sales are maintained until patents expire or competition is carried out either from competitors or planned cannibalization.

Product supply chain activities. Sample preparation, process research, process development, process design and plant construction occur simultaneously with other development activities (see Figure 2.2). Initial focus is on preparing sufficient sample material for various animal/human studies. Once the launch prospects appear promising, the emphasis changes to developing a process for commercialization. This includes a pilot plant which provides data for plant design as well as the larger quantities of material needed for Phase III clinical trials. During Phase III clinical trials, the new plant is designed or other arrangements for product manufacture are carried out. Once Phase III trials are successful, a new plant is needed, existing facilities must be expanded.

2.2.2 An example as a guideline

Typical features

In this section, the example taken from Blau et al. [2004] is considered as a reference and will serve as a guideline of this presentation. Nine potential new drugs have been identified as lead molecules by the discovery function and are candidates for entering the new product development pipeline. From the flow diagram shown in Figure 2.2, historical data from a major pharmaceutical company were used to represent the parameters with a triangular possibility distribution. The data concerning both duration and cost of each phase are presented in Table 2.1. For example, the time required for Phase I testing ranges from minimum (min) of 225 days to a maximum (max) of 375 days with a most likely (ML) value of 300 days; it can be represented by a triangular distribution as shown in Figure 2.3. Costs are not distributed between manpower and equipment/clinical costs. This type of detailed physical resource-based data is generally available and could be used for future resource planning but is beyond the scope of this work. In the last column of Table 2.1, for example, the maximum resources available are specified for each activity. If all the leads are advanced at the same time, the resource levels are exceeded in almost all of the activities. The challenge, therefore, is to propose a strategy, which will mitigate risk while carrying attractive expected financial, since the resources are rarely available to develop all these projects at once.

Additional data are provided in Table 2.2. The components of and trends in the costs of pharmaceutical innovation are analyzed in what follows. The example illustrates the interest to diversify a portfolio of new products in order to minimize vulnerability to competitor's products.

How to obtain economic and technical data

One of the largest components of the overall cost of bringing a new drug to the market is the cost of product development. Cost of product development can account for as much as 30% to 35% of the total cost of bringing a new drug to the market [Dimasi et al., 2003, Suresh and Basu, 2008].

Activity	Duration (days)			Cost ($MM)			Total available
	Min	ML	Max	Min	ML	Max	resources $MM
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 upIII	250	350	450	35	40	45	100
Mature sales	250	350	450	46	53	60	150

Table 2.1: Data for Nine Drug Candidates

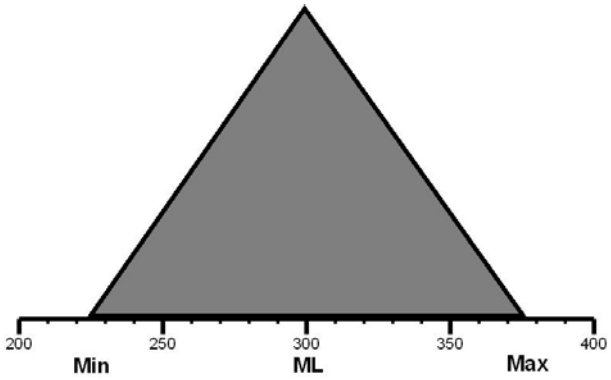


Figure 2.3: Triangular distribution example

The technical success factors are generally available from researchers while sales/marketing personnel can provide estimates for the sales expected if the compound reaches the marketplace.

For instance, there are three different diseases treated by the nine leads. It would be preferable to treat as many different diseases as possible to minimize the impact of new drugs from competitors sources. Although the drug candidates have never undergone any actual testing beyond the discovery stage, it is often possible to extract subjective probability estimates of their anticipated performance [Morgan and Henrion, 1990, Nutt, 1998] from feedback experience with similar drugs. This remark is also valid for capital cost estimation for manufacturing a candidate drug from the structure of the molecule and the chemical or biological process used to manufacture discovery samples. Finally, market research studies and forecasting practices [Cooper et al., 1999, Kahn, 2002] provide price and sales estimates for the product at some launch date in the future. Of course, in all three of these situations, the uncertainty in the estimates is quite large, ranging from 50 to 100 percent of the most likely values.

Product	Disease	Success probabilities			Capital cost ($MM)			Mature sales ($MM)			Degree of
name	type	Phase I	Phase II	Phase III	Min	ML	Max	Min	ML	Max	difficulty DoD
D1	III	0.9	0.3	0.9	40	50	60	2675	2800	2950	5
D2	I	0.85	0.2	0.85	20	30	40	1850	1900	1975	2
D3	I	0.95	0.35	0.95	30	45	60	3000	3300	3500	8
D4	II	0.87	0.22	0.8	28	34	40	2000	2250	2500	9
D5	II	0.97	0.36	0.99	38	40	46	1200	1690	2200	3
D6	I	0.83	0.18	0.86	50	60	70	2500	2830	3000	7
D7	I	0.94	0.4	0.94	65	75	90	1800	2150	3000	1
D8	II	0.86	0.2	0.88	60	65	90	1400	1600	1850	4
D9	II	0.98	0.34	0.92	52	62	72	2750	2870	2900	10

Table 2.2: Success probability, capital cost, mature sales and degree of difficulty

Data source. Based on Blau et al. [2004], cost and duration are taken from a major pharmaceutical company historical data which are used for the simulation. Other practices as market research studies and forecasting are considered too. Capital estimation is carried out by engineers from their experience accumulated from the treatment of similar cases. In the same way, success probabilities are an important parameter considered in NPD evaluation [Morgan and Henrion, 1990]. It is important to note that success probabilities are linked to every single product. Success probabilities can be calculated from information about drugs approved by the Food and Drugs Administration or the European Medicines Agency [Reichter, 2001]. An example about how success probabilities can be calculated is applied for monoclonal antibodies (mAbs) which are nature's biological warheads, able to target and help eliminate foreign or abnormal agents from the body. Table 2.3 shows data ordered by year and kind of mAbs.

Data were collected from surveys of sponsoring companies, and from public documents. The percentage of completion is the percentage of products that have been discontinued and approved, providing an indication of how far trials have progressed. A low value will inevitably reduce the accuracy of the estimated success rates for that class of mAbs. The percentage of success is the percentage of mAbs that successfully completed trials and were approved by the US FDA, e.g., in Table 2.3 for years 1992-1994, the total number of mAbs was 41, the number of mAbs discontinued was 23 and the number of mAbs approved was 5; then, the percentage of completion is: (number of mAbs discontinued + number of mAbs approved)/total number of mAbs (23+5)/41= 0.68. For the percentage of success, this is calculated as follows: number of mAbs approved/(Number of mAbs approved + Number of mAbs discontinued) 5/(23+5)=0.18.

Success rates (Table 2.4) for phase transitions were calculated as follows: the number of products that completed a given phase and entered the next phase, divided by the total number of products that entered the first phase and did not remain in that phase (i.e., all products entering the phase minus those that remained), e.g., drugs entering (de) to the phase 1:2, drugs completing (dc) the phase 1: 1; the success rate (sr)=dc/de.

Degree of Difficulty. There is a relationship between activity times and costs in Table 2.1 for specific drug candidates. For example, new drugs from a class of chemistries products would require activity resource levels closer to the maximum of the triangular time and cost distributions than those familiar to a company. 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 of this thesis is on project selection and sequencing rather than resource planning, the analysis can be

Initiation of clinical trials (years)	Total number of mAbs	Number of mAbs discontinued	Number of mAbs Approved	% completion	% success
1980-1982	2	1	1	100	50
1983-1985	9	8	0	89	0
1986-1988	33	29	2	94	6
1989-1991	34	29	2	91	6
1992-1994	41	23	5	68	18
1995-1997	33	12	0	36	0
1998-2000	34	2	0	6	0
All mAbs (1980-2000)	186	104	10	61	9
Murine mAbs	49	34	1	71	3
Chimeric mAbs	23	13	4	74	24
Humanized mAbs	59	15	5	34	25

Table 2.3: Success rates by year and by product

Initiation of clinical trials (years)	Total number of mAbs	Number of mAbs discontinued	Number of mAbs Approved	% completion	% success
1980-1982	100	50%	100%	100%	100%
1983-1985	89	67%	50%	50%	0%
1986-1988	94	58%	47%	57%	50%
1989-1991	91	64%	40%	29%	100%
1992-1994	68	85%	55%	NA	NA
1995-1997	36	77%	55%	NA	NA
1998-2000	6	90%	NA	NA	NA
Murine mAbs	71	77%	52%	45%	33%
Chimeric mAbs	74	86%	40%	80%	100%
Humanized mAbs	34	84%	72%	75%	100%

NA=Not applicable.

Table 2.4: Success rates by transition phase

simplified by using a single value of DoD ranging from 1 (very easy) to 10 (very difficult). Table 2.2 lists DoD for a set of new product candidates. The reported DoD values are used as follows: (1) The minimum and maximum of the triangular distribution remain the same as the values shown in Table 2.1 for all the drug candidates; while (2) the most likely value of the distribution is proportional to DoD. If DoD is one, for example, the most likely value is set equal to the minimum of the triangular distribution while the maximum remains the same. Conversely, if DoD is 10 the most likely value is set to the maximum while the minimum remains the same.

The data concerning the example must not be considered in absolute values but represent yet common features observed in this kind of industrial activities.

2.2.3 Data analysis

Basic data

Duration. From the data related to operation duration for the considered fifteen stages (see Table 2.1), a triangular distribution has been deduced for representing uncertainty. Let us consider for instance the FHDP activity with a minimum value of 300, a mid value of 400 and a maximum value of 500: 9 possible values are generated (taking into account the number of products to evaluate) from the triangular distribution.

Manufacturing cost. As for duration, data for costs are represented by a triangular distribution.

Resources. For each stage, there is a limited level of available resources for developing drugs projects. This means that the number and order in which drugs projects are initiated then condition the success of a sequence. It is assumed that this constrained resource is renewed after a project has been treated in a step.

Disease type. Three disease types are considered (I, II and III) in Table 2.2. Disease of type I involves drugs D2, D3, D6 and D7. Disease of type II is relative to drugs D4, D5, D8 and D9. Disease of type III only involves drug D1.

Success probability. Success probability represents uncertainty related to every drug in stages Phase I, II and III. Given that these stages concern trials, a drug can be rejected or returned for its reprocessing because of inconvenient results in patients. It must be observed that on the one hand, probabilities for phases I and III do not exhibit a big difference between them and are even identical for some drugs. On the other hand, phase II success probability is lower than success probability for phases I and III (this corresponds to the drug test in patients with a disease and outcomes can not be always as desired or expected).

Capital cost. Capital cost is represented by a triangular distribution. A value for the capital cost for a drug will fluctuate between its upper and lower limits established by using a triangular distribution.

Mature sales. As for capital cost, sales are represented by a triangular distribution to model the uncertainty: market behavior is not defined or known in advance due to competition between drugs designed by the same or other companies.

Product dependencies

Product development is generally influenced by the other products considered in the pipeline and by competitor products Roberts [1999]. In some instances, it may be advantageous to manage a candidate drug for early development despite unattractive financial and low technical success probabilities: the gained experience will provide a knowledge base to forecast the success better for dependent drug candidates later in the product sequence.

Traditionally, four frequently occurring types of dependencies are taken into account:

Financial return dependencies. Competition between similar products reduces the profit for each one because of a smaller part of market is gained: this is true for products coming either from the same company or from other ones. This dependency gives a relation for sales taking into account the quantity of products that will arrive until the stage Mature Sales (MS). For a given sequence for the products for the same disease, if two drugs arrive until the stage MS, sales for each drug will be 0.85 of the values for sales reported in Table 2.2 for every drug. If three drugs arrive until stage MS, sales will be 0.75 of values presented in Table 2.2. Finally, if four drugs arrive until stage MS, sales will be 0.60 of values in Table 2.2(See Table 2.5).

Technical dependency. This kind of dependency modifies the success probability. 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 50%. 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 10%. 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 pharmaceutical managers. This explains why it has not been taken into account for modelling.

Manufacturing cost dependency. Manufacturing cost dependencies occur when the combined cost of a development activity for two similar drug candidates is less than the sum of the cost of the individually considered projects. This has been taken into account in this way: For any sequence of drugs for Disease I, the 1st drug uses full capital shown in Table 2.5, the 2nd drug in the sequence uses 1/2 of its individual capital, the 3rd drug uses 1/3 of its capital cost, while the 4th drug uses 1/ 4 of its capital shown in Table 2.5.

Resource dependency. Learning effects frequently lead to resource dependencies. A common example occurs when the development times are reduced when taking into account the experience gained by developing two functionally similar drug types one after one.

All these dependencies are summarized in Table 2.5.

Disease	Drug	Type of dependency		
Disease I	D2, D3, D6, D7	Financial dependency	Drug number in the pipeline	Percentage reduction of matures sales
			2	0.85
			3	0.75
			4	0.6
		Technical dependency	Percentage that decreases the probability of success if the first drug in a sequence fails	Percentage that increases the probability of success if the first drug in a sequence succeeds
			50	10
		Manufacturing cost dependency	Number of drugs	Proportion of capital that is used
			1	1
			2	1/2
			3	1/3
			4	1/4
		Resources dependency	Difficulty reduction of 20% due to learning curve experience	
Disease II	D4, D5, D8, D9	Financial dependency	Total market for drugs of disease II is set at 9000 millions dollars	
		Technical dependency	Percentage that decreases the probability of success if the first drug in a sequence fails	Percentage that increases the probability of success if the first drug in a sequence succeeds
			50	10
		Manufacturing cost dependency	Number of drugs	Proportion of capital that is used
			1	1
			2	1/2
			3	1/3
		Resources dependency	Difficulty reduction of 20% due to learning curve experience	
Disease III	D1	No dependency		

Table 2.5: Dependency analysis for the treated example

2.3 Conclusions

This short chapter aims at presenting the typical formulation of New Product Development in a pharmaceutical industrial context: an example serves for illustration purpose. Basically, three stages are involved in the life cycle of a pharmaceutical product: discovery, development and commercialization. Our idea is now to use the potential of discrete event simulation to model the series of decision points along the drug development pathway. For example, at the end of each phase of clinical trials the probability of clinical success resulted in go/no-go decisions. The goal is now to model the pharmaceutical enterprise portfolio by using the principles of discrete event simulation and this is examined in Chapter 3. For this purpose, an object-oriented model structure previously developed for batch plant scheduling and design is extended to embed the case of product management, which is particularly adequate for reuse of both structure and logic.

Chapter 3

Development of a discrete event simulator for the NPD process

3.1 Introduction

The goal of this chapter is to model the various paths and the precedence relations between the activities involved in New Product Development by discrete event simulation principles used in previous works for batch plant design [Bérard et al., 2003a,b]. The problem of evaluating and selecting which new products to develop and then of sequencing or of scheduling them is not a trivial task due to dependencies between products both in the market place and in the development process itself. Discrete event simulation is a common tool used to understand how a system works and would work when changes are implemented, without incurring in expensive trials. This technique has thus been chosen and it must be highlighted that experience has been gained in our research group about DES for several years (see for instance the work of Bérard et al. [1999] about batch plant modelling in the pharmaceutical industry). DES development has shown that its internal structure and its performances were well fitted for the problem under consideration. This is why DES has been considered as the basic tool for the present work, even if some modifications and adjustments are necessary to obtain an efficient tool to tackle the NPD pharmaceutical projects of this study.

Some investigations [Blau et al., 2004, Rajapakse et al., 2005, 2006, Varma et al., 2008] used commercial simulation software tools based on discrete event simulation (CSIM [1], Simulator Extended Industry Suite V5, GenSight software[2], ...) with specific advantages. However, limitations of graphically based simulators to interact with other applications, has forced us to develop our own simulators.

This explains our main motivation to use and adapt the simulator previously developed in the group, taking into account our background on this topics since 1992 (defense of ten PHD thesis, more than 22 international publications on the subject). The C++/implemented DES can be easily modified for modelling the NPD problem considered here. This constitutes the first step of a general framework for managing NPD projects that will consider the integration of this simulation tool in a more general-purpose simulation-optimization perspective.

This chapter involves three sections:

- The extension of the DES previously developed for batch plant scheduling to the NPD problem is first presented;
- A typical simulation analysis is then performed using the example derived from [Blau et al., 2004] involving nine drugs and three target diseases. The drugs are first analyzed independently via the so-called bubble chart. Then, the interest of simulation is justified for drug sequences. The influence of capacity limitation is also highlighted.
- Another example is also considered to show the capability of the model to take into account various situations [Rajapakse et al., 2006] and to demonstrate its interest as a stand-alone decision aid tool.

[1] http://www.mesquite.com
[2] http://www.gensight.com/Project-Portfolio-Management/Overview/Home.htm

3.2 From batch plant scheduling (BPS) and design to NPD management

In a DES, a process is described as it evolves with time and changes take place only a finite number of times, i.e. event occurrence date. The DES was developed using C++ object-oriented language, according to the approach proposed by Bérard et al. [1999] (Figure 3.1). It must be emphasized that object-oriented (OO) techniques have received a lot of attention in recent years and the use of OO techniques are becoming increasingly common. The power of object oriented techniques lie in the ability to produce "modular" code (known as classes) that can be "easily" modified and reused. The ability to contain software complexity into classes and to be able to realistically represent entities from the real world in software makes OO techniques ideally suited to simulation which is inherently complex.

In Bérard et al. [1999], a four layer framework was proposed 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 (see Table 3.1).

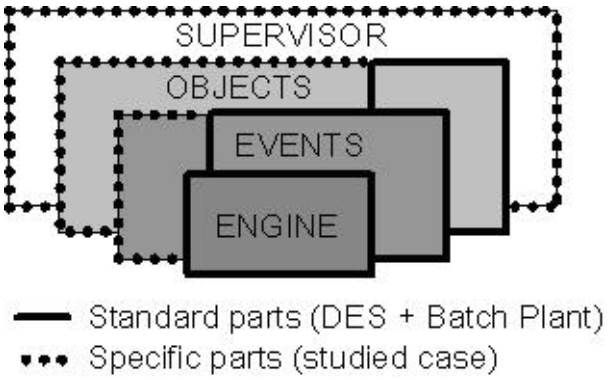


Figure 3.1: DES Framework.

In what follows, the description of the basic design of the DES and its operation principles are presented. Following the classical terminology used in object-oriented approaches, the main so-called objects of the DES will be described. To treat a particular problem, specific objects should be derived from this basic structure.

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

Table 3.1: Terminology in BPS and NPD project problems

3.2.1 Model classes

Following the classical terminology used in object-oriented approaches, the main so-called objects of the DES are described. 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.

As previously mentioned, 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.

Engine, *Event* and *Agenda* classes are shown in Figure 3.2 with their associated attributes and methods.

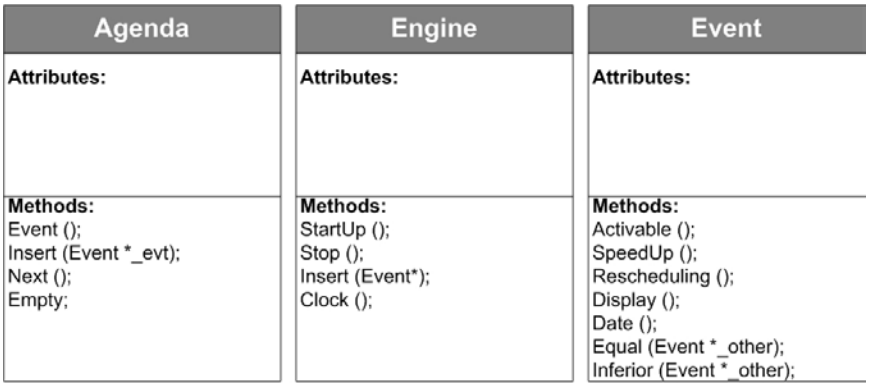


Figure 3.2: *Agenda*, *Engine* and *Event* classes

Equipment class is the basic class for activity modelling. *Product* Class is another basic representation in the system (Figure 3.3).

Loading and *Release* classes inherit from *Event* class taking into account that activities as loading and release are carried out before and after a stage is released. A class *stage* represents either a facility or a resource with name and duration as attributes (Figure 3.4). The relationship between

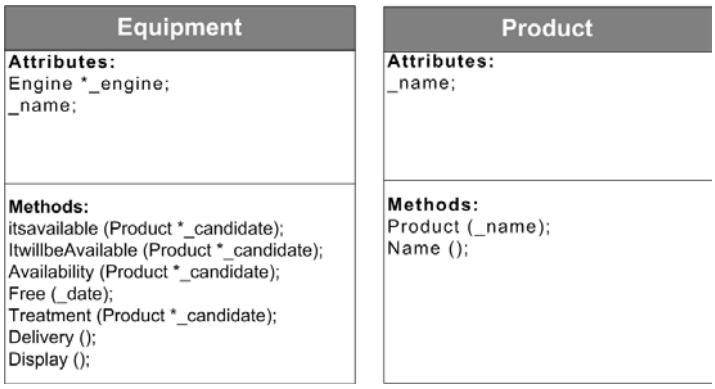


Figure 3.3: *Equipment* and *Product* Classes

classes and system activities is represented in Figure 3.5. For a better representation of the links between the conceptual and simulation models, tasks are identified by a number. More details are given in Table 3.2. Let us note at that level that NPV corresponds to the classical Net Present Value criterion.

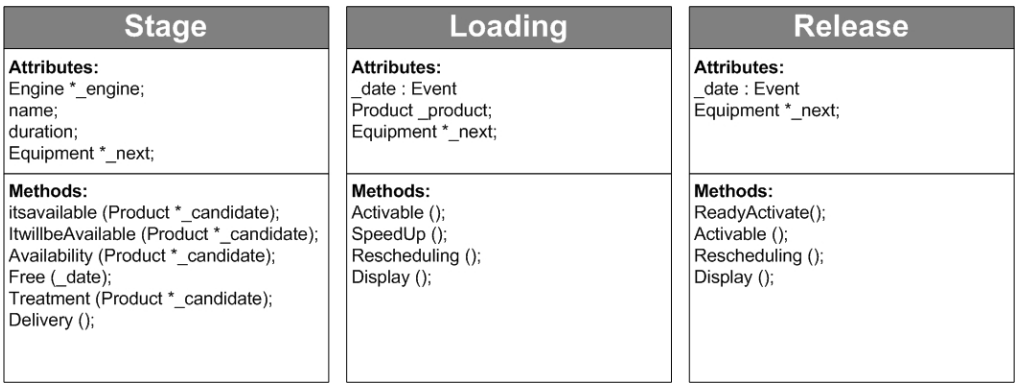


Figure 3.4: *Stage*, *Loading* and *Release* Classes

The main classes presented in Figures 3.2, 3.3 and 3.4 are the core of the DES; it must be emphasized that class *Stage* (see figure 3.4) can take into account simultaneously several projects when its capacity is sufficient.

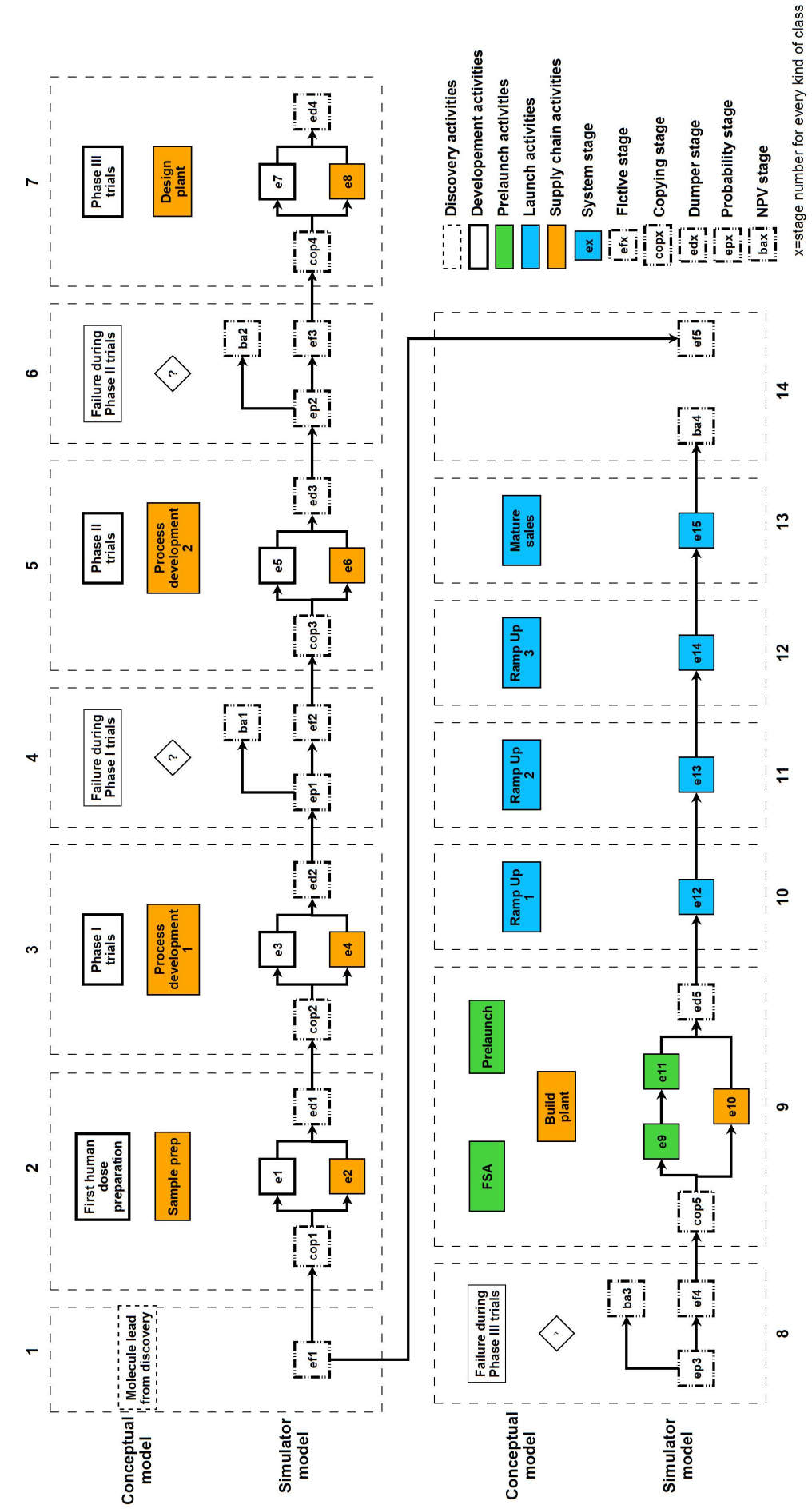


Figure 3.5: Relationship classes -system activities.

Section	Conceptual model	Simulator model	Description
1	Molecule lead from discovery	ef1	This stage is the starting activity (launch) of a sequence of products or a product (release time). A product not considered in a given sequence goes to the stage ef5.
2	First human dose preparation-Sample preparation	cop1, e1, e2, ed1	Four classes are involved: cop1, is used for "'copying"' a product (this means that a product needs parallel activities through Classes e1 and e2 (with associated name, processing time, capacity, etc...). The class ed1 is the result of the parallel activities.
3	Phase I trials-Process development 1	cop2, e3, e4, ed2	id. see Section 2
4	Failure during phase I trials	ep1, ba1, ef2	Three classes are involved: Class ep1 involves a success probability which is compared with the randomly generated value; computation of NPV is performed in case of failure (if a product class ba1) and is allocated to class ef2 in case of success.
5	Phase II trials Process development 2	cop3, e5, e6, ed3	id. see Sections 2 and 3
6	Failure during phase II trials	ep2 ba2 ef3	id. see Section 4
7	Phase III trials Design plant	cop4, e7, e8, ed4	id. see Sections 2, 3 and 5
8	Failure during phase III trials	ep3, ba3, ef4	id. see Sections 4 and 6
9	First Submission for Approval-Prelaunch-Build plant	cop5, e9, e10, e11, ed4	Decomposition principle identical to the previous sections
10	Prelaunch 1	e12	Only one activity is carried out in e12
11	Prelaunch 2	e13	Idem
12	Prelaunch 3	e14	Idem
13	Matures sales	e15	Idem
14	None	ba4 ef5	Computation of the total NPV. A product that has not been considered is allocated to ef5.

Table 3.2: Relationship conceptual-simulation model.

3.2.2 Additional classes for NPD modelling

Stage class
The *Stage* class (Figure 3.6) considers the treatment of several products or projects, with a capacity as attribute and some adequate methods for information management. More precisely, Attribute *Equipment _previous* concerns the previous equipment in the path; Attribute *TabTemps*b is a table that contains information about product duration by stage. Attribute *_capacity* is relative to the capacity of a given stage, viewed here from a financial viewpoint. Attribute T*abSortie _sort* concerns data that will be used in the following stages. Specific methods have been implemented for capacity management when too many projects compete for the same resource.

Figure 3.6: Modified *Stage* class

StageN class
This class is considered as an auxiliary class where a product goes through before going to the next step. The attributes and methods are identical to Class *Stage* (Figure 3.7).

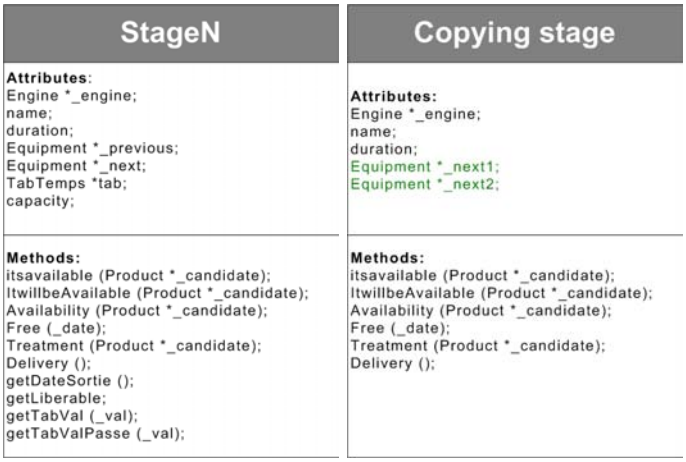


Figure 3.7: StageN and Copying classes

Copying class
This class is another auxiliary class that creates a copy of a product/project that has been imple-

mented for the treatment of parallel tasks. Basic parameter are involved in this duplication step (Figure 3.7).

StageDepA class
This class computes the waiting time of a product/project when several parallel tasks are involved with different operating times. Attributes and methods for this class are almost similar to these of a basic Stage class (Figure 3.8).

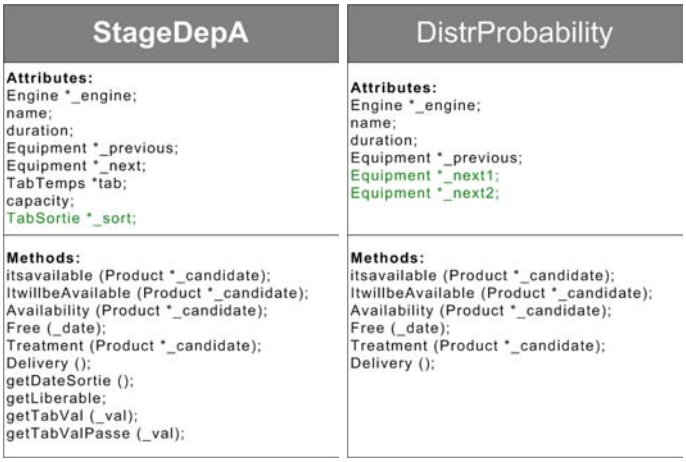


Figure 3.8: StageDep and DistrProbability classes

DistrProbability class
This class considers failure probability for each product. This value is compared with a randomly generated value. If this random number is less or equal to the failure probability value, the product will follow the remaining steps. Besides, the product is eliminated (Figure 3.8).

GeneratorRN class
The *GeneratorRN* class (Figure 3.9) generates random numbers that are used in *Distrprobability* class and also manages dependence relationships. Attributes for this class are *min, max, maxnom* and *nomSeq*: *min* and *max* define the bounds of the interval for numbers to generate; *maxnom* attribute is used for defining the maximum of random numbers to generate; *nomSeq* attribute defines the number of sequences to be evaluated by the simulator. Dependence relationships are defined by products in a sequence and are precomputed by this class before the dynamic process begins. The randomly generated values are also managed in this class.

NPV (Net Present Value) class
The *NPV* class (Figure 3.9) computes NPV when either a product goes out the pipeline (use of a *DistrProbability* class) or has completed its processing along all the stages. Relevant attributes for this class are related to cost and market information. The global NPV computation adds all the NPV of each product contributing to a sequence.

These significant classes have been used for system modelling. The simulation procedure is now described below.

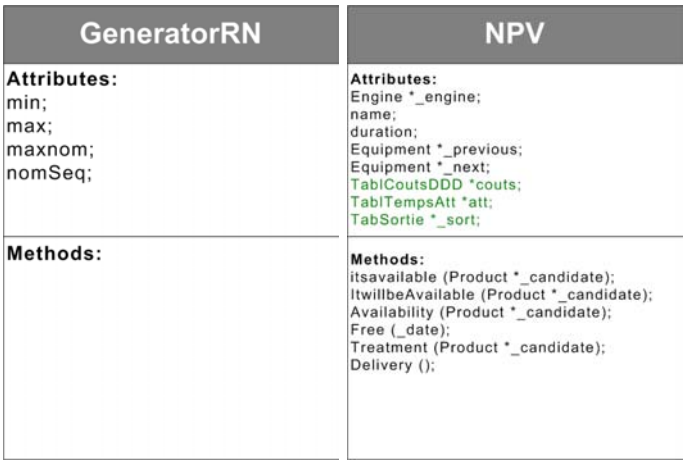


Figure 3.9: Generator and NPV classes

3.3 Simulator validation: application to the 9-drug problem of Blau et al. [2004]

The use of the DES model can be performed in a 3-stage way, as depicted in Figure 3.10:

- At first, it is interesting to study the individual behavior of each drug to evaluate its potential contribution to Net Present Value;
- Second, this preliminary phase can help to study the potential of each drug by studying the compromise between risk and potential gain. A so-called "bubble chart" representation is used for this purpose. Its principle will be explained in what follows.
- Finally, the interest of the model is much more underlined as far as the interactions between products are studied.

3.3.1 Simulation for each drug

In this subsection and the three following ones, the simulations are carried with resource capacities as described in Section 2.2.2 and Table 2.1.

Due to failure probabilities at Phase I, Phase II and Phase III trials, corresponding respectively to steps 4, 5 and 8 of the schematic process shown in Figure 3.5, the problem is by definition a stochastic one. So each simulation is repeated a large number of times (300), selecting random sampling values from a uniform random number generator for defining failure or success. If the generated number is less than a given probability value (see Table 3.5), the drug succeeds at the considered Phase, otherwise it fails and is stopped. Probability distributions for various economic and risk indicators, as well as statistical parameters, can be deduced from results gathering. For each drug launched in the process, the Net Present Value (NPV) is computed by cumulating each of its intermediate value.

The computation of the Net Present Value can be summarized as follows:

$$NPV_{drug} = \sum_m \frac{-c_m + r_m}{(1+i)^{d_m}}$$

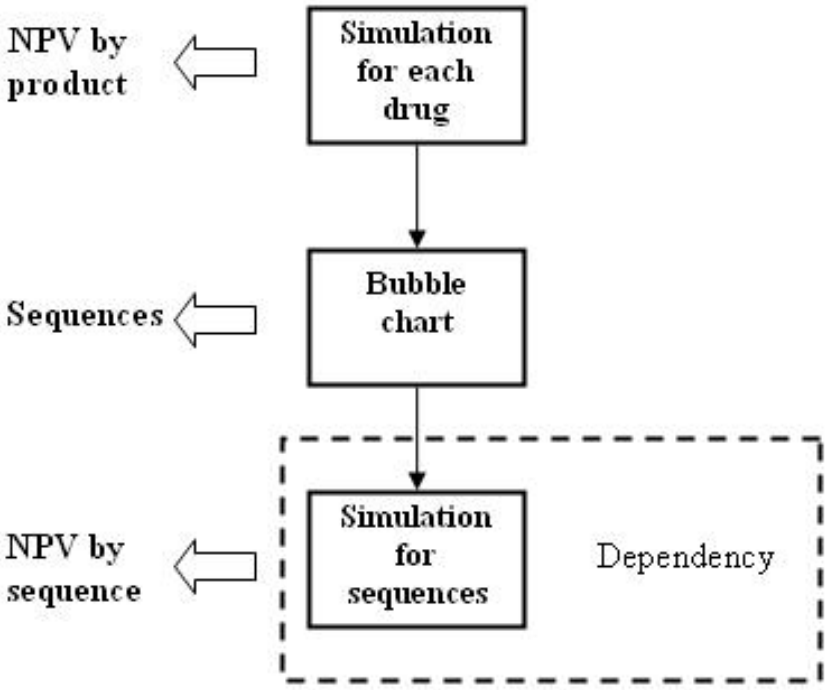


Figure 3.10: Summary of the simulation steps

Where, c_m is the cost for a given drug m, r_m is the revenue by sales for a drug m, d_m is the makespan for a sequence and i is the interest rate considered in this work.

A reward/loss ratio (A) is obtained by dividing the mean reward (mean of positive values of NPV) by the mean loss (mean of negative values of NPV). This ratio A gives the attractiveness of the project under consideration and can be used to refine the decision concerning drugs to be launched.

Figures 3.11 to 3.13 represent relative frequency for NPV values for each of the nine drugs. In all cases, there is a higher frequency for negative values (loss) than for positive values (reward). This due to the values of failure probabilities (mainly in Phase II) which entails the stop of the drug before it reaches the mature sales stage.

In Figures 3.11 to 3.13, the mean NPV corresponds to the mean value of NPV obtained from each of the 300 simulations. When a drug fails at one of the trial phase, its NPV has a negative value. The only occurrence of a positive NPV corresponds to a product which succeeds at all phases. So the NPV distribution presents two modes: one for negative NPV, another for positive NPV. This distribution being bimodal, there is no real solution corresponding to the global mean value. This is valid for all the treated examples in the manuscript. When considering the mean NPV in the remaining chapters of this manuscript, we must be aware that this represents only as a statistical parameter indicating if real solutions tend toward negative or positive values.

To represent the NPV trends more clearly, Figure 3.14 shows the maximal and mean values for the NPV related to each drug and can be used to compare the behavior of the various drug. As mentioned above, another item for choosing a strategy is the attractiveness A defined as the ratio of mean positive NPV value (reward) and mean negative NPV value (loss) reported in Table 3.3. The ranking (last line of Table 3.4) is established by performing a trade-off between max NPV, min NPV and ratio A. The best solution is drug 7, followed by drugs 1, 3, 5; drug 4 is the worst solution.

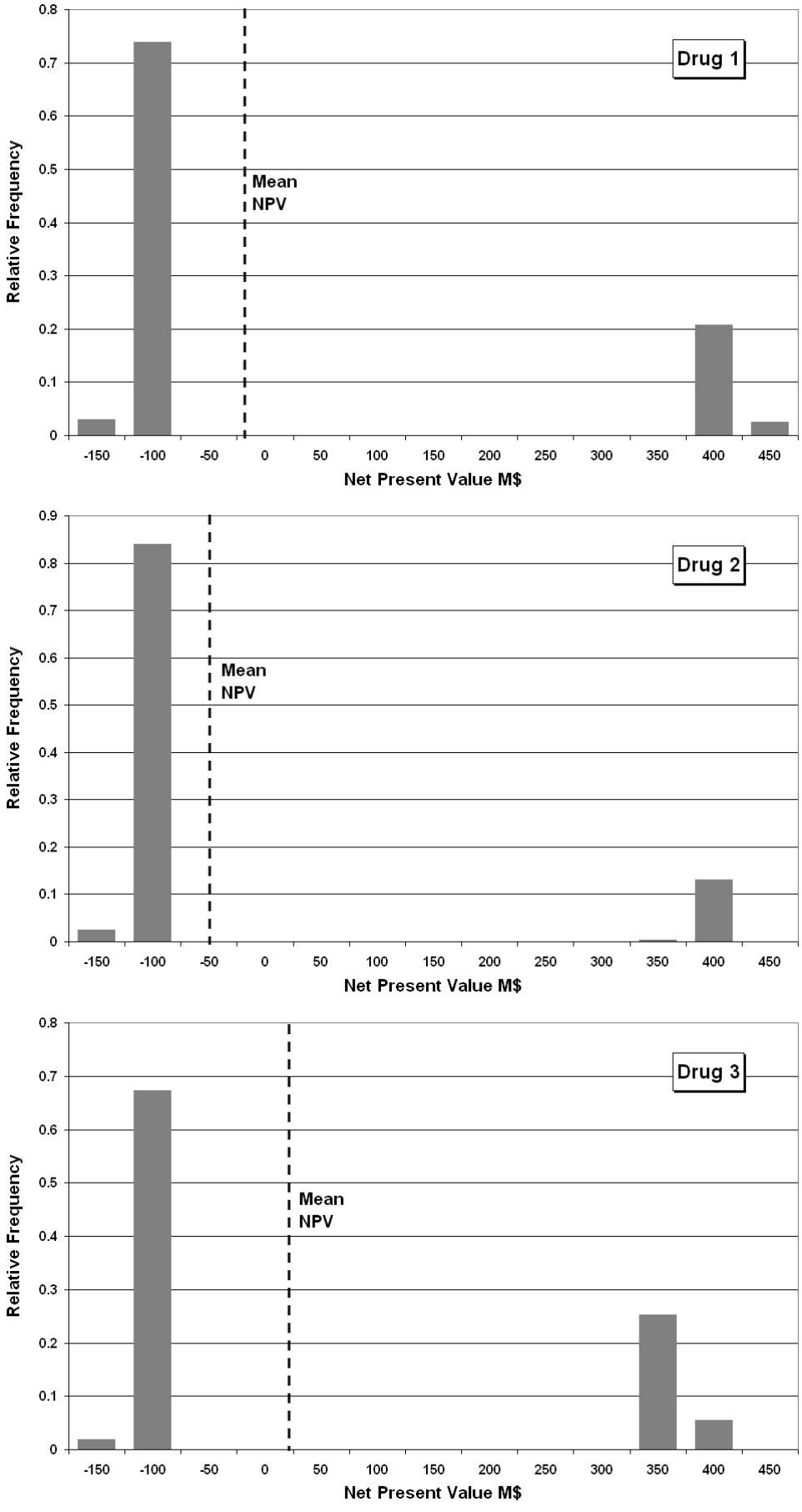


Figure 3.11: Relative frequencies for products-projects 1, 2 and 3

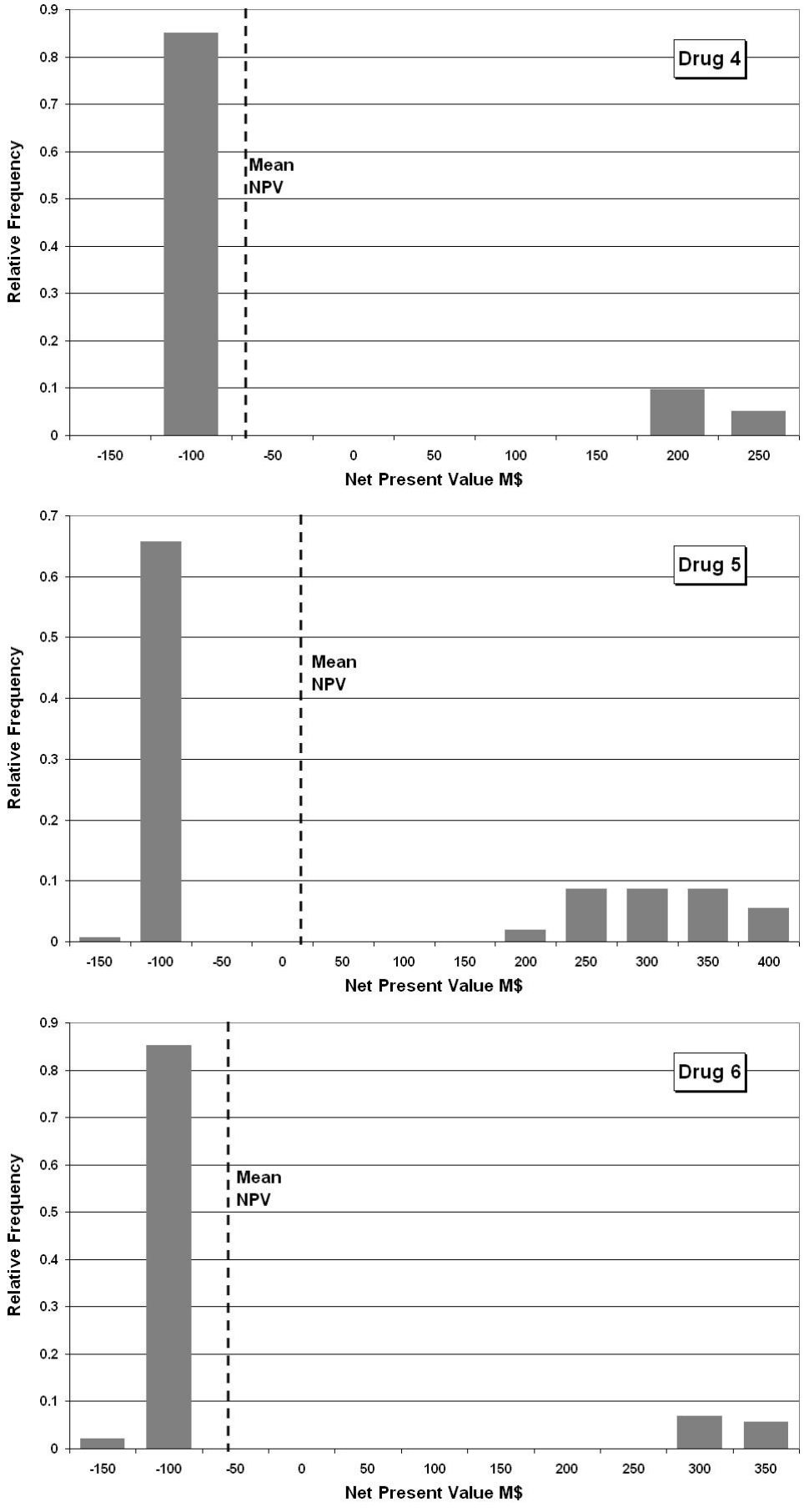


Figure 3.12: Relative frequencies for products-projects 4, 5, and 6.

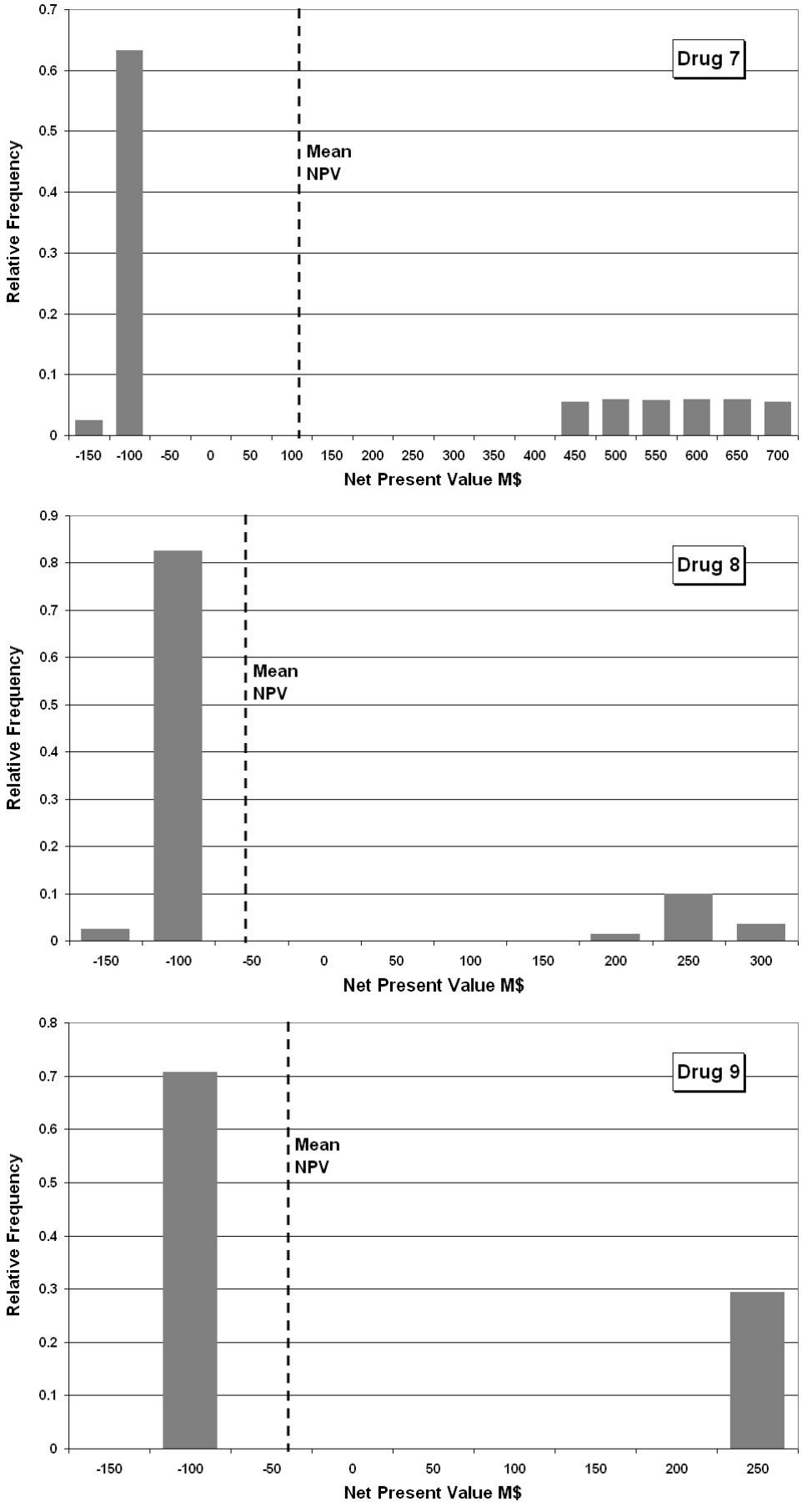


Figure 3.13: Relative frequencies for products-projects 7, 8 and 9.

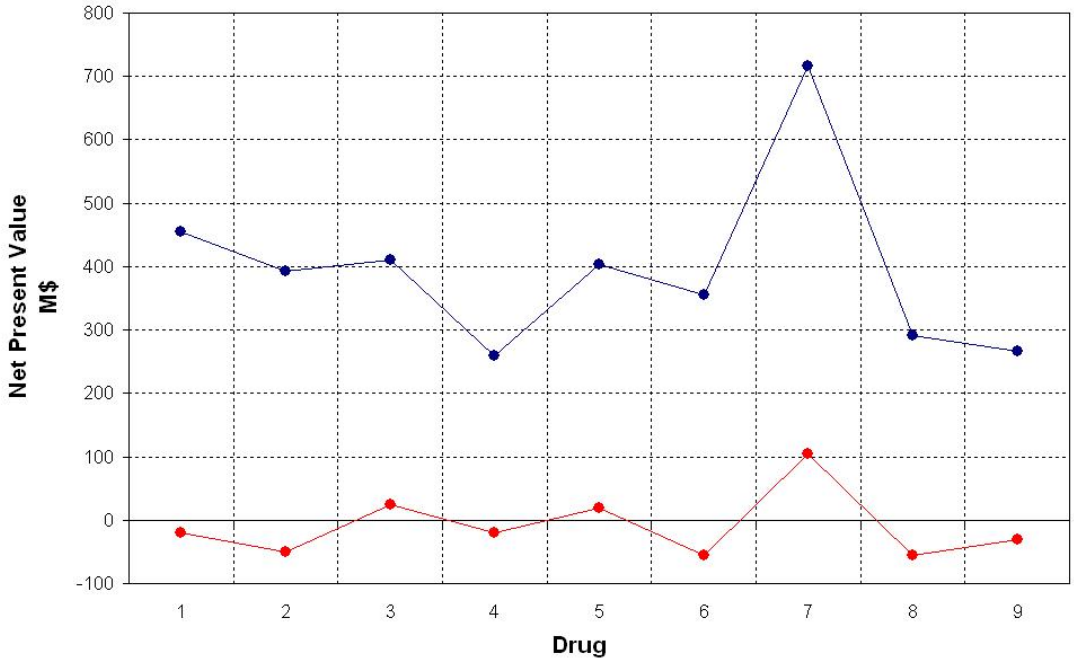


Figure 3.14: Max and Mean values for the NPV by product

Drug	1	2	3	4	5	6	7	8	9
Pos	404	400	358	217	314	323	575	254	250
Neg	102	101	101	100	100	101	102	101	100
A	3.96	3.96	3.54	2.17	3.14	3.20	5.64	2.51	2.50

Table 3.3: Attractiveness for each drug

Drug	1	2	3	4	5	6	7	8	9
Max NPV	455	393	410	259	403	355	717	291	267
Mean NPV	-20	-50	24	-20	20	-55	105	-55	-30
A	3.96	3.96	3.54	2.17	3.14	3.20	5.64	2.51	2.50
Rank	2	5	3	9	4	6	1	8	7

Table 3.4: Drug ranking according to simulation

3.3.2 Bubble chart ranking

The bubble chart is a graphical technique used in pharmaceutical field for ranking projects (see Figure 3.15). Each drug is plotted in a bubble format with capital (investment) cost as diameter, according to its success probability (x-axis) and attractiveness A (y-axis). Obviously, the best solutions are located onto the upper-right side of the bubble chart, and must have a diameter as small as possible.

The same example as in the previous section is treated here. The ratio A is reported in the last line of Table 3.3. Assuming independent success probabilities for Phases I, II and III, the overall success probability is the product of the three previous ones (see Table 3.5). For defining the bubble

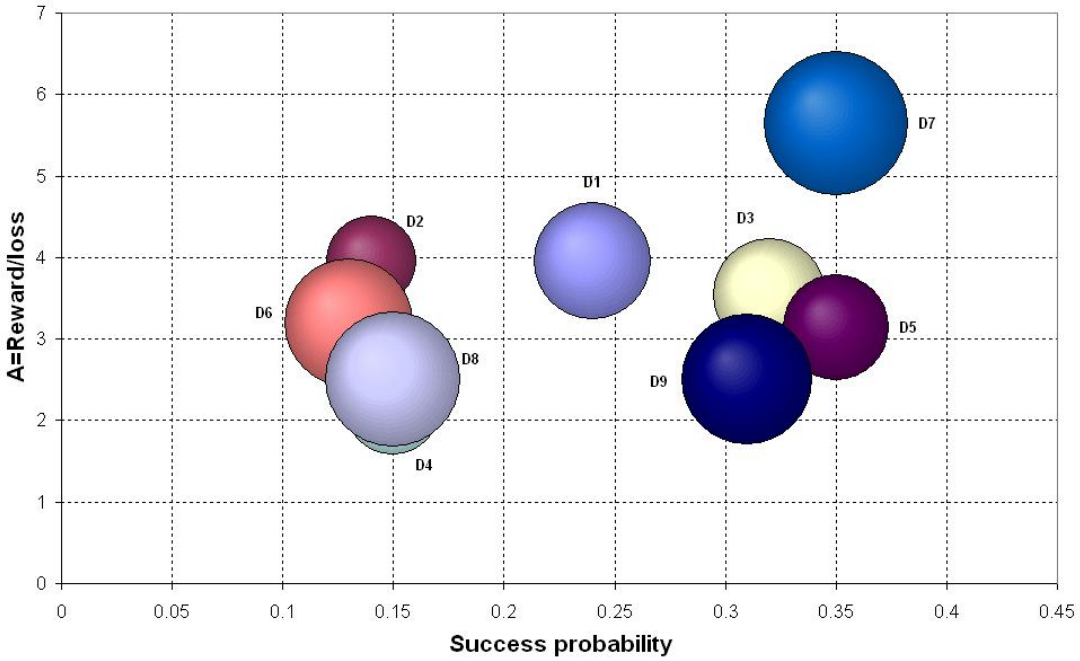


Figure 3.15: Bubble chart for all product- projects

Product	Phase 1	Phase 2	Phase 3	Overall probability
1	0.9	0.3	0.9	0.24
2	0.85	0.2	0.85	0.14
3	0.95	0.35	0.95	0.32
4	0.87	0.22	0.8	0.15
5	0.97	0.36	0.99	0.35
6	0.83	0.18	0.86	0.13
7	0.94	0.4	0.94	0.35
8	0.86	0.2	0.88	0.15
9	0.98	0.34	0.92	0.31

Table 3.5: Success probability by product

size, the middle value of the capital cost is used in this case according to Blau et al. [2004]. The capital cost, computed following Blau et al. [2004], related to each drug is reported in Table 3.6.

Clearly, the bubble chart shows two groups of solutions: the good ones constituted by drugs 3, 5, 7 and 9, the bad ones involving drugs 2, 4, 6 and 8 and an intermediate one drug 1.

Except for rank one, where it is difficult to conclude among drugs 7, 5 and 3 since they have practically the same success probabilities, drug 7 is better regarding ratio A, but drug 5 and 3 are better when considering their capital cost. Finally, we have chosen to rank drug 7 in first position followed by drug 3. The drug ranking is given in Table 3.7, where it can be highlighted that drug 7 is ranked first for the two types of ranking (simulation and bubble chart), followed by the group 3 and 5, the pair 8 and 4 being the worst solutions. So, the two types of ranking procedures give the same trends.

Drug	1	2	3	4	5	6	7	8	9
Capital cost (M$)	50	30	45	34	40	60	75	65	62

Table 3.6: Capital cost for each drug

Drug	1	2	3	4	5	6	7	8	9
Rank	5	6	2	8	3	7	1	9	4

Table 3.7: Drugs ranked according to the bubble chart

3.3.3 Weighted attractiveness

In the two previous sections, attractiveness is used as a criterion for ranking solutions. However, this ratio does not take into account the frequencies of positive and negative values of NPV. That is why the weighted attractiveness WA defined by (see Table 3.8) can be used:

$$WA = [(Freq. > 0\ values) * (Mean > 0\ NPV)]/[(Freq. < 0\ values) * (Mean < 0\ NPV)]$$

The ranking according to simulation (established by performing a trade-off between max NPV, mean NPV and ratio WA) is reported on the last line of Table 3.9. The best solution is drug 7, followed by drugs 3, 1, 5; drug 4 is the worst solution. The rankings with ratios A and WA only differ for drugs 1 and 3 which are permuted.

Drug	1		2		3		4		5	
Pos-Freq	404	0.24	400	0.13	358	0.30	217	0.16	314	0.32
Neg-Freq	102	0.76	101	0.87	101	0.70	100	0.84	100	0.68
WA	1.24		0.59		1.52		0.41		1.48	
Drug	6		7		8		9			
Pos-Freq	323	0.13	575	0.33	254	0.16	250	0.30		
Neg-Freq	101	0.87	102	0.67	101	0.84	100	0.70		
WA	0.48		2.79		0.48		1.07			

Table 3.8: Weighted attractiveness for each drug

The new bubble chart is presented in Figure 3.16 with ratio WA instead of ratio A. For the best solutions 7, 3, 5 and 9, the two charts are globally the same. Chart 3.16 exhibit more concentrated worst solutions than in the previous case. The ranking reported in Table 3.7 is still valid when ratio WA is used.

Drug	1	2	3	4	5	6	7	8	9
Max NPV	455	393	410	259	403	355	717	291	267
Mean NPV	-20	-50	24	-20	20	-55	105	-55	-30
WA	1.24	0.59	1.52	0.41	1.48	0.48	2.79	0.48	1.07
Rank	3	5	2	9	4	6	1	8	7

Table 3.9: Drug ranking according to simulation and weighted attractiveness

In conclusion, it can be emphasized that even if the weighted attractiveness ratio gives more accurate information, ratio A and WA lead to the same conclusions concerning the drug ranking of the considered example both by simulation and bubble chart.

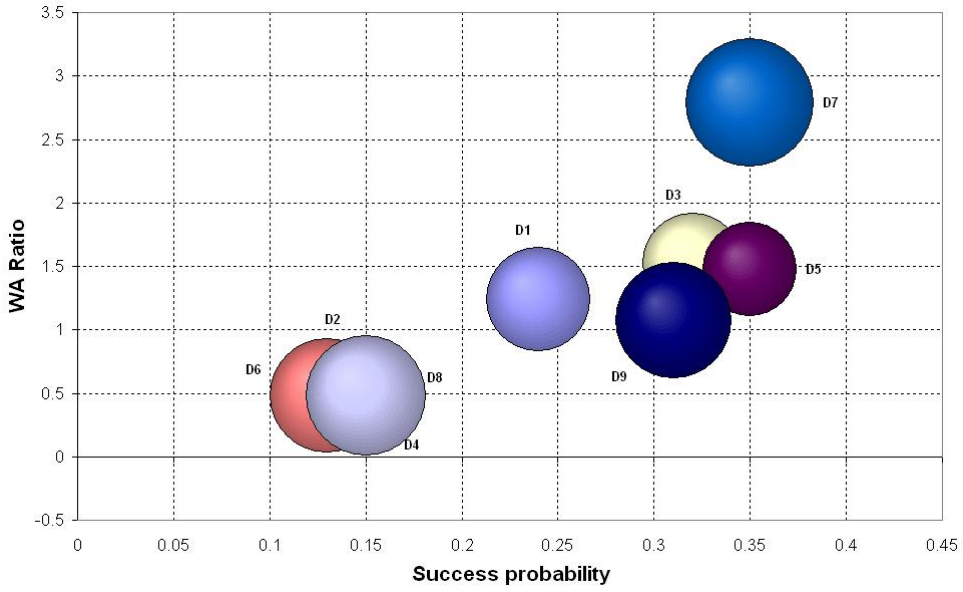


Figure 3.16: Bubble chart according to weighted attractiveness

3.3.4 Sequence simulation

Ranking with positivity probability

Contrary to the two previous cases where only one drug was released into the pipe, this section is related to sequence simulation, where a sequence is defined by a set of drugs launched in a given order into the pipe. The ten sequences of Table 3.10 were randomly defined. For each row of this table, every number is the launching rank; a zero means that the drug is not launched. Let us note that sequence 3 involves drug 7, the best one according to the previous studies in "pole position". In sequence 4, the two best solutions 7 and 3 are in positions 1 and 2. As in the two previous studies, each simulation is repeated 300 times, selecting random sampling values from a uniform random number generator for defining successes or failures.

Sequence	P1	P2	P3	P4	P5	P6	P7	P8	P9
1	6	0	4	7	1	0	5	3	2
2	3	9	4	7	1	8	5	6	2
3	5	2	7	8	3	6	1	4	9
4	4	0	2	0	6	5	1	0	3
5	4	1	0	0	0	2	3	5	6
6	4	2	1	3	5	0	0	0	0
7	5	0	4	0	2	3	6	1	0
8	2	3	0	7	1	6	0	5	4
9	1	0	0	0	4	5	0	3	2
10	1	0	3	0	4	0	6	2	5

Table 3.10: Sequences to consider

For each sequence, frequencies by drug and by phases of the pipe are displayed in Figures 3.17 to 3.20. The higher failure frequencies are obtained for stage 2. stages 1 and 3 have frequencies located in the same range of values. This is due to the success probability values which are quite the same for Phase I and III, while the success probability of Phase II is much lower (see Table 3.5).

The mean for the NPV related to each sequence is plotted in Figure 3.21. Another measure of performance commonly used is constituted by the positivity probability defined by the number of times the NPV was positive divided by the total number of runs; it is reported in Table 3.11. The ranking of sequences (see Table 3.12) is established by performing a trade-off between mean NPV and positivity probability. The best solutions are sequences 4, 7 and 10, and the worst are sequences 2 and 9.

Sequence	1	2	3	4	5	6	7	8	9	10
PP	0.40	0.35	0.38	0.42	0.33	0.35	0.36	0.35	0.38	0.40

Table 3.11: Positivity Probability (PP)

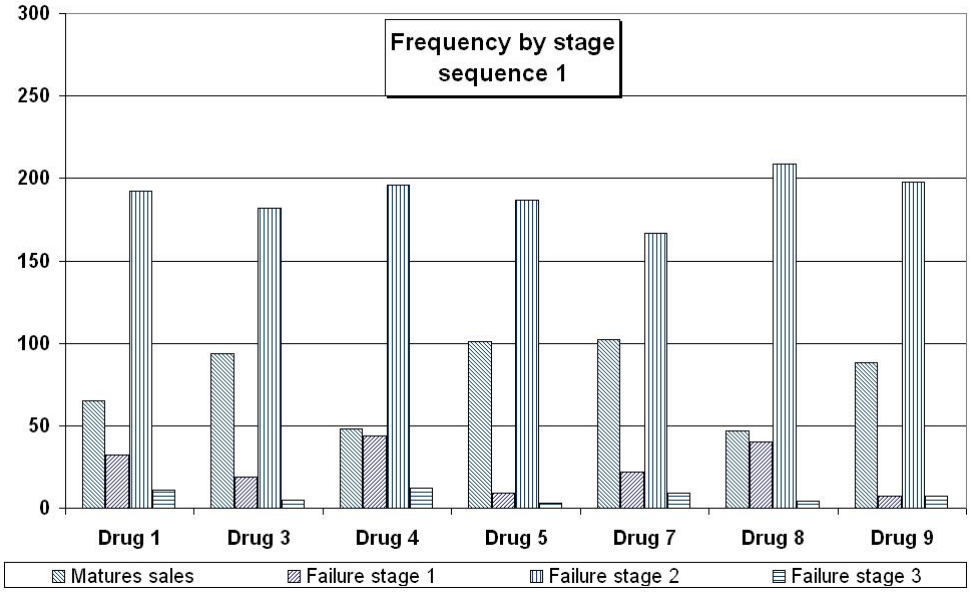
Sequence	1	2	3	4	5	6	7	8	9	10
Rank	5	10	7	1	8	4	2	6	9	3

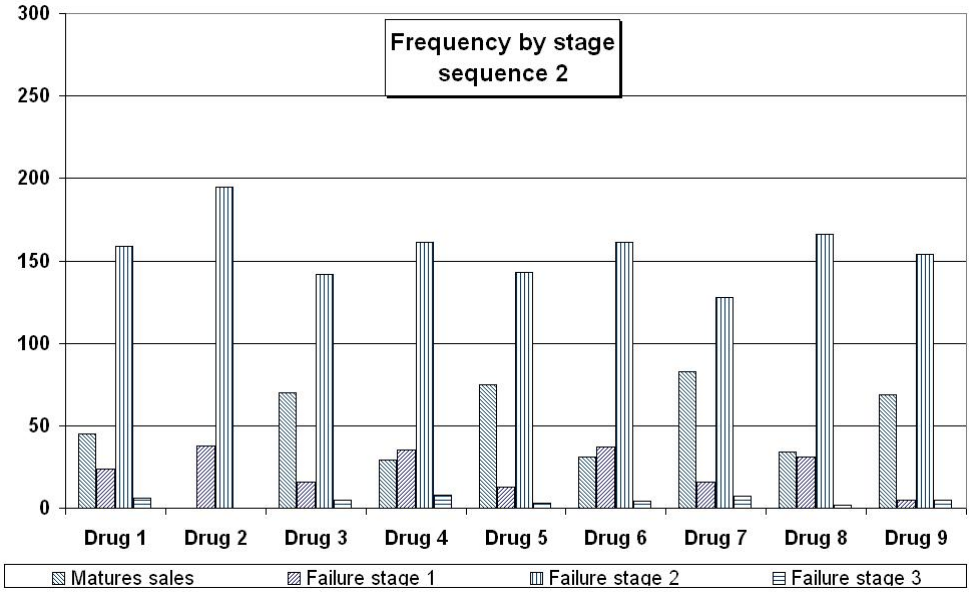
Table 3.12: Ranking of the ten sequences

The best solution (sequence 4) involves drugs 7 and 3 in the first and second positions. These solutions are now identified as the best ones, when drugs were studied individually. On the other hand, sequence 7 where drug 8, one of the worst solution identified in the previous studies is in the pole position, in second rank. So it seems difficult to correlate the sequence performances with the ones of each drug. To refine this conclusion, two other sequences (7, 3, 1, 5) and (6, 9, 8, 4) corresponding to the best and worst results from individual studies were simulated. For sequence 11 (7, 3, 1, 5), the mean NPV and the positivity probability are respectively 569.58 M$ and 0.62; the results for sequence 12 (6, 9, 8, 4) are -121.19 M$ and 0.23. The ranking of the 12 sequences is reported in Table 3.13. Concerning the positivity probability, sequence 11 gives the best result and sequence 12 the worst one. From the NPV, sequence 12 arrives in last position, while sequence 11 has the fourth rank. With regard to the two items, sequence 11 can be classified in third position with sequence 10, and sequence 12 in the last position. As a conclusion, it can be emphasized that when a sequence involves drugs with good (respectively bad) individual performances, its ranking is good (respectively bad).

Sequence	1	2	3	4	5	6	7	8	9	10	11	12
Rank	6	11	8	2	9	5	3	7	10	4	1	12

Table 3.13: Sequence ranking





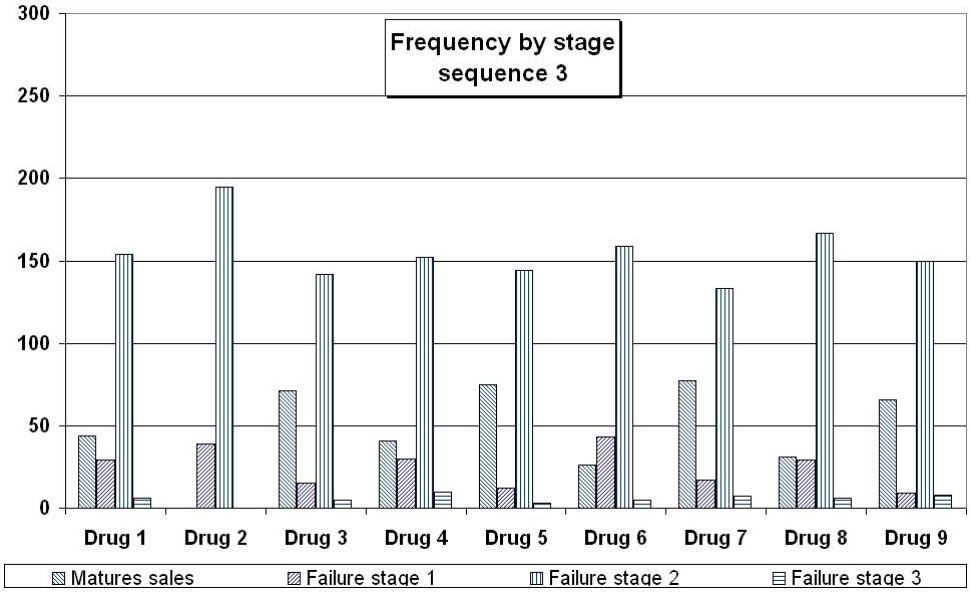
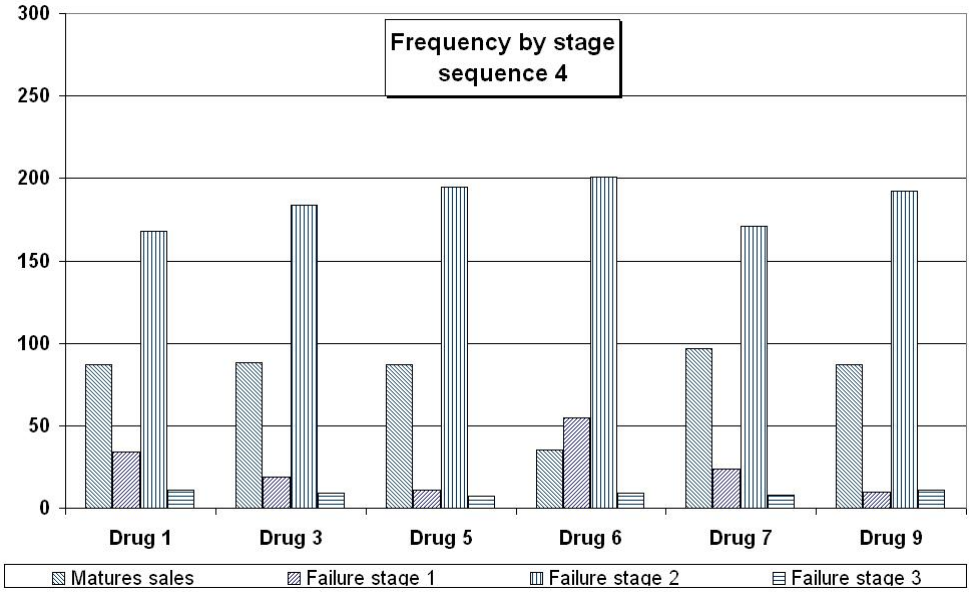
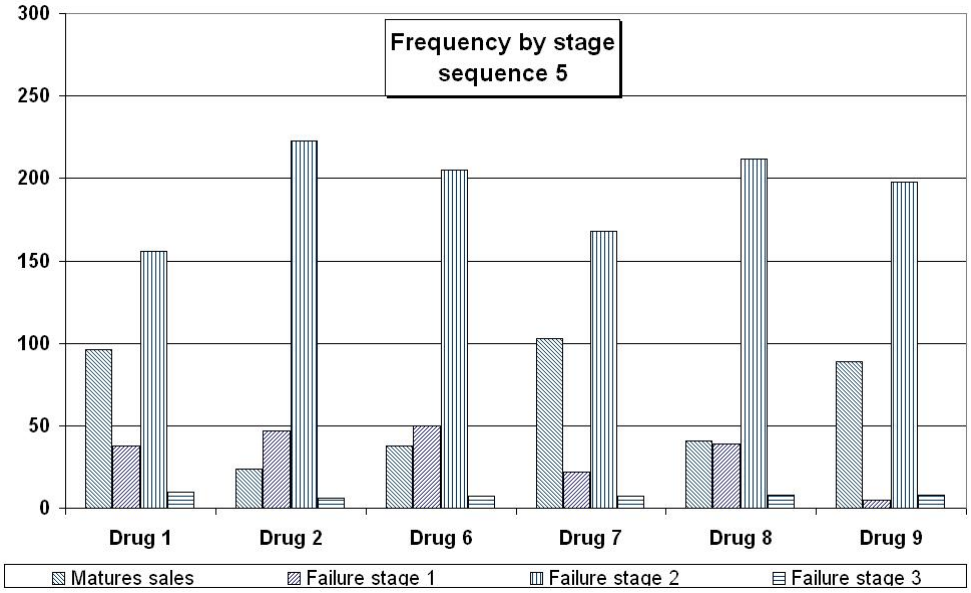


Figure 3.17: Relative frequencies for sequences 1, 2 and 3.





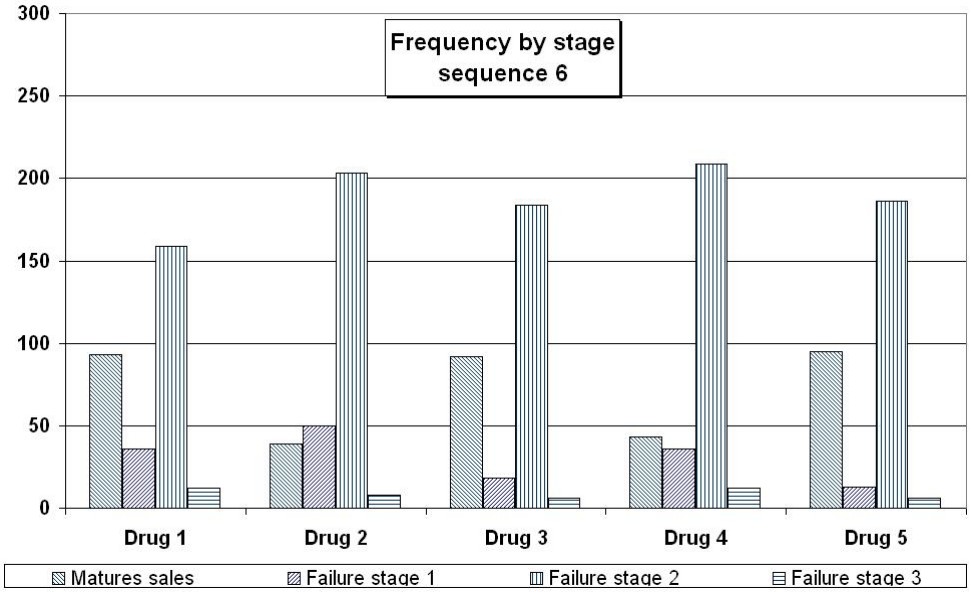
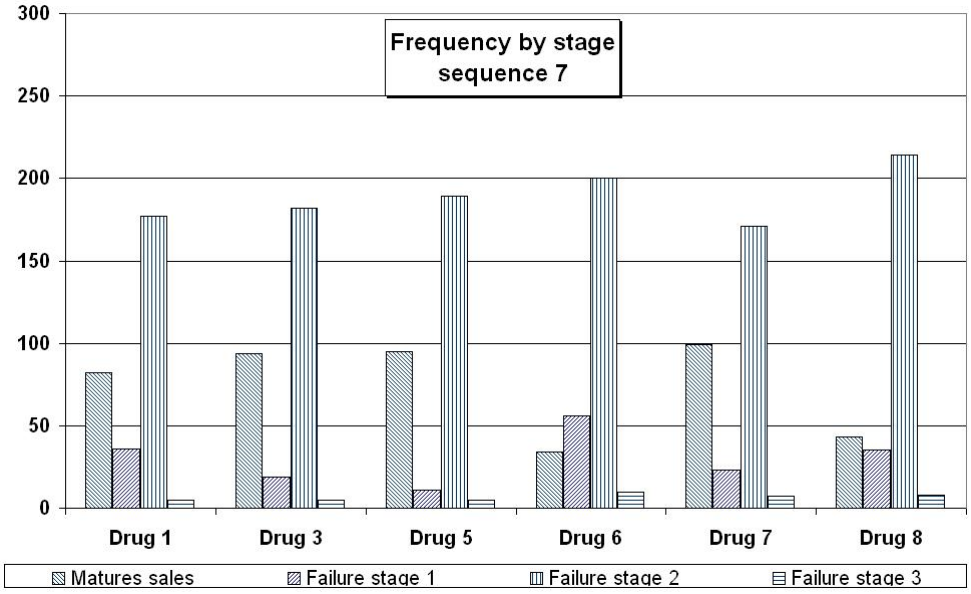
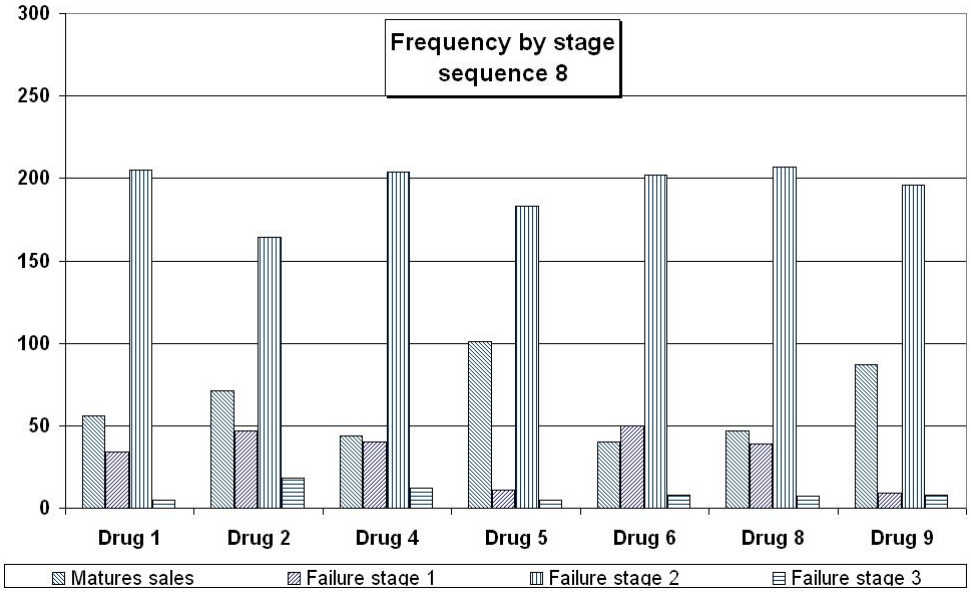


Figure 3.18: Relative frequencies for sequences 4, 5 and 6.





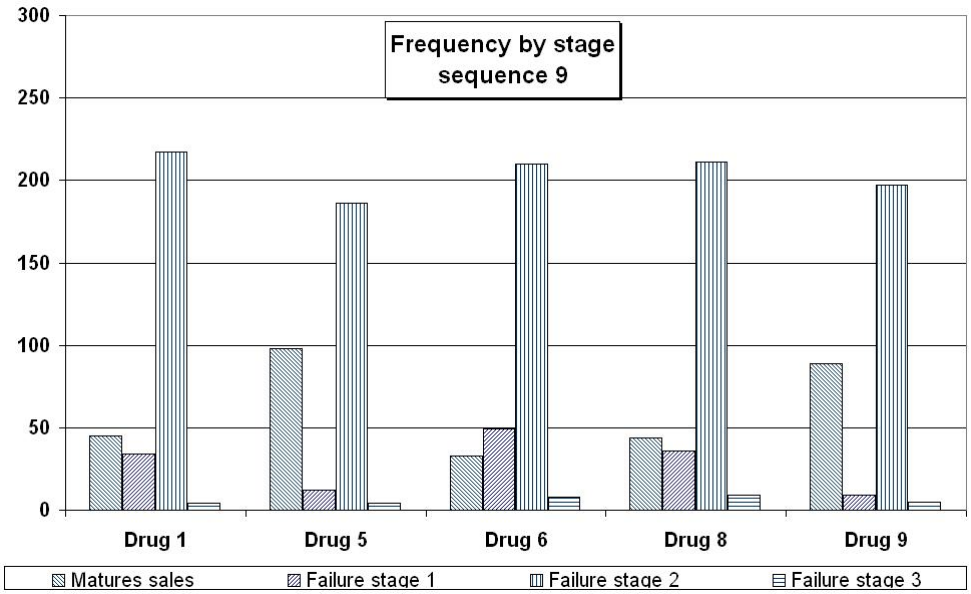


Figure 3.19: Frequency for sequences 7, 8 and 9

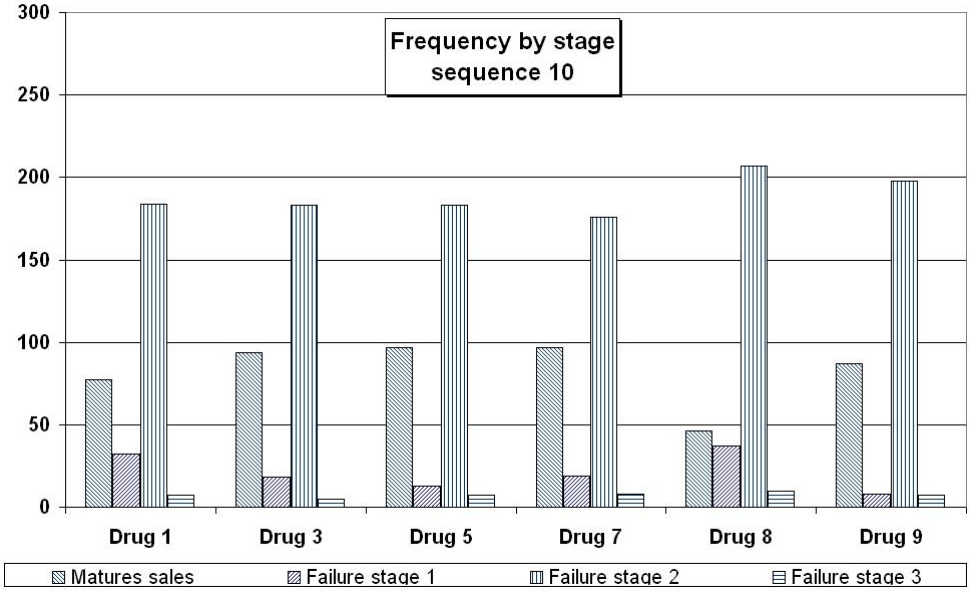


Figure 3.20: Frequency for sequence 10

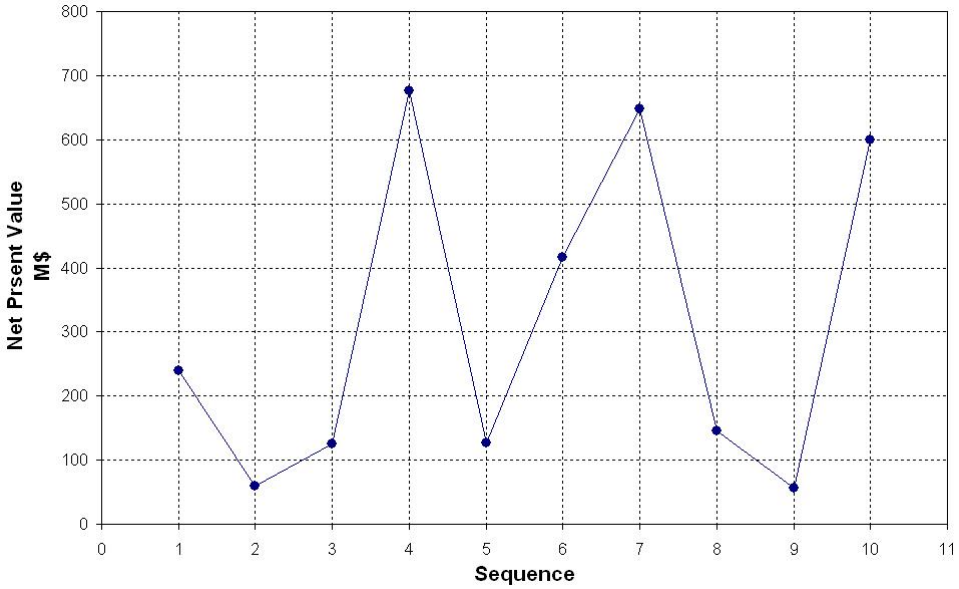


Figure 3.21: Mean values for the NPV by sequence.

Ranking with Weighted Attractiveness

As in the previous case, the weighted attractiveness (WA) (see Table 3.14) is used instead of positivity probability (PP) for ranking sequences. Only three sequences according to the previous ranking are studied: the best one (4), a good one (11) and the worst one (12). The ranking of sequences is now established by performing a trade-off between mean NPV and WA. Obviously, the ranking remains the same.

Sequence	4		11		12	
Pos-Freq	2046	126	1398	186	978	69
Neg-Freq	301	174	862	114	449	231
WA	4.92		2.65		0.65	

Table 3.14: Weighted attractiveness for each sequence

3.3.5 Conclusion

On this example, the DES gives correct trends, but a strict comparison with the solutions obtained by Blau et al. [2004] was not able to be carried out as a result of the lack of some data in this article.

Concerning the release of only one drug into the pipe, the simulation with statistical indicators like mean NPV and attractiveness gives the same results as the bubble chart ranking method either with attractiveness or weighted attractiveness.

When sequences of drugs are launched into the process, the mean NPV can be considerably increased compared to the one obtained for only one drug (see Figures 3.14 and 3.21). However, the NPV may vary strongly from a sequence to another (Figure 3.22). Sequences are ranked according to two pairs of indicators (mean NPV PP) and (mean NPV WA): the obtained results are identical. Sequences with all the drugs launched (2 and 3) give bad results, and the best ones are obtained for sequences 4, 7, 10 and 11 involving respectively 6, 6, 6, and 4 drugs. Furthermore, when a sequence involves drugs with good (respectively bad) individual performances, its ranking is good (respectively bad).

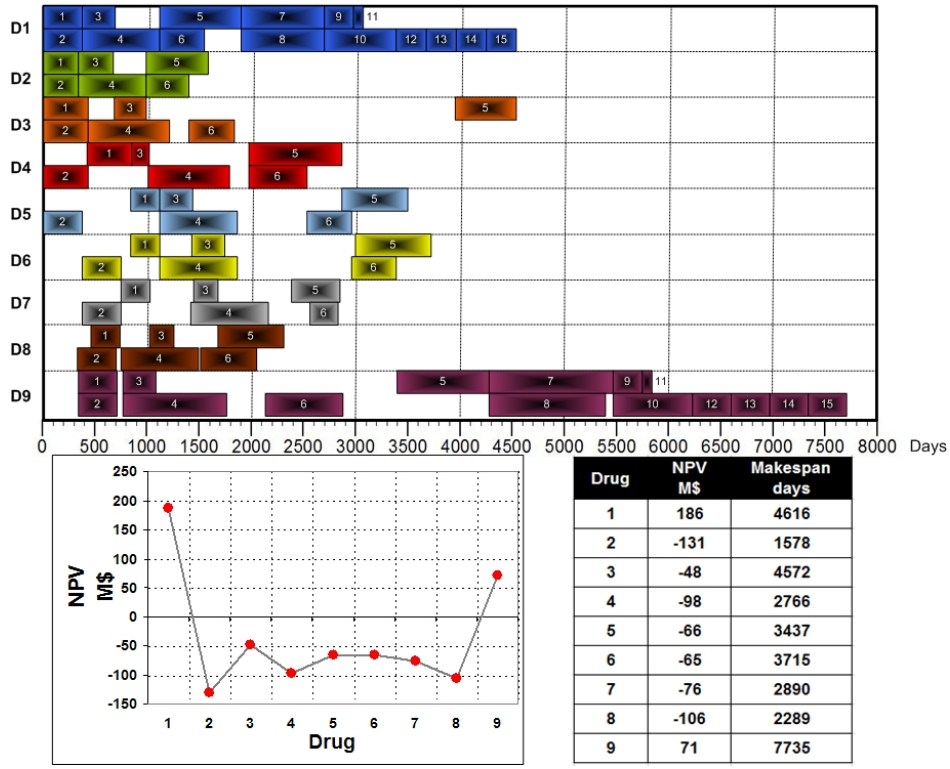
Taking into account the combinatorial nature of the problem, all the sequences cannot be exhaustively evaluated. As it is shown in Chapter 5, there are 951,744 potential solutions for the problem under consideration. If each sequence is simulated 300 times, requiring a mean CPU time of 10 seconds, the total CPU time for an exhaustive enumeration would be 2700 days! So, the only way to obtain optimal or quasi-optimal solutions is to implement efficient scanning algorithms, as it is presented in Chapter 5.

3.4 Resource capacity management

For analyzing the impact of changing resource capacities at each stage of the system, two cases are now considered. First, the case which will be considered throughout this work unless other conditions are explicitly mentioned considers a limited capacity by stage (see section 2.2.2): resource is reallocated only once a drug gets out from the stage; this study has been carried out in the previous section. In the new case, there is no limited capacity for stages (unlimited capacity). As in the previous cases, the simulations are repeated 300 times.

First the sequence (1, 2, 3, 4, 5, 6, 7, 8, 9) is launched in that order into the pipe. For the sake of illustration, a particular sequence was selected amongst the 300 sequences simulated. The results are ported in Figure 3.22. Only two drugs (1 and 9) completed all the processes with positive NPV. In the Gantt chart located on the top, some spaces between tasks represent waiting times for an available resource before going to the next step. Drug 9 being launched in the last position, its waiting time is the most important because some processing steps are occupied by drugs launched before, so its makespan is the highest one. The NPV of this solution is M$ -335.25, with a mean makespan of 3733 days.

The same sequence was studied again without limitation on resource capacity (see Figure 3.23). In that case, drugs do not wait for a resource considering its capacity. As shown on the Gantt chart, four drugs (1, 5, 7 and 9) have now positive NPV values. Because of unlimited resources, the makespan globally decreases. The new NPV of this solution is M$ 113.57, with a mean makespan of 3100 days. Obviously, performances increase when constraints are relaxed.

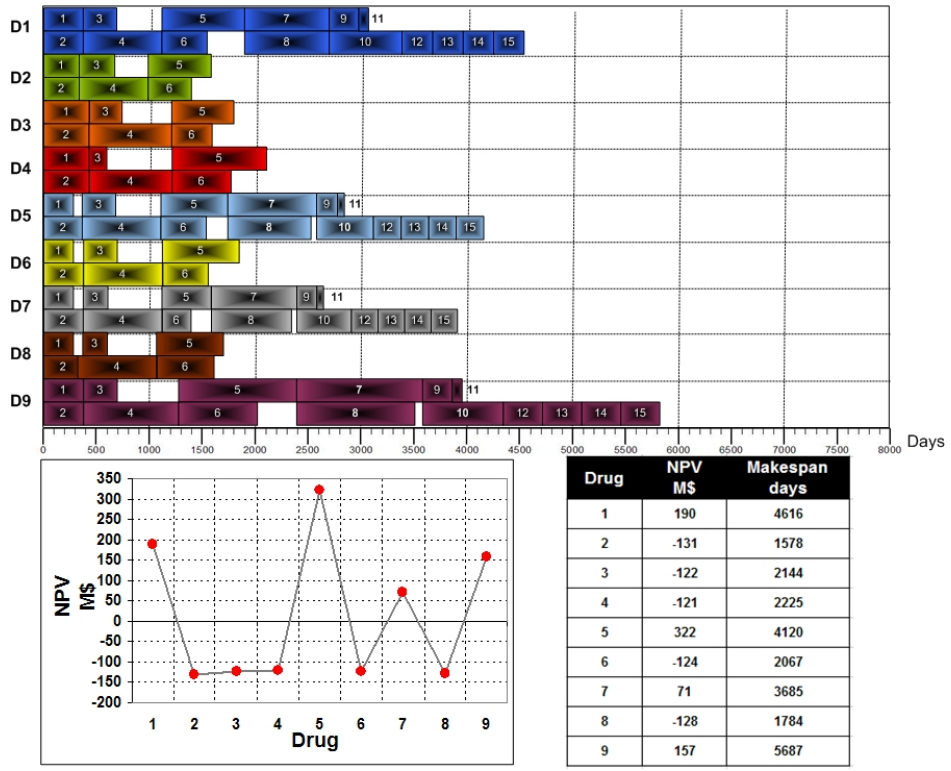


Drug	NPV M$	Makespan days
1	186	4616
2	-131	1578
3	-48	4572
4	-98	2766
5	-66	3437
6	-65	3715
7	-76	2890
8	-106	2289
9	71	7735

Figure 3.22: Simulation with capacity constraints

A statistical analysis is now performed on sequences 4, 11 and 12 defined in the previous subsection (the best one, a good solution and the worst one), and sequence 13 defined by drugs 1, 2, 3, 4, 5, 6, 7, 8, 9 launched in this order. The results obtained from 300 simulations with and without limitations on capacities are respectively reported in Tables 3.15 and 3.16, where for each sequence:

- Mean NPVpos in the mean value of positive NPV;
- Npos is the number of positive NPV obtained;
- Mean NPVneg in the mean value of negative NPV;
- Nneg is the number of negative NPV obtained;



Drug	NPV M$	Makespan days
1	190	4616
2	-131	1578
3	-122	2144
4	-121	2225
5	322	4120
6	-124	2067
7	71	3685
8	-128	1784
9	157	5687

Figure 3.23: Simulation without capacity constraints

Sequence	$MeanNPV_{pos}$	N_{pos}	$MeanNPV_{neg}$	N_{neg}	*MeanNPV*	*PP*
4	2046	126	-301	174	685	0.42
11	1398	186	-862	114	539	0.62
12	978	69	-449	231	-121	0.23
13	1040	101	-606	199	-52	0.33

Table 3.15: Results with capacity limitation

- Mean NPV is the global mean, computed for all negative and positive values of NPV;
- PP is the positivity probability defined in sub-section 3.3.4.

It can be observed in Tables 3.15 and 3.16 that when the system becomes unconstrained, positive terms (NPV and frequencies) and negative NPV globally increase while numbers of negative NPV decrease. The mean NPV and the positivity probability increase too, except for the last case concerning the positivity probability. With regard to the variations of all these terms, it can be emphasized that limitations on system capacity plays an important role on its performances. However, the unlimited capacity does not change sequence ranking.

Sequence	$MeanNPV_{pos}$	N_{pos}	$MeanNPV_{neg}$	N_{neg}	Mean NPV	PP
4	2120	144	-280	156	872	0.48
11	1512	216	-840	84	853	0.72
12	1048	105	-422	195	92	0.35
13	1240	87	-483	213	17	0.29

Table 3.16: Results without capacity limitation

3.5 Simulator Validation. Application to Rajapakse et al. [2006] problem

3.5.1 Problem description

The case study is an adaptation of the example presented in Rajapakse et al. [2006]. It involves a portfolio of six potential monoclonal antibodies (MAbs) that are ready for clinical development. Monoclonal antibodies are nature's biological warheads, able to target and help eliminate foreign or abnormal agents from the body.

The goal is to use again the discrete event simulation to model the series of decision points along the drug development pathway. In this new example, at the end of each phase of clinical trials, the probability of clinical success results in go/no-go decisions. Further decision points include those made at each manufacturing stage which could be carried out in-house or by a contract manufacturer (CMO, contract manufacturer organization); this was determined as a function of the available capacity at a particular time. Figure 3.24 shows the main decisions and the options available for a drug process development, established from the work of Rajapakse et al. [2006]. Even if the CMO option is considered, the model is simpler than in the example of Blau et al. [2004]. The main activities, Phase I, Phase II and Phase III, are involved but a delay and a failure due to technical reasons are added. So, before stage Phase I, a delay in material deliverance is considered, followed by the clinic trials for finalizing with a failure probability or transition between Phase I and Phase II for the outcomes from the Phase I trials. These activities are repeated for the Phase II but after the transition, a failure probability is considered. This failure is due to technical reasons and is considered only in this part of the system. For Phase II, a delay in material deliverance, Phase II trials and a transition probability are considered. Another failure probability is taken into account; it is related to the fact that a drug must be approved before arriving to the market by the related department of public health. A set of potential products are given, each of which must undergo a set of testing tasks. Each task has an associated duration, cost, and probability of success. Given the income for each product as a function of the time of product introduction, the problem is to model tasks scheduling while computing the traditional Net Present Value (NPV) economic criterion.

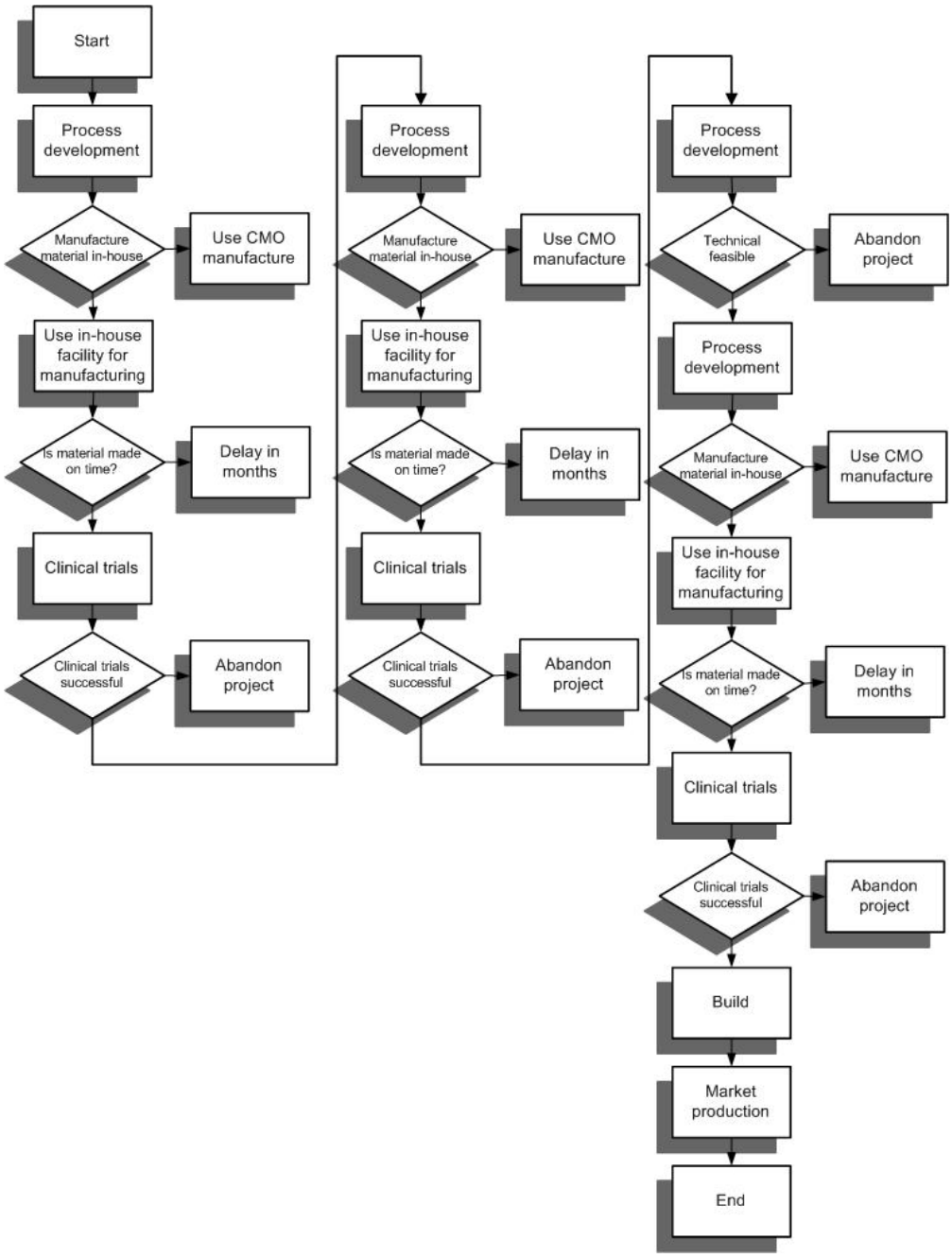


Figure 3.24: General diagram of the drug development process.

3.5.2 Project management by Discrete event simulation

For implementing the DES, the relationships between the classes and system activities are shown in Figure 3.25. Table 3.17 presents general information about the six drugs in the portfolio that are ready for clinical development.

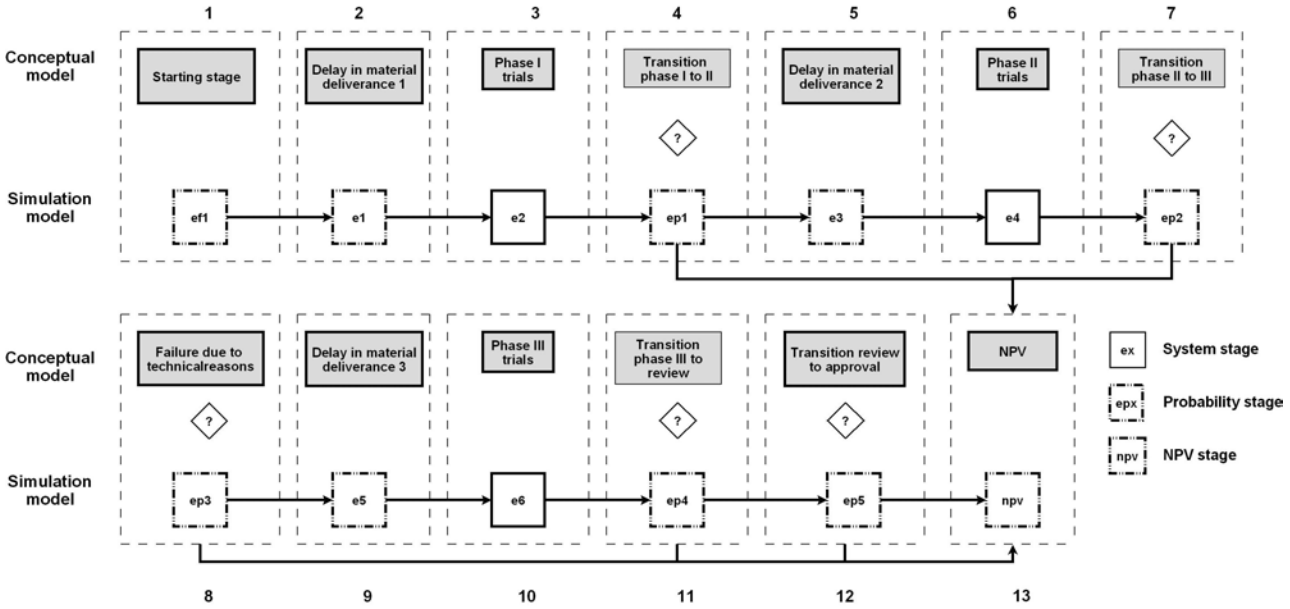


Figure 3.25: Relationship between classes-system activities.

Drug candidate	MAb type	Market share	Uncertainty
A	Chimeric	High	Medium
B	Humanised	Low	Very low
C	Chimeric	Medium	Medium
D	Murine	Very high	High
E	Chimeric	Low	Medium
F	Chimeric	Low	Low

Table 3.17: General information about drugs in the portfolio

3.5.3 Problem data

Given the finite level of resources available, the simultaneous development of the six products can not be carried out. The interest of the simulation approach is to provide decision-makers with an explicit view of the best product mix in order to satisfy performance indicators. Probabilities were assigned to reflect the level of uncertainty related to the considered drugs (see Table 3.17). They are representative of the kind of information available from experts in the pharmaceutical industry. The probability distributions assigned to key risk factors for drugs relevant with differing levels of uncertainty are given in Table 3.18.

The phase transition probabilities reported in Reichter [2001] from a large data collection and analysis for 186 antibodies entering clinical studies were considered for this case study and are shown in Table 3.19.

The probability of failure due to technical reasons was set again according to the uncertainty level assigned to a particular drug candidate (Table 3.20). Drug B was modeled as a new indication for an existing drug and hence there was no technical uncertainty associated with it.

	High uncertainty drug		Medium uncertainty drug		Low uncertainty drug	
	Value	Probability	Value	Probability	Value	Probability
Ph I Development time (Months)	12	20	9	20	6	10
	18	30	12	50	9	70
	24	50	15	30	12	20
Ph II Manufacturing time (Months)	12	20	9	20	6	10
	18	40	30	30	9	70
	24	30	50	50	12	20
Ph III Clinical trial time (Months)	18	20	12	20	12	10
	24	30	18	50	15	70
	36	50	24	30	18	20
Market manufacturing cost (COG) ($ g-1)	750	10	1125	30	1200	20
	1000	20	1500	50	1600	70
	1250	70	1875	20	2000	10
Product yield (%)	30	40	30	30	30	10
	50	30	50	50	50	20
	70	30	70	20	70	70
Delay in material delivery (Months)	0	20	0	50	0	70
	3	30	3	30	3	20
	6	50	6	20	6	10

Table 3.18: Risk factors and probability distributions for drugs with levels of uncertainty

Monoclonal antibody type	Phase I to II (%)	Phase II to III (%)	Phase III to review (%)	Review to approval (%)
Murine MAbs	77	52	45	33
Chimeric MAbs	86	40	80	100
Humanised MAbs	84	72	75	100

Table 3.19: General information about drugs in the portfolio

Drug candidate	Probability of failure due to technical reasons at Phase III (%)	Market value ($M)
A	20	523
B	0	427
C	10	410
D	30	1029
E	5	31
F	15	49

Table 3.20: General information about drugs in the portfolio

Three scenarios are considered based on level of resources for drug development (see Table 3.21). Scenario 1 considers all six drugs in the portfolio and unlimited levels of resources, thus establishing a base case against which to compare the results from resource-constrained simulations. Next, two resource levels of 500M$ and 750M$ were defined as constraints with a lower number of drugs to be developed.

All the considered sequences are presented in Table 3.22 where 20 sequences are related to three drugs and 15 to four drugs and only one to six products. For each sequence number 1 indicates a drug present in the sequence and 0 indicates a drug absent in the sequence.

Resource level (M$)	Number of drugs per sequence	Number of possible sequences
Unconstrained	6	1
500	3	20
700	4	15

Table 3.21: Resources levels and drugs in the portfolio

	Sequence																	
Product	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
B	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0
C	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1
D	0	1	0	0	1	0	0	1	1	0	1	0	0	1	1	0	1	1
E	0	0	1	0	0	1	0	1	0	1	0	1	1	1	0	1	1	0
F	0	0	0	1	0	0	1	0	1	1	0	0	0	0	1	1	0	1

	Sequence																	
Product	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
A	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1
B	0	0	1	1	1	1	1	1	0	0	0	0	1	1	1	1	0	1
C	1	0	1	1	1	0	0	0	1	1	1	0	1	1	1	0	1	1
D	0	1	1	0	0	1	1	0	1	1	0	1	1	1	0	1	1	1
E	1	1	0	1	0	1	0	1	1	0	1	1	1	0	1	1	1	1
F	1	1	0	0	1	0	1	1	0	1	1	1	0	1	1	1	1	1

Table 3.22: Sequences considered for simulation

3.5.4 Simulation results

As in the previous cases, each sequence simulation is repeated 300 times, where successes or failures at trial phases are determined by generating random values. The economic objective function is the mean NPV computed over the 300 simulations. Besides this criterion, Rajapakse et al. [2006] define two other indicators: the former Risk1 is the frequency of negative NPV obtained by simulating 300 times each sequence (it is the same type of indicator than the positivity probability defined at section 3.3), and the latter Risk2 is the standard deviation of the NPV. The results obtained for the 36 sequences defined in Table 3.22 are displayed in Figure 3.26 (respectively 3.27) for the pair of objectives (NPV-Risk1) (respectively, NPV-Risk2).

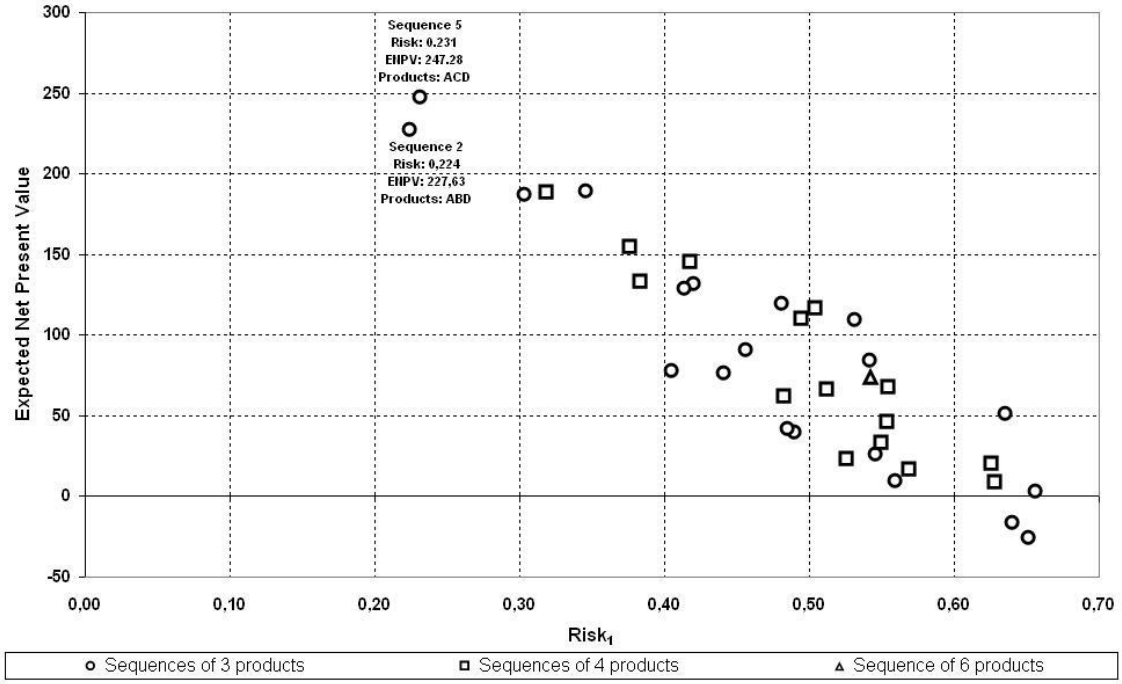


Figure 3.26: Simulation outcomes for ENVP Risk1.

Sequence	Drug	ENVP	$Risque_1$
5	A C D	247.28	0.231
2	A B D	227.63	0.224
36	All products	74.29	0.542

Table 3.23: Sequences with ENPV and Risk1

For objectives (NPV-Risk1), two sequences of three products are particularly attractive for the decision maker: sequence 5 (A, C, D) and sequence 2 (A, B, D). Furthermore, Figure 3.28 shows that diversifying the portfolio is not recommended for generating profits: sequences with four, five and six drugs exhibit globally low performances. For illustration, NPV and Risk1 are reported in Table 3.23 for the two best sequences (2 and 5) and one worst, sequence 36, with six products. This behavior can be explained by the high number of product failures at trial phases, which is directly linked to the number of products reaching the commercialization phase.

For a given risk (Risk1 or Risk2), Rajapakse et al. [2006] define the so-called 'Efficient Frontier', which is a set of preferred solutions. The Efficient Frontier is made up by sequences located on the left frontier of the set of points of the graphs NPV vs. Risk. This is a set of solutions with minimal risk. For Risk1, the two first elements of the Efficient frontier are sequences 5 and 2. For Risk1, sequence 5 has the best NPV but a higher risk than sequence 2.

The interpretation of Figure 3.27 is less easy. In the studied example, the Efficient Frontier contains four sequences, they are listed in Table 3.24.

In Rajapakse et al. [2006] work, solutions belonging to the Efficient Frontier for Risk2 are 1, 2, 12 and 13 (see Table 3.25). Sequence 2 has the highest NPV but the highest Risk1 too. On the other hand, sequences 12 and 13 present the lowest NPV and Risk1. The sequence 1 shows intermediate values for NPV and risks. All these sequences are constituted of three drugs.

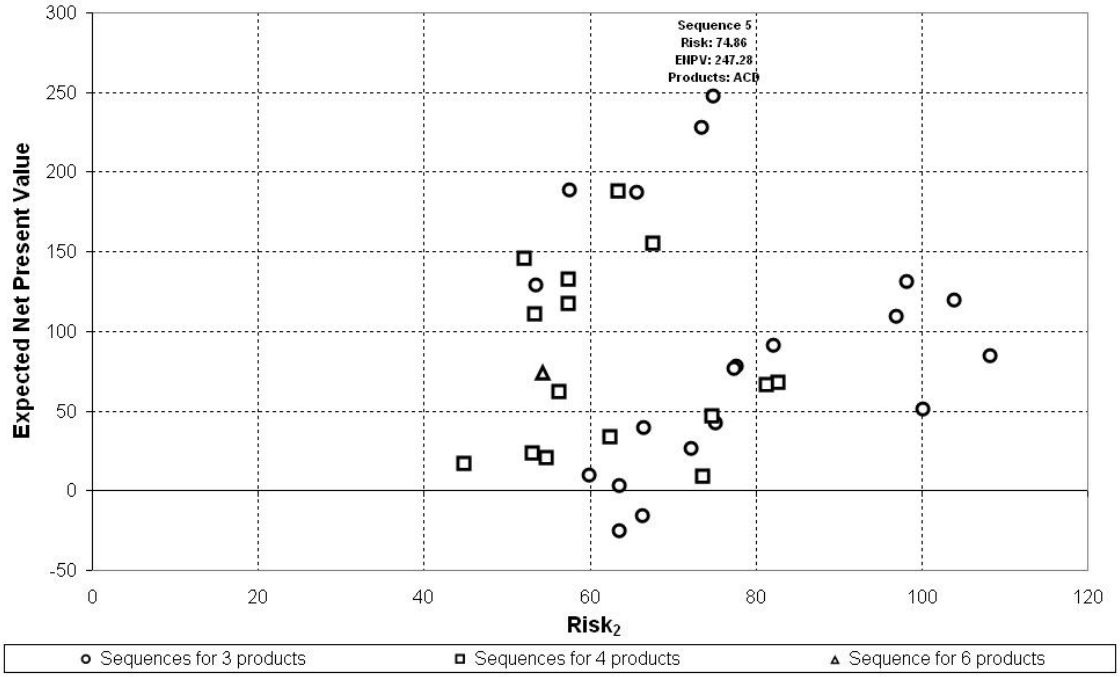


Figure 3.27: Simulation outcomes for ENVP Risk2.

Sequence	Drug				NPV	$Risk_2$
5	A	C	D	-	247.28	74.86
8	A	D	E	-	188.96	57.57
27	A	C	D	E	145.43	52.13
29	A	C	E	F	16.58	44.9
36	All products				74.29	54.28

Table 3.24: Sequences on Efficient Frontier

For results from the DES developed in this work for the same system, the solutions on the Efficient Frontier are sequences 5 and 2 (Table 3.23) for Risk1, with 3 drugs per sequence. Considering Risk 2 (Table 3.24), sequences containing three and four drugs are located on the Efficient Frontier. Sequences with three drugs are riskier but with a better NPV. The sequence with six drugs is one of the worst taking into account either Risk1 or Risk2. The same conclusions were obtained in Rajapakse et al. [2006] work.

Results from Rajapakse et al. [2006] and from this work are constituted mainly of three antibodies. Difference in drugs between solution sequences is due to uncertainty in values considered for each antibody. For example, for drug D with the best value for sales (Table 3.20), uncertainty for its development is considered as high (Table 3.17), while for drug B uncertainty is defined as very low, but sales are at least two times lower that drug B.

According to Rajapakse et al. [2006], Risk2 for a sequence is the standard deviation of NPV computed over the 300 simulations. However, as indicated in Section 3.3, the population being bimodal, if the global mean value does not correspond to existing sequences, it can be nevertheless used as a statistical indicator. A statistical interpretation of the global standard deviation is more difficult to carry out. The global standard deviation always takes very large values because it takes into account all the positive and negative terms, whose range is important. So, conclusions obtained

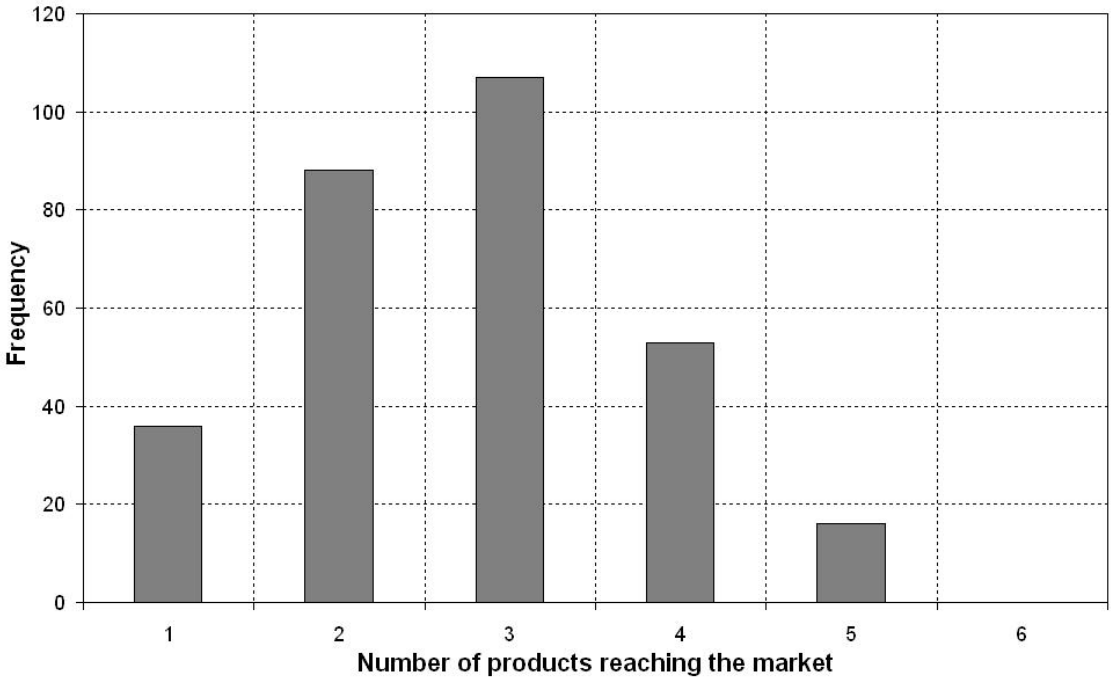


Figure 3.28: Number of products reaching the market.

Sequence	Antibodies
1	ABC
2	ABD
12	BCE
13	BCF

Table 3.25: Sequences on Efficient Frontier for Risk2 [Rajapakse et al., 2006]

from Risk2 can be unreliable. It would have been more judicious to use a weighted mean (by the number of values) of NPV corresponding to positive and negative values.

3.6 Conclusions

In the first part of this chapter, a Discrete Event Simulator (DES) previously developed in C++ in our research group for Batch Plant Scheduling (BPS), is adapted to the New Product Development (NDP) problem considered in this study. C++ programming has the ability to produce, with the notion of class, modular code that can be easily modified. Among the four layers of the DES, layers 'Engine' and 'Events' are the same for the two types of problems. However, some existing objects and classes must be adapted to the NPD problem, and more specific classes have to be designed.

Then, the DES is evaluated on a recently published work concerning the portfolio management of nine drugs for treating three diseases, four drugs for disease I, four drugs for disease II, and only one for disease III. At least one drug per disease must be produced. In all the numerical studies, the problem being by definition stochastic, each simulation is repeated a large number of times (300), selecting random sampling values from a uniform random number generator for defining failure or success for drugs at trial phases. In this evaluation section, the simulations are carried out with limitation on resource capacities as defined in Chapter 2. On this example, the DES gives correct trends, but a strict comparison with the solutions obtained by Blau et al. [2004] was not able to be performed as a result of the lack of some data in this article.

First, each of the nine drugs is launched alone into the pipe in order to evaluate individual performances by means of three ranking methods. The first one is based on the mean NPV and the attractiveness A defined as the ratio of mean positive NPV to mean negative NPV. Then, the classical bubble chart ranking commonly used in pharmaceutical field, is implemented. Each drug is plotted in a bubble format, with attractiveness vs. success probability, the diameter of a given bubble being its investment cost. Previous both studies are performed again, by using the weighted attractiveness WA instead of A. For computing WA, numerator and denominator of A are weighted by the number of positive, respectively negative, NPV values. It can be observed that even if the weighted attractiveness ratio gives more accurate information, ratios A and WA lead to the same conclusions concerning drug ranking for the considered example both by simulation and bubble charts.

Different sequences of drugs (12) are then investigated. In that case, the mean NPV can be considerably increased compared to the one obtained for only one drug. However, the NPV may vary strongly from a sequence to another. Sequences are ranked according to two pairs of indicators (mean NPV-PP) and (mean NPV -WA), and the results are identical (PP is the positivity probability defined as the ratio of the number of positive NPV to the total of runs per sequence (300)). Sequences with all the nine drugs launched together give bad results, and the best ones are obtained for sequences involving respectively six, six, six and four drugs. Furthermore, when a sequence involves drugs with good (respectively bad) individual performances, its ranking is good (respectively bad). Taking into account the highly combinatorial nature of the problem (about one million of sequences), all the sequences cannot be exhaustively evaluated. So, the only way to obtain optimal or quasi-optimal solutions is to implement efficient scanning algorithms, as it will be presented in Chapter 5.

All the studies of the present work are carried out under constraint on limitation of resource capacities. This constraint is relaxed and an example with no limited capacity for stages (unlimited capacity) is studied for four sequences: the best one, a good solution and the worst one of the previous section and a last one involving the nine drugs 1 to 9, launched in that order. The Gantt charts displayed for the last sequence show that some waiting times for an available resource disappear for the unconstrained problem, leading to better performances concerning both the mean NPV and the makespan. For the three other sequences, performances also increase when the problem becomes unconstrained, but the ranking remain the same as in the limited resource capacity case.

The DES is evaluated again on another recently published work [Rajapakse et al., 2006] concerning the portfolio management of six monoclonal antibodies. Thirty six sequences involving three, four and six antibodies are evaluated by simulation (300 runs) and compared on the basis of pair criteria (mean NPV-Risk1) and (mean NPV-Rik2), where Risk1 is the frequency of negative NPV obtained by carrying out 300 simulation runs by sequence (it is the same type of indicator as the positivity probability defined in the previous example), and Risk2 is the standard deviation of the NPV. In both cases, a sequence with three drugs is the best one and the sequence involving all the antibodies exhibits bad results.

According to Rajapakse et al. [2006], Risk2 for a sequence is the standard deviation of NPV. However, as indicated in the Blau et al. [2004] example, the population being bimodal, if the global mean value does not correspond to existing sequences, it can be nevertheless used as a statistical indicator. A statistical interpretation of the global standard deviation is more difficult. So a sequence ranking based on Risk2 can be unreliable.

Finally, it must be emphasized that several criteria, i.e., the Net Present Value of a sequence, its associated risk (equivalently measured by an attractiveness ratio or by the so-called positivity probability) and the makespan are important to evaluate a sequence quality and must be considered simultaneously for decision making.

Chapter 4

Imprecision modelling with interval analysis

4.1 Introduction

Uncertainty, imprecision and multiple criteria are important factors to take into account in decision making. This applies particularly in the problem of portfolio selection problem, as it typically involves multiple objective functions (net present value, risk and makespan) including uncertainties and imprecision in several parameters, as described in the previous chapter.

This chapter is devoted to imprecision modelling involved in the NPD problem. A first solution was to use interval bounds to model some imprecise parameters associated with a probability distribution within a Monte Carlo framework. The concept of Degree of Difficulty, the so-called DoD was also used to reflect the more or less difficulty to carry out a process task. In this chapter, the objective is to investigate alternative approaches to represent imprecision in order to determine the final strategy that could be then selected at the optimization step.

This chapter is organized as follows: first, a literature review on imprecision modelling is proposed. The principles of interval arithmetic that has been studied in addition to a probabilistic method are then briefly recalled, with a special focus to the operations involved in the implementation within the DES simulator. Then, typical results are presented and analyzed. Guidelines are then provided for the optimization phase.

4.2 Imprecision modelling

4.2.1 Introduction

As extensively discussed in Klir and Wierman [1999], uncertainty can be considered by the result of some information deficiency: the information to form the basis of a certain model may be incomplete, imprecise, fragmentary, not fully reliable or vague ... As a result, these various information deficiencies are associated with different types of uncertainty, which can be modeled by different well-established theories. Let us cite for instance classical set theory, fuzzy set theory, probability theory, possibility theory ... In the problem considered in this chapter, our analysis is restricted to imprecision modelling, which is prevalent in the problem definition and which occurs due to partial lack of information. 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. The latter class includes interval computation and fuzzy set theory. The fuzzy theory has indeed been considered as a powerful alternative for several years, with many applications in various fields Buckley and Hayashi [1998], Yang et al. [2000], Kuroda and Wang [1996]: the idea is to quantify uncertain model parameters by using fuzzy numbers and to trace the propagation of the uncertainties through the system by using fuzzy arithmetic. The advantage of this approach over the method of interval computation is to express parametric imprecision by fuzzy numbers, representing the more or less possible values of some parameters.

4.2.2 Probability methods

According to the theory of probability, the uncertain model parameters are represented by random variables and quantified by probability density function. The computation of the probability density function of the model outputs is then performed in a numerical way by using Monte-Carlo methods: this means that the models are evaluated for a large number of combinations for the parameter

values, generally randomly according to the predefined distribution. However, the application of probability theory assumes that historical data exist for each unknown parameter involved in the process. This may be not true in the case of New Product Development since this phase is by definition at a preliminary stage. Moreover, probability measures are not well-suited to express the intrinsic fuzziness of natural language that is used to verbally quantify imprecise information. Yet, they have been used in the simulation study of the NPD problem. The results have been presented in detail in the previous chapter.

4.2.3 Interval modelling and fuzzy methods

Interval modelling

Interval modelling is a very popular and elementary non-probabilistic uncertainty model. It is generally applied in this form:

$$X = [x_l, x_r] = \{x \in \mathbb{R} \| x_l \leq x \leq x_r\}$$

Intervals represent an appropriate model to mathematically describe uncertainty in those cases where only a possible value range between crisp bounds x_l and x_r is known for the uncertain quantity and no additional information concerning variations, fluctuations, value frequencies, preferences, etc. between interval bounds is available nor any clues as how to specify such information, respectively.

The respective limitation of information may be associated with a lack of knowledge, imprecision, or vagueness.

A sound theoretical basis has been developed for the mathematical treatment of interval-valued quantities Alefeld and Herzberger [1983], Moore [1966]. This represents the fundamental means for extending engineering computations to dealing with intervals. Generally, interval analysis connotes the mapping of interval input quantities X_i to interval result quantities Z_j,

$$\{X_1, ..., X_n\} \rightarrow \{Z_1, ..., Z_m\}$$

In this mapping, the dependency between crisp input values $x_i \in X_i$ and crisp result values $z_j \in Z_j$ from respective intervals is given by a deterministic algorithm,

$$f : \underline{x} \rightarrow \underline{z}, \underline{x} = (x_1, \ldots, x_i, \ldots, x_n),$$

$$\underline{z} = (z_1, \ldots, z_j, \ldots, z_m), x_i \in X_i, z_j \in Z_j$$

This deterministic algorithm may be referred to as the mapping model (see Figure 4.1).

The major beneficial feature of interval modelling concerns the enabling of best and worst case studies in absolute terms at a reasonable numerical cost as a result of the direct search for these cases. In interval analysis, an envelope is obtained for the results which definitely includes all possibilities resulting from the full range of input uncertainty.

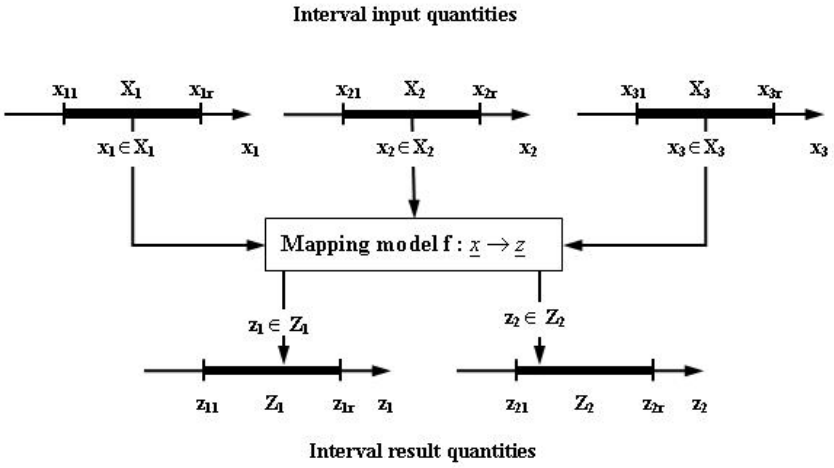


Figure 4.1: Interval analysis scheme

This feature further provides a basis for considering coarsely specified parameters in design problems and for deriving decision margins that are completely verified.

Also, an indication of sensitivities and robustness, respectively, may be gained by evaluating the width of input and result intervals with respect to one another Moens and Vandepitte [2007]. Altogether, a variety of useful new insights are obtained via interval modelling within the process of engineering analysis and design.

A limitation of the interval model is, however, its binary treatment of information. An element either belongs to the interval, or it does not belong to the interval. A gradual assignment of elements to the interval or a weighting of elements within the interval, respectively, cannot be accounted for. Consequently, a degree of confidence that a particular event occurs - as needed, for instance, in safety assessments - cannot be deduced with the aid of interval quantities alone. However, useful results can be obtained by including interval quantities in computations based on other uncertainty models. For example, safety assessments may be extended by an interval formulation of limit states.

The developments in interval modelling represent a sound basis for numerical processing of uncertainty within the framework of extended and generalized uncertainty models (see Muhanna et al. [2007], for instance). The formulation of these uncertainty models with a structure that contains the interval as a component makes the interval model a special case.

Fuzzy set modelling

A direct generalization and enhancement of the interval model is a fuzzy set. This represents an extension of the interval by a component of gradual assignment; the interval internal values $x \in [x_l, x^r]$ are assessed or weighted with the aid of membership values $\mu(x)$, from a continuous scale. The semantics of these membership values $\mu(x)$, may generally be categorized in three groups to express similarity, preference, or uncertainty.

This semantics of $\mu(x)$ is relevant, for example, in the fields of fuzzy control, regression analysis.

An understanding of membership degrees as preference utilizes a numerical expression of the intensity of being in favor of the associated elements x of the fuzzy set. The membership values $\mu(x)$

serves also as the basis for subsequent considerations. An association with possibility theory exists, insofar, as the membership values $\mu(x)$ may be understood as the degree of possibility, with which the underlying uncertain quantity may take on the associated values x. That is, the membership values reflect a subjective assessment. They express a degree of subjective confidence that particular values x actually occur. Also, $\mu(x)$ may be understood as the degree of compliance of relevant values x with an underlying condition for being assigned to the uncertain quantity.

The membership scale is usually normalized with the bounds zero for no membership and unity for full membership. Specifically, a normalized fuzzy set is described by

$$\widetilde{X} = \{(x, \mu(x))|x \in \mathbb{R}, 0 \leq \mu(x) \leq 1\}$$

The concept of fuzzy sets in its present form was initially formulated about forty years ago (see Zadeh [1965]), after which considerable developments have been reported and summarized in comprehensive books such as Dubois and Prade [1980], Zimmermann [1992], Bandemer and Gottwald [1995], Dubois and Prade [1986]. In basic mathematical literature, the connection of fuzzy sets $\widetilde{X}_i$ to produce fuzzy results $\widetilde{Z}_i$ which is referred to as fuzzy analysis

$$\left\{\widetilde{X}_1, \ldots, \widetilde{X}_n\right\} \rightarrow \left\{\widetilde{Z}_1, \ldots, \widetilde{Z}_m\right\}$$

is solved with the aid of the extension principle. This includes both the mapping model f for computing result values $z_i \in \widetilde{Z}_j$ from input values $x_i \in \widetilde{X}_j$ and a rule for specifying the membership values of the result values.

In the processing of fuzzy quantities through engineering computations, the original form of the extension principle has been revealed as being unsuitable for implementation in numerical algorithms. Thus, an alternative method of solution for fuzzy analysis has been developed based on α-discretization and use of α-cuts. This method yields results equivalent to those from the extension principle with the application of the min-max operator. For a selected α-level $\alpha_k \in (0, 1]$ the α-level set

$$X_{\alpha k} = x \in \widetilde{X}|\mu(x) \geq \alpha_k,$$

is obtained from the fuzzy set $\widetilde{X}$. If $\widetilde{X}$ is a convex fuzzy set, as can generally be assumed in engineering applications, its α-level are intervals $X_{\alpha_k} = [X_{\alpha kl}, X_{\alpha kr}]$ which are assessed in terms of a minimum membership. Vice versa, this enables the representation of the fuzzy set $\widetilde{X}$ of its α-level sets,

$$\widetilde{X} = \{(X_{\alpha k}, \mu(X_{\alpha k}))|\mu(X_{\alpha k}) = \alpha_k, \forall \alpha_k \in (0, 1]\}$$

4.2.4 Selection of the imprecision modelling technique: interval modelling vs. fuzzy set concepts for NPD problem

As presented in Chapter 2, the NPD process has been modeled via a discrete event simulator. The integration of fuzzy set theory in discrete event system simulation in order to cope with the

representation of qualitative uncertainty has been proposed in Nguyen and Le [1997], Grieco et al. [2003], Azzaro et al. [1997]. In Nguyen and Le [1997], fuzzy and temporal logics are combined to establish a temporal logic-based simulation system that is capable of handling possibilistic values of both system state variables and event occurrence times. In Grieco et al. [2003], the problem of processing fuzzy data within a discrete event simulation process is discussed and new methods, able to avoid time paradox problems, are proposed.

Previous works in our research team were devoted to the development of a discrete event simulation model of an industrial production system using fuzzy concepts to represent uncertainties in the performance of people (time and duration of their intervention). The work has been focused in semiconductor manufacturing but can be applied to other kinds of batch processes presenting similar features. A solution to the problems related with the management of fuzzy uncertainty in discrete event simulation is proposed. Recently, fuzzy uncertain durations have been considered in Zhang et al. [2005]. The fuzzy ranking measure is merged with an activity scanning simulation algorithm for performing fuzzy simulation time advancement and event selection for simulation experimentation.

Another investigation concerned batch plant design with imprecise demands modeled by using fuzzy concepts. A new approach to the design problem was proposed, based on a multiobjective genetic algorithm, taking into account simultaneously maximization of the net present value and two other performance criteria, i.e. the production delay/advance and a flexibility criterion. The methodology provides a set of scenarios that are helpful to the decision maker at product development stage. Besides, a hybrid selection method Pareto rank-tournament was proposed and showed a better performance than the classical Goldberg's wheel, systematically leading to a higher number of non-dominated solutions Dietz et al. [2008].

Although many types of fuzzy sets [Zadeh, 1965] have been used to describe uncertainties, triangular and trapezoidal fuzzy sets (Figure 4.2) are very often used in the applications (e.g., fuzzy controllers and managerial decision-making) because the parameters defining them can be easily specified in linguistic terms [Bojadziev and Bojadziev, 1997]. In our previous works on batch plant scheduling, trapezoidal fuzzy sets are applied to describe uncertain activity duration.

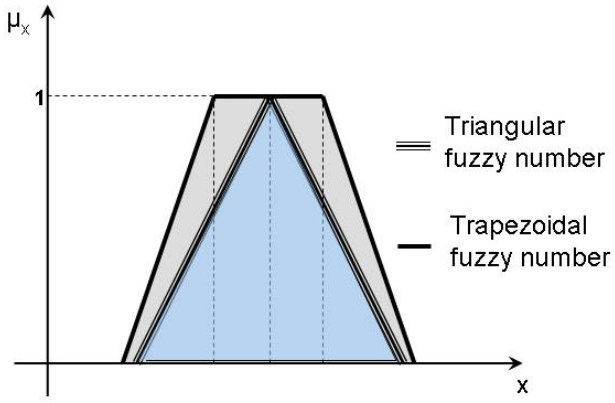


Figure 4.2: Triangular and trapezoidal fuzzy numbers

Fuzzy ranking problem usually deals with determining the best ordering procedure fitting a fixed objective. In our work, a ranking algorithm involved in the simulation environment to represent any feasible system evolution was used.

It must be yet emphasized that data gathering (a four-tuple for each duration) was a hard task in a production environment. This is all the more true for NPD management problem for which the

degree of imprecision may be higher. Hence subjectivities in selecting distributions and estimating related parameters are unavoidable. This explains why only an interval analysis has been performed and preferred to a fuzzy method. It must be also highlighted that the imprecision through probability success parameters have not been taken into account in this imprecision analysis.

4.3 Combining discrete event simulation (DES) and interval analysis (IA)

4.3.1 Introduction

The combination of interval analysis with discrete event simulation to handle subjectivity, vagueness or imprecision in estimating activity duration and capital costs in the NPD problems is presented in what follows. In particular, the application to the control of interval-time advancement and event selection for the simulation experiment of discrete event simulation is described. Illustrations on the interval analysis (IA) with discrete event simulation and an example that compares the IA discrete event simulation with the traditional Monte-Carlo based simulation are also provided. IA discrete event simulation focused on considering time values as intervals instead of crisp numbers. In other words, processing time and arrival time of entities are influenced by ill-defined uncertainty and event occurrence is represented by interval analysis. From a theoretical point of view, no difference exists between classical and interval analysis simulation because only the formalism used to represent variables (e.g. event occurrence time) changes.

- Initiation of a start event or activation of an activity, where the start time of an activity is determined; $T^S(i) = max_{j=1,\ldots,J} T^A(i,j)$, where J is the total quantity of the entities (i.e. resources and logical dependencies) required by activity i; T^s is the start time of activity i; $T^A(i,j)$ is the available time of entity j at activity i.
- Determination of the due time of an end event, which should equal the start time of the current activity plus its activity duration; $T^{EE}(k) = T^S(i) + D(i)$, where $T^{EE}(k)$ is the due time of end event k, recorded in the end event list; $D(i)$ is the duration activity i.
- Simulation advancement and end event selection, that is, the simulation is updated from the current time to the time of next one or more end events that will happen. The earliest and the end events that are due at the updated simulation time will be selected for initiation; $m\_Now = min_{k=1,\ldots,K} T^{EE}(k)$, where m_Now is the simulation time and K is the number of end events recorded in the end event list.

When intervals are used to represent activity durations, all the times in the above operations become intervals. Hence the above operations can be expressed in interval arithmetic.

4.3.2 Interval arithmetic

As with fuzzy theory, intervals theory has a well-developed specific arithmetic where main arithmetic operations are defined on real arithmetic operations and applied to interval bounds.

Basic operations on intervals

Given $a, b \in \mathbb{R}$, where $a < b$. Then, the subset $\{x \in \mathbb{R} | a \leq x \leq b\}$ is an interval of real values or simply an interval that will be represented by $\mathrm{X} = [a, b] = \{x \in \mathbb{R} | a \leq x \leq b\}$.

Let $A = [a, b]$ and $B = [c, d]$ be two interval numbers. The basic arithmetic operations of addition, subtraction, multiplication, and division of these two interval numbers are defined as follows:

$$[A] + [B] = [a, b] + [c, d] = [a + c, b + d]$$

$$[A] - [B] = [a, b] - [c, d] = [a - d, b - c]$$

$$[A] \times [B] = [a, b] \times [c, d] =$$
$$[min\,(a \times c, a \times d, b \times c, b \times d)\,, max\,(a \times c, a \times d, b \times c, b \times d)]$$

$$[A] \div [B] = [a, b] \times \left[\frac{1}{d}, \frac{1}{c}\right] si\ 0 \notin [c, d]$$

$$Max\,(A, B) = \frac{a + b + |a - b|}{2}$$

$$Min\,(A, B) = \frac{a + b - |a - b|}{2}$$

4.3.3 Illustration examples

For the sake of illustration, let us consider $A = [5, 12]$ and $B = [7, 15]$:

Addition (Figure 4.3)

$$[A] + [B] = [5, 12] + [7, 15] = [5 + 7, 12 + 15]$$

$$[A] + [B] = [12, 27]$$

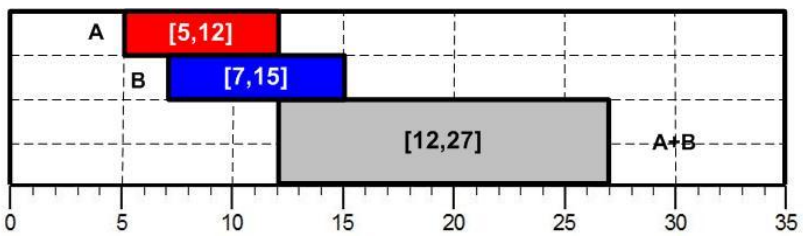


Figure 4.3: Addition of A and B

Subtraction (Figure 4.4)

$$[A] - [B] = [5, 12] - [7, 15] = [5 - 15, 12 - 7]$$
$$[A] - [B] = [-10, 5]$$

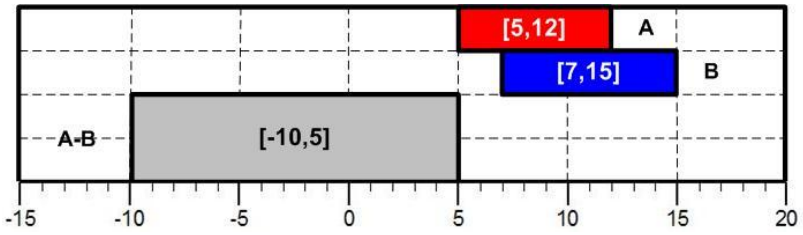


Figure 4.4: Subtraction of A and B

Multiplication (Figure 4.5)

$$[A] \times [B] = [5, 12] \times [7, 15] =$$
$$[min\,(5 \times 7, 5 \times 15, 12 \times 7, 12 \times 15)\,, max\,(5 \times 7, 5 \times 15, 12 \times 7, 12 \times 15)]$$

$$[A] \times [B] = [5, 12] \times [7, 15] = [min\,(35, 75, 84, 180)\,, max\,(35, 75, 84, 180)]$$
$$[A] \times [B] = [35, 180]$$

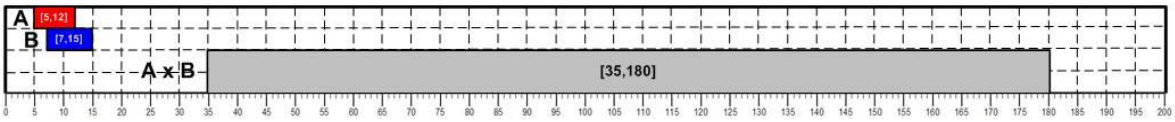


Figure 4.5: Multiplication of A and B

Division (Figure 4.6)

$$[A] \div [B] = [5, 12] \times \left[\frac{1}{15}, \frac{1}{7}\right] = \left[\frac{5}{15}, \frac{12}{7}\right]$$
$$[A] \div [B] = [0.333, 1.714]$$

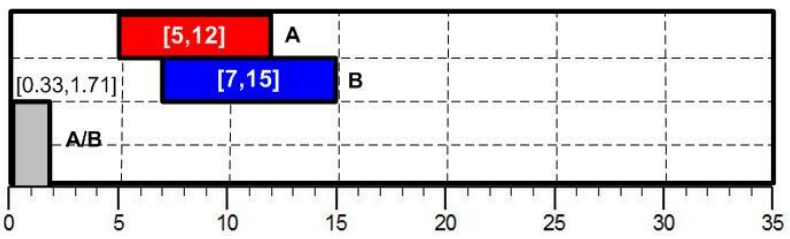


Figure 4.6: Division of A and B

Maximum

$$Max\,(A,B) = (12,15) = \frac{12+15+|12-15|}{2} = 15$$

Minimum

$$Min\,(A,B) = (5,7) = \frac{5+7-|5-7|}{2} = 5$$

The DES model has thus been adapted to take into account imprecision modeled by interval concepts. The main concepts are now presented.

4.4 Extension of the Discrete event simulator for NPD formulation with interval analysis

4.4.1 Duration and cost modeled as intervals

Imprecise data embedded in the NPD problem used in the previous chapter concern, on the one hand, duration and operating cost for each stage, and mature sales and capital cost for each drug, on the other hand. All these parameters have been represented by a triangular representation, based on the experience of pharmaceutical managers; for the second group of parameters related to each drug, the concept of Degree of Difficulty (DoD) was used to reflect the more or less difficult way to develop a drug. This DoD was defined based on the work presented by Blau et al. [2004]. Let us recall that it takes on values ranging from one (low difficulty) to ten (high difficulty). The resources required for a drug candidate with a degree of difficulty of five represent an average value. The degree of difficulty is then used to scale the resource cost distributions in a linear fashion. The example which serves here as an illustration is the 9-case drug portfolio presented and analyzed in detail in the previous chapter. The aforementioned parameters have been designed as intervals $A = [a, b]$ as follows:

- For duration and operating cost for each stage, the bounds of the interval A have been taken equal to the average value of the initial data ±4%;
- For mature sales and capital cost for each drug, the bounds of the interval A have been taken equal to the value deduced from DoD ±4%.

The value of ±4% has been considered first to examine the effect of imprecision propagation along the pipeline process and to verify that this is not redhibitory with result interpretation and decision making.

This procedure leads to the following interval parameters relative to drug 1 for the considered example (see Table 4.1).

4.4.2 Parallel activities

Although the NPD problem is mainly linear, some parallel stages are involved (see the conceptual model diagram in Figure 3.5 of Chapter 3) for which the following stage can start only when all parallel stages have finished.

Drug 1		
Stage	Interval value for duration (Days)	Interval value for cost (M$)
First human dose preparation	[384 ,416]	[1.9 ,2]
Sample preparation	[384,416]	[1.9,2]
Phase I	[288,312]	[76,83]
Process development 1	[768,832]	[9.4,10.5]
Process development 2	[768,832]	[9.4,10.5]
Phase II	[480,520]	[76,83]
Phase III	[744,806]	[190,210]
Design plant	[720,780]	[10.48,9.5]
FSA	[360,390]	[19,20]
Built plant	[720,780]	[59,64]
Prelaunch	[96,104]	[47,52]
Launch I	[336,364]	[11.4,12.6]
Launch II	[336,364]	[21,23]
Launch III	[336,364]	[41,38]
Mature sales	[336,364]	[50,55]

Table 4.1: Example of interval data design for drug 1

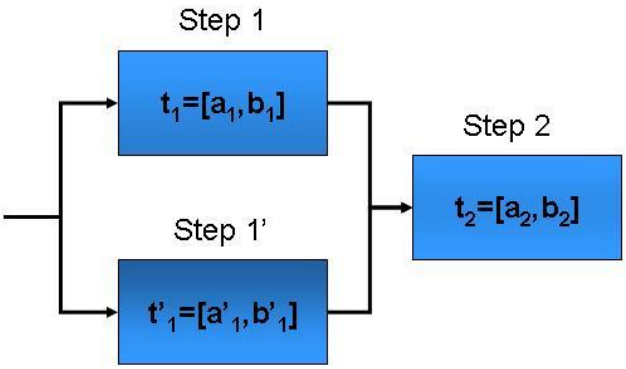


Figure 4.7: Parallel steps with interval durations

Let us consider two parallel stages P_1 and $P_1^{'}$ and their associated end dates $t_1 = [a_1, b_1]$ and $t_1^{'} = \left[a_1^{'}, b_1^{'}\right]$ (See Figure 4.7).

The earliest date t_2 at which the following stage P_2 can begin can be computed as follows:

$$t_2 = max[(a_1, b_1), (a_1^{'}, b_1^{'})] = [max(a_1, a_1^{'}), max(b_1, b_1^{'})]$$

The waiting time WT of one parallel item can thus be computed in the following way.

Either,

$$WT = max\ (t_1, t_1^{'}) - t_1 = max\ (0, t_1^{'} - t_1)$$

or,

$$WT = max(t_1, t_1^{'}) - t_1^{'} = max(t_1 - t_1^{'}, 0)$$

$$t_1 - t_1^{'} = (a_1, b_1) - (a_1^{'}, b_1^{'}) = (a_1 - b_1^{'}, b_1 - a_1^{'})$$

Finally,

$$WT = max[(0,0), (a_1 - b_1^{'}, b_1 - a_1^{'})] = [max(0, a_1 - b_1^{'}), max(0, b_1 - a_1^{'})]$$

4.4.3 Net Present Value computation

The computation of the Net Present Value can be summarized as follows:

$$NPV_{seq} = \sum\nolimits_{drug} \sum\nolimits_{m} \frac{-(c_m^-, c_m^+) + (r_m^-, r_m^+)}{(1+i)^{d_m^-, d_m^+}}$$

Where,

NPV_{seq}	is the Net Present Value of a given sequence
m	represents a step of the NPD pipeline relative to a drug
$[c_m^-, c_m^+]$	represents the costs (interval) for a drug at step m
$[r_m^-, r_m^+]$	represents the sales (interval) for a drug at step m
$[d_m^-, d_m^+]$	represents the duration interval for a drug at step m
i	is the actualization rate

In this expression,

$$(1+i)^{d_m^-, d_m^+} = [(1+i)^{d_m^-}, (1+i)^{d_m^+}]$$

4.5 Simulation example

The simulation example which serves as a test bench (see Chapter 3) is investigated here in an interval-based fashion. The objective is to compare the results obtained from both analysis approaches : (a)- using interval discrete event simulation (b)- using traditional simulation with stochastic parameters. The idea is to see if the interval approach is sound in the NPD problem, meaning that it does not induce too large uncertainties at the end of the simulation of the whole process, which is often considered as a major drawback for both fuzzy and interval-based approaches. Consequently, under such information, it is difficult for the decision maker to conclude.

4.5.1 Ideal simulation for each drug (success probabilities equal to 1)

Some preliminary simulation runs were performed considering a success probability equal to unity at each critical step of the NPD pipeline (Phases I, II and III). The results are displayed in Figure 4.8 and in Table 4.2. Each bar represents the interval NPV contribution to the global NPV at each step of the process. It must be highlighted that the spread of uncertainty is highest for the step relative

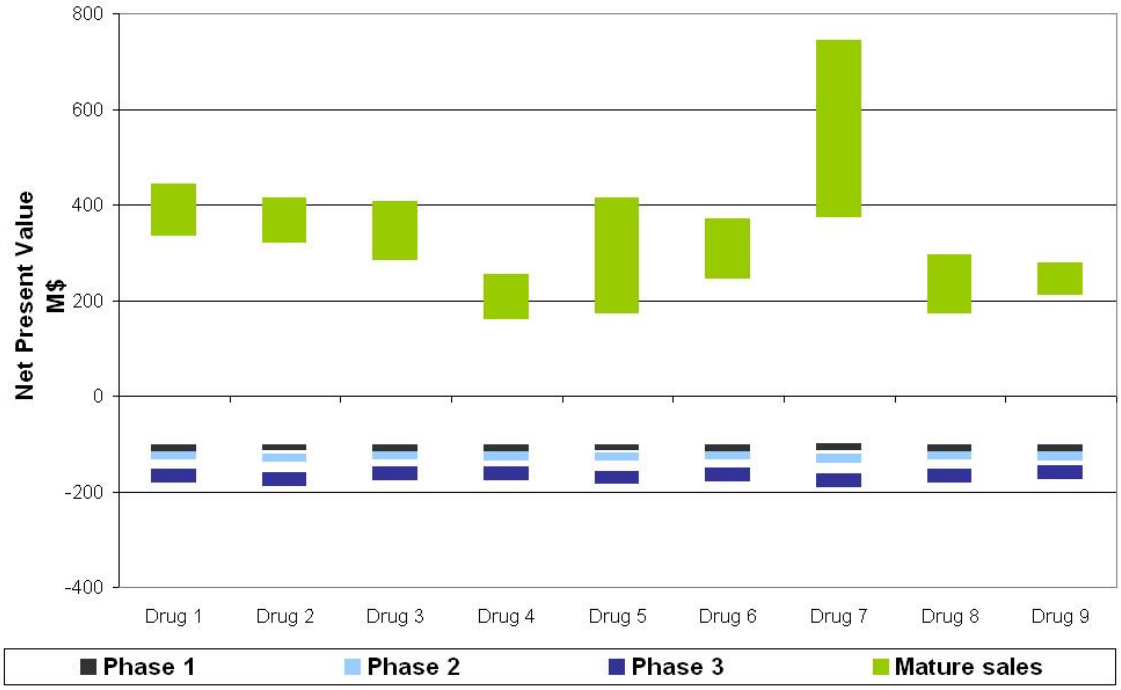


Figure 4.8: Interval NPV contribution to the global NPV at each step of the process for each drug

to sales involved at the final stage of the process: this can be explained by the cumulated uncertainty influence of time and cost at the latest stage of the process.

Drug	Phase 1 NPV (M$)	Phase 2 NPV (M$)	Phase 3 NPV (M$)	Mature sales NPV (M$)
1	[-111,-107]	[-125,-119]	[-168,-159]	[343,428]
2	[-110,-105]	[-132,-126]	[-179,-169]	[328,405]
3	[-111,-107]	[-119,-113]	[-157,-149]	[287,380]
4	[-111,-107]	[-117,-111]	[-154,-146]	[157,232]
5	[-110,-106]	[-130,-123]	[-175,-166]	[179,404]
6	[-111,-107]	[-121,-115]	[-161,-152]	[249,350]
7	[-109,-105]	[-134,-128]	[-182,-171]	[382,737]
8	[-110,-106]	[-127,-121]	[-172,-162]	[181,286]
9	[-111,-107]	[-114,-109]	[-150,-142]	[210,251]

Table 4.2: Interval NPV contribution to the global NPV at each step of the process for each drug

The results are also compared with those obtained with the probability approach. The interval results relative to each phase (I, II, II and mature sales) are positioned with the mean NPV deduced from the probability approach NPV_{prob} (red dots) (see Figures 4.9, 4.10, 4.11 and 4.12). It can clearly be observed that NPV_{prob} corresponds approximately to the average value of the interval bounds. Figure 4.12 also indicates the lower and upper values of the set of all the values generated by the stochastic approach. The results of the probabilistic approach obtained in the previous chapter are summarized in Table 4.3.

Drug	Phase 1 NPV (M$)	Phase 2 NPV (M$)	Phase 3 NPV (M$)	Mature sales NPV (M$)
1	-109	-122	-163	384
2	-107	-129	-173	364
3	-109	-116	-153	331
4	-109	-114	-150	193
5	-108	-126	-170	286
6	-109	-118	-156	297
7	-107	-131	-176	550
8	-108	-124	-167	231
9	-109	-112	-146	229

Table 4.3: NPV for 9 drugs obtained from the probabilistic approach

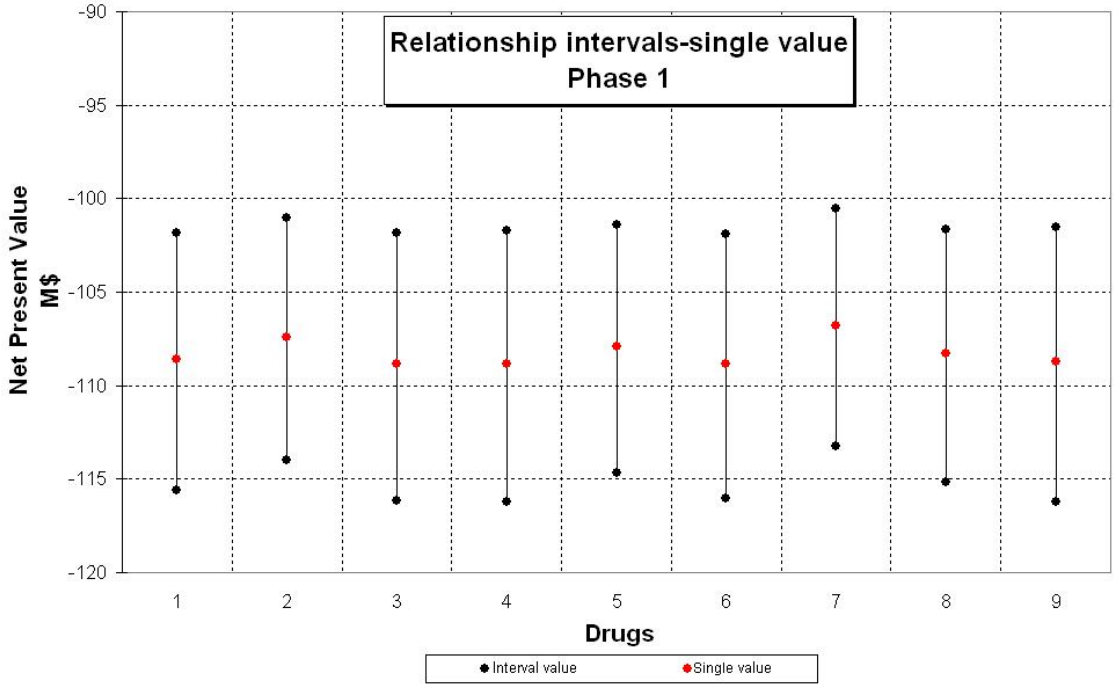


Figure 4.9: Comparison between interval and probabilistic approach for each drug at phase I

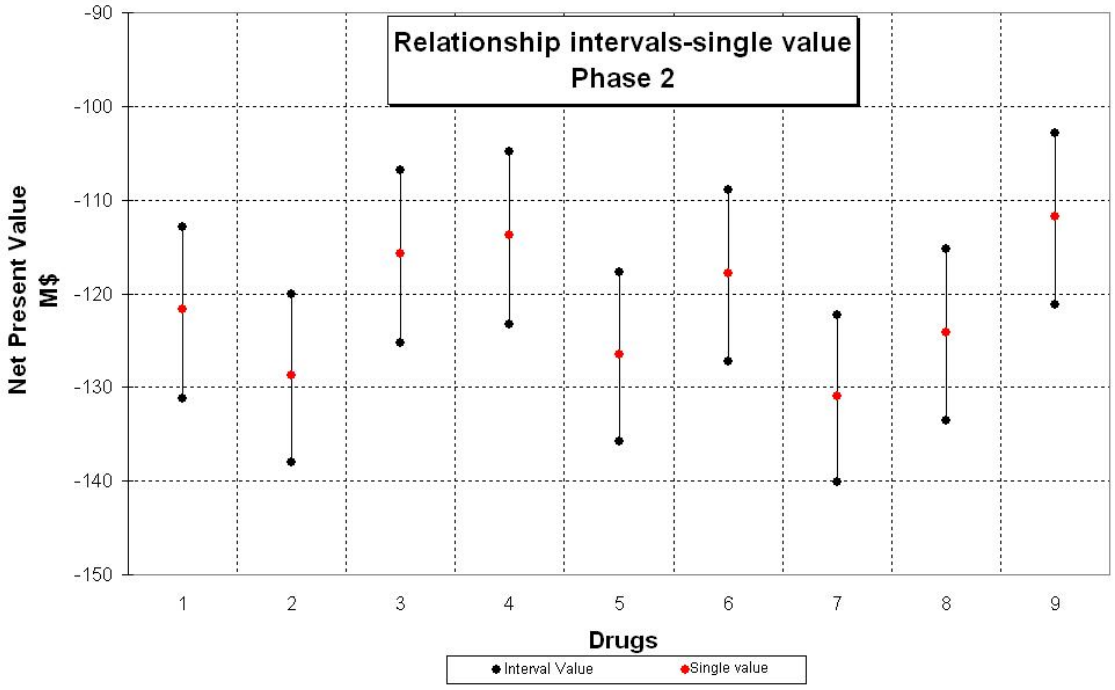


Figure 4.10: Comparison between interval and probabilistic approach for each drug at phase II

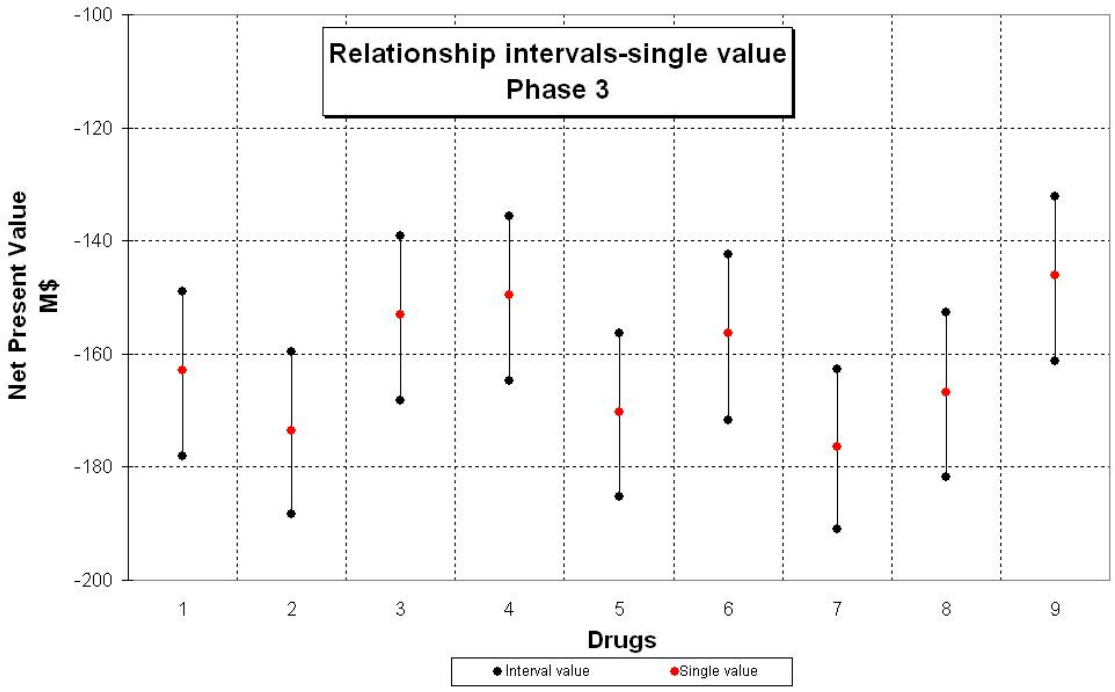


Figure 4.11: Comparison between interval and probabilistic approach for each drug at phase III

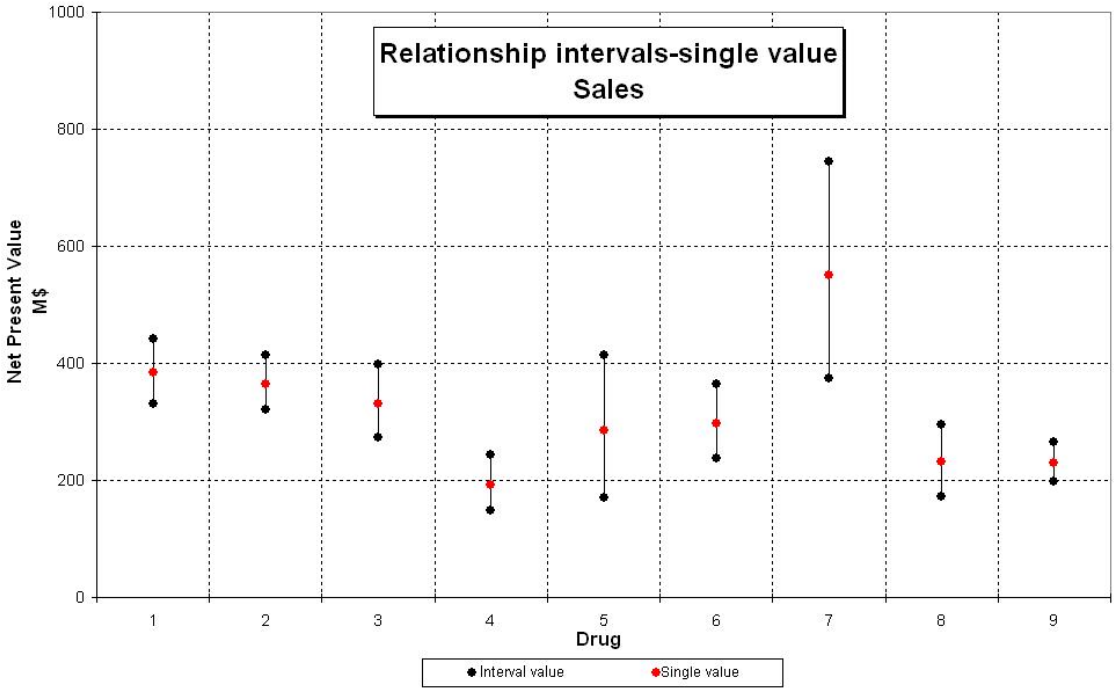


Figure 4.12: Comparison between interval and probabilistic approach for each drug at mature sales

4.5.2 Ideal simulation for 10 sequences (success probabilities equal to 1)

In this section, the 10 sequences that have been previously investigated in the previous chapter have been considered for simulation with interval data. The results are displayed in Figure 4.13, with a special emphasis to Phases I, II, II and mature sales, where the respective contribution of each drug to the portfolio of the sequence is illustrated. Such diagrams are imported to detect the major sources of uncertainty in a given sequence. For example, it can be seen that drug 7 contributes largely to the uncertainty.

4.5.3 Simulation for 10 sequences with success probabilities

In this section, the simulation of the 10 sequences has been carried out taking into account the success probability. Since this parameter has not been modeled with an interval, a sampling set of simulations has been performed. Of course, this hybrid method combining interval and probabilistic approaches is time consuming, but the objective here is to see if the interval concepts need to be investigated in more details. As previously, 300 simulations were run. The lower (respectively upper) bound of the resulting interval was taken equal to the minimal (respectively maximal) value of the lower (respectively maximal) bounds over the 300 simulation interval values. The results are presented in Figure 4.14 and Table 4.4. A distinction is performed according to the positive and negative part of the interval (P and N-intervals), in order to mimic what happens with the bimodal distribution when considering a probabilistic approach. The same approach as aforementioned is used to build the two intervals corresponding to a sequence. Let us mention that when an interval overlaps negative and positive values, it has been allocated to the negative section. A first comment can be made concerning the mean values obtained from the probabilistic approach. They all are located in the interval. Figure 4.14 also indicates the average value obtained over all values, either positive or

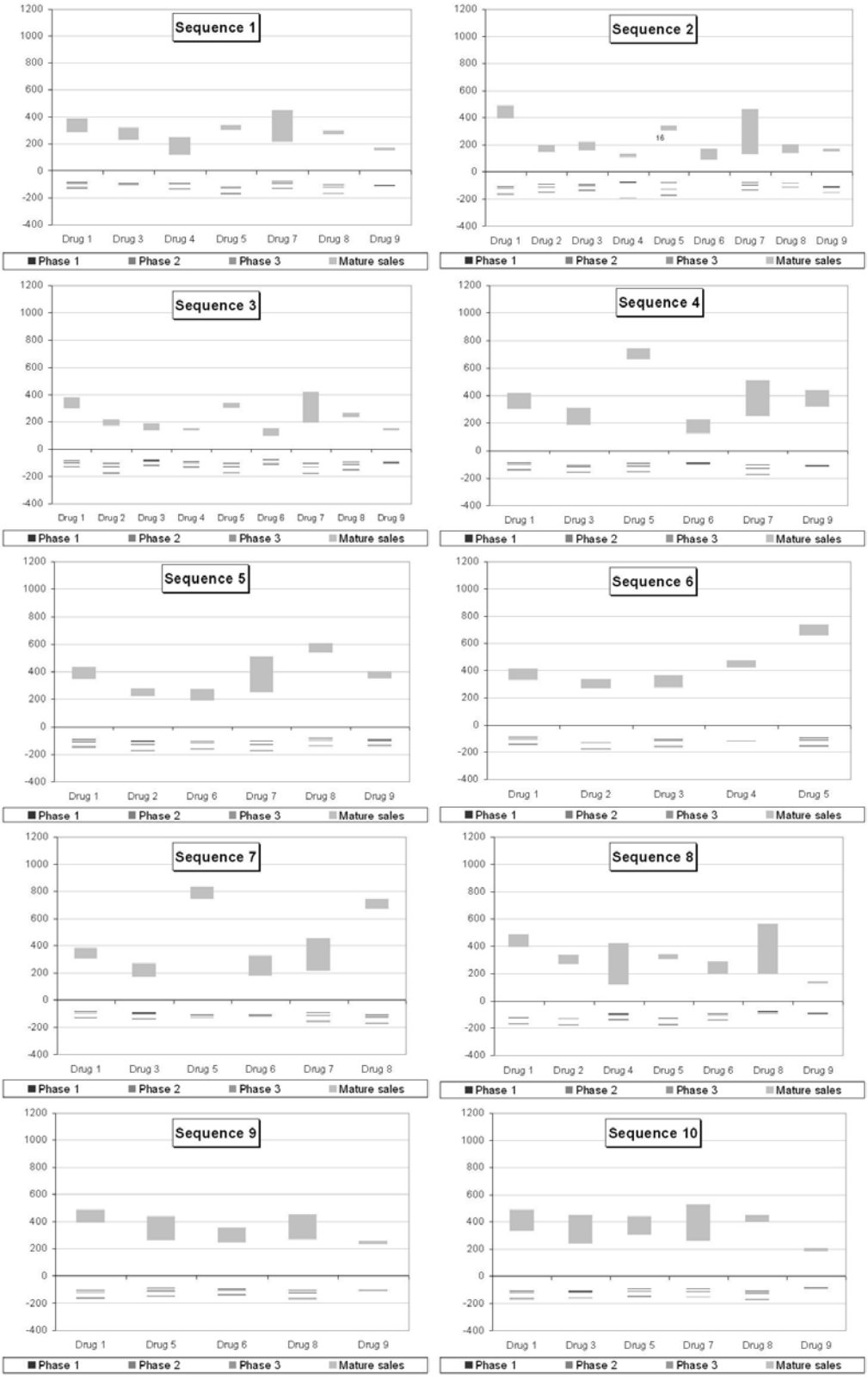


Figure 4.13: Evaluation of 10 sequences with an interval-based approach

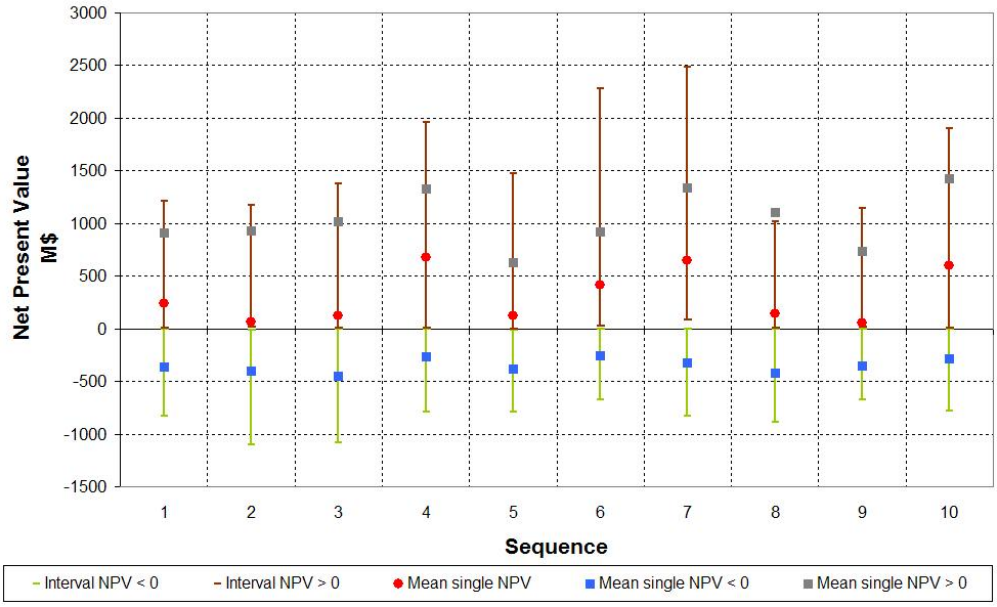


Figure 4.14: P-Intervals and N-Intervals for each sequence relative to NPV

negative. For sequence 8, it can be seen that the average is slightly higher that the upper bound of the so-called positive interval. This situation is not paradoxical since the spread of imprecision has not been considered equivalent for both methods. A lower spread was considered for intervals in order to examine the consistency of this approach for NPD problems.

Sequence	NInterval for NPV		Mean NPV <0 (prob.)	PInterval for NPV		Mean NPV >0 (prob.)
1	-793	-3	-358	11	1111	911
2	-996	-62	-406	8	1092	926
3	-1009	-34	-447	28	711	1018
4	-704	-4	-267	1	1583	1333
5	-751	-2	-378	38	1619	626
6	-638	-29	-257	126	1242	916
7	-936	-1	-324	1	2870	1342
8	-796	-14	-417	49	1051	1105
9	-634	-4	-351	160	1076	732
10	-900	-84	-282	4	2042	1422

Table 4.4: Comparison between interval-based and probabilistic approaches for 10 sequences with success probabilities (M$)

From all these simulation results, it can clearly be seen that the propagation of uncertainty along the pipeline process is too important and disturbs decision making. The interest of the probabilistic approach is to clearly highlight the two-peaked distribution which is typical of new drug development. Such information is lacking for the intervalbased approach. Moreover, the concept of risk may be difficult to appreciate. For all these reasons, the examination of the interval-based approach has not been more investigated. Even if this result may be considered as negative, we found it important to report this study since either interval or fuzzy based approaches are attractive candidates to account for imprecision. At this level of presentation, it must be said that the typical nature of the problem makes their application difficult.

4.6 Conclusions

The aim of this chapter was to compare various approaches to model the uncertainty embedded in the NPD problem. Several levels of uncertainty can be identified: on the one hand, they are associated with cost and activity durations, on the other hand, they concern the success probabilities involved at three steps of the NPD pipeline (the socalled Phases I, II and III). Two approaches have been used, a classical probability approach and an interval-based one. The former implies to carry out simulations many times to consider a representative sampling of the problem. The latter may be attractive for the optimization phase of the NPD problem that will be tackled in the following part of this work. 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 intervalbased optimization method as an outer loop of the discrete event simulation model for NPD has not been developed. The following chapter now addresses the optimization of the product portfolio.

Chapter 5

Multiobjective optimization strategies for the NPD process

5.1 Introduction

This chapter divided into three sections is particularly devoted to the bicriteria and tricriteria optimization of the New Product Development (NPD) problem. The previous chapters have shown that the NPD problem involves multi-stage decisions under uncertainty. The recurrent key issues are:

- What are the projects to develop once target molecules have been identified?
- In what order?
- Which is the level of resources to assign?

Chapter 2 was specifically related to the modelling of new product development pipeline. 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. As previously shown, this kind of problem 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 1 is first devoted to 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. Section 2 then applies the selected NSGA II algorithm to the treated case. Section 3 analyses and discusses the test bench examples and provides with some guidelines for the treatment of new cases.

5.2 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.

5.2.1 General multiobjective optimization problem formulation

A general multiobjective design problem is expressed by next equations.

$$f(x) = (f_1(x), f_2(x), \cdots, f_k(x))^T$$

$$s.t.\ x \in S$$

$$x = (x_1, x_2, \cdots, x_n)^T$$

where $f_1(x), f_2(x), \cdots, f_k(x)$ are the k objective functions, $(x_1, x_2, \cdots, x_n)$ are the n optimization parameters and $S \subset R^q \times N^r : q + r = n$ is the solution or parameter space.

Obtainable objective vectors $F(x)|\, x \in S$ are denoted by Y, so $F : S \mapsto Y$, S is mapped by **F** onto $\mathbf{Y} \rightarrow \mathbf{F}$ is non−linear and multi−modal and S might be defined by nonlinear constraints containing both continuous and discrete variables. $f_1^*, f_2^*, \ldots, f_k^*$ will be used to denote the individual minima of each respective objective function, and the utopian solution is defined as $F^* = (f_1^*, f_2^*, \cdots, f_k^*)^T$. It simultaneously minimizes all objectives, it is an ideal solution that is rarely feasible.

Figure 5.1 provides a visualization of the nomenclature. In this formulation, minimize F(x), lacks clear meaning as the set F(x) for all feasible x lacks a natural ordering, whenever F(x) is vector−valued. In order to determine whether $F(x_1)$ is better then $F(x_2)$, and thereby order the set F(x), the subjective judgment from a decision-maker is needed.

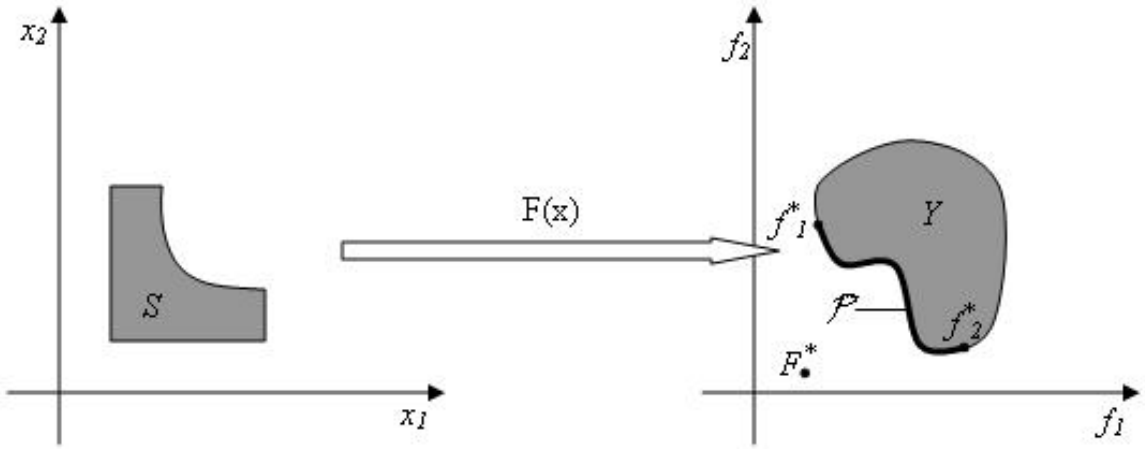


Figure 5.1: Parameter/solution and attribute space nomenclature for a two dimensional problem with two objectives.

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:

$$\forall i \in \{1, 2, \ldots, k\} : f_i(x) \leq f_i(y) \; and \; \exists j \in \{1, 2, \ldots, k\} : f_i(x) < f_i(y) \tag{5.1}$$

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 et al. [2004], 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.

5.2.2 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 and Grossmann, 1998], supply chain design [Guillén et al., 2006], and multi product [Ravemark and Rippin, 1998] or multipurpose [Dedieu et al., 2003] batch plant design or retrofitting [Montagna and 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 and 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, as illustrated in Figure 5.2. 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.

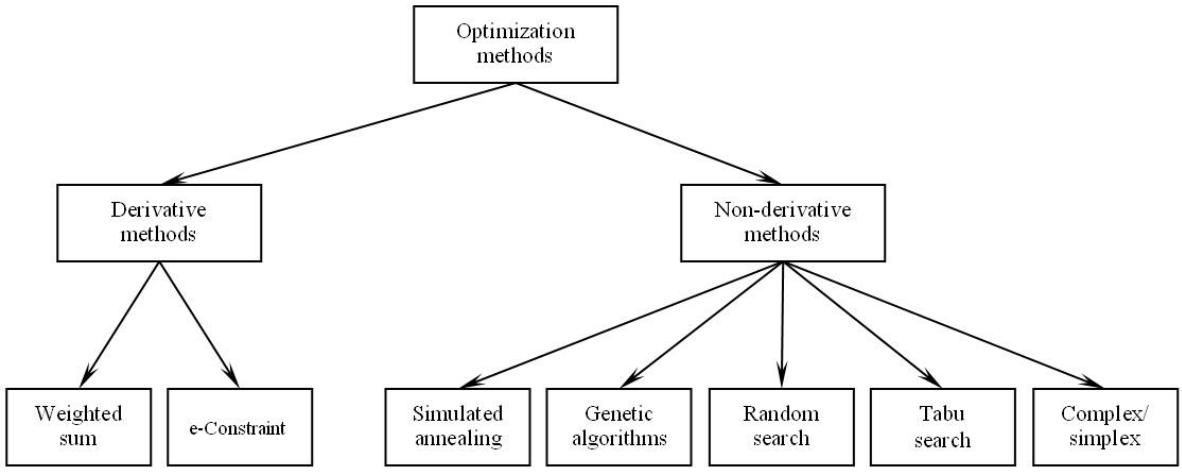


Figure 5.2: A classification of optimization methods in derivative and non-derivative methods with examples of some common non-derivative methods.

5.2.3 Multiobjective optimization road map

As mentioned earlier real engineering design problems are usually characterized by the presence of many conflicting objectives that the design has to fulfill. Therefore, it is natural to look at the engineering design problem as a multiobjective optimization problem (MOOP). References to multiobjective optimization could be found in Deb [2004],Fonseca and Fleming [1993],Rowe et al. [1996],Zitzler and Thiele [1999] and with engineering applications in Deb and Goel [2000],Gen and Cheng [2000].

As most optimization problems are multiobjective to there nature, there are many methods available to tackle these kind of problems. Generally, the MOOP can be handled in three different ways, as shown in Figure 5.3, depending on when the decision-maker articulates his or her preference on the different objectives, never, before, during or after the actual optimization procedure.

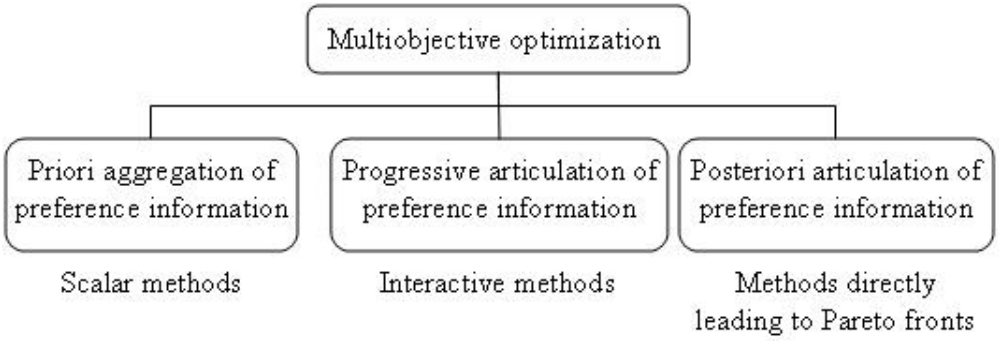


Figure 5.3: Ways to perform multiobjective optimization.

Priori articulation of preference information. These methods involve the most easy and perhaps most widely used method, i.e. the weighted-sum approach Steuer [1986], goal programming, ϵ-constraint approach, lexicographic approaches.

Progressive articulation of preference information. These methods include the STEM-Method or STEP-method Benayoun et al. [1971], and Steuer method.

Posteriori articulation of preference information. There are a number of techniques which enables to first search the solution space for a set of Pareto optimal solutions and present them to the decision-maker. The big advantages with these type of methods is that the solution is independent of the decision maker's (DM) preferences. The analysis has only to be performed one, as the Pareto set would not change as long as the problem description is unchanged. However, some of these methods suffer from a large computational burden. Another disadvantage might be that the DM has too many solutions to choose from. There are however methods that support in screening the Pareto set in order to cluster optimal solutions, see Morse [1980], Rosenman and Gero [1985].

In the following, a set of approaches is presented which are particularly attractive for engineering problems.

5.2.4 Genetic algorithms

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 70's. 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, $x_1, ..x_n$, 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, depicted in Figure 5.4. 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.

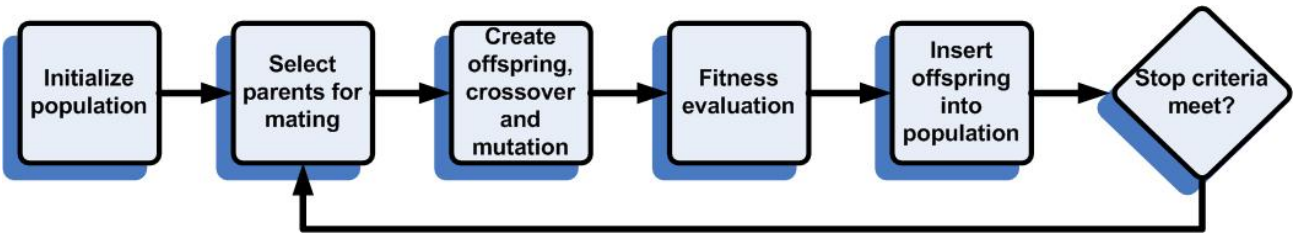


Figure 5.4: A simple genetic algorithm.

The popularity of genetic algorithms has grown tremendously under recent years and they have been applied to a wide range of engineering problems Dietz et al. [2005], Yoshikawa and Terai [2005], Deb and Srinivasan [2005], Altiparmak et al. [2009]. 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 et al. [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.

5.2.5 Short notes on comparisons of the different non-derivative methods

As already mentioned, there is no simple answer to which optimization methods is the best for any given problem.

In most comparison studies, the way how the different methods have been tuned to fit that particular problem is a recurrent issue. Comparative studies of different types of non-derivative methods could be found in for instance Borup and Parkinson [1992], Jansson [1994],Mongeau et al. [1998], Hajela [1999]. An interesting question that one should keep in mind when comparing different methods are the times spent on optimizing the different methods before they are compared. If a method is five percent faster then another one, but takes three times as long to implement and parameterize, it might not be worth the effort.

GAs seems to be most suitable to handle multi modal function landscapes and to identify multiple optima in a robust manner. GAs are however associated with a high computational cost.

Since 1975, many evolutionary procedures appear. For example, one can cite genetic algorithms [Holland, 1975], simulated annealing [Kirkpatrick et al., 1983], artificial immune systems [Farmer et al., 1986], ant colonies [Dorigo, 1992, Monmarché et al., 2009], particle swarms [Kennedy and Eberhart, 1995], artificial bee colonies [Nakrani and Tovey, 2004] and artificial neural networks [Ang et al., 2007]. All these algorithms can be adapted to the multiobjective case, and as it can be observed in the list of references proposed by Coello [2009]. The two most popular methods in the chemical engineering field are MOGA (MultiObjective Genetic Algorithm, see Konac et al. [2006]), and MOSA (MultiObjective Simulated Annealing, see Shu et al. [2004], Smith et al. [2004], Bandyopadhyay et al. [2008]). None of these two methods is perfect and selecting one depends on the requirements of the particular design situation considered. From the literature survey [Veldhuizen and Lamont, 2000, Branke et al., 2004, P.J. Turinsky and and Abdel-Khalik, 2005, Mansouri et al., 2007] it appears that MOGA is generally preferred to MOSA.

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. For an overview on genetic algorithms in multiobjective optimization, see Fonseca and Fleming [1995a]. Literature surveys and comparative studies on multiobjective genetic algorithms are also given in Coello [1996], Bäck et al. [1997], Tamaki et al. [1996], Zitzler and Thiele [1999].

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

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. Each individual objective is designated as the selection metric for a portion of the population. However, it is reported that the method tends to crowd results at extremes of the solution space, often yielding poor coverage of the Pareto frontier.

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. Crowding techniques, dominance, and diploid to maintain diversity in the population were also added.

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. Maintaining diversity is a required quality of multiobjective methods, often implemented, [Harik, 1995, Grueninger and Wallace, 1996].

Pareto based approaches Goldberg [1989] introduced non-dominated sorting to rank a search population according to Pareto optimality. 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 [1995b, 1998] each individual is ranked according to a degree of dominance. The more population members that dominate an individual, the higher ranking the individual is given. An individual's ranking equals the number of individuals that it is dominated by plus one. Individuals on the Pareto front have a rank of 1 as they are non-dominated. 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 tournaments are used to select individuals for the next generation. For binary tournaments, a subset of the population is used as a basis to assess the dominance of the two contestants. If one of the contestants is dominated by a member in the subset but the other is not, the non-dominated one is selected to survive. If both or neither are dominated, selection is based on the niche count of similar individuals in the attribute space. An individual with a low niche count is preferred to an individual with a high count to help maintain population diversity. Zitzler and Thiele [1999] developed a multiobjective genetic algorithm called the strengthen Pareto evolutionary algorithm (SPEA). SPEA uses two populations, P and P'. Throughout the process, copies of all non-dominated individuals are stored in P'. Each individual is given a fitness value, f_i, based on Pareto dominance. The fitness of the members of P' is calculated as a function of how many individuals in P they dominate.

  The individuals in P are assigned their fitness according to the sum of the fitness values for each individual in P' that dominate them plus one. Lower scores are better and ensure that the individual spawns a larger number of offspring in the next generation. Selection is performed using binary tournaments from both populations until the mating pool is filled. In this algorithm, fitness assignment has a built-in sharing mechanism. The fitness formulation ensures that non-dominated individuals always get the best fitness values and that fitness reflects the crowdedness of the surroundings.
- 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. Thereafter, the non-dominated individuals in the remaining population are identified and scored lower than the lowest one of the previously ranked individuals. Sharing is then applied to this second set of non-dominated individuals and the procedure continues until the whole population is ranked.

  Sharing is performed in the parameter space rather than in the attribute space. This means that the score of an individual is reduced according to how many individuals there are with

similar parameters, regardless of how different or similar they might be based on objective attributes.

Over the years, the main criticisms of the NSGA approach have been as follows:

1. High computational complexity of non dominated sorting;
2. Lack of elitism;
3. 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 and 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.

Constraints 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:

$$\left.\begin{array}{l} g(x) \leq c1 \\ r(x) < c2 \\ h(x) = c3 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} constr1(x) = c1 - g(x) \geq 0 \\ constr2(x) = c2 - r(x) > 0 \\ constr3(x) = c3 - h(x) = 0 \end{array}\right. \tag{5.2}$$

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:

$$\begin{array}{r} \overline{constr1}(x) = ((c1 - g(x))_1, \cdots, (c1 - g(x))_{n1}) = (contr1(x)_1, \cdots, contr1(x)_{n1}) \geq 0, \\ \overline{constr2}(x) = ((c2 - r(x))_1, \cdots, (c2 - r(x))_{n2}) = (contr2(x)_1, \cdots, contr2(x)_{n2}) > 0, \\ \overline{constr3}(x) = ((-|c3 - h(x)|)_1, \cdots, (-|c3 - h(x)|)_{n3}) = (contr3(x)_1, \cdots, contr3(x)_{n3}) = 0, \end{array} \tag{5.3}$$

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. The conversion of Eq. (5.2), that is a classical representation of constraints set, to Eq. (5.3) representation constitutes the first step of an unified formulation of constrained-optimization problems. In practice, due to round-off error on real numbers, the equality constraint *constr3* was modified as follows.

$$\begin{array}{c} \overline{constr3}(x) = (-|c3 - h(x)|_1 + \epsilon_1, \cdots, -|c3 - h(x)|_{n3} + \epsilon_{n3}) = \overline{contr3}(x) + \bar{\epsilon} \\ \bar{\epsilon} = (\epsilon_1, \cdots, \epsilon_{n3}), \forall i \in \{1, \cdots, n3\}, \epsilon_i \in R \end{array}$$

$\bar{\epsilon}$ 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.

From Eq.(5.2) and (5.3), 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, from Eq. (5.2), a multiobjective maximization problem. From a theoretical point of view, a constrained multiobjective optimization problem can be formulated as a two-step optimization problem. The first step implies the comparison of constraint satisfaction degrees between two solutions, using the Pareto's domination definition of Eq. (5.1), but a more simple solution consists in comparing the sum of values of violated constraints only, as in NSGA II algorithm of Deb et al. [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.

Constrained multiobjective optimization is important from the point of view of practical problem solving, but little attention has been paid so far. Constraint handling was the focus on some previous works developed in our reseach team [Ponsich et al., 2008].

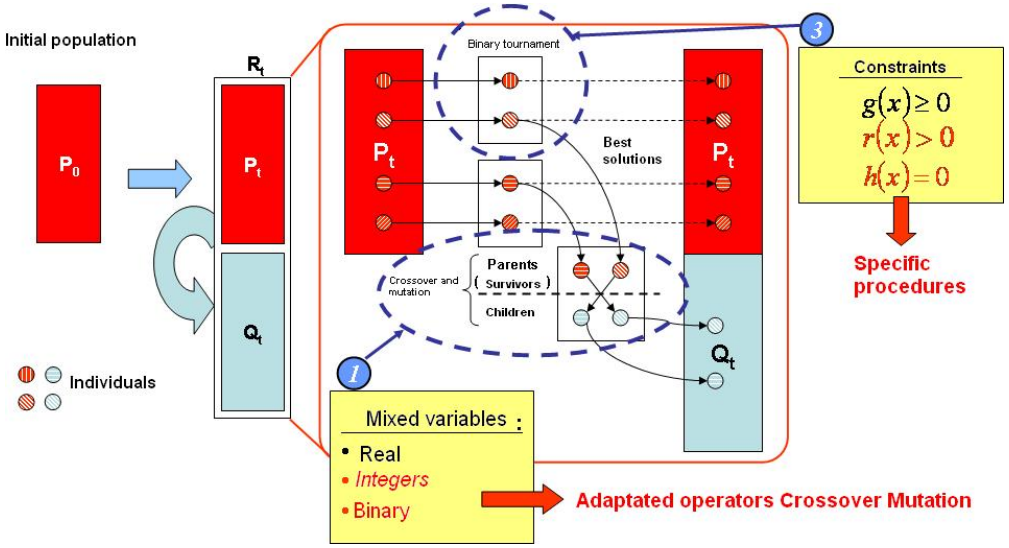
On four problems chosen from the literature [Deb et al., 2002], NSGA II has been compared with another recently suggested constraint-handling strategy proved to be more efficient. These results lead us to apply NSGA II to our problem.

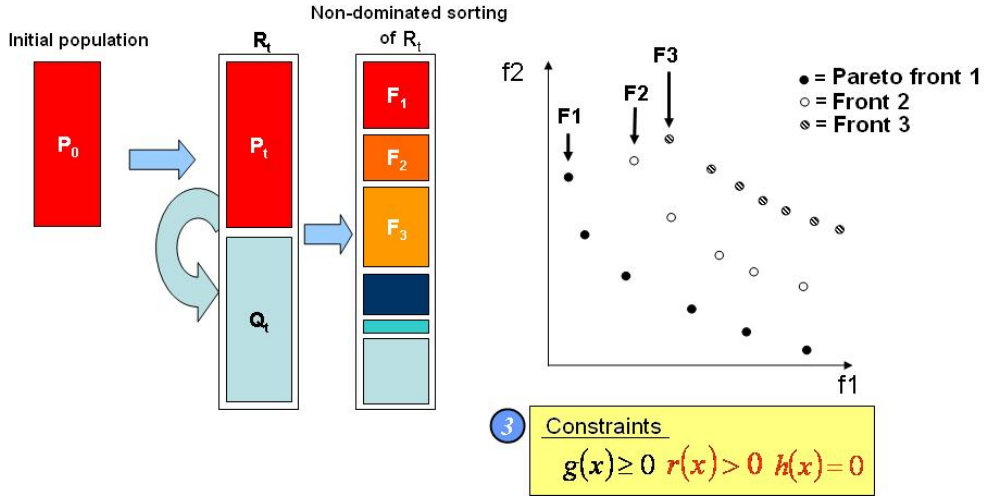
For all these reasons, NSGA II has been chosen as the multiobjective strategy. Its principles are now recalled.

5.2.6 Principles of Non-Sorted Genetic Algorithm II (NSGA II)

Initially, a random parent population P_o of size N is created. The population is sorted based on the non domination principle. Each solution is assigned a fitness (or rank) equal to its non domination level (1 is the best level, 2 is the next-best level, and so on). Thus, maximization of fitness is assumed. At first, the usual binary tournament selection, recombination, and mutation operators are used to create an offspring population Q_o of size N. Since elitism is introduced by comparing current population with the previously best found non dominated solutions, the procedure is different after the initial generation. The step-by-step procedure illustrated in Figure 5.5 shows that NSGA II algorithm is simple and straightforward. First, a combined population $R_t = P_t \; U \; Q_t$ is formed. The population R_t is of size 2N. Then, the population is sorted according to non domination.If the size of F_1 is smaller then N, we definitely choose all members of the set F_1 for the new population P_{t+1}. The remaining members of the population P_{t+1} are chosen from subsequent non dominated fronts in the order of their ranking. Thus, solutions from the set F_2 are chosen next, followed by solutions from the set F_3, and so on. This procedure is continued until no more sets can be accommodated. Say that the set F_l is the last non dominated set beyond which no other set can be accommodated. In general, the number of solutions in all sets from F_1 to F_l is greater than the population size.

To choose exactly population members, we sort the solutions of the last front using the crowded-comparison operator in descending order and choose the best solutions needed to fill all population slots. The new population P_{t+1} of size N is now used for selection, crossover, and mutation to create a new population Q_{t+1} of size N. It is important to note that we use a binary tournament selection operator but the selection criterion is now based on the crowded-comparison operator. Since this operator requires both the rank and crowded distance of each solution in the population, these quantities are calculated while forming the population P_{t+1}, as shown in the following algorithm.





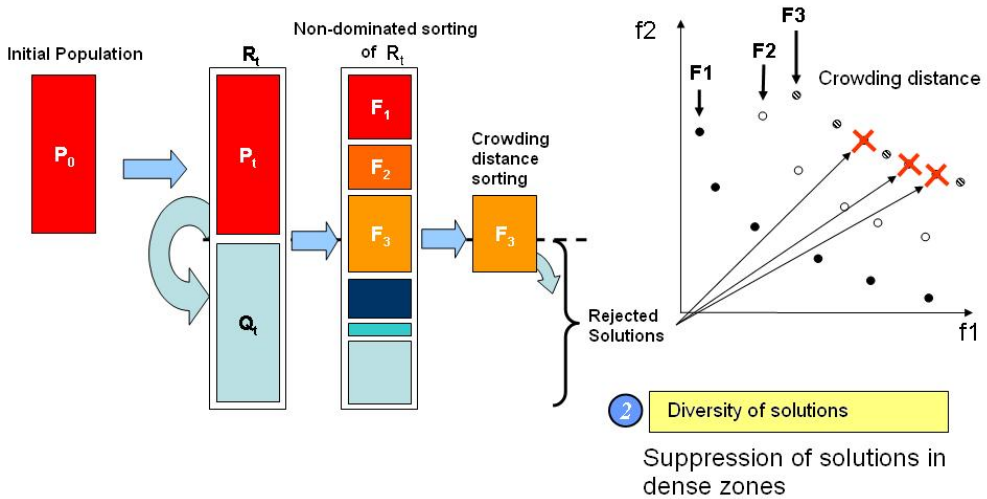


Figure 5.5: NSGA II Reproduction

5.2.7 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 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:

$$\begin{aligned} Card(S) = \; & (n_{d_1} + n_{d_2} + \ldots + n_{d_{M_D}})! \\ & + \sum_{p=1}^{n_{d_1}+n_{d_2}+\ldots+n_{d_{M_D}} - M_D} (n_{d_1} + n_{d_2} + \ldots + n_{d_{M_D}} - p)! \times (\textstyle\sum_{i,j,\ldots,k} C^i_{n_{d_i}} C^j_{n_{d_j}} \ldots C^k_{n_{d_k}}) \\ & i + j + \ldots + k = n_{d_1} + n_{d_2} + \ldots + n_{d_{M_D}} - p; \\ & i \leq n_{d_1}, j \leq n_{d_2}, \ldots, k \leq n_{d_{M_D}} \end{aligned}$$

The combinatorics related to an example involving drugs A and B for the disease I, C and D for the disease II and F for disease III is presented in Table 5.1.

The total number of possibilities can be computed as follows:

$$5! + (4!)(2C_2^1C_1^1) + (3!)(2C_2^1C_1^1) = 5! + (4!)4 + (3!)4 = 240$$

This can be applied to the example which serves as a test bench (see Chapter 2) involving 3 diseases (4 drugs for d_1; 4 drugs for d_2; 1 drug for d_3).

$$\begin{aligned} Card(S) = \; & (9)! + (8)!(2C_4^3C_4^4C_1^1) + (7)!(2C_4^4C_4^2C_1^1 + 2C_4^3C_4^3C_1^1) + (6)!(2C_4^3C_4^2C_1^1) \\ & + (5)!(2C_4^2C_4^2C_1^1) + (4)!(2C_4^1C_4^2C_1^1) + (3)!(2C_4^1C_4^1C_1^1) \end{aligned}$$

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.

Number of drugs released				
5	4		3	
5!	4!	A C D E B C D E A B C E A B D E	3!	A C E A D E B C E B D E

Table 5.1: Example of combinatorial calculation

5.3 Implementation of the NSGA II key procedures for NPD modelling

The methodology used for solving 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 NSGA II 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.

5.3.1 Coding, crossover and mutation

Sequence generation. A sequence is modeled 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 Pi 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 the studied examples and has thus been selected in this work. Figure 5.6 illustrates the used coding representing

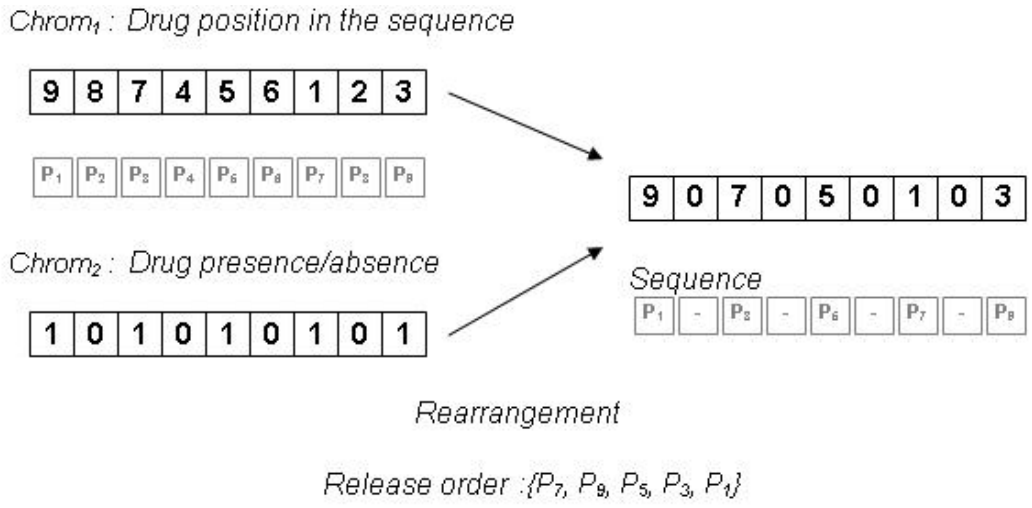


Figure 5.6: Coding for generating a sequence

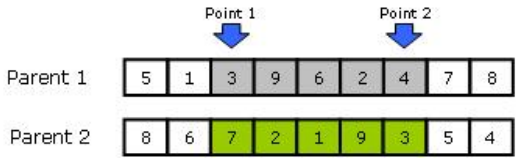
a solution that is then evaluated by the simulator for 5 products. Crossover and mutation have been carried out by specific procedures for each type of chromosome.

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.

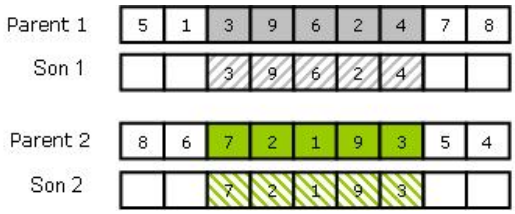
For this purpose, a crossover operator with respect to genotype constraints without clone generation in the offspring genetic code has been carried out, the so-call MPX operator (Maximal Preservative X) [Andersson, 1999]. Its principle is recalled below.

MPX Crossover (Maximal Preservative X): This crossover has been proposed by Mülhenbein [9] and [10] for the traveling salesman problem. This operator's idea is to insert a segment of parent's chromosome in the chromosome of the other parent so that the resulting crossover is closer to his parents. It is a two-point crossover. The two sons are obtained in a symmetrical manner. The following example illustrates this process:

The zone of crossover lies between the positions Point 1 and Point 2 .



First, the crossover zone of $parent_1$ ($parent_2$) is copied into son_1 (son_2).



Then, the son's gene that is not in the zone of crossover is completed in the following way:

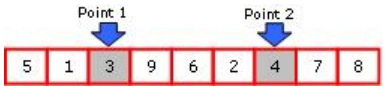
The i^{th} gene of $parent_2$ is copied on the i^{th} gene of son_1.

Otherwise, the i^{th} gene of the $parent_1$ is copied on the i^{th} gene of son_1 if it does not create any duplicates. If the two previous cases cannot apply, the i^{th} gene of the son_1 receives a gene of the crossover zone of the $parent_2$ (the first one is not taken).

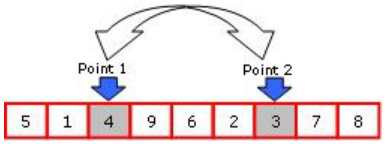
Mutation We choose a mutation that randomly permutes two genes of a chromosome. This operator is applied to the individuals derived from crossover with an adapted rate (preferably 0.5). Then, we place the new offspring in a new population P_{i+1}.

For mutation, two steps are involved

1. Two points are calculated



2. The gene value in locus 1 is copied into locus 2, and conversely.



5.3.2 Optimization parameters

The optimization parameters are presented in Table 5.2.

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

Table 5.2: Optimization parameters

5.3.3 Optimization criteria

The optimization criteria are evaluated by use of the previous developed discrete event simulator,involving:

- 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:

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

- 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 on the Positivity Probability defined in section 3.3:

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

- the makespan of a sequence is computed from the average makespan of the samples:

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

$$Where$$
$$W \rightarrow number\ of\ drugs\ in\ the\ sequence$$
$$n \rightarrow number\ of\ runs\ by\ sequence;$$
$$dur \rightarrow Makespan\ for\ a\ sequence;$$

5.3.4 Constraints

It must be emphasized that the resource constraints have already been taking 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:

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

For the example, 3 diseases and 9 molecules have been considered.

- Disease M1 : Products P2, P3, P6, P7 ;
- Disease M2 : Products P4, P5, P8, P9 ;
- Disease M3 : Product P1 ;

This can be illustrated for the example which serves as a guideline for the proposed methodology. These constraints involve at least one gene value equal to 1 for chromosome Chrom2 in the loci corresponding to the genes of the products of the given disease g_i, that is:

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

5.4 Result presentation and discussion

The case study results are 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.

5.4.1 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 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 hours for this study with 9 drugs for 3 diseases.

5.4.2 Bicriteria Optimization *Net Present Value-Risk*

A first study concerns the Net Present Value and Risk to optimize simultaneously. The general behavior of the whole population (16000 individuals) is displayed in Figure 5.7.

The improvement of the procedure can be observed through the evolution of the population along the progression of the algorithm through the generations 40, 80, 120, 160 and 200, as presented in Figures 5.8 to 5.9 where only an optimization run is illustrated. As might be expected, random initialization of the population has resulted in a dispersion of the various solutions. It can be seen in this example that the convergence of the algorithm can be tracked visually. 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 :

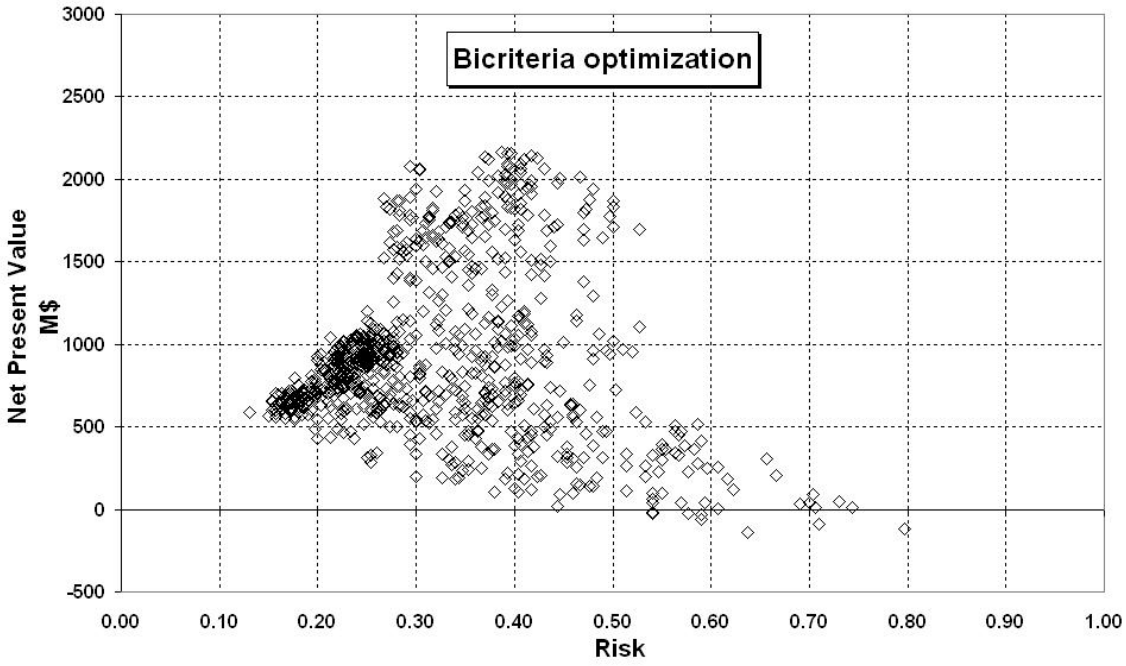


Figure 5.7: Whole population for the case Net Present Value/Risk

- 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 5.3 (4 drugs) and 5.4 (6 drugs).

As it can be seen in Figure 5.10, 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 an 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 individuals of the first generation are positioned at the lower part of the figure, with a lot of solutions exhibiting poor performances (high values for risks and negative values for net present values). This behavior is strongly improved along the generations: for the first generations, it can be observed that both criteria are improved simultaneously (a gain of 2000 $ for NPV whereas the worst value for risk is improved by 30%). Then, the risk ratio seems stabilized and from generation 160, only the NPV criterion seems slightly improved till the end of optimization.

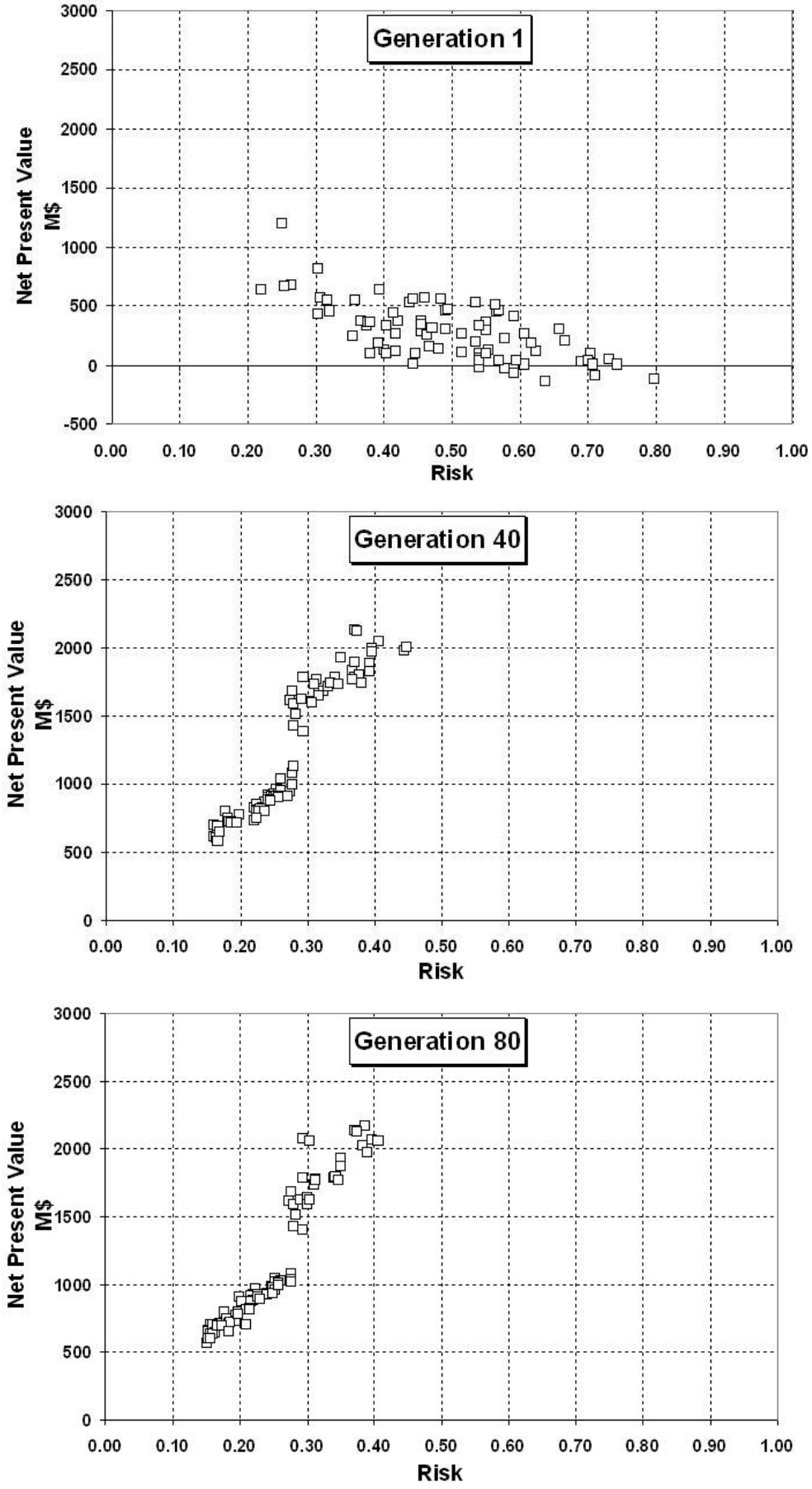


Figure 5.8: Behavior for generations 1, 40 and 80. Net Present Value/Risk case

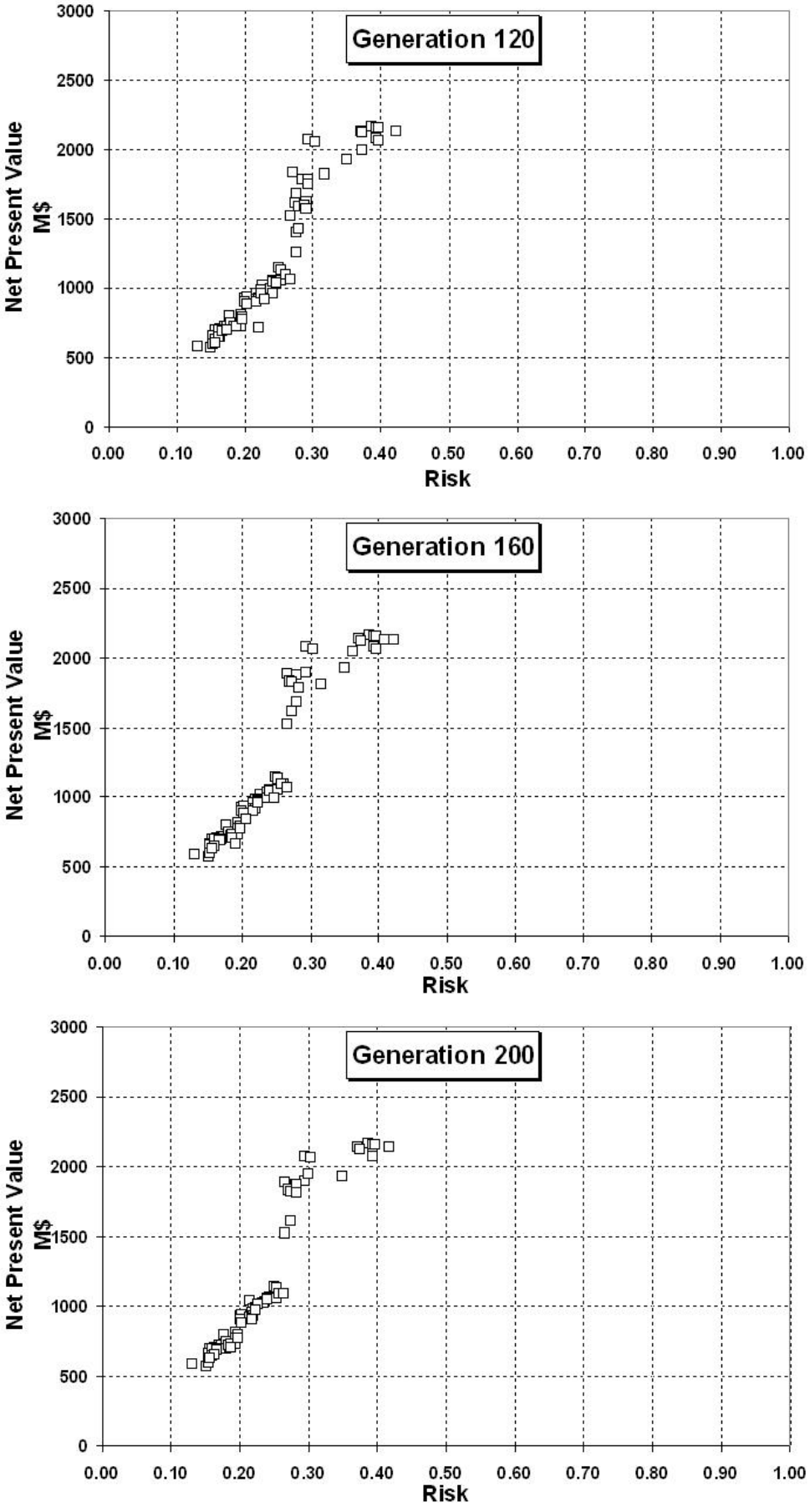


Figure 5.9: Behavior for generations 120, 160 and 200. Net Present Value/Risk case

Solutions	NPV (M$)	Risk (%)	Makespan by simulation	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

Table 5.3: Net Present Value and Risk for 4-drug sequences (9 drug-portfolio optimization)

Solutions	NPV (M$)	Risk (%)	Makespan by simulation	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 5.4: Net Present Value and Risk for 6-drug sequences (9 drug-portfolio optimization)

The union of the Pareto fronts obtained from the optimization runs, can be visualized in Figure 5.10 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. In these figure, are illustrated sequences for 4 drugs (P1, P2, P5, P7) but, other sequences for 4 drugs were found containing drugs P1, P3, P5 and P9 from later optimizations. That's why drugs P1, P3, P5, P9 are evaluated by simulation.

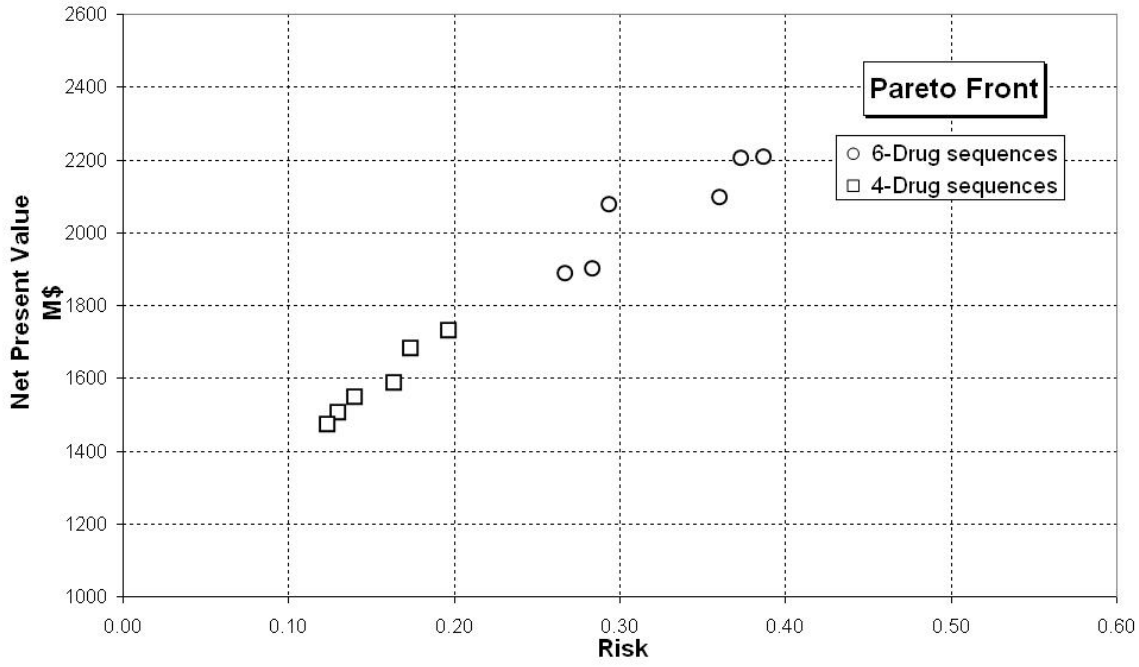


Figure 5.10: Pareto Front. Net Present Value/Risk

For these two distinct behaviors, 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 the so-called attrition phenomenon occurring in NPD problems. It is also observable that the positioning of the solutions found by NSGA II in the objective space and even their occurrence within the optimized population is not so intuitive from the bubble charts presented in Chapter 3 (Figures 3.15 and 3.16). 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 behavior without the use of a numerical tool. Let us recall that in Chapter 3 (section 3.3.4), ten sequences randomly selected plus two other ones were simulated for fixed release orders. For the 12 sequences evaluated by simulation, the same behavior 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). 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.

An additional analysis has thus been performed to refine the search for 4 and 6-product sequences.

Since two groups of products for a 4-product sequence are systematically found after the Pareto front procedure of the last generation, a validation step was performed with these two groups of products with enumerative simulations of the 24 possible sequences. As it was mentioned before, the 4-drug sequences found by the optimization runs are now evaluated by simulation: they are relative to sequences involving on the one hand drugs 1, 2, 5, 7 and on the other hand drugs 1, 3, 5, 9.

The objective is to see if the release order for the non-dominated solutions can be found by simulation. The solutions obtained by simulations are then ranked by a Pareto-sort procedure.

The comparison of both Pareto fronts is illustrated in Figure 5.11. The non-dominated individuals lie between 1308 and 1726 M$ for the NPV (respectively between 13-23 % for the Risk). These values exhibit the same order of magnitude as the values for the case of Figure 5.10 (see also Table 5.3). The values for NPV and Risk are presented in detail in Table 5.5. The same sequences as those previously found for the 9-product optimization are found, which validates the optimization procedure. This also validates the coding phase, for which an individual is not represented by a unique chromosome.

Solutions	Net Present Value (M$)	Risk (%)	Release order			
1	1726	23	P2	P7	P5	P1
2	1683	17	P7	P5	P2	P1
3	1654	17	P5	P7	P2	P1
4	1581	17	P7	P5	P2	P1
5	1563	15	P5	P2	P7	P1
6	1454	14	P5	P7	P2	P1
7	1308	13	P7	P5	P2	P1

Table 5.5: Net Present Value and Risk for 4-drug sequences (4-drug portfolio simulation)

The same approach was carried out for the P1, P3, P5, P9-drug set.

For sequence 1, 3, 5, 9, the NPV ranges between 573 and 1016 M$ (respectively between 14 and 28% for Risk), as it is reported in Table 5.6. The Pareto fronts for the two 4-drug sequences evaluated by simulation are displayed in Figure 5.12, where it can be observed that sequence P1, P2, P5, P7 exhibits always better performances according to both criteria, NPV and Risk.

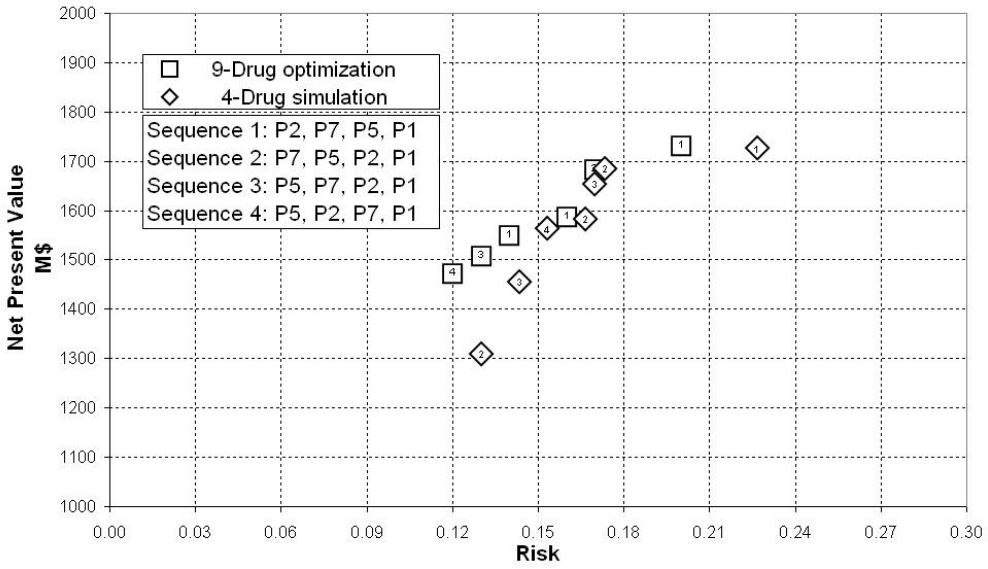


Figure 5.11: Comparison between 9-drug portfolio optimization and 4 drug simulation (1, 2, 5, 7)

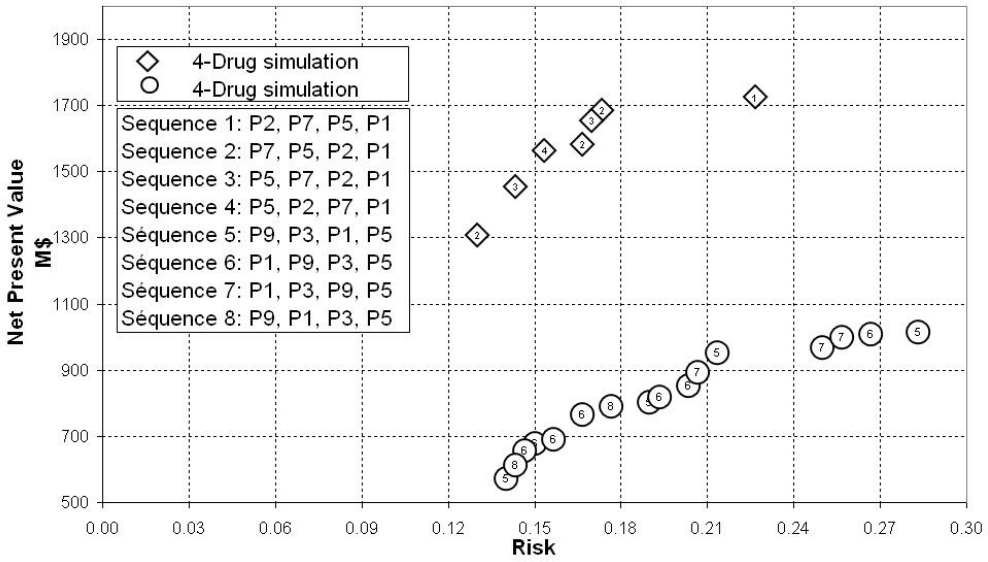


Figure 5.12: Comparison between and 4 drug portfolio simulation (P1, P2, P5, P7) and (P1, P3, P5, P9)

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

Solutions	Net Present Value (M$)	Risk (%)	Release order			
1	1016	28	P9	P3	P1	P5
2	1008	27	P1	P9	P3	P5
3	1000	26	P1	P3	P9	P5
4	968	25	P1	P3	P9	P5
5	953	21	P9	P3	P1	P5
6	892	21	P1	P3	P9	P5
7	851	20	P1	P9	P3	P5
8	819	19	P1	P9	P3	P5
9	801	19	P9	P3	P1	P5
10	791	18	P9	P1	P3	P5
11	764	17	P1	P9	P3	P5
12	689	16	P1	P9	P3	P5
13	679	15	P1	P9	P3	P5
14	657	15	P1	P9	P3	P5
15	612	14	P9	P1	P3	P5
16	573	14	P9	P3	P1	P5

Table 5.6: Net Present Value and Risk for 4-drug sequences.(4-drug portfolio simulation)

order to examine how the number of products evolves naturally along the generations. The results are presented in Table 5.7 for generations 1, 40, 80, 120, 160 and 200. It can be clearly observed that the sequences with a low number of products are favored in the optimization process.

	Number of drugs in the sequence								
Generation	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

Table 5.7: Evolution of the number of drugs by sequence

Figure 5.13 presents the typical evolution for some generations. It must be observed that the initial generation exhibits solutions that are not as dispersed as those obtained with a random initialization procedure. Once more, the optimization procedure seems efficient to improve both criteria. Figure 5.14 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 5.8 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.

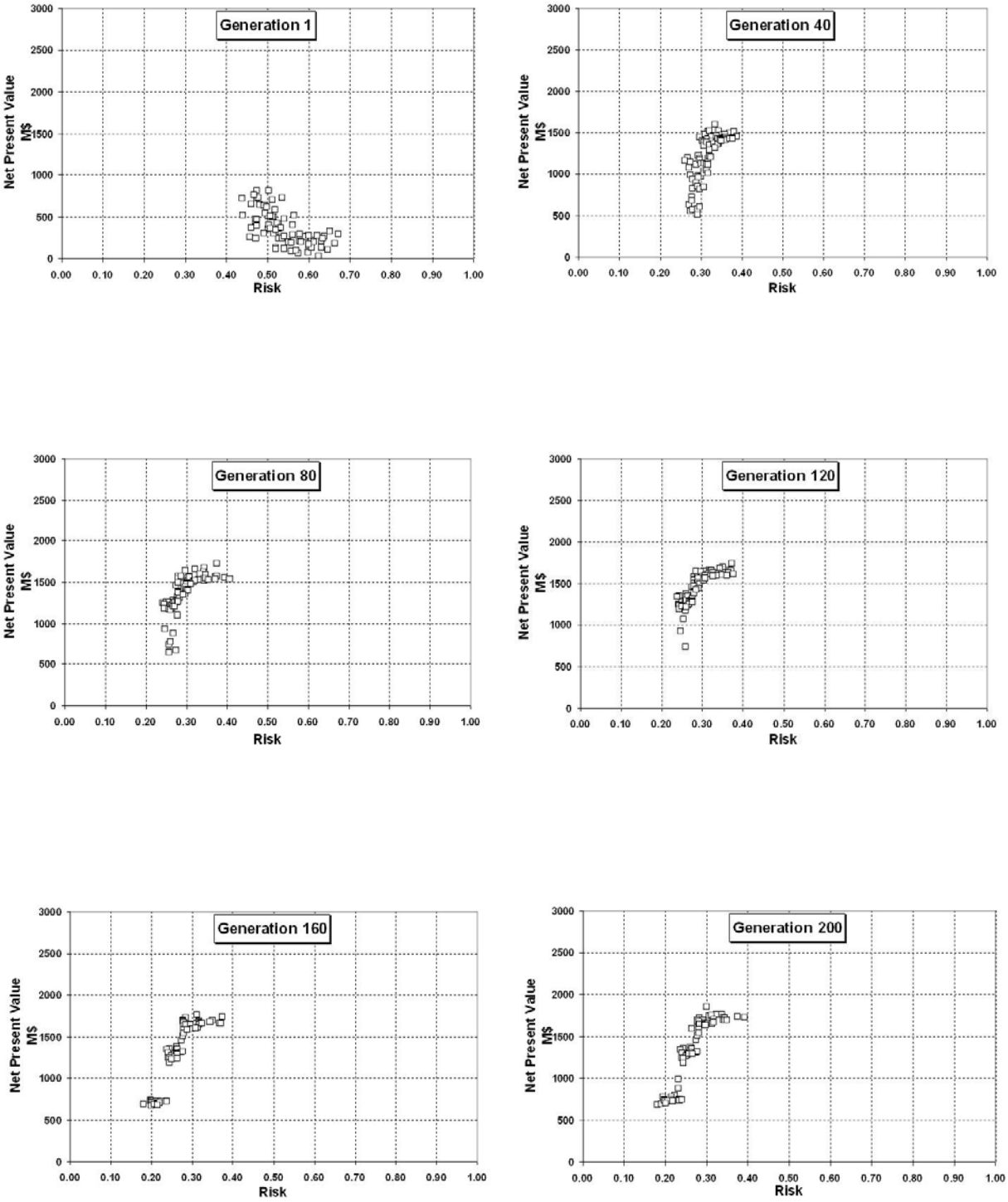


Figure 5.13: Evolution for the optimization Net Present Value/Risk for 9-drug initialized portfolio

Solutions	Net Present Value (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	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	P7	P3	P6	P5	P2	P1	
9	1722	28	P1	P7	P6	P5	P2	P3	
10	1696	28	P7	P1	P6	P5	P2	P3	
11	1355	24	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

Table 5.8: Results of Net Present Value and Risk for 9-drug initialized portfolio (Pareto front solutions)

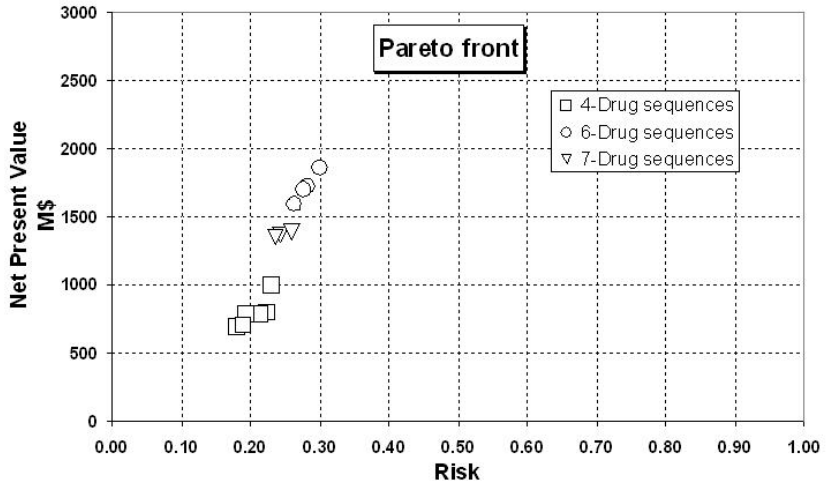


Figure 5.14: Pareto front for the optimization Net Present Value/Risk for 9-drug initialized portfolio

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). The results exhibit bad performances for both criteria as compared with those previously obtained (Figure 5.15).

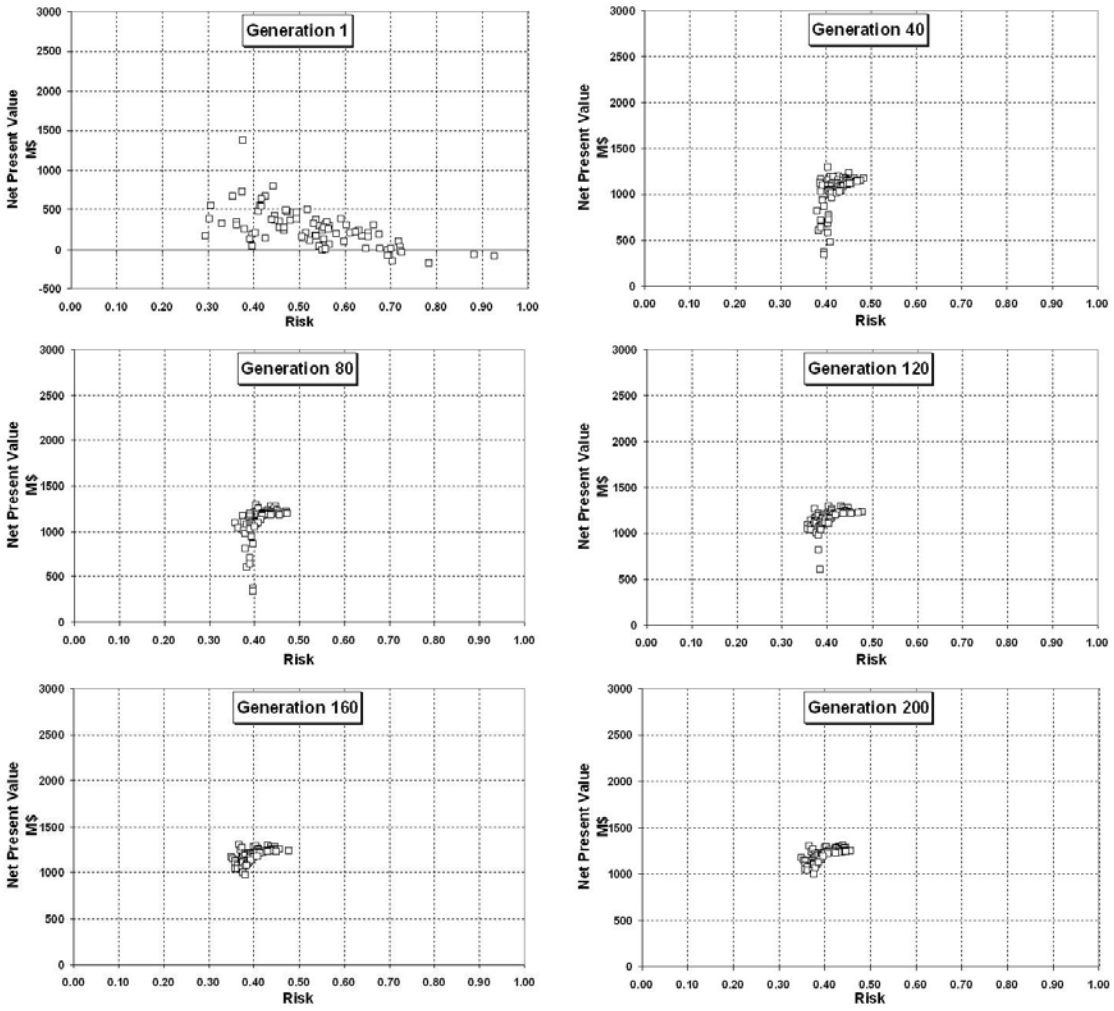


Figure 5.15: Evolution for the optimization Net Present Value/Risk for 9-drug optimized portfolio

This is confirmed by the Pareto front (Figure 5.16) 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 5.9. 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.

Solutions	Net Present Value (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

Table 5.9: Results of Net Present Value and Risk for 9-drug optimized portfolio (Pareto front solutions)

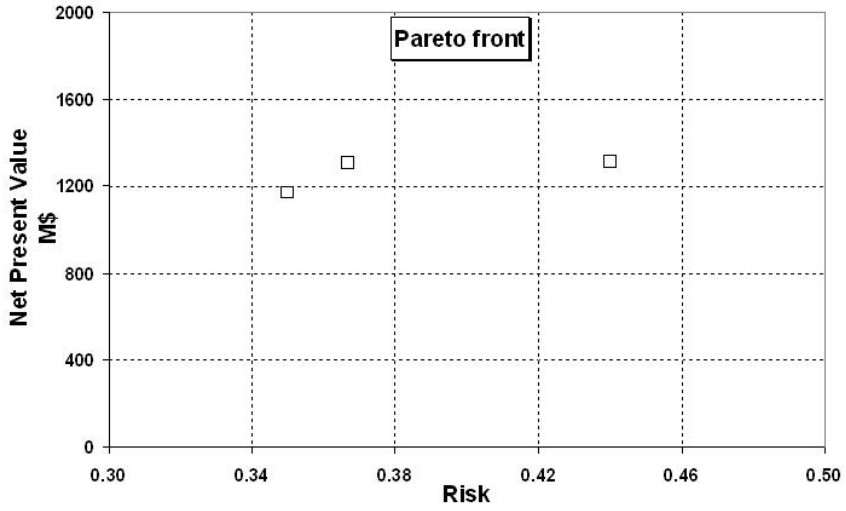


Figure 5.16: Pareto front for the optimization Net Present Value/Risk for 9-drug optimized portfolio

Bicriteria optimization (Net Present Value-Risk) for a 6-drug portfolio

The study concerning the 6-drug portfolio identified as a good candidate (P1, P2, P3, P5, P6, P7) was carried by optimization instead of analyzing the enumerative simulation pathway. The evolution is presented in Figure 5.17 for generations 1, 40, 80, 120, 160 and 200 and the Pareto front. The numerical values are presented in Table 5.10. A first comment is that the previous 4-drug portfolio which has been identified as a potential candidate is a subset of the 6-drug solutions.

Solutions	Net Present Value (M$)	Risk (%)	Release order					
1	2555	38	P3	P5	P6	P7	P2	P1
2	2512	35	P3	P5	P6	P7	P2	P1
3	2476	31	P3	P5	P6	P7	P2	P1
4	2340	29	P3	P5	P6	P7	P2	P1
5	2330	28	P3	P5	P6	P7	P2	P1
6	1658	26	P3	P5	P7	P6	P2	P1
7	1647	25	P3	P5	P7	P6	P2	P1
8	1553	24	P3	P5	P7	P6	P2	P1

Table 5.10: Net Present Value and Risk for 6-drug sequences (6-drug optimization)

Two sequences are found several times (P3, P5, P7, P6, P2, P1) and (P3, P5, P6, P7, P2, P1) with the same products and only a permutation between products 6 and 7, they have also been identified in the 9-drug optimization (Figure 5.18).

5.4.3 Bicriteria optimization *Net Present Value-Makespan*

The second optimization study is based on Net Present Value-Makespan (expressed in days). As for the NPV-Risk optimization, the general behavior is illustrated in Figure 5.19 at the top.

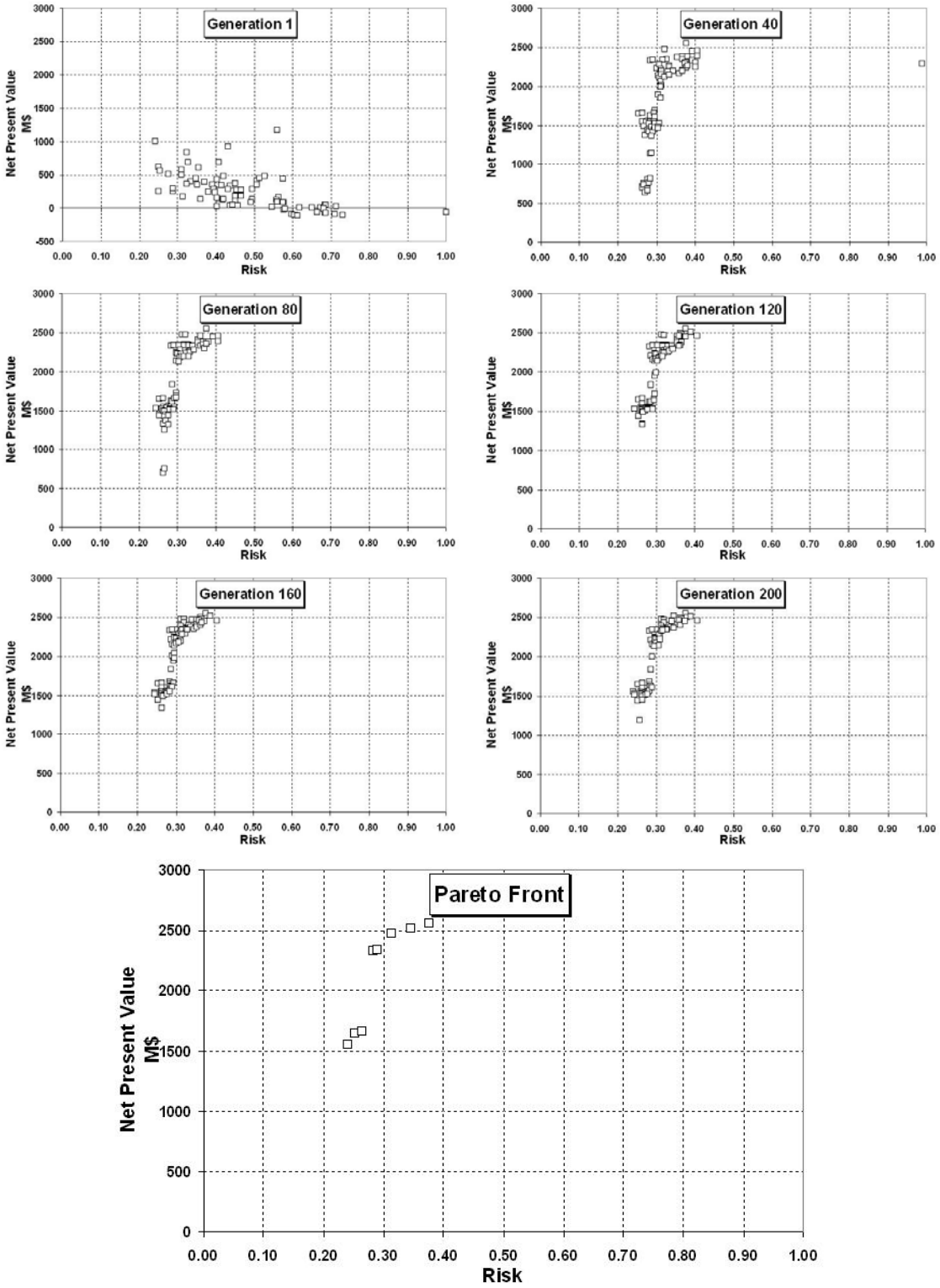


Figure 5.17: Optimization results for Net Present Value/Risk criteria for 6 drugs (P1, P2, P3, P5, P6, P7) Pareto Front

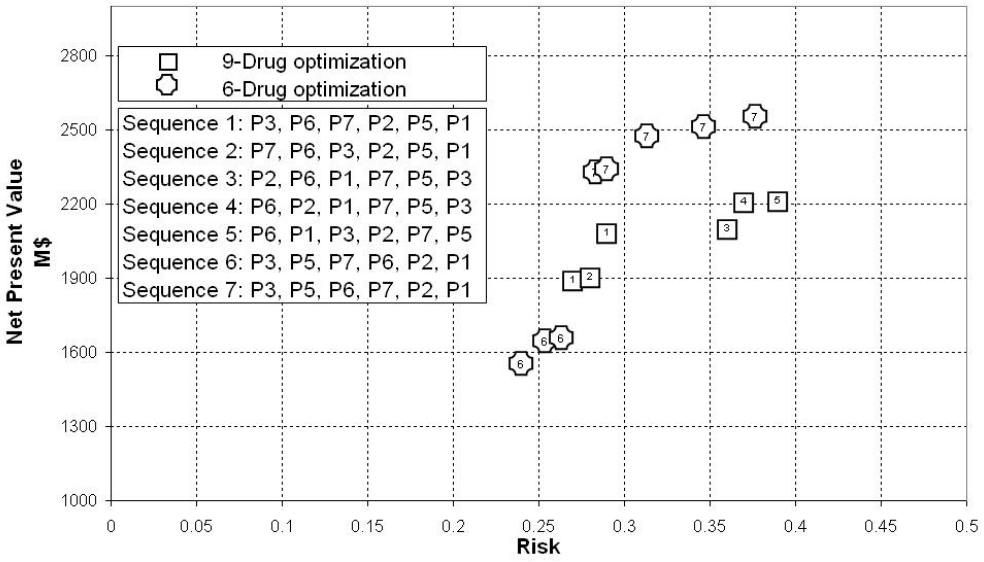


Figure 5.18: Relationship 9 and 6 drug optimization (P1, P2, P3, P5, P6, P7)

The improvement of the strategy along the generations is presented in Figures 5.19, 5.20 and 5.21. After generation 80, the behavior is almost stabilized and there is a slight (respectively negligible) improvement for NPV (respectively for duration).

The results relative to this pair of criteria are presented in Figure 5.22. 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 5.11 and 5.12).

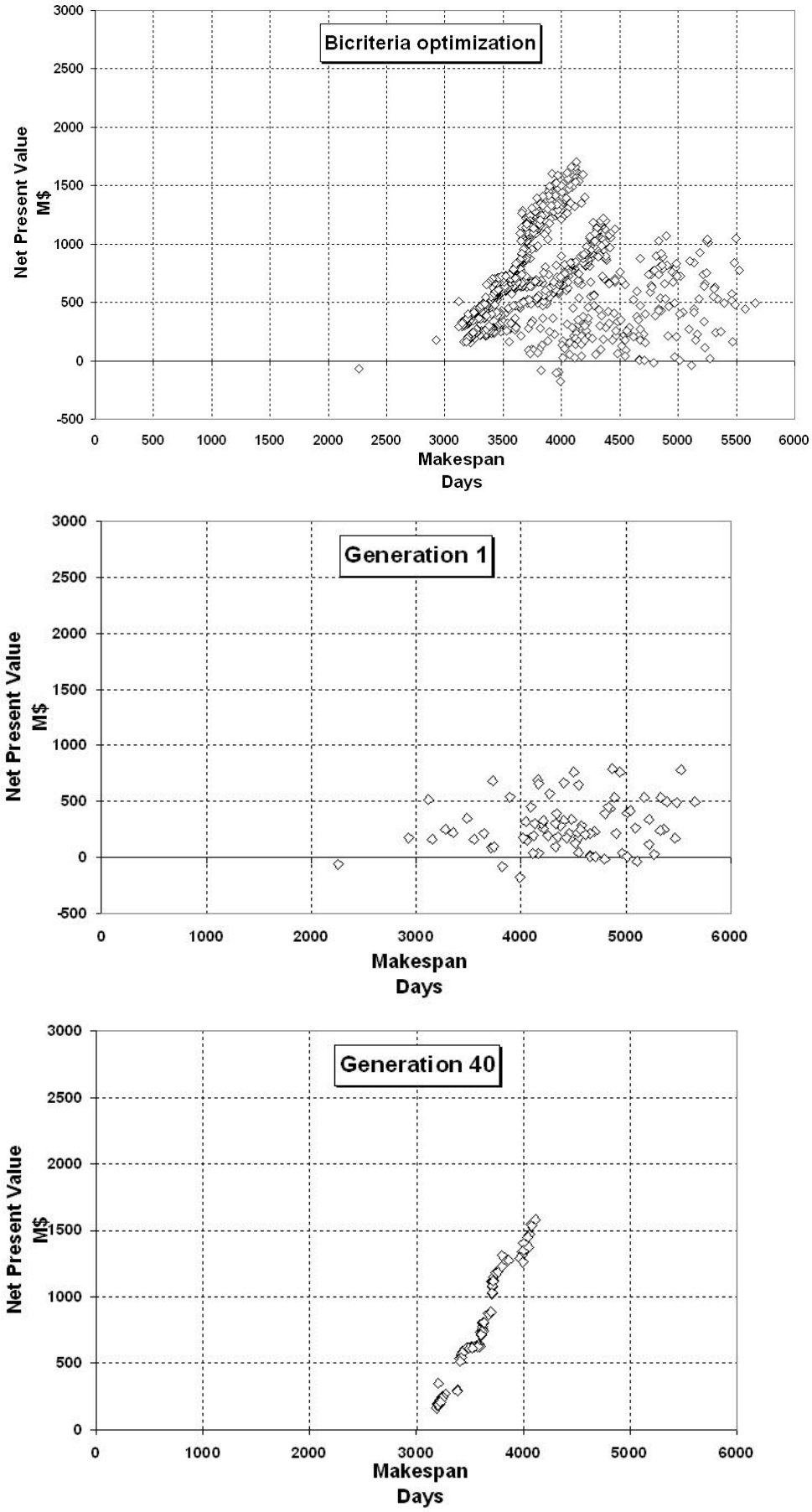


Figure 5.19: Whole population and behavior for generations 1 and 40. Net Present Value-Makespan case

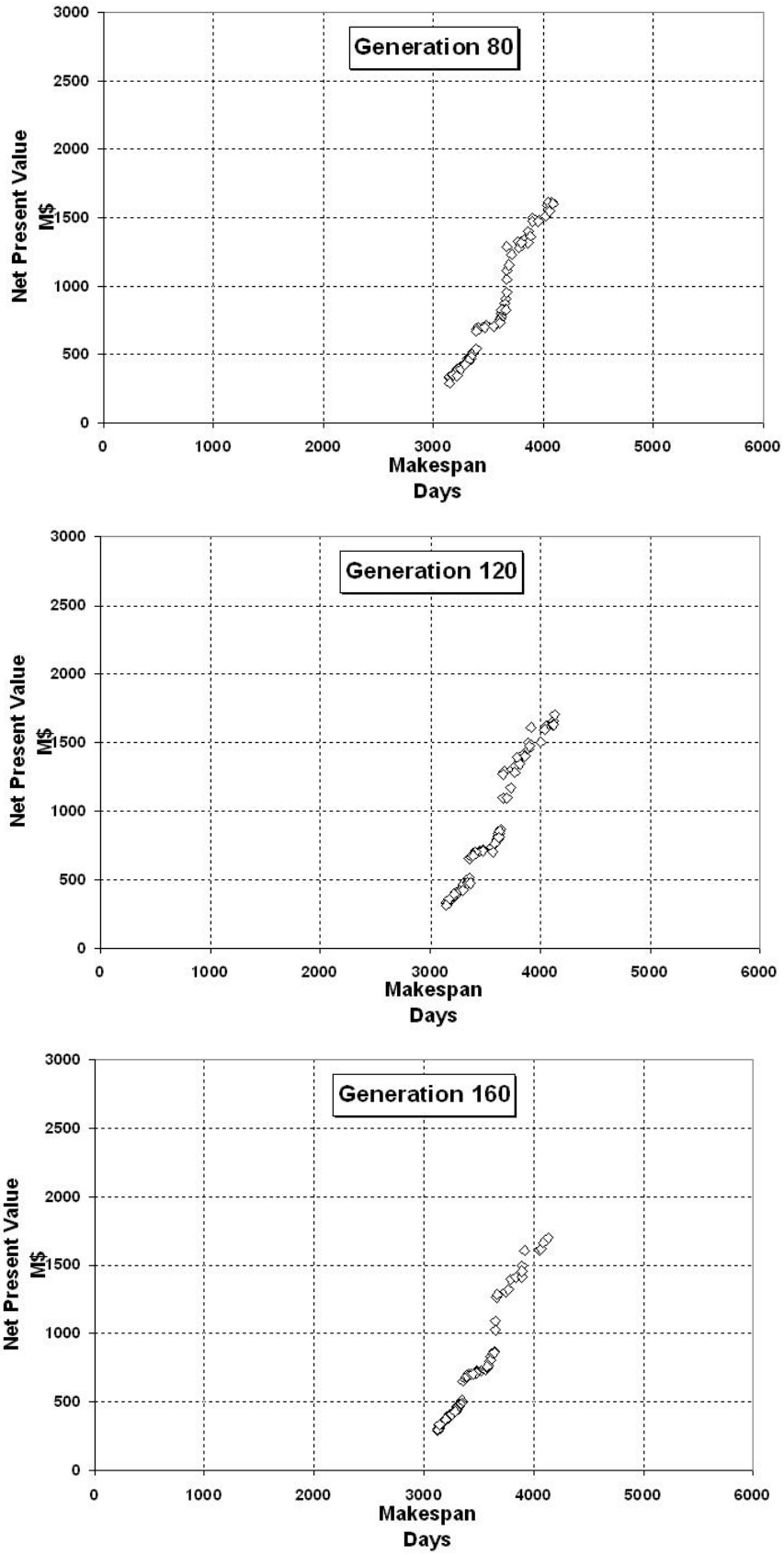


Figure 5.20: Behavior for generations 80, 120 and 160. Net Present Value-Makespan case

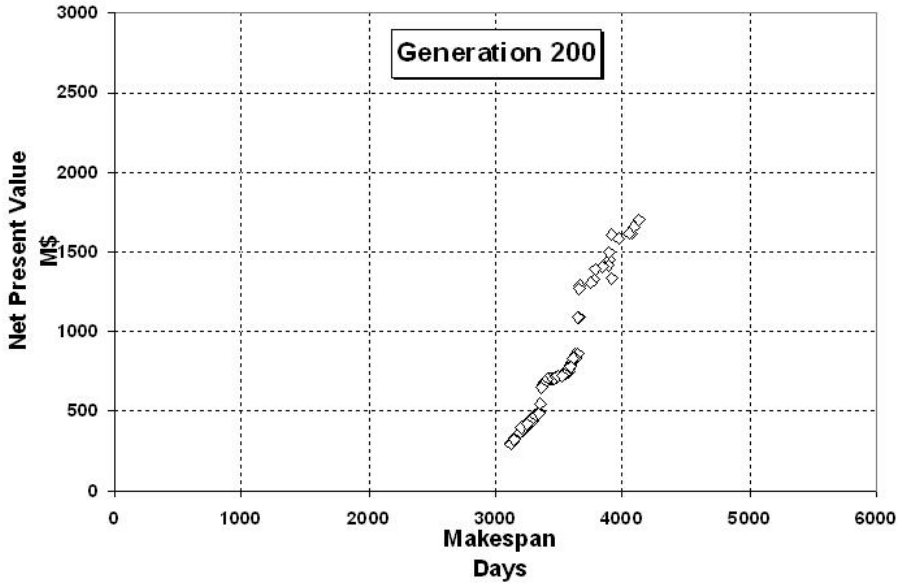


Figure 5.21: Behavior for generation 200. Net Present Value-Makespan case

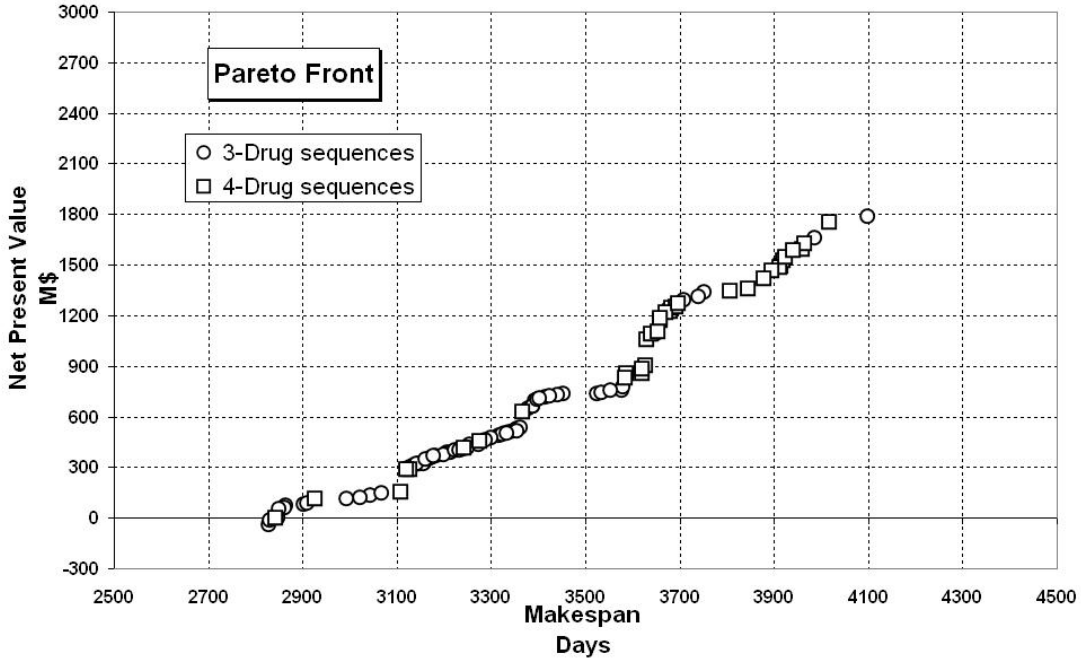


Figure 5.22: Pareto Front. Net Present Value/Makespan

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 5.11: Net Present Value and makespan for 3-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

Table 5.12: Net Present Value and makespan for 4-drug sequences

5.4.4 Bicriteria optimization *makespan-risk*

As before, all the individuals evaluated along the generations are presented in Figure 5.23. The progression of the algorithm is illustrated in Figures 5.24 and 5.25. As for the previous cases (NPV-Risk,NPV-Duration), an important improvement is observed from generation 1 to generation 40.

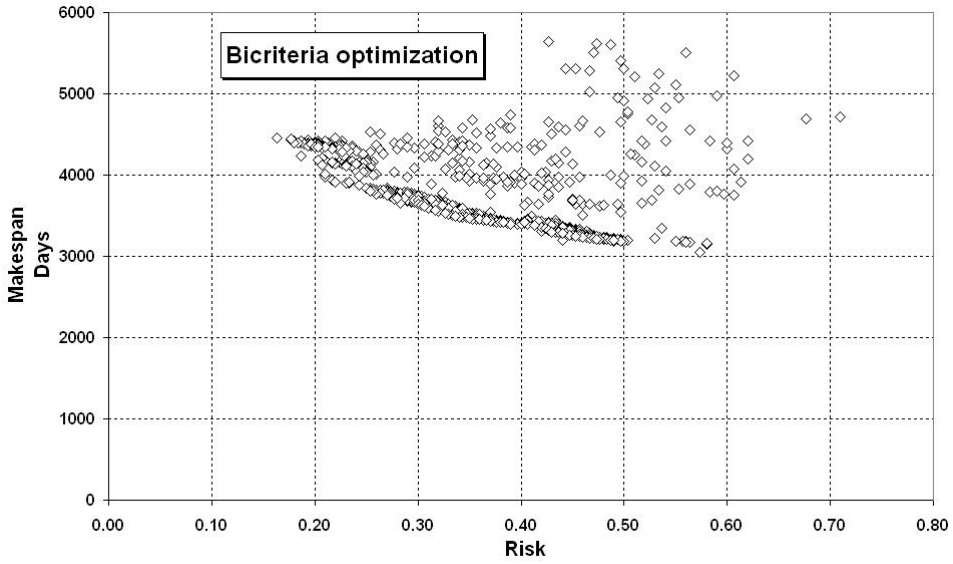


Figure 5.23: Whole population for the case Risk/Makespan

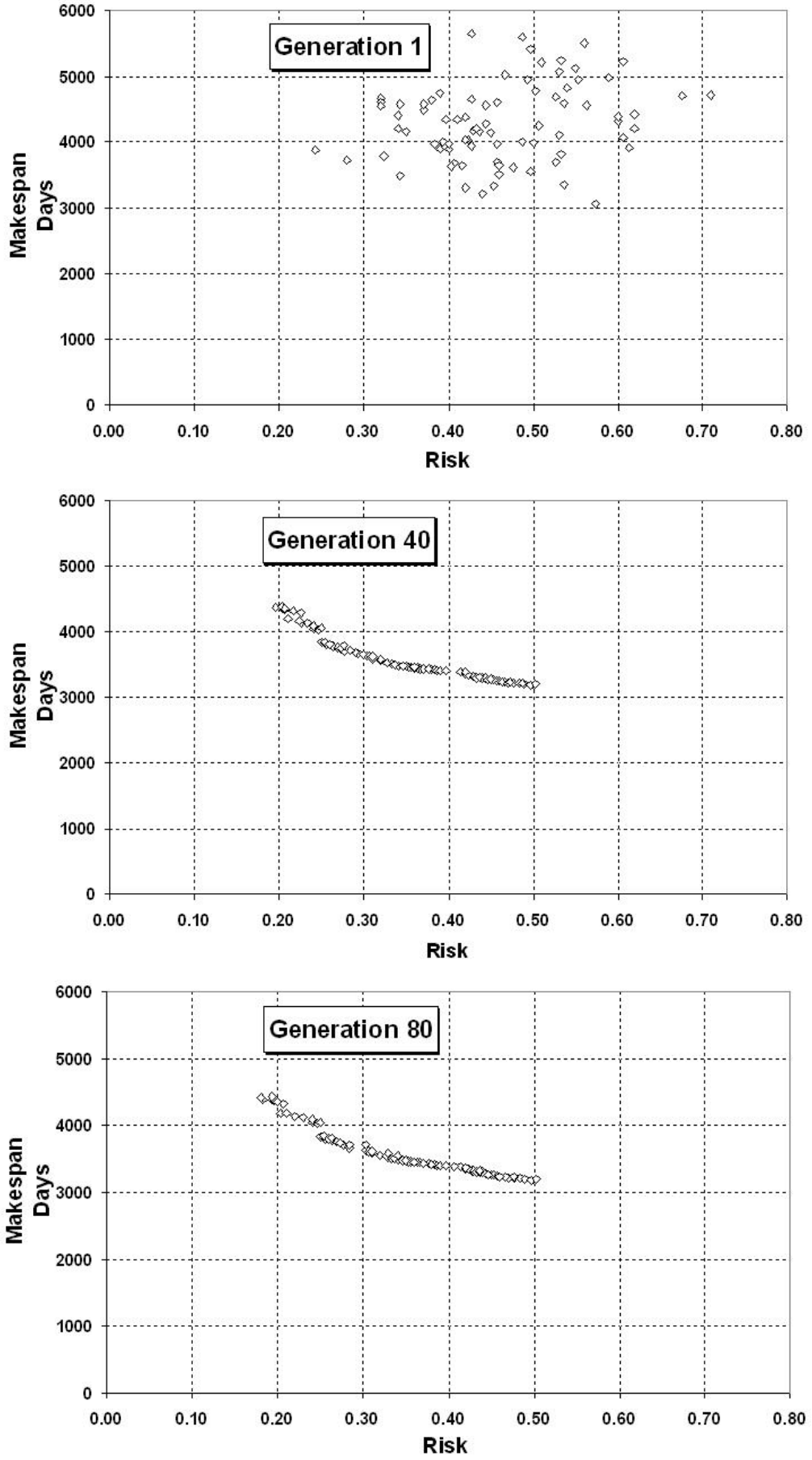


Figure 5.24: Behavior for generations 1, 40 and 80. Risk/Makespan case

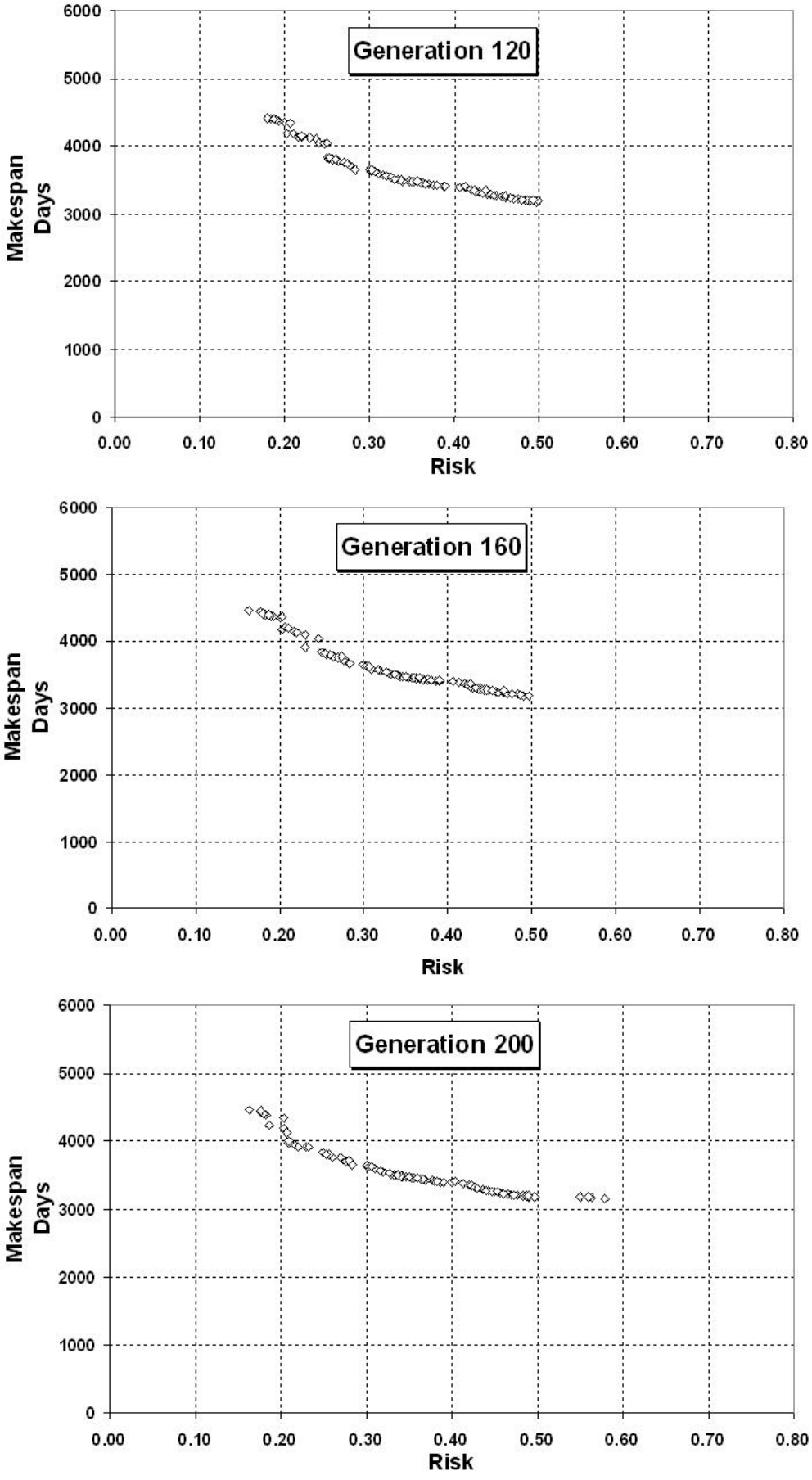


Figure 5.25: Behavior for generations 120, 160 and 200. Risk/Makespan case

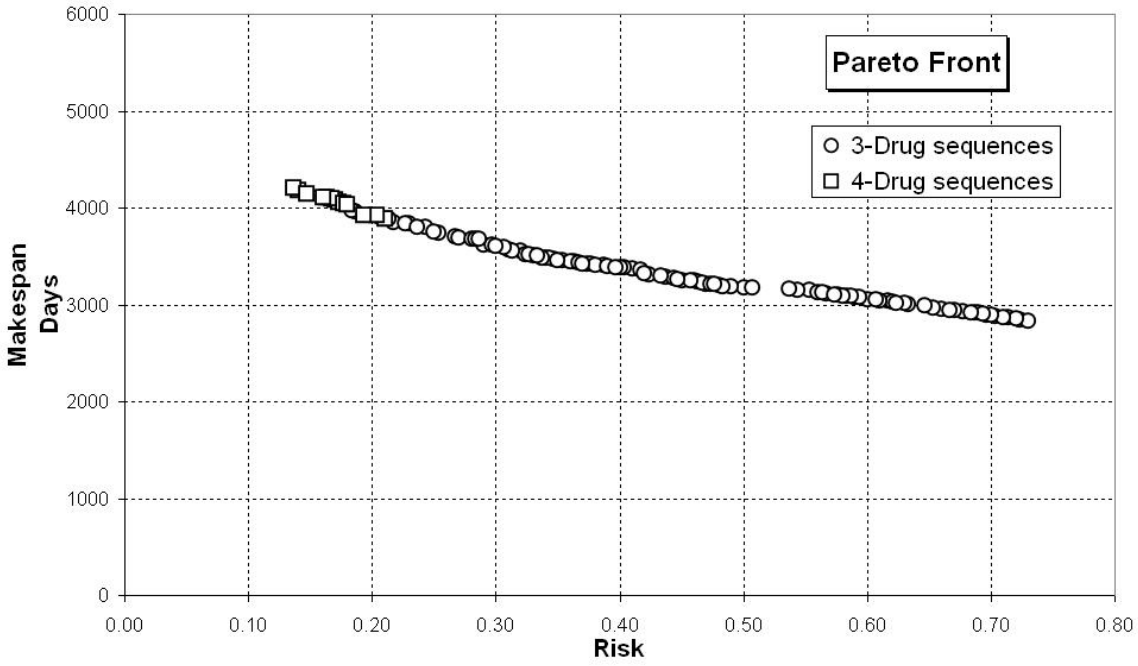


Figure 5.26: Pareto Front Makespan/Risk.

The results of the bicriteria optimization makespan-risk are presented in Figure 5.26. 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 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 5.13 and 5.14).

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

Table 5.13: Risk and Duration for 3-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

Table 5.14: Risk and Duration for 4-drug sequences

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 favorably 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.

5.5 Tricriteria optimization *NPV-Duration-Risk*

5.5.1 Study presentation

In this study, the tricriteria optimization is performed NPV-Duration-Risk. To make the interpretation easier, the results relative to a given pair of criteria are presented as a projection on a 2D-axis (see Figures 5.27, 5.28 and 5.29). 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. Figure 5.27 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 riks values between 35-70 %, where the higher the risk, the lower the NPV. This phenomenon can be now attributed to a strong decrease in makespan which is optimized simulatneously 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.

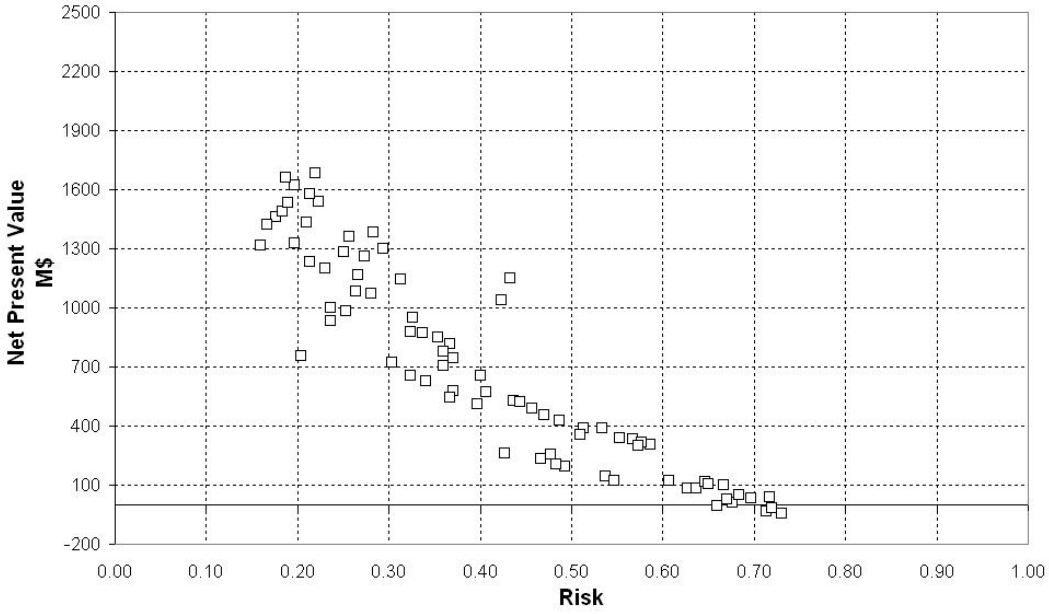


Figure 5.27: Tricriteria solutions projection for *NPV Risk*

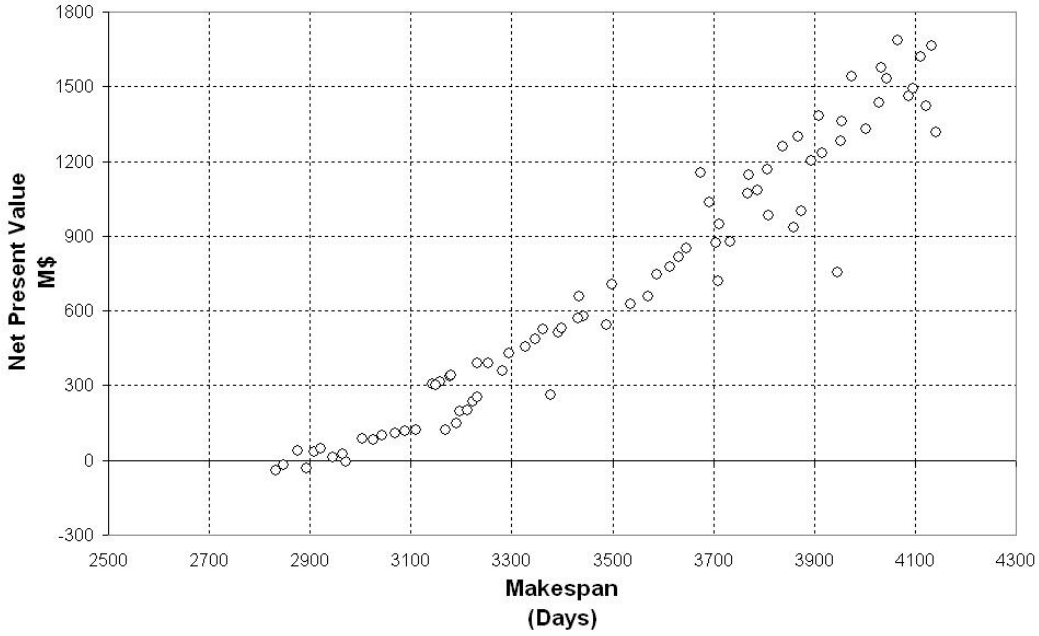


Figure 5.28: Tricriteria solutions projection for *NPV Duration*

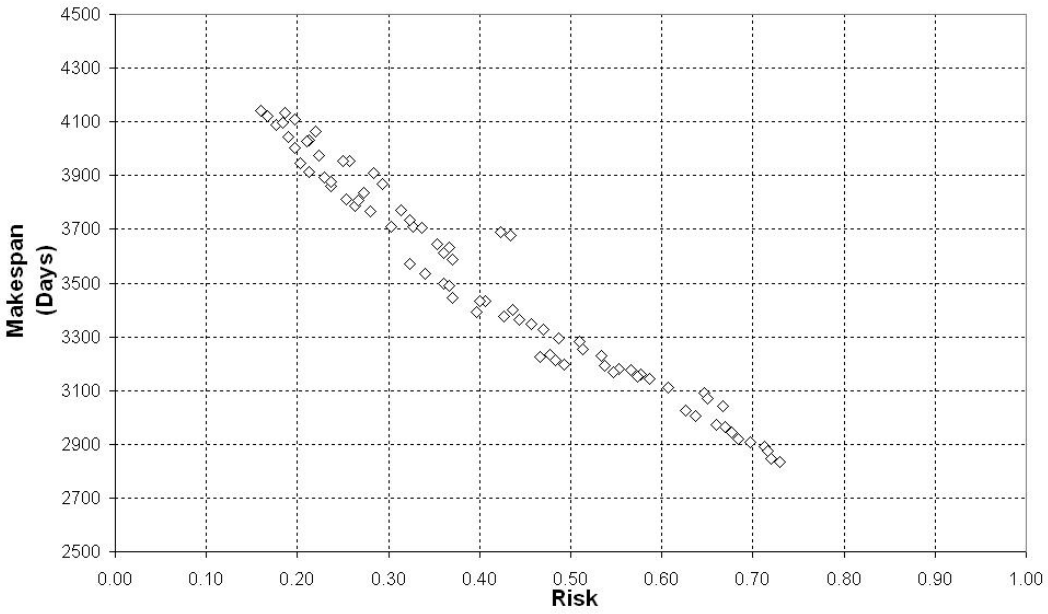


Figure 5.29: Tricriteria solutions projection for *Risk Duration*

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 5.15 all constitute potential candidates.

Solutions	Risk (%)	Net Present Value (MM)	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

Table 5.15: Some interesting solutions from tricriteria optimization

A closer examination of the solutions presented in Table 5.15 is presented in Figures 5.30 and 5.31 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 in Chapter 3 has shown that they exhibit a bimodal behavior. 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.

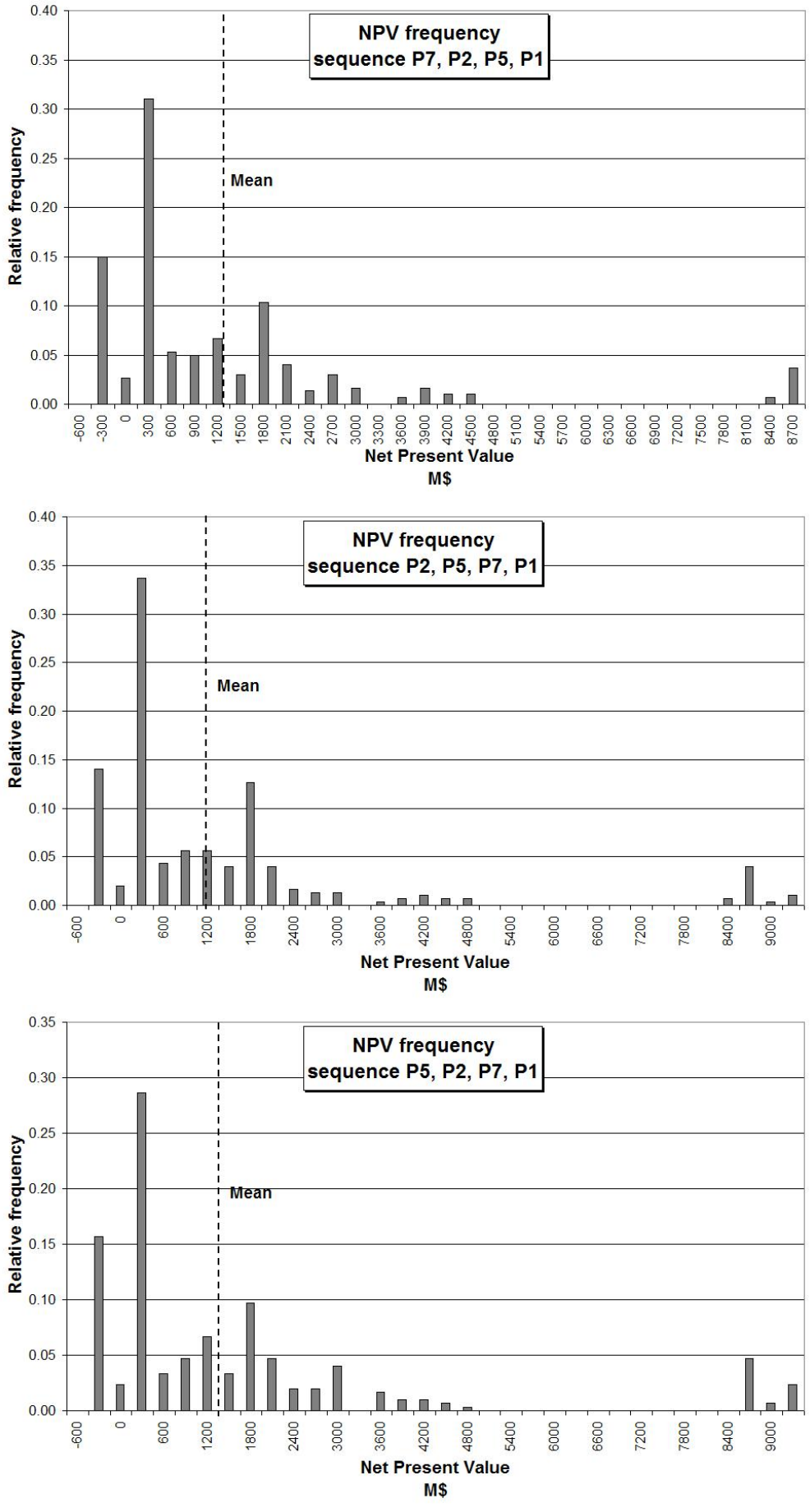


Figure 5.30: Frequency and behavior for non−dominated tricriteria optimization

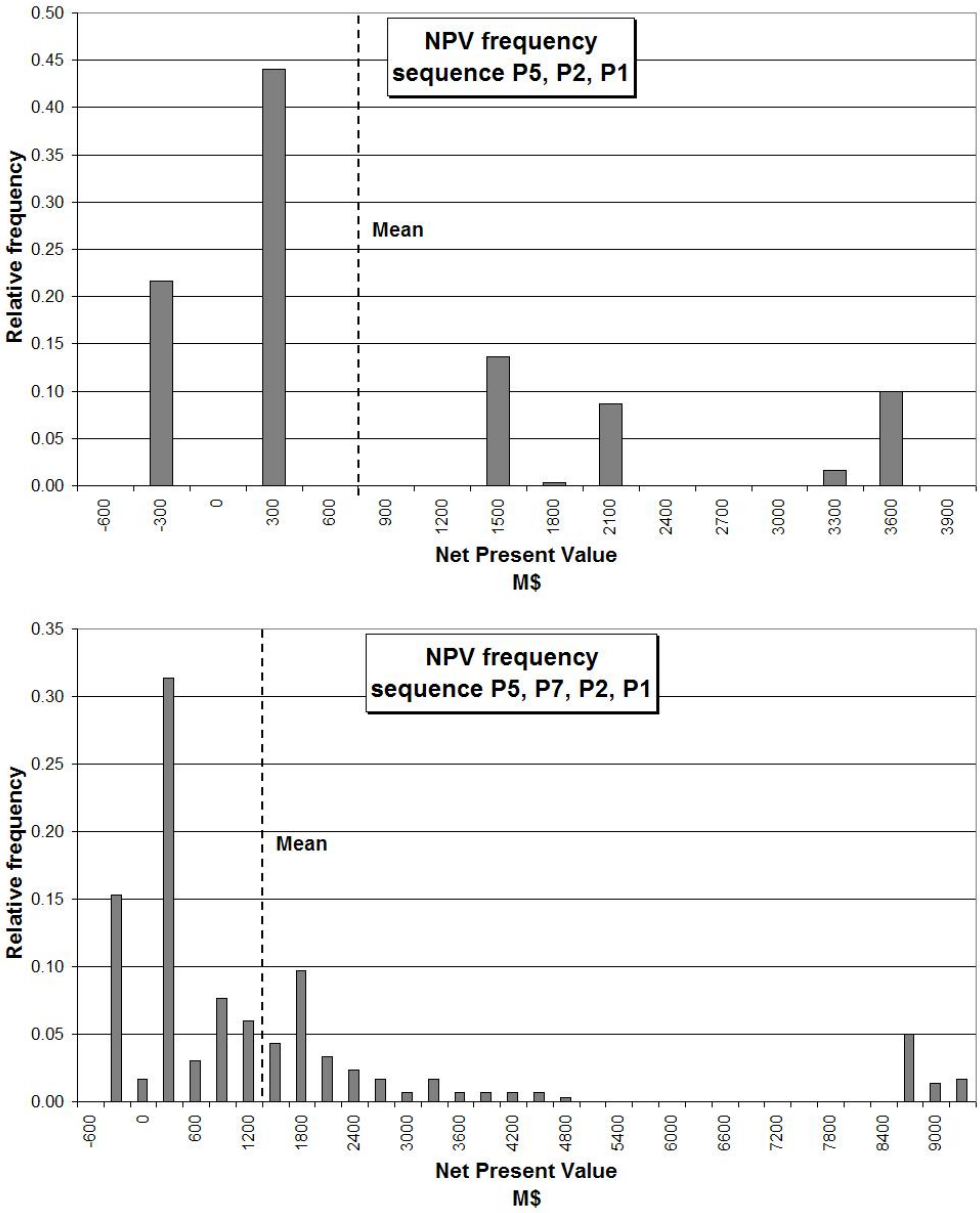


Figure 5.31: Frequency and behavior for non−dominated tricriteria optimization

Solutions 1 and 2 are equivalent from the makespan criteria, but differ significantly from the risk and NPV criteria. Solution 2 exhibits some peaks higher than 9000 M $. The decision maker has to decide if the higher risk induced by Solution 2 is justified. Solution 3 must be investigated if the time criterion is important to consider at that level, even if risk and NPV are lower than for solutions 1 and 2. Solutions 4 and 5 show the same order of magnitude for mean NPV and durations (risk is higher for solution 4). Yet, Solution 4 concerning the 3−drug sequence has poor performances (no peak higher than 3600 M$) 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.

5.5.2 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 NSGA II, 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 as TOPSIS will allow to select a sequences according to the decision maker preferences. This method is as an alternative to ELECTRE. 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].

5.6 Optimization of a 20-drug portfolio

A larger case study involving 20 drugs (called the "20−drug case") is proposed to evaluate the impact of various scheduling and resource allocation policies on pipeline performance.

Tables 5.16 and 5.17 show activity and drug data for this case. The DES was adapted to this case. It must be said that all the stages involved in the pipeline network remain exactly the same as in the 9-drug case. Data that will be updated or modified are relative to:

- Duration by stage;
- Cost by stage;
- The number of drugs by disease;
- Available resources;
- Capital cost by drug;
- Sales by drug;
- Success probabilities;
- No dependency is considered between drugs.

The main difference is that the four types of interdependencies analyzed in Blau et al. [2004] and presented in Table 2.5 in Chapter 2, that are financial, technical, manufacturing and resource based are not taken into account. The only dependency considered here is relative to the capacity constraint for each step. It must be also mentioned that capital costs are not any more represented by a triangular distribution(this representation is yet still valid for sales): this means that the concept of degree of difficulty is not considered here because stage duration and cost are not identified as triangular distributions but as crisp values.

Activity	Duration (days)	Cost (M$)	Available resources M$
FHDP	400	80	752
Sample prep	800	2	20
Phase I	300	80	752
Phase II	500	80	755
Phase III	775	200	2000
Process develop I	800	10	34
Process develop II	750	10	61
Design Plant	750	10	94
FSA	375	20	109
Prelaunch	100	50	140
Build Plant	750	62	120
Ramp up I	350	12	58
Ramp up II	350	22	82
Ramp up III	350	40	108
Matures sales	400	53	548

Table 5.16: Data for 20 drugs

Drug	Success probabilities			Capital cost (M$)	Mature sales (M$)		
name	Phase I	Phase II	Phase III		Min	ML	Max
D1	0.9	0.3	0.9	50	1675	1800	1950
D2	0.85	0.2	0.85	30	850	900	975
D3	0.95	0.35	0.95	45	2000	2300	2500
D4	0.87	0.22	0.8	34	1000	1250	1500
D5	0.97	0.36	0.99	40	200	690	1200
D6	0.83	0.18	0.86	60	1500	1830	2000
D7	0.94	0.4	0.94	75	800	1150	2000
D8	0.86	0.2	0.88	65	400	600	8500
D9	0.98	0.34	0.92	62	1750	1870	1900
D10	0.9	0.4	0.92	62	800	1000	1200
D11	0.9	0.45	0.92	60	400	500	600
D12	0.98	0.2	0.92	65	1200	1400	1600
D13	0.9	0.45	0.92	70	200	400	600
D14	0.98	0.25	0.92	62	800	1000	1200
D15	0.90	0.40	0.92	65	500	700	900
D16	0.98	0.15	0.92	70	1200	1500	1800
D17	0.9	0.35	0.92	65	600	800	1000
D18	0.98	0.3	0.92	62	900	1100	1300
D19	0.9	0.5	0.95	60	300	500	700
D20	0.98	0.2	0.9	60	1100	1400	1700

Table 5.17: Success probabilities, capital cost and mature sales

Figure 5.32 shows the reward/risk bubble plot for the 20−drug case.

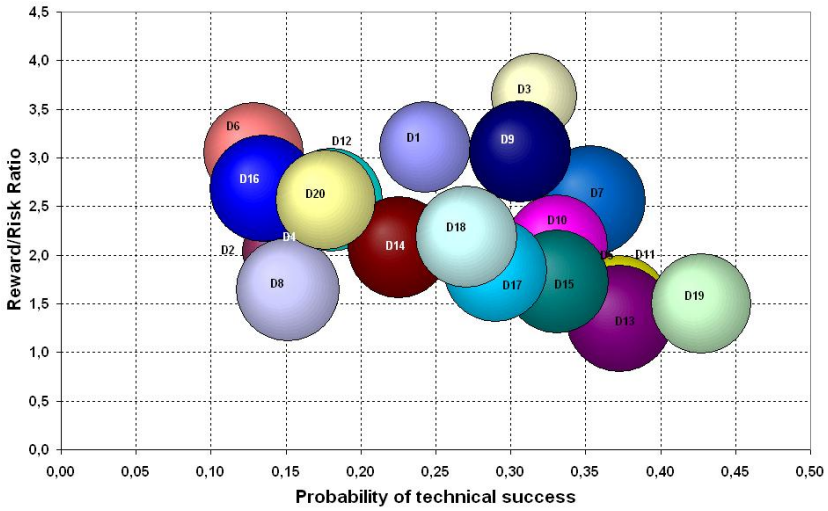


Figure 5.32: Bubble chart for the 20drug portfolio case

The bubble chart is used as before for having an idea about the best drugs according to *Reward/Risk* ratio, probability of technical success and capital cost.

A first optimization run was carried out with imposing that one drug targets one disease, meaning that the sequences must be of constant size (20 drugs). The induced optimization problem is to find the order in which the drugs must be released in the pipeline, thus leading to a scheduling problem with a high combinatorial aspect (20! possibilities, about 2.4310^{18}). The same GA parameters as those used for the 9-drug case have been selected (see Table 5.2). For this kind of problem, the bubble chart gives no direct answer.

The results are not satisfying at all, since no positive solutions for NPV were found.

This is why the following optimization runs involve as previously both the number of drugs in the portfolio and their order in the sequence as decision variables. Due to the high computational time induced by the problem combinatorics, only the NPV-risk pair of criteria has been investigated. The evolution of the generations is presented in Figures 5.33 and 5.34. It can be shown that the first generation exhibits very poor solutions. Yet, the optimization procedure leads rapidly to better solutions: from the 40th generation, individuals present positive values for the Net Present Value and lower values for the Risk. After this generation, a radical change in values for optimization criteria is no more observed. In Figure 5.35, the NPV (respectively risk) ranges between 99 and 111 M$ (respectively between 27 and 31 %). The sequences that have been identified after the Pareto procedure are presented in Table 5.18. From the constellation of possibilities, the optimization procedure turns out ot be very elitist.

The results also suggest that R&D managers need to consider resource sharing between competing drug projects as a viable alternative to the standard decision-making norm of just prioritizing projects for resource allocation. In summary, this last result calls for the development of tools that can monitor and track the impact of resource allocations on activity durations as well as the ones that yield optimal resource allocations with respect to financial and cycle time criteria.

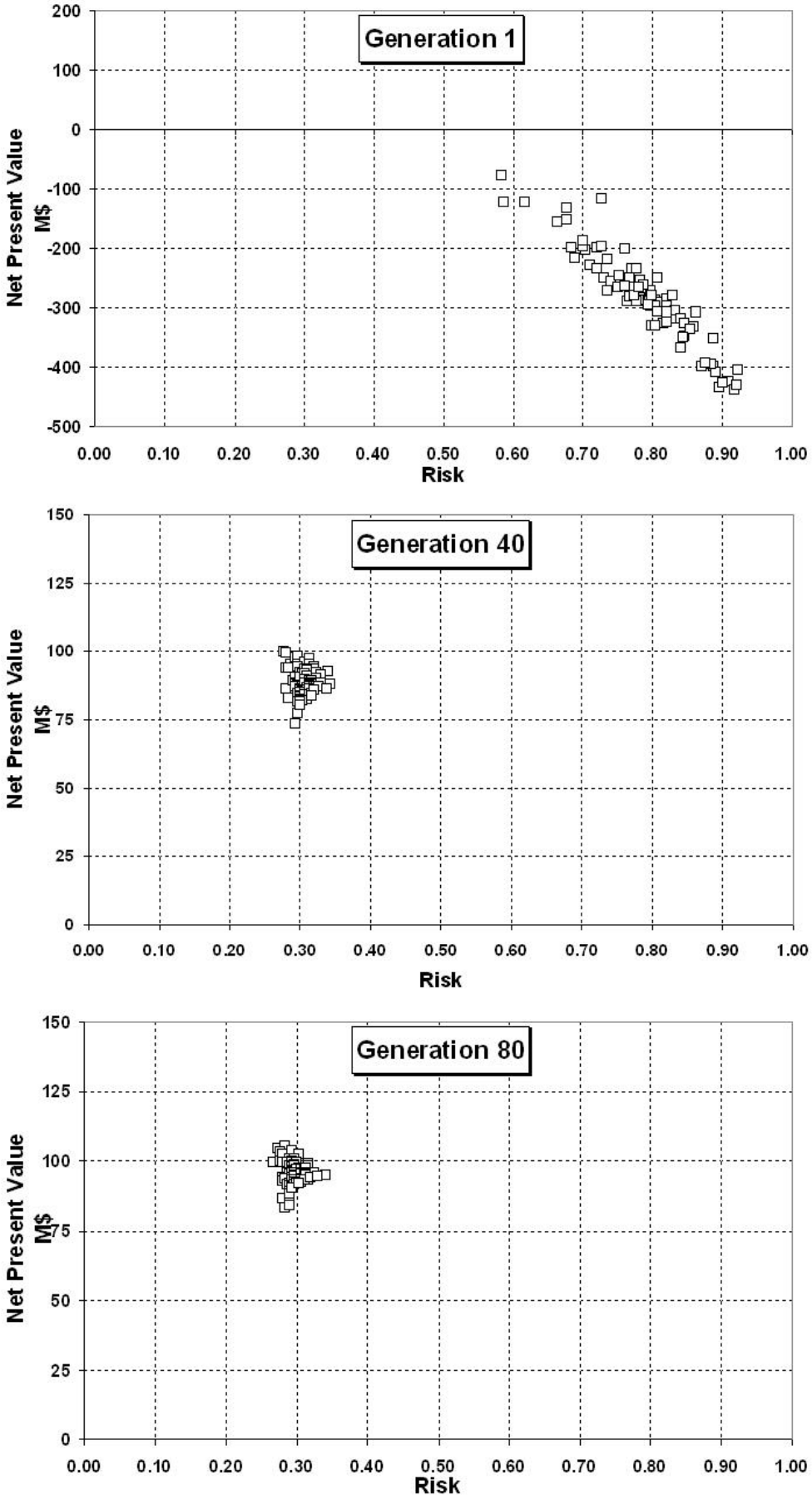


Figure 5.33: Behavior for generations 1, 40 and 80. Optimization of Risk/NPV for the case or the 20-drug case

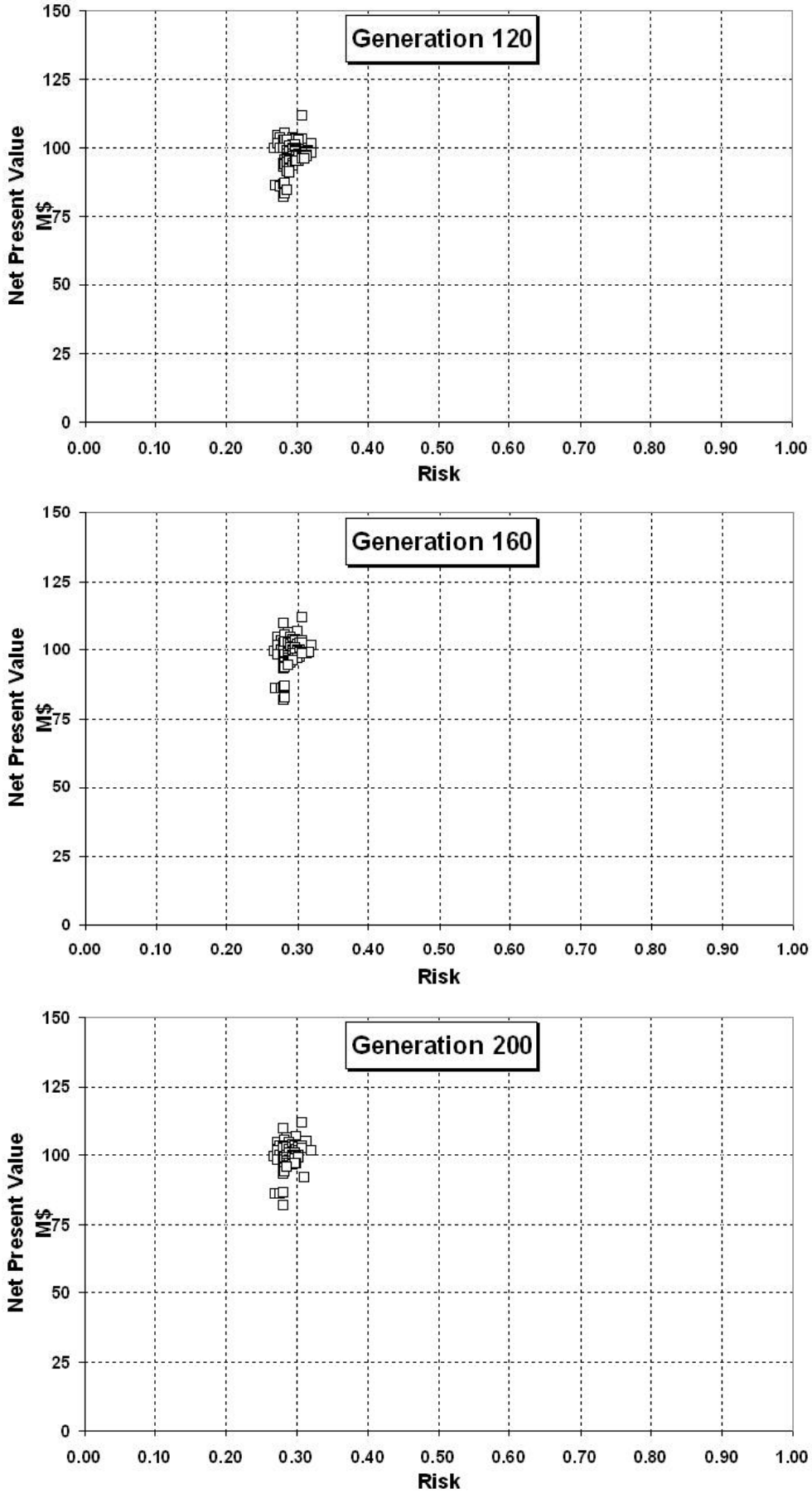


Figure 5.34: Behavior for generations 120, 160 and 200. Optimization of Risk/NPV for the 20-drug case

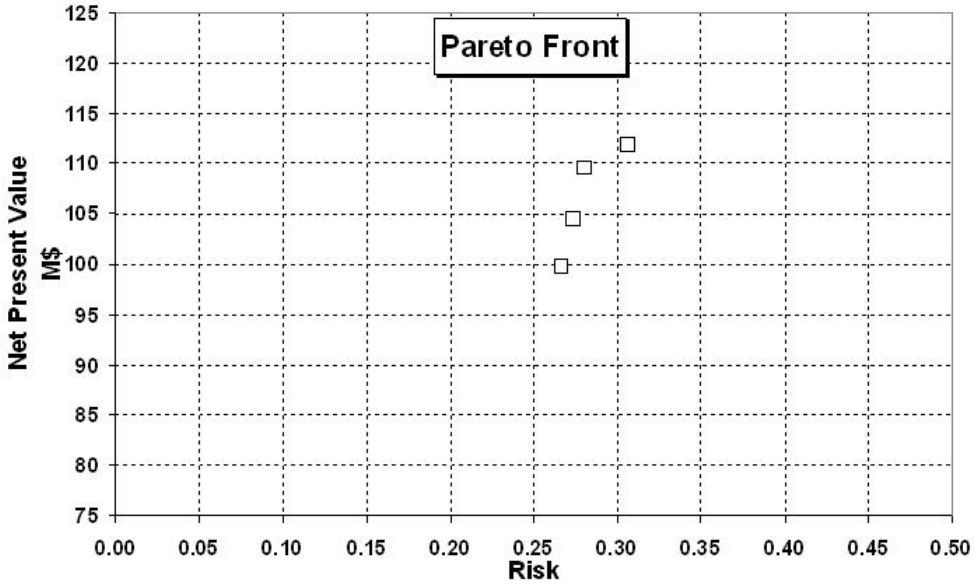


Figure 5.35: Pareto Front Risk/NPV for the 20-drug case

Sequence	NPV (M$)	Risk (%)	Makespan by simulation	Release order		
1	109	28	4592	P3	P9	P7
2	111	31	4596	P3	P7	P9
3	104	27	4571	P3	P9	P7
4	99	27	4601	P7	P9	P3

Table 5.18: NPV and Risk for sequences for the 20-drug case

5.7 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 can not 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.

Chapter 6

Conclusions and Perspectives

6.1 Conclusions

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.

The proposed approach combines discrete event stochastic simulation with multiobjective genetic algorithms to optimize the highly combinatorial portfolio management problems facing pharmaceutical businesses.

The typical formulation of New Product Development in a pharmaceutical industrial context was first presented with a case study for illustration purpose. This analysis shows the necessity to account for the interdependencies between products along the pipeline. The R&D pipeline consists of four stages of development Early Stage Development, Phase I, Phase II and Phase III, before plant design and drug commercialization. Each stage is associated with several development activities related by precedence arcs. From this description it is clear that each stage requires significant physical resources such as manpower, manufacturing and testing equipment resources. At any point in time the pipeline has drugs in each development and compete for resources. 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. 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.

Hence, we focused on developing a simulation decision support tool that uses probabilistic data in the form of durations of activities, resource requirements (modeled 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. Two examples illustrate and validate the methodology.

Before considering a portfolio of products, it is interesting to examine the behavior of each individual drug candidate. Using net present value (NPV) with an internal rate of return of 15 percent as the economic criterion, the behavior of each drug can be simulated by using the discrete event simulator. 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 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. A first solution was to use interval bounds to represent some imprecise parameters associated with a probability distribution within a Monte Carlo framework. The concept of Degree of Difficulty initiated by Blau et al. [2004] was also used to reflect the more or less difficulty to carry out a process task. An alternative approach based on interval analysis was investigated in order to determine the final strategy that could be then selected at the optimization step.

The former implies to carry out simulations many times to consider a representative sampling of the problem. The latter may be attractive for the optimization phase of the NPD problem that is then tackled. Both approaches have been illustrated by a numerical example. The results obtained by the interval-based approach turn out to 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.

New Product Development (NPD) problem can clearly be viewed as multiobjective problem with multi-stage decisions under uncertainty. The recurrent key issues are:

- What are the projects to develop once target molecules have been identified?
- In what order?
- Which is the level of resources to assign?

Several criteria, the Net Present Value of a sequence, its associated risk and time to market must be optimized simultaneously. Among the different multiobjective optimization methods that may be

used, 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. Among the various GAs, a discussion was then performed to select the most appropriate variant, NSGA II which has been adapted to the treated case for taking into account both the number of products in a sequence and the drug release order. From a 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 NSGA II, 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.

6.2 Perspectives

The development of sophisticated optimization and decision support tools is needed to help explore and analyze alternatives among the constellation of drug candidates for the NPD process, and predict actions for the operation of the supply chain so as to yield overall optimum economic performance.

- *Multiobjective optimization and multi-criteria decision making (MCDM).* This work has shown that a multiobjective optimization framework is needed which is particularly greedy in computational time, even if the induced computation time is particularly short as compared to the time scale of the involved phenomena : it must be emphasized that simulation and optimization are the only way to tackle NPD issue that are difficult to appreciate by an intuitive manner. An improvement would be to consider directly other optimization criteria that would be more appropriate to account for the bimodal distribution of the NPV: it would be interesting to simultaneously maximize the expected positive net present value (EPNPV), which is defined as the expected value over the positive axis of the NPV distribution, minimize the negative part of the distribution, minimize the risk defined as the number of positive occurrences of NPV over the total number of samples and minimize NPD makespan. Of course, this would require a systematic procedure to help decision aid from the Pareto set generated at the end of the multiobjective procedure. It is well-known that the number of compromise solutions increases with the number of objectives to consider. The use of a multi-criteria decision making (MCDM) approach can thus be suggested: Analytical Hierarchy Process (AHP), ELECTRE, PROMETHEE, TOPSIS may be powerful candidates. A review is proposed in Pirdashti et al. [2009].
- *Computational Grid.* Solving the NPD problem requires significant computational effort. Advances in algorithms and modelling must go hand-in-hand with advances in toolkits that enable

algorithms to harness computational resources. One promising approach that has emerged over the last decade is to deliver computational resources in the form of a computational grid, which is a collection of loosely-coupled, (potentially) geographically distributed, heterogeneous computing resources. For instance, Grid'5000, is a nation-wide infrastructure for research in grid computing (see http://www.grid5000.org). It is designed to provide a scientific tool for computer scientists similar to the large-scale instruments used by physicists, astronomers, and biologists. Grid'5000 is a large-scale experimental tool, with a deep reconfiguration capability, a controlled level of heterogeneity and a strong control and monitoring infrastructure. The Grid'5000 platform is distributed over nine sites in France has been used for some optimization runs of this application. Its use needs to be generalized and to such problems. The gridification will be particularly interesting since NPD application is parallel and multi-parametric.

- *Uncertainty challenge.* Uncertainty is a critical issue in supply chain operations, namely for NPD. Furthermore, it is complicated by the fact that the nature of the uncertainties can be quite different (e.g., success probabilities, activity cost and duration). How to better account for variations in order to effectively handle the effect of uncertainties (e.g., demands, success probabilities)? Major issues here are the development of novel, meaningful and effective tools.

- *Integration of strategic and tactical decision support strategies in NPD.* This work did not discuss the dynamic nature of the portfolio. As selected projects evolve from early phases towards completion, more data on projects are collected. Moreover, new project opportunities may arise which must be evaluated and added to the portfolio that must be updated. An extension of this work would be to carry out strategic and tactical levels reconciliation.

- *Extensions to other industrial fields.* Although illustrated for the pharmaceutical industry, this work can be extended readily to any industry such as aerospace, semiconductors, agrochemical, or biotechnology that is regulated highly and loses new product candidates during the development process.

To conclude, the proposed framework is just a piece of the New Product Development optimization puzzle.

Chapter 7

Nomenclature

7.1 Nomenclature

Acronym	Meaning
A	Attractiveness ratio
BPS	Batch Plant Scheduling
CDR	Complementary Determining Region
COG	Cost Of Goods
CSIM	Commercial discrete-event simulation software
DES	Discrete Event Simulator
DM	Decision Maker
DoD	Degree of Difficulty
ENPV	Expected Net Present Value
EPNPV	Expected Positive Net Present Value
EWO	Entreprise−Wide Optimization
FDA	Food and Drug Administration
FHDP	First Human Dose Preparation
FIFO	First In First Out
FSA	First Submission for Approval
GA	Genetic Algorithm
IA	Interval Analysis
mAbs	Monoclonal antibodies
MeanNPVpos	Mean Net Present Value Positive
MeanNPVneg	Mean Net Present Value Negative
MINLP	Mixed Integer Nonlinear programming
MOGA	Multiobjective Genetic Algorithm
MOOP	Multiobjective Optimization Problem
MOSA	Multiobjective Simulated Annealing
NLP	Nonlinear Programming
NPD	New Product Development
NPGA	Niched Pareto Genetic Algorithm
NPV	Net Present Value
NSGA	Non−Sorted Genetic Algorithm
PAES	Pareto−Archived Evolution Strategy
PP	Positivity probability
PSE	Process Systems Engineering
RCPSP	Resource Constrained Project Scheduling Problem
ROV	Real Options Valuation
SPEA	Strength Pareto Evolution Strategy
WA	Weighted Attractiveness

Bibliography

A. A. Aguilar-Lasserre, C. Azzaro-Pantel, L. Pibouleau, and S. Domenech. Enhanced genetic algorithm-based fuzzy multiobjective strategy to multiproduct batch plant design. *Anal. and Des. of Intel. Sys. using SC Tech*, 41:590–599, 2007.

G. Alefeld and J. Herzberger. *Introduction to interval computations*. Academic Press, 1983.

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

J. Andersson. On engineering system design - a simulation optimization approach. Master's thesis, Licentiate thesis, Department of Mechanical Engineering Linköping University, 1999.

J.H. Ang, C.K. Goth, E.J. Teoh, and A.A. Mamun. Multiobjective evolutionary recurrent neural netwoks for system identification. In *CEC 2007, IEEE*, 2007.

C. Azzaro, P. Floquet, L. Pibouleau, and S. Domenech. A fuzzy approach for performance modeling in batch plant: Application to semiconductor manufacturing. *IEEE Transaction on Fuzzy Systems*, 5:338–357, 1997.

H. Bandemer and S. Gottwald. *Fuzzy sets, fuzzy logic, fuzzy methods, with applications*. Wiley, 1995.

S. Bandyopadhyay, S. Saha, U. Maulik, and K. Deb. A simulated annealing based multiobjective optimization algorithm: Amosa. *IEEE Trans Evolutionary Computation*, 12:269–283, 2008.

T. Bäck, D. Fogel, and Z. Michalewicz. *Handbook of evolutionary computation*, chapter Multicriterion decision making. IOP Publishing Ltd and Oxford University Press, 1997.

R. Benayoun, J. de Montgolfier, J. Tergny, and O. Laritchev. Linear programming with multiple objective functions: Step method (stem). *Mathematical Programming*, 1:366–375, 1971.

L.T. Biegler and I. E. Grossmann. Retrospective on optimization. *Comp. and Chem. Eng*, 28: 1169–1192, 2004.

G. E. Blau, J. F. Pekny, V. A. Varma, and P. R. Bunch. Managing a portfolio of interdependent new product candidates in the pharmaceutical industry. *Journal of Product Innovation Management*, 21(4):227–245, 2004.

G. Bojadziev and M. Bojadziev. *Fuzzy Logic for Business, Finance and Management*. World Scientific Publishing Co., 1997.

L. Borup and A. Parkinson. Comparision of four non-derivative optimization methods on two problem contaning heuristic and analythic knowledge. In *ASME Advances in Design Automation*, 1992.

J. Branke, K. Deb, H. Dierolf, and M. Osswald. *Parallel Problem Solving from Nature VIII*, chapter Finding knees in Multi-objective Optimization, pages 722–731. Springer Berlin / Heidelberg, 2004.

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

F. Bérard, C. Azzaro-Pantel, S. Domenech, L. Pibouleau, and P. Floquet. Combinatorial study of multipurpose batch plant planning. In *Process Systems Engineering 2003, 8th International Symposium on Process Systems Engineering*. Elsevier, 2003a.

F. Bérard, C. Azzaro-Pantel, L. Pibouleau, and S. Domenech. A production planning strategic framework for batch plants. In *European Symposium on Computer Aided Process Engineering-13, 36th European Symposium of the Working Party on Computer Aided Process Engineering*. Elsevier, 2003b.

J.J. Buckley and Y. Hayashi. Applications of fuzzy chaos to fuzzy simulation. *Fuzzy sets and systems*, 99:151–157, 1998.

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

R.J. Calantone, C.A. Di Benedetto, and J.B. Schmidt. Using the analytic hierarchy process in new product screening. *Journal of Product Innovation Management*, 16:65–76, 1999.

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

C. A. Coello. *An empirical study of evolutionary techniques for multiobjective optimization in engineering design*. PhD thesis, Tulane University, 1996.

C. A. Coello Coello. Evolutionary multi-objective optimization: some current research trends and topics that remain to be explored. *Frontier of computer science in China*, 3:18–30, 2009.

S. Coldrick, P. Longhurst, P. Ivey, and J. Hannis. An r&d options selection model for investment decisions. *Technovation*, 25:185–193, 2005.

A.M. Connor and D.G. Tilley. A tabu search method for the optimization of fluid power circuits. *Journal of Systems & control engineering*, 212:373–381, 1998.

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

T.E. Copeland and V. Antikarov. *Real options: A practitioner's guide*. Texere, 2001.

K. Deb. Single and multi-objective optimization using evolutionary computation. Technical report, Department of mechanical engineering Indian Institute of Technology Kanpur, 2004.

K. Deb and T. Goel. Multi-objective evolutionary algorithms for engineering shape desing. Technical report, Department of mechanical engineering Indian Institute of Technology Kanpur, 2000.

K. Deb and A. Srinivasan. Innovization: Innovative design principles through optimization. Technical report, Department of mechanical engineering Indian Institute of Technology Kanpur, 2005.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. ISSN 1089-778X.

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

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

A. Dietz, A. Aguilar-Lasserre, C. Azzaro-Pantel, L. Pibouleau, and S. S. Domenech. A fuzzy multiobjective algorithm for multiproduct batch plant: Application to protein production. *Computers and Chemical engineering*, 32:292–306, 2008.

J. A. Dimasi, R. W. Hansen, and H. G. Grabowsky. The price of innovation: new estimates of drug development costs. *Journal of healthh economics*, 22:151–185, 2003.

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

M. Dorigo. *Optimization, learning and natural algorithms*. PhD thesis, Politecnico di Milano, Italy, 1992.

D. Dubois and H. Prade. *Fuzzy sets and systems theory and applications*. Academic Press, 1980.

D. Dubois and H. Prade. *Possibility theory*. Plenum Press, 1986.

J.D. Farmer, N. Packard, and A. Perelson. The immune system, adaptation and machine learning. *Physica D*, 22:187–204, 1986.

C. M. Fonseca and P. J. Fleming. Genetic algortihm for multiobjective optimization: formulation, discusion and generalisation. *Proc. of the 5th Int Conf. on genetic algorithm*, 00:416–423, 1993.

C. M. Fonseca and P. J. Fleming. An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3:1–18, 1995a.

C. M. Fonseca and P. J. Fleming. Multiobjective genetic algorithms made easy: Selection, sharing and mating restriction. In *IEE/IEEE International Conference on Genetic Algorithms in Engineering Systems*, 1995b.

C. M. Fonseca and P. J. Fleming. 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, 1998.

M.P. Fourman. Compaction of symbolic layout using genetic algorithms. In *Conference on Genetic Algorithms*, 1985.

M. Gen and R. Cheng. *Genetic Algorithms and Engineering Optimization*. Wiley-Interscience publication, 2000.

D. Goldberg. *Genetic Algorithms in Search and Machine Learning*. Addison Wesley, 1989.

A. Grieco, F. Nucci, and A. Anglani. Representation of fuzzy time variables in discrete event simulation. *Integrated Computer-Aided Engineering*, 10:305–318, 2003.

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

I. E. Grossmann and A. W. Westerberg. Research challenges in process systems engineering. *AIChE Journal*, 49:1700–1703, 2000.

T. Grueninger and D. Wallace. *Multi-modal optimization using genetic algorithms*. Massachusetts Institute of Technology, Cambridge, 1996.

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

P. Hajela. Nongradient methods in multidisciplinary design optimization - status and potential. *Journal of Aircraft*, 36:255–265, 1999.

G. Harik. Finding multimodal solutions using restricted tournament selection. In *Sixth International Conference on Genetic Algorithms*, 1995.

J. H. Holland. *Adaptation in Natural and Artificial Systems, an introductory analysis with application to biology, control and artificial intelligence.* Complex Adaptive Systems, 1975.

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

J. Horn and N. Nafpliotis. Multiobjective optimization using the niched pareto genetic algorithm. Technical report, Illinois Genetic Algorithm Laboratory, Dept. of General Engineering, University of Illinois at Urbana-Champaign, Urbana, USA, 1993.

W.F. Jacob and Y.H. Kwak. In search of innovative techniques to evaluate pharmaceutical r&d projects. *Technovation*, 23:291–296, 2003.

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

A. Jansson. *Fluid Power System Design - A Simulation Approach.* PhD thesis, Department of mechanical engineering, Linköping University, 1994.

K. B. Kahn. A critical look at technological innovation typology and innovativeness terminology: a literature review. *Product Innovation Management*, 19:110–132, 2002.

S. Kaplan and B J. Garrick. On the quantitative definition of risk. *Risk Analysis*, 1:11–27, 1981.

J. Kennedy and R. Eberhart. Particle swarm optimization. In *IEEE International Conference on Neural Networks*, pages 1942–1948, 1995.

S. Kirkpatrick, C.D. Gellat, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220: 671–680, 1983.

G. J. Klir and M. J. Wierman. *Uncertainty-Based information: Elements of Generalized Infirmation Theory.* Physica-Verlag, 1999.

A. Konac, D.W. Coit, and A.E. Smith. Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System*, 91:992–1007, 2006.

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

M. Kuroda and Z. Wang. Fuzzy job shop scheduling. *International journal of productions economics*, 44:45–51, 1996.

C.H. Lin and P.J. Hsieh. A fuzzy decision support system for strategic portfolio management. *Decision Support Systems*, 38:383–398, 2004.

C.H. Lin, B. Tan, and P.J. Hsieh. Application of the fuzzy weighted average in strategic portfolio management. *Decision Sciences*, 36:498–511, 2005.

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

S.A. Mansouri, S.H. Hendizadeh, and N. Salmassi. Bicriteria two-machine flowshop scheduling using metaheuristics. In *GECCO '07*, 2007.

C.T. Maravelias and I.E. Grossmann. Simultaneus planning for new product development and batch manufacturing facilities. *I & EC Research*, 40:6147–6164, 2001.

H. M. Markowitz. *Portfolio Selection*. Cowles foundation, 1959.

D. Moens and D. Vandepitte. Interval sensitivity theory and its application to frequency response envelope analysis of uncertain structures. . *Comput Methods Appl Mech Eng*, 196:21–24, 2007.

M. Mongeau, H. Karsenty, V. Rouzé, and J.-B. Hiriart-Urruty. Comparision of public-domain software for black-box global optimization. Technical report, Université Paul Sabatier Toulouse, France, 1998.

N. Monmarché, F. Guinand, and P. Siarry. *Fourmis artificielles, Volume 1 Des bases de l'optimisation aux applications industrielles*. IC2, 2009.

J. M. Montagna and A. R. Vecchietti. Retrofit of multiproduct batch plants through generalized disjunctive programming. *Math. Comp. Model*, 38:465, 2003.

R. E. Moore. *Interval analysis*. Prentice Hall, 1966.

M. G. Morgan and M. Henrion. *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambrige University Press, 1990.

J. N. Morse. Reducing the size of the nondominated set: pruning by clustering. *Computers & Operations Research*, 7:55–66, 1980.

R.L. Muhanna, R.L. Mullen, and H. Zhang. Interval finite element as a basis for generalized models of uncertainty in engineering mechanics. *Reliab Comput*, 13:173–194, 2007.

S. Nakrani and S. Tovey. On honey bees and dynamic server allocation in internet hosting centers. *Adaptative Behaviour*, 12:223–240, 2004.

D.P. Newton, D.A. Paxson, and M. Widdicks. Real r&d options. *International Journal of Management Reviews*, 5-6:113–130, 2004.

Q. Nguyen and T. Le. A fuzzy discrete-event simulation model. In *Proceeding of Australia-Pacific Forum on Intelligent Processing and manufacturing of Materials*, 1997.

P. C. Nutt. How decision makers evaluate alternatives and the influence of complexity. *Management Science*, 44:1148–1166, 1998.

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

P.M. Keller P.J. Turinsky and and H.S. Abdel-Khalik. Evolution of nuclear fuel management and reactor operational aid tools. *Nuclear Engineering and Technology*, 37:79–90, 2005.

K.L. Poh, B.W. Ang, and F. Bai. A comparative analysis of r&d project evaluation methods. *R&D Management*, 31:63–75, 2001.

A. Ponsich, C. Azzaro-Pantel, S. Domenech, and L. Pibouleau. Constraint handling strategies in genetic algorithms application to optimal batch plant design. In *Chemical Engineering and Processing: Process Intensification 10th French Congress on Chemical Engineering*, volume 47, pages 420–434, 2008.

A. Rajapakse, N. J. Titchener-Hooker, and S. S. Farid. Modelling of the biopharmaceutical drug development pathway and portfolio management. *Computers & Chemical Engineering*, 29(6): 1357–1368, May 2005. ISSN 0098-1354.

A. Rajapakse, N. J Titchener-Hooker, and S. S Farid. Integrated approach to improving the value potential of biopharmaceutical R&D portfolios while mitigating risk. *Journal of Chemical Technology & Biotechnology*, 81(10):1705–1714, 2006.

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

J. M Reichter. Monoclonal antibodies in the clinic. *Nature biotechnology*, 19:819–822, 2001.

Peters W. Roberts. Product innovation, product-market competition and persistent profitability in the u.s. pharmaceutical industry. *Strategic Management*, 20:655–670, 1999.

M.J. Rogers, A. Gupta, and C.D. Maranas. Real options based analysis of optimal pharmaceutical research and development portfolios. *Ind Eng Chem*, 41:6607–6620, 2002.

M. A. Rosenman and J. S. Gero. Reducing the pareto opimal set in multicriteria optimization (with application to pareto optimal dynamic programing). *Engineering Optimization*, 8:189–206, 1985.

J. Rowe, K. Vinsen, and N. Marvin. Parallel gas for multiobjective functions. In *Proceedings of the 2NWGA*, 1996.

L.P. Santiago and T.G. Bifano. Management of r&d projects under uncertainty: A multidimensional approach to managerial flexibility. *IEEE Transactions on Engineering Management*, 52:269–280, 2005.

J. Schaffer. Multiple objective optimization with vector evaluated genetic algorithms. In *1st Int. Conf. on Genetic Algorithms*, 1985.

C.W. Schmidt and I.E. Grossmann. Optimization models for the scheduling of testing tasks in new product development. *Industrial and Engineering Chemistry Reseach*, 35:3498–3510, 1996.

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

J. F. Shapiro. *Modeling the Supply Chain*. Duxbury, 2001.

L.S. Shu, S.J. Ho, S.Y. Ho, J.H. Chen, and M.H. Hung. A novel multi-objective orthogonal simulated annealing algorithm for solving multi-objective optimization problems with a large number of parameters. *LNCS*, 00:737–747, 2004.

K.I. Smith, R.M. Everson, and J.E. Fielsen. Dominance measure for multi-objective simulated annealing. In *Proceeding of the 2004 IEEE Congress on Evolutionary Computation*, pages 23–30, 2004.

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

R. Steuer. *Multiple criteria optimization: theory, computation and application*. John Wiley & Sons, Inc, 1986.

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

D. Subramanian, J.F. Pekny, G.V. Reklaitis, and G.E. Blau. Simulation-optimization framework for stochastic optimization of r&d pipeline management. *AIChE Journal*, 49:96–112, 2003.

P. Suresh and P. K. Basu. Improving pharmaceutical product development and manufacturing: Impact on cost of drug development and cost of goods sold of pharmaceuticals. *Journal of Pharmaceutilca Innovation*, 3:175– 187, 2008.

H. Taboada and D. Coit. Data mining techniques to facilitate the analysis of the pareto-optimal set for multiple objective problems. In *Industrial Engineering Research Conference*, 2006.

H. Tamaki, H. Kita, and S. Kobayashi. Multi-objective optimization by genetic algorithms: a review. In *IEEE International Conference on Evolutionary Computation*, 1996.

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

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

V. A. Varma, J. F. Pekny, G. E. Blau, and G. V. Reklaitis. A framework for addressing stochastic and combinatorial aspects of scheduling and resource allocation in pharmaceutical R&D pipelines. *Computers & Chemical Engineering*, 32(4-5):1000–1015, April 2008. ISSN 0098-1354.

D. Van Veldhuizen and G. Lamont. Multiobjective evolutionary algorithms: analyzing the sate-of-the-art. *Evolutionary Computation*, 8:125–147, 2000.

X. Wan, J.F. Pekny, and G.V. Reklaitis. 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*, 2006.

J. Wang and W.L. Hwang. A fuzzy set approach for r&d portfolio selection using real options valuation model. *The international journal of management science*, 35:247–257, 2007.

D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Trans. EVol. Comput*, 1:67, 1997.

Y. Yang, X. Xu, and W. Zhang. Design neural networks based fuzzy logic. *Fuzzy sets and systems*, 114:325–328, 2000.

M. Yoshikawa and H. Terai. Hybrid genetic algorithm engine for high-speed floorplanning. In *Circuit Theory and Design, 2005. Proceedings of the 2005 European Conference*, 2005.

L.A. Zadeh. Fuzzy sets. *Inform. and Control*, 8:338–353, 1965.

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

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

H. Zhang, C. Tam, and H. Li. Modeling uncertain activity duration by fuzzy number and discrete-event simulation. *European Journal of Operational Research*, 164:715–729, 2005.

H.J. Zimmermann. *Fuzzy set theory and its applications*. Kluwer Academic Publishers, 1992.

E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *IEEE Transaction on evolutionary computation*, 3:257–271, 1999.

E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algortihms: Empirical results. *Evolutionary Computation*, 8:173–195, 2000.