UNIVERSIDADE TÉCNICA DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

"Make-to-Stock vs. Make-to-Order in Glass Manufacturing"

NUNO MANUEL ROSA DOS SANTOS ÓRFÃO

(Licenciado em Engenharia Mecânica)

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA E GESTÃO DE TECNOLOGIA

Doutor CARLOS FILIPE GOMES BISPO (Orientador)
Doutor PAULO MANUEL CADETE FERRÃO (Co-Orientador)

Doutor MANUEL FREDERICO TOJAL DE VALSASSINA HEITOR (Presidente)
Doutor FERNANDO MANUEL LOBO PEREIRA
Doutor CARLOS FILIPE GOMES BISPO
Doutor PAULO MANUEL CADETE FERRÃO

Fevereiro de 2001

To my parents, Lurdes and Manuel

my sister, Paula

and Cláudia.

ACKNOWLEDGEMENTS

Many people have contributed their share to the development of this thesis; without them, the achievement of this aim would have been even harder. I am very grateful to them, for all their understanding, inspiration and support in the last few years, especially when one is willing to let go of everything.

I must necessarily apologize to those who I may have forgotten. It is understandable that, after this “journey”, my brain is no longer what it used to be... Thank you very much to them as well. For everything.

Firstly, I would like to thank Professor Carlos Bispo. His role extended well beyond what can be expected from an adviser. He was someone who was able to show me the right direction at the right time, so that I could see the light at the end of the tunnel before it was too late. From the very first days, when we were trying to define the problem, having Operations Management classes, testing the code and being at “marathon-like” meetings, his support was always unconditional and definitely crucial to this project.

To the board of Professors who co-ordinate this master’s degree, in particular to Professor Manuel Heitor, I owe the chance of attending the Master in Engineering and Technology Management, the knowledge I acquired in his classes and his absolutely essential support throughout the months when I desperately tried to finish this thesis. Going back to the first year of this master’s degree, I must also remember all those who, during the time when we had to attend classes, made it possible for this experience to become both pleasant and beneficial at all levels. Apart from those I have already mentioned, I would like to refer to Professor António Caetano in particular, for having made it possible for us to get to know concepts and experiments that are only very seldom to be found in Engineering Schools and which turned out to be of the utmost importance within the context of today’s organizations. I would also like to thank Professor Paulo Ferrão for all his support.

João, our friendship has been growing since we started being flat-mates in Quinta da Luz, 7th floor. Together, we have been through a great deal of things, which will, some day, remind us of the good old days. To you, I owe all the help in the initial stage of the simulator’s development. You were always able to understand my somewhat cloudy algorithms and my reasoning (which was sometimes not that crystal clear). It is not easy at all to put the importance of your help down on paper. Thank you very much for everything, Johnny.

To my colleagues and friends of the 1998 MEGT course, I am also very indebted for the times we spent together. Especially, I recall the nights before every Thursday, Friday and Saturday, when we had to make our presentations (weren’t we supposed to have classes only on these days?!) and, of course, when we went out together, desperately trying to keep ourselves mentally healthy ☺ !! To César Branco and the board of administration of Atlantis, I express my gratitude for the data that have largely enriched this project.

Many people at the ESTG had a closer contact with this project. Firstly, I wish to thank my students, who could not have been more understanding in times of additional work. My thanks to all my colleagues and friends of the Mechanical Engineering Department at the ESTG for their co-operation as well, in particular to Doutor João Ramos for all his support, to Doutor Carlos Neves for the tips on *.exe & co., to Carlos for lots of things, to Nuno Martinho for the derivatives and so on. Also, to Dina, Sofia, Judite, Alves, Mateus, Baptista, Lima, P. Martinho and to those who I unwillingly forgot to mention.

I must also express my gratitude to my friends Filipe, Gil, Rui, Hugo, C&P Sousa, Távora, C&F Alves and A&A Manso, for the strength they have given me throughout the different stages of this project. To my little sister Paula, to my cousin Rodolfo and to João Albino: thank you for your companionship during the time when I had to attend classes in Lisbon (and of course, thanks for the roof).

To my parents and, again, to my little sister, I am most grateful for everything they have done for me throughout my whole life.

Finally, I cannot finish without mentioning the person who probably had the closest contact with this period and who certainly felt its consequences more intensely than anyone else. My absence, along with my bad-tempered nature, did not make you feel happy all the way. I will be eternally grateful to you for your understanding. Thank you so much for everything, Cláudia.

ABSTRACT

This thesis addresses the problem of computing the optimal parameters for production control policies in the glass manufacturing industry and provides a framework of analysis related with the structure of the production policies. We consider a multi-product, multi-stage, and capacitated discrete-time production-inventory system with random yield. Random demand occurs in each period. The optimal parameters for a given production control policy are determined in order to minimize the expected costs or reach a given service level. Three different production strategies are discussed: *Make-to-Order* (MTO), *Make-to-Stock* (MTS), and *Delayed-Differentiation* (DD).

We use real data (processing times, random yield factors, etc) from a glass manufacturing company, providing simultaneously the model validation and the evaluation of the relative performance of the different strategies.

The approach used to analyze this problem will be simulation based optimization. Simulation will be used as a tool to obtain estimates of the objective function value and gradient with respect to the parameters that describe the control policy. The gradient estimates are obtained through *Infinitesimal Perturbation Analysis* (IPA).

KEYWORDS

- Capacitated Inventory Systems;
- Alternative Production Strategies;
- Random Yield;
- Glass Manufacturing;
- Infinitesimal Perturbation Analysis;

RESUMO

Esta tese analisa o problema da definição de parâmetros óptimos para diferentes estratégias de produção no âmbito da indústria do cristal, proporcionando simultaneamente um enquadramento genérico de análise da estrutura de estratégias de produção. A análise considera um modelo discreto de controlo de inventário composto por múltiplas máquinas em série, de capacidade finita e taxa de produção aleatória. O sistema processa múltiplos produtos que, por sua vez, estão sujeitos a procura aleatória. Os parâmetros óptimos de determinada estratégia de produção são calculados por minimização da função custo ou por manutenção de um nível de serviço estabelecido. As estratégias de produção consideradas são: *Produção-para-Stock*, *Produção-por-Encomenda* e *Diferenciação-Retardada*.

São utilizados dados de produção (cadências, taxas de rejeição, etc) de uma unidade produtiva nacional permitindo, além da validação do modelo, a comparação do desempenho relativo de estratégias de produção alternativas.

A abordagem utilizada para análise do problema consiste em optimização baseada em simulação. Esta técnica permite estimar o valor da função objectivo e respectivo gradiente em ordem às variáveis de decisão da política de controlo em causa. A estimação de gradientes é baseada na *Análise de Perturbações Infinitesimais* (IPA).

PALAVRAS CHAVE

- Controlo de Inventário;
- Estratégias de Produção Alternativas;
- Taxa de Produção Aleatória;
- Indústria de Cristal;
- Análise de Perturbações Infinitesimais;

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The purpose of this thesis is to define, under several different demand scenarios and based on a cost trade-off evaluation, which are the best production strategies for each business environment, on the specific area of glass manufacturing processes. Two extreme strategies may be exercised: *Make-To-Order* (MTO) and *Make-to-Stock* (MTS). From these two strategies one can derive a composite one, which consists on the establishment of intermediate stocks containing semi-processed products: a *Delayed Differentiation* (DD) strategy.

Presently, there is a significant lack of production decision support systems and even the most advanced information systems do not work with such kind of tools. Frequently, on the commercial management software, the tool available for inventory control is typically the Economic Order Quantity (EOQ). Out of the assumptions of this model, two of them can significantly restrict its application to more complex industrial processes. First, the demand is assumed to be constant and deterministic, and second, each item is treated independently of other items. That is, benefits from joint review or replenishment do not exist or are simply ignored, as well as the impact on the overall system.

Managers are often confronted with decisions on whether or not to hold inventories. If the demand realization turns out to be smaller than the available finished products, then some processing cost have unnecessarily been incurred. But if the demand realization turns out to be larger than the available finished products, then some customers might balk and their demand (and probably the future ones) could be lost.

The cost of producing to stock is usually less expensive than in a rush job, when customers are waiting for the conversion of raw materials into finished products. Moreover, under an MTO policy one may lose the advantages of economies of scale (working in batches) and the learning effects. These losses tend to increase the cost under an MTO policy. Of course, decisions of this type go beyond the area of production planning and inventory management. One needs to consider other relevant issues such as marketing, quality and environmental considerations.

The main motivation of this thesis is related with the actual generalized glass industry managerial strategy of predominantly producing-to-order and, occasionally, if there is excess capacity, stock the products with the highest levels of demand. That is, operating close to 100% resource utilization. Those production decisions are not based on solid arguments, but rather based on the experience and intuition. Therefore, we intend to understand the reasons and driving force of such decision. To do this, we develop a model that supports the production decision process, in order to validate the production strategy just referred. Also, we intend to define alternative production strategies, eventually more appropriate to different business environments.

1.1 SUMMARY OF CONTRIBUTIONS

This section presents a brief summary of the main contributions of this thesis, the objective of which being to analyze the problem of glass manufacturing production system working under different production strategies.

The numerical study presented illustrates not only the strength of Infinitesimal Perturbation Analysis (IPA) as an optimization tool, but also the impact of different production strategies on system's response. This will be measured in terms of average total cost, in-house costs, and products' lead-time. Instances of the main conclusions are listed below.

i) An MTS strategy incurs the lowest average total cost and presents the best lead-times;

ii) An MTO strategy accounts for the lowest in-house costs, while presenting the highest lead-times among all strategies;

iii) A composite strategy could be a relevant option if we intend to significantly improve the lead-times relative to the MTO, with a slight increase of in-house costs;

iv) It is not possible to define *the right strategy* without understanding the business context.

At last, a set of operational contributions can be summarized as:

i) Definition of a framework to deal with multi-product, multi-stage, capacitated, random yield production processes, facing random demand;

ii) The use of the Infinitesimal Perturbation Analysis methodology to find the optimal parameters of the pre-defined strategies, which showed to be a technique with great potential.

iii) The development of a simulation optimization package to analyze production systems with the features listed in *i)*.

1.2 ORGANIZATION OF THE THESIS

This thesis is composed of five chapters and appendices. The present chapter, Chapter 1 – **Introduction**, presents a brief description of the glass manufacturing business context and a reference to the problem of make-to-stock or make-to-order. In the next chapter, Chapter 2 – **Literature Review**, is presented a review of relevant literature, focusing on the issues of multi-product and multi-stage production systems, random yield, and infinitesimal perturbation analysis.

In Chapter 3 – **Model Definition –**, the basic model will be presented. First, the glass manufacturing process will be described. Then, the dynamic equations governing the system are established and the derivatives of the state variables with respect to the parameters defining the control policies are presented. This chapter continues with the definition of the performance measures, production strategies, and production policy. The IPA validation is discussed on the remaining section of this chapter.

Chapter 4 – **Experimental Study –**, provides the set of experimental data obtained with the simulation package. Different production strategies are tested and the results compared in order to get some insights on the system's response.

The last chapter, Chapter 5 – **Conclusions & Future Research**, presents a summary of the thesis and discusses some topics for future research.

Three appendices are provided at the end of the thesis. The first two, Appendix A – **Description of the Simulation & Optimization Software** and Appendix B – **The SimulGLASS User-Interface Windows**, present a description of the developed simulation optimization package. Appendix C – The **SimulGlass Package**, provides a cd-rom with a ready to install version of the software package and an *Acrobat* file with the full version of the thesis.

Chapter 2

Literature Review

2.1 Inventory Models

Almost all mathematical inventory models are designed to address two fundamental problems: when should a replenishment order be placed, and how much should the order quantity be. Assumptions concerning demand distribution, cost structure, and physical characteristics of the system are the main settings defining the model complexity. Naturally, when uncertainty is present, the established approach consists on optimizing expected values of performance measures.

Most of the real inventory control problems involve multiple products. However, single product models are able to simulate the main features of the larger problems. This is the reason why single product models dominate the literature and are often used to provide guidelines in solving real size problems.

As referred in [Bispo, 1997], the inventory control problems are concerned with decisions regarding when and how much to produce or order so that it is possible to satisfy an external demand. The aspects that need to be considered for model design are the number of stages of production, the number of products, and the demand process, characterized in terms of its stochastic nature. The characterization of the production process in terms of operation times, availability, reliability, and number of decisions is also an important issue.

2.2 SINGLE PRODUCT, SINGLE AND MULTIPLE-STAGES MODELS

In [Graves et al., 1993], the authors describe the different types of inventory models for a single product and a single location. The classification is made in terms of three key variables that determine the structure and complexity of the model: *demand*, *costs*, and *other physical aspects*.

The *demand* variable is classified into the following types: Deterministic and Stationary – the simplest model to assume is constant and known demand, like the Economic Order Quantity (EOQ) model; *Deterministic and time varying* – those changes that can be forecasted in advance are called systematic otherwise they are unsystematic; *Uncertain* – the demand distribution is known but its realization cannot be predicted in advance, e.g., there is historical data from which it is possible to estimate the demand distribution. With new products the demand uncertainty could be assumed but it is necessary an estimation of the probability distribution; *Unknown* – the demand distribution is unknown. In this situation it is normal to assume one distribution for the demand and optimize the parameters using Bayes theorem, at each new observation.

The *cost* variable can be divided as follows: averaging versus discounting – when the time value of money is an important issue, a discount rate must be considered; *Structure of the order cost* – the simplest assumption is considering a cost proportional to the number of items necessary. However, assuming a cost with both fixed and variable components is much more realistic; In*ventory costs* – both holding and shortage costs should be considered.

The *other physical aspects* that may be considered are: the *lead-time*, *backordering assumptions*, the *review process*, and *changes that occur in the inventory during storage*.

The *lead-time* is defined as the amount of time from the instant of a replenishment order takes place until it arrives. It is also a measure of the system's response time. The simplest assumption is to consider zero lead-time. However, this only makes sense in situations where the replenishment order time is short compared with the time between reorder instants.

Backordering assumptions are also important to make assumptions about system reactions under shortage inventory situations, e.g., when demand exceeds supply. In these situations the most common assumption is to backorder all the excess demand, which is represented by a negative inventory level. On the other hand, depending on system characteristics, it is also common to assume that all excess demand is lost. A compromise situation where both backorder and lost sales are present is also possible.

The *review process* can be continuous or periodic. Continuous review means that the level of inventory is known at all times and reorder decisions can be made at any time, while periodic review means that the stock level is known only at discrete points and reorders are only possible at pre-determined points corresponding to the beginning of periods.

Changes that occur in the inventory during storage – Usually inventory items do not change their characteristics during the stock period but radioactive materials or volatile liquids may experience some losses. Fixed life inventories, such as food, are assumed to have constant utility until the expiration date is reached. Obsolescence may also affect inventories since their useful lifetime may not be predictable in advance. Such is the case of fashion products.

Despite the fact that models with deterministic and stationary demands seem quite restrictive, they are very important since results are robust with respect to the model parameters, for instance demand rate and cost. The *Economic Order Quantity* (EOQ) is a tremendous example of such property. The results provided for these models are also a fine starting point for more complex models. The literature provides a broad spectrum of variations of the above model, which includes quantity discounts, demand with trends, or perishability, among other issues.

Under the single-product, single stage scenario, the stochastic demand models are probably the most analyzed by current research. The newsboy problem is the basis for most discrete time stochastic inventory models. This single product, single stage, and single period model considers random demand, and charges a unit holding cost for holding stock at the end of the period, a unit shortage penalty cost for lost sales, and a unit ordering cost for the items purchased at the beginning of the period. The optimal policy is of the base stock type. If the initial inventory is under the called *order up to level*, the decision should be to order the difference between that level and the initial inventory. The literature presents some very interesting extensions of this model, namely those considering positive order lead time, lead time uncertainty, or batch ordering.

As referred in [Graves et. al, 1993], on the next level of complexity one can find the dynamic models with positive set-up costs. The optimal policy is an (s, S) policy. The level S is defined as the value that minimizes the cost function G, while s satisfies $s<S$, that is, $G(s)=G(S)+K$. If the initial inventory x is above level s then $G(x)<G(S)+K$ and it is optimal not to order. On the contrary, if initial inventory x is under the level s, the optimal decision is order up to S.

On a continuous time basis, all transactions are monitored systematically and thus inventory ordering decisions can be made as soon as these transactions occur. Therefore, such a system will be more responsive than the periodic review system, where inventory ordering actions can only happen at specific review times. However, one needs to trade-off the costs between the profits achieved with such responsiveness and the continuous-time monitoring system cost, finding the optimal (s, S) values that minimize long run average costs.

[Kapuscinski and Tayur, 1996] analyzed a single product, single-stage capacitated production-inventory model under stochastic and periodic demand. For the finite-horizon, the discounted infinite-horizon, and the infinite-horizon average cases, the authors showed that the optimal policy is of base-stock, or order-up-to type. They used a simulation based optimization method based on infinitesimal perturbation analysis (IPA) to find the optimal policy parameters. Several properties of the optimal policy are discussed, such as the effect of capacity, demand, and penalty cost on the stock-levels. Additionally, some relevant issues related with the optimal solutions are numerically tested.

The survey realized by [Bispo, 1997] refers that, for single machine and single product systems, base stock policies are optimal in a variety of settings. The optimal control policy for single period systems, multiple period finite horizon problems, and multiple period infinite horizon problems, maintains its structure whenever the machine capacity is boundless, deterministic or even stochastic.

The work of [Tayur, 1996] provides some insights into some more complex issues and extends the analysis to single product, serial systems. Once more, a simulation based optimization used IPA as a tool to find the optimal values for the decision variables.

2.3 MULTI-PRODUCT, MULTI-STAGE MODELS

Approaches to the planning, control, and scheduling of automated manufacturing systems originates from at least two distinct disciplines: Operations Research and Control Theory. From a general perspective, these two disciplines have some common characteristics. Both are concerned with decision-making based upon some analytical model which represents the system behavior with a limited degree of accuracy.

2.3.1 Operations Research Perspective

The multi-product, multi-stage models are natural evolutions of the above models. They try to simulate with a high level of accuracy the real production systems. An interesting result was obtained by [Clark and Scarf, 1960]. They showed that for multiple machines in series with single product and no capacity bounds, the base stock policies are optimal in terms of multi-echelon inventory. At a given machine, the echelon-inventory is defined as the sum of inventory downstream from that machine to the last one. For the finite horizon problem the optimal policy is defined by a critical number to order up to for each of the echelon inventories. Later research has extended the previous result to the infinite horizon problem. Despite of the remarkable previous results, for the multi-stage and capacitated models producing single or multiple products, there are very few results concerning optimal production policies. These results usually refer sub-optimal base-stock variant policies proposed as heuristics to find the sub-optimal problem decision variables.

The work presented by [Glasserman and Tayur, 1995], which contributed decisively to this thesis development, considers capacitated, multi-echelon systems producing a single product and operating under base-stock policies. The determination of the optimal base-stock values is achieved by means of an optimization procedure based on estimation of the derivatives with respect to base-stock variables through infinitesimal perturbation analysis. The authors showed that those estimates converge to the correct value for infinite-horizon and infinite-horizon discounted and average cost criteria, are easy to implement, and can also be computed with respect to parameters of demand and yield distribution. Also, the authors presented some numerical examples with very useful simulation details and results. The clear sensitivity analysis makes the final contribution.

Another contribution on multiechelon inventory systems is provided on [Graves, 1996]. The author considers an uncapacitated distribution system consisting of M inventory sites serving more than one site but receiving items from only other site, with deterministic delivery times, and stochastic demand. Each site in the system places orders according to an order-up-to policy. A computational study is used to better understand how to define the order-up-to levels in a two-echelon system, and it is assumed that the policy is constrained in order to achieve a given service level. Hence, for each demand scenario and order policy, are presented the two order-up-to levels, that verify a desired service level (probability of stock-out and fill rate) with the minimum inventory amount. Finally, they conclude that, first, for the minimum-inventory stocking policy, the central warehouse (CW) will stock out during an order cycle with a very high probability, since the base stock at CW is less than the expected system demand for the time from when the CW orders to the time the retail site place their last order. Second, the total safety stock in the system is less sensitive to under-stocking the CW than to over-stocking. Namely, if the CW order-up-to level is changed to achieve a conventional service level, one can notice a significant inventory increase. Third, the CW base stock level appears insensitive to the number of retail sites, which suggests that the case of retailers with non-identical demand rates would show similar behavior. Finally, the general behavior of the optimal inventory policy seems not sensitive to the service criterion. Although, it depends on the specifics of the service criterion.

2.3.2 Control Theory Perspective

The control theory discipline is an alternative approach to address with the planning, control, and scheduling problems of manufacturing systems. As we move up in the production control hierarchy, the time scale increases, and the aggregate models can be used, taking the form of differential equations. This mathematical form provides techniques to determine a feedback control law that looks at the present inventory, machine status, and product demand to determine the present production rate.

The works of [Kimemia, 1982] and [Kimemia and Gershwin, 1983] brought an important contribution to this area. They addressed the problem of controlling a production system with multiple machines and multiple part types, each subject to deterministic demand rates, where the machines are prone to failures. A multilevel hierarchical control algorithm was proposed, involving a stochastic optimal control problem at the top level. Next the authors note that, for each feasible region and for the stationary problem, there is a fixed buffer level, denominated *hedging point*, above which the optimal production rates are zero. In other words, the hedging point can be used to determine when to release a part into the system and how to set the production rates. The interpretation of the hedging point value is that of a base stock. Nevertheless, it should be stressed that, since the formulation only takes into account the production surplus for end products, nothing can be said about the values of the internal buffers.

[Akella and Kumar, 1986] show that the hedging point strategy is the optimal solution for a single-product manufacturing system with two-machine states (up and down). They also use the same aggregate-production model and assume that demand is constant. [Sharifnia, 1988] derives equations for the steady-state probability distribution of the surplus level in a single-product manufacturing system with multiple-machine states, when the hedging point strategy is used. This work is important because there is an arbitrary number of machine states' corresponding to different failure modes of the system. The author shows that for each machine state, the cost-to-go function reaches its minimum at a hedging point.

Other papers have continued the original work of Kimemia by means of addressing increasingly complex systems and models, such as including the internal buffers in the analysis for multiple products and considering manufacturing systems with re-entrant flow, [Bai and Gershwin, 1994], [Bai and Gershwin, 1995], and [Bai and Gershwin, 1996].

In summary, the area of flow rate control has produced one of the most comprehensive and complete methodologies to deal with production planning and control for manufacturing systems. It involves long-term decisions (determining safety levels for the production), mid-term decisions (release of new parts into the system) and short-term decisions (scheduling parts into the available machines). The approach relies on the sound theory of optimal control and benefits from its elegant results, generating control policies that are functions of the system state. However, there are some issues which remain to be dealt with in a more satisfactory way. The only source of randomness considered is machine breakdowns. Demand variance is not taken into account, as well as the effects of random yield and processing time uncertainty, in order to determine the long-term safety levels. Also, there is no effort made to determine how capacity should be shared, nor on studying the impact of different dynamic capacity schemes have on the system performance measures.

2.4 MAKE-TO-STOCK VS. MAKE-TO-ORDER

The decision of stock or not to stock a product can be influenced by several factors. As referred in [Silver et al., 1998] these factors include the system cost (forecasting activities, file maintenance, etc.) per unit time of stocking an item, the unit variable cost of the item both when it is bought for stock and when it is purchased to meet each demand transaction. It includes also the cost of temporary backorder associated with each demand when the item is not stocked, the fixed setup cost associated with the replenishment in each context, the carrying charge (including the effects of obsolescence), which, together with the unit variable cost, determines the cost of carrying each unit of inventory per period of time. Finally, it accounts also for the frequency and magnitude of demand transactions and for the replenishment lead-time. It is important to examine the replenishment lead-time and certify that customers are willing to wait the additional transportation time. If not, the cost of a temporary shortage should also be included in the analysis.

2.5 MODELING ISSUES

There are several tools to model and analyze discrete-event dynamic systems: Queuing Theory; Markov Chains; Petri nets and Simulation. This last tool, in addition with the perturbation analysis technique, are the two complementary methods used in this thesis for the implementation and optimization of a production process.

The problem of planning, control, and scheduling activities in a manufacturing facility has received considerable attention from the operations research and industrial engineering communities over the last 25 years. These main issues are concerned with decision-making based upon some model. These models represent the system behavior with a limited degree of accuracy, and they depend heavily on the issues being studied.

There are several key modeling issues that complicate the control of an automated manufacturing facility. Any strategy or policy for operating such a system must be able to handle: uncertainty in product demand knowledge (at all levels of the production hierarchy); finite and random manufacturing capacity; random machine failures and repair rates. The knowledge of total product demand is based on actual orders that are only known for a specific period into the future plus some forecasted value obtained from prior experience as well as seasonal and cyclic variations. Uncertainty in product demand makes it difficult to set manufacturing capacity and rates of production.

If manufacturing capacity were excessive, the control problem would become trivial. There would be sufficient machines to do all jobs at the expected time. However, excess capacity would incur higher costs. Capacity is not just the number of machines in the shop floor. The true capacity is related to the sources of uncertainty in the manufacturing system. One source of uncertainty is the reliability of the machines. Machines are prone to fail at random times, and the time to repair is also a random variable. This problem is strongly related with the plant maintenance program and procedures.

Examples of other sources of uncertainty that affect capacity are worker and material absence; variations in the quality of the raw materials, which may affect production yield; variation in machine processing times caused by different operator experience or quality of raw materials. These types of uncertainty must be considered when estimating system capacity. The models and control algorithms should also be able to handle issues like *setup times* and *finite buffer sizes*. When a machine switches from making one part to another, the setup of the machine must be physically changed. This takes a certain amount of time to accomplish and is called the *setup time*, which in many situations is a random variable.

As referred in [Desrochers, 1990] there are several performance criteria that are used to determine optimal production control policies. Among these are: satisfy demand accurately; meet the due date; minimize the production costs; and minimize the total time required to complete all the jobs.

Additionally, [Graves et al., 1993] referred that there are several other important issues that must be considered in model design: the time unit (months, weeks, days or shifts); the time horizon (one month, one year, etc); level of aggregation of the products (one aggregate product, a few product categories, or full detail); level of aggregation of the production resources (one plant or multiple resources); frequency of re-planning (every basic time unit or less); number and structure of production plans (e.g., a one-year plan, a three-month plan, and a one-week plan, the last one released to the plant).

Time Unit and Time Horizon

The aggregate production planning is usually modeled using weeks or months as the time unit. The shorter the unit the more complex the problem that needs to be solved at the detailed level. In most real situations there is a range of values for the planning horizon. If seasonality exists it should be used a plan horizon of one year or more.

Products Aggregation Level

The aggregation approach is to focus attention on major cost sources, defining a plan that can be implemented in a simple and economic way. To do so the model structure considers relatively large costs and important resources. The appropriate level of aggregation depends on the cost structure, the production line, and stationarity of the production process. Most of the models described in the literature assume *one aggregate product*.

Facilities Aggregation Level

Many aggregate planning models define the facility as a single resource. Some authors consider the work force as a single resource and assume that all levels of the work force can fit within the plant.

Resources like types of labor, work centers, raw material availability may also be considered.

Frequency of re-planning

Although plans are made over an acceptable planning horizon, re-planning occurs often and the plans are used in what is usually termed as a *rolling* mode.

In practice, re-planning is done all the time because of changes in the data (forecast revisions, modified schedules, machine breakdowns, etc.). If re-planning is too frequent, then the system may become unstable due to excessive nervousness, as mentioned in the literature (see [Graves et al., 1993]).

Levels of plans before implementation

Industrial companies often have several plans, at different levels of aggregation, rather than just one aggregate plan and a detailed schedule.

The levels of aggregation depend on the existing number of plans, and the choices depend on the situation. The development of an appropriate set of models is related with the hierarchical production planning.

2.6 RANDOM YIELD

In the literature the predominant emphasis has been on measures to deal with demand uncertainty. However, in a complex production and inventory system there are several other sources of uncertainty. One can divide them into two main categories: *capacity uncertainty* and *yield uncertainty*, according to the different ways by which they influence the outputs. Under a variable production capacity scenario, the realized capacity in a given time period could constrain what can be actually produced. If the planned production quantity is greater than the realized capacity, only part of it can be processed. On the other hand, the presence of random yield causes a random portion of processed items to be defective. Although both categories of uncertainty referred above are often present simultaneously in a production process, variable capacity is outside the scope of this thesis.

The implications of yield losses are typically in the form of the costs of surplus due to good yield and scrap or rework and the costs of not satisfying the total demand (shortage costs), which are then incorporated into the objective function that drives the definition of lot sizing policies. Other consequences of random yield are in the reduction of production capacity and delayed output and the consequent late deliveries. This may cause starvation of downstream processes/operations, and thereby affect the production capacity.

[Bispo, 1997] suggests that when it comes to studying systems with random yield one needs to consider essentially three different domains: modeling yield, controlling systems with a given yield structure, and improving yield. The modeling domain is concerned with identifying the more suitable method to approximate the random yield process as a stochastic one. The second domain deals with determining production decisions given that random yield exists and is unavoidable. Finally, improving yield is related with quality control issues and aims to improve production processes in order to increase yield or monitoring the output at critical stages to prevent that the process turns out of control. This last area is also outside the scope of this research.

A notable overview of developments in the context of optimal inventory control in the presence of random yield is provided by [Yano and Lee, 1995]. The authors described the research to date on lot sizing in the presence of random yield, discussing issues related with the yield uncertainty, modeling of costs, and performance. The issues presented on the following sub-sections are strongly based on the referred paper.

2.6.1 Modeling of Yield

The modeling of systems with random yield should consider the following topics: modeling of costs, modeling of yield uncertainty, and measures of performance. The modeling of some costs, such as setup and shortage costs, usually are not affected by the presence of random yields. Variable unit costs and inventory holding costs must be modeled depending on the scenario. Hence, variable costs should be defined as a function of the appropriate quantity (input, output, ordered or received quantity), while inventory holding costs often depend upon the timing and nature of the inspection process.

[Yano and Lee, 1995] present five different ways of modeling yield uncertainty. The simplest model of random yield assumes that the number of good units in a batch of size Q follows a binomial distribution with parameters Q and p, where p is the probability of generating a good output from one unit of input. Such model is suitable for systems in control for long durations. A second way to model yield uncertainty is to define the distribution of the fraction of good units, often referred to as yield rate or stochastically proportional yield. This model is applied for large batch sizes, or when the variation of the batch size between production runs tends to be small. The third modeling approach considers that the distribution of the fraction of good items changes with the batch size. It involves specifying the distribution of the time until a repetitive process starts to make defective parts and becomes out of control. Situations where the failure is the result of state deterioration of the production system during a production run are examples of application of this model. Another modeling approach is applied for systems with the fraction of acceptable items stochastically increasing with the length of the production run. This happens when the process setup involves trial and error in setting values that influence decisively the output quality. At last, the fifth approach assumes that yield uncertainty is the consequence of random capacity, the main source of which might be unreliable equipment. In such systems the output quantity is the minimum of the input quantity and the available capacity.

The performance measures for lot sizing with random yield have as dominant criterion the minimization of expected costs. However, [Yano and Lee, 1995] highlight the fact that, with few exceptions, constraints on various measures of service have not been considered yet.

In summary, models for lot sizing decisions in presence of random yield should consider a consistent characterization not only of the yield loss process and the distribution of yield losses, but also of the inspection process and its effect on timing and costs. An objective function and constraints that capture the consequences of yield losses is also a critical element of the modeling process.

2.6.2 Production Decisions in the presence of Random Yield

The literature in continuous-time review models contemplates exclusively single-stage models, and almost all require rather strong assumptions regarding stationarity of yield distributions, demands, and costs.

As mentioned in [Bispo, 1997] even the model with stochastic demand and random capacity retains the base stock structure of the optimal policies. However, the presence of random yield in production systems neutralizes many of the nice structural properties of inventory control policies, namely the base stock structure. Many base stock policies are simultaneously *order-point* and *order-up-to* policies. Order-point policies are characterized by a point (or set of points) defined in terms of initial inventory, above which it is optimal not to order. Order-up-to policies are identified as those where, when the optimal decision is to order, the optimal quantity is such that the ending inventory is a particular inventory point (or set of points). Many systems with random yield preserve the order-point property but lose the order-up-to property. Intuitively, one could expect that it might be estimated based upon the amount produced multiplied by the inverse of the expected random yield. However, the following references show that such policy is not correct.

[Gerchak et al., 1988] analyzed the finite-horizon problem with constant costs and stationary demand. Yield was assumed invariant with the production quantity and stationary over time. Using as performance measure the profit maximization, they concluded that the order point does not depend on the yield distribution. Moreover, the amount to order depends not only on the expected random yield but also on the second moment of the random yield. In fact, the optimal quantity increases as the expected yield decreases but it decreases as the variance increases. It sounds reasonable since, with a high yield variance, large batch orders have a high probability of wasting a big amount of material, whereas smaller batch orders induce smaller absolute waste. For the multiple period problem they have demonstrated that myopic policies are not generally optimal, and that order-up-to policies are not optimal, that is, the optimal production quantity is not necessarily a linear function of beginning-of-period inventory.

In [Henig and Gerchak, 1990] the single stage, single period, finite-horizon and infinite-horizon models with general production, holding and shortage cost structures, are analyzed. They verified that, under an assumption of stochastically proportional yield, there exist critical order points for both finite and infinite horizon problems. These critical order points are such that no order should be placed if the on-hand inventory level is above the critical order point; otherwise, an order should be placed. However, they derived a significantly complicated function of the system's parameters, based on Taylor series, for the order quantity. Additionally, the critical order point for the infinite-horizon problem with stationary demand and costs is also stationary and greater than or equal to the one obtained in the model with no losses. This implies that one can acquire an advantage from producing farther ahead than one would without random yields. They also provide an interesting literature review on random yield problems covering continuous and periodic review models.

Another interesting discussion on random yield effects is provided in [Grosfed-Nir and Gerchak, 1996]. A general framework for modeling and analyzing Multiple Lot-sizing Production to Order (MLPO) single stage problems is discussed. The relevance of the MLPO problem can be explained not only under the scope of this thesis, but also mainly by the tremendous growth of production to order of relatively small volumes of custom-made items in recent years. The paper also derives a method for computing the mean and variance of the cost, and presents a numerical example where such cost structure is tested for several policies. The authors used a "motivating example", with a yield structure characterized by a fixed probability for the process to become out of control, for each unit produced. All units completed before such occurrence are good; otherwise are defective. They referred several properties of the optimal solution that seem so intuitive that one would expect them to hold for all reasonable yield patterns, and researchers had indeed occasionally assumed them in their models. However, they proved that often it may not be the case. The first property usually assumes that the optimal run size N_D increases with demand D. Their example shows clearly that this property does not always hold. Intuitively, beyond some value of N, the chances of the process not getting out of control are so low that even with a high demand level D, it is better to save on variable costs by starting with a smaller run, since incurring several setups is practically certain. The other interesting result contradicts the idea that the optimal lot size N_D is always greater than demand *D*, since selecting an *N* that is larger that *D* does not increase the probability that *D* units will be good.

A different approach, related with Material Requirements Planning (MRP) systems, is presented in [New and Mapes, 1985]. The paper derives four different strategies to deal with random yield. The authors concluded that the best policy for make-to-stock situations under continuous production is to adjust the production quantity for the average yield rate and use fixed buffer stocks. For the make-to-order situations, they suggest a modification to the above policy to allow multiple production runs. For products with infrequent demand or for a single custom order the strategy recommended is the use of service level measures. At last, for multiple-order custom products, they propose using safety time in order to make possible the production division into smaller batches, which enable more frequent inspection and reduces the probability and size of overruns.

[Karmakar and Lin, 1986] consider a multi-period problem with both random demand and yield. The objective is to allocate resources (consumable and renewable) to *N* products to minimize the sum of expected variable production, inventory holding, shortage, regular time, overtime, capacity acquisition, and capacity retirement costs. They develop methods to determine upper and lower values of the objective function.

2.7 INFINITESIMAL PERTURBATION ANALYSIS

Many systems found in manufacturing environments can be described as discrete-event systems. These events, which may occur at either deterministic or stochastic times, push the system from state to state throughout time by finishing and initiating a sequence of activities. Machine failures and repairs, sudden changes in demand profile, and machine starvation or blockages are all events occurring at discrete instants in time. The occurrence of these events has an impact on the dynamic response of the manufacturing system, given that the beginning or conclusion of one of these events can propagate from machine to machine, changing its performance. The interaction of these events throughout time is a critical issue of discrete-event dynamic systems.

2.7.1 The Key Features of Perturbation Analysis

The Perturbation Analysis (PA) theory had its origins with [Ho et al., 1979]. They established the flavor of perturbation analysis, focusing on the buffer storage optimization in a production line, based on a gradient technique. However, as referred in [Desrochers, 1990] the PA theory had its accidental beginning in 1977 while researchers from Harvard University were working at the FIAT 131 engine production-line's monitoring system. This research leads to an efficient technique for calculating gradients, or sensitivities, in a serial-transfer line. The gradient for a discrete-event dynamic system is related to the perturbation of events and can be interpreted as the addition of one unit on the buffer size b_i, which allows machine M_i to produce the same number of parts in less time. This effect represents a local gain in production, and tends to propagate pieces downstream in the line, since M_i feeds b_i. Correspondingly, buffer vacancies can be propagated upstream. The status of the other machines will determine if this propagation will result in a line gain. For example, if the last machine has an extremely high failure rate that gain may not occur. Then, one is interested in computing the sensitivity represented by the rate of *gain in production line* by *Buffer size increase at each buffer location* and then allocate the buffer size at each location to maximize the performance index. The system-performance measures are statistically compared through experimentation. Performing experiments on discrete-event simulation models involves a continuous procedure of executing a set of instructions on the computer and, each simulation run evaluates a single scenario.

In a manufacturing system, for example, we may be interested in the effect of changing the processing time of one machine on the overall throughput of the system. To analyze such effect it would be necessary to run several experiments with the model at different processing times. Each simulation experiment would consist of changing one parameter and then estimating a measure for system performance. The main problem with this method is that each parameter must be analyzed individually in order to isolate its effect on the system. This procedure becomes very complicated as the number of parameters becomes significant, resulting in a prohibitively large number of experiments. Suppose there are N machines and N buffers. If we use extensive simulation it would be necessary $N+1$ simulations in order to compute the N components of the gradient vector. On the other hand, if we use PA the N gradients can be determined from one single simulation. This procedure gives the possibility to derive sensitivity estimates at each model simulation, giving to the analyst a significant help on the decision process of which parameters it is necessary to adjust in order to improve performance.

The implementation of perturbation analysis techniques involves the installation of a mechanism within the simulation model to monitor the sample path as it evolves. With the simulation of the unperturbed (nominal) path, perturbation analysis tries to forecast what would have happened if the values of the parameters had been changed.

See also [Ho, 1988] for some intuitive explanations of what perturbation analysis is and why does it work. One of the leading works of the perturbation analysis on general queuing networks was developed by [Cassandras and Ho, 1983]. The authors have introduced the concept of similarity and, on the basis of an appropriate state-space representation, showed that perturbation equations around a nominal trajectory can be used to predict behavior by observing only one single sample realization for some classes of discrete-event dynamic systems, such as queuing networks and production systems.

2.7.2 Finite and Infinitesimal Perturbations

The sensitivity of a transfer line production to a buffer level change is an example of *finite perturbations* in which the rules of perturbation propagation are used on parameters that take discrete values. On the other hand, there are several continuous parameter changes that could influence a series of discrete events. By way of illustration, if the repair rate could be increased at machine M_i, it would also result in a local production gain at M_i which, depending on the status of other machines, may propagate downstream to the end of the line. Therefore, *infinitesimal perturbations* are used to obtain the sensitivity of a performance measure with respect to continuous parameter changes.

In spite of the fact that there is a deficit of theoretical foundation for the finite perturbation analysis algorithms, the infinitesimal perturbation branch had an interesting development and during the 80's there was already significant theory supporting it. In the specific context of Infinitesimal Perturbation Analysis (IPA) [Suri and Zazanis, 1988] were responsible for the first developments on this area. For an M/G/1 queuing system, they considered the sensitivity of the mean flow time to a parameter of the arrival or service distribution. Among other issues, they showed that a perturbation analysis algorithm implemented on a single sample path of the system gives strongly consistent estimates of the sensitivity.

The main issue of IPA is that we can make small changes to certain input parameters (φ) of a system without altering the sequence in which events occur. This means that, for any given simulation and for certain parameters, a change in one parameter ($\Delta\varphi$) can be made small enough, like an infinitesimal change ($\delta\varphi$), such that the event times are shifted without any alteration in the respective order of occurrence. [Johnson and Jackman, 1989] referred this as the "deterministic similarity" assumption. The authors showed that IPA is an accurate technique for sensitivity analysis for serial transfer lines and that the assumption of deterministic similarity is a sufficient rather than a necessary condition. Such feature is due to the linear nature of the IPA estimate. Nevertheless, as the experimental results indicated, discontinuities do exist in the performance measures functions, and hence the accuracy of IPA estimation for finite perturbation is not guaranteed. Consequently, the main limitations of the infinitesimal perturbation analysis are related with the incorporation of such a technique into a general-purpose simulation language.

2.7.3 Perturbation Generation and Propagation

The issue of perturbation generation and propagation is addressed in [Johnson and Jackman, 1989] in a very simple manner. Two separate issues must be considered in the implementation of an IPA algorithm: *perturbation generation* and *perturbation propagation*.

As was already mentioned, a small change ($\Delta\theta$) in a parameter (θ) has a smooth effect on the performance measure function [$f(\theta)$]. Perturbation generation is nothing but the difference (Δx_i) between the nominal (θ_n) and perturbed (θ_p) value of the parameter. For example, if we are interested in perturbing the mean of a parameter that follows an exponential distribution, we could simply convert the same random seed for the two means (μ_n, μ_p) and subtract them to find the difference (Δx_i). Because we are interested in the instantaneous gradient or derivative of the performance measure function, and since we are considering infinitesimal changes ($\delta\theta$) of the parameter (θ), this process can be substantially simplified. One never needs to specify the perturbed mean or calculate a finite difference (Δx_i) since the information necessary to calculate the gradient is contained in the nominal value of the service time ($dx/d\mu = x/\mu$). Such property turns IPA into an efficient method since it requires little additional effort in addition to the computation necessary to simulate the system.

CHAPTER 3

MODEL DEFINITION

3.1 INTRODUCTION

The model presented in this thesis considers a multi-product, multi-stage capacitated process, with product splitting, random yield, and insignificant set-up times, facing uncertain demand. At a given process stage s, the inventory available of product p can be used to process m different sub-products. Whenever such a stage exists, it is said that a product split occurs. The framework used is periodic review, capacitated, multi-product, production-inventory system working under an echelon base stock policy. To be precise, given a particular product p and stage s, it is necessary to add all inventory downstream from that stage to compute the echelon inventory. If the echelon inventory falls below the corresponding base stock value, the production decision will be equal to that difference, provided there is sufficient upstream inventory and enough capacity.

Under the framework of discrete time inventory control, the problem now addressed analyzes several different production policies, which are tested in perfect and random yield production systems. The system performance is measured in terms of costs and lead-times. One may consider discrete time as a valid approach for this problem, since decisions are made every production shift, inducing a periodic review structure on the production decisions.

3.2 THE GLASS MANUFACTURING PROCESS

The model presented on this thesis simulates the hand-made glass production process, which may be divided in two main areas – the hot area and the cold area – consisting on a set of sequential operations. Figure 3.1 shows a general layout for the glass manufacturing process. Usually there are seven different product families (“Centrifugado”, “Belga”, “Frascaria”, “Marisa Fina”, “Marisa Grossa”, “Monomóldica” and “Multimóldica”) depending on the product type and used technique, some of them being more technological-intensive and others labor-intensive.

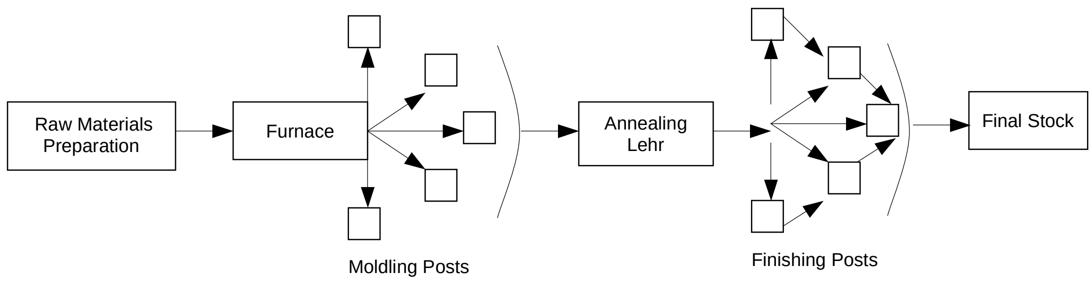

Figure 3.1 – Glass Manufacturing Production Process

As soon as the several operations were perfectly identified, it was possible to identify a typical working sequence. Remark that at the finishing posts the products always follow the same operations sequence. However, there are some products, which do not need to perform all the operations. In these situations the products move immediately downstream to the next relevant stage. To improve the understanding and identification of the products flow, and for an easy model implementation, a simplified version of the process was developed. Figure 3.2 presents such version.

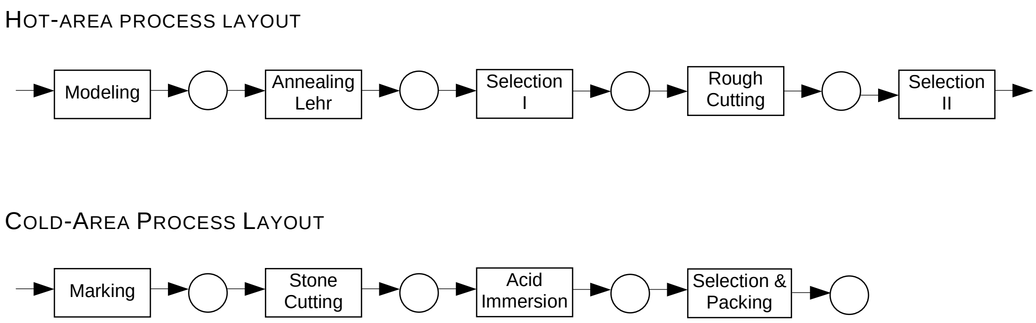

Figure 3.2 – Glass manufacturing process layout (model version)

I. Hot-Area Process Stages

After the raw materials preparation, the glass is melted on the furnace according to specific parameters (temperature, time, pressure, etc.) in order to achieve the correct mechanical and visual properties. The melted glass is taken out of the furnace to the several molding posts. Some of these posts may be manual or automatic. Each manual post consists of, apart from the molding tools, a team of two to twelve workers, one of which is the post chief. The number of people in each team changes according to the product family. The features of the production process in this area may become much more complex, increasing the cycle time, depending on mold types, product color, product weight, finishing, warming and cooling processes.

As far as the capacities of the furnace, molding posts and annealing lehr, they all are finite. However, the furnace's capacity (800ton./day) ought to be the main constraint to the process. The furnace needs to work on a continuous mode, in order to avoid a possible failure. Therefore, it is valid to assume that melting glass is available at any time for the modeling stage, or in other words, the furnace capacity is several times larger than the total modeling stage capacity, excluding the possibility of a starvation scenario. Before and after the rough cutting posts, the articles go through an inspection stage (Selection I and Selection II).

II. Cold-Area Process Stages

After being taken out of the hot-area's last stage, most of the articles need to go through a set of finishing process (Marking, Stone Cutting, Acid Immersion, and Selection and Packing). The number of people working in each post/machine is usually one or two. Also, there are people whose task is carrying the articles from one post to another. The posts capacity is finite and the setup times of the different posts are negligible when compared to the operation processing time.

3.3 PRODUCTION STRATEGIES

This work deals with three different production strategies, all supported on the existence of a set of base stock variables, *Z*, defined for each echelon inventory variable. The echelon inventory of a stage is defined as the sum of local inventories from that stage down to the last stage of production. The three strategies are: Make-to-Order strategy, Make-to-Stock strategy, and Delayed-Differentiation strategy.

Under an MTS strategy, the production decision for product *p* at stage *s* will depend, simultaneously, on the difference between the base stock level and present inventory, the inventory available at the previous buffer, the available capacity, and the production limit.

With an MTO strategy the products are only produced after demand is realized and all the base stock levels are set equal to zero. That is, an MTO strategy is an MTS strategy with the *Z* levels set to zero.

One can describe a DD strategy as a combination of a make-to-stock policy and a make-to-order strategy. In order to decrease the product's manufacturing lead-time one can create an intermediate stock of products with the same baseline geometry. On the hot-area stages the production strategy is an MTS one. After demand is realized, the products are subtracted from the intermediate stock and produced under an MTO strategy, during all cold-area stages. In other words, a DD policy is an MTS strategy with the cold area *Z* levels set to zero.

Under this thesis context one should distinguish two concepts: policy and strategy. The first concept is related with the production decision making procedure, which is based on a comparison between the weighted shortfall, the available inventory and capacity, and the production limit. Then, the production policy can execute several different management strategies such as the strategies under analysis, depending on the setting of some of the control parameters.

3.4 THE BASIC MODEL

The remaining of this chapter describes the main structure of a production process model implemented to analyze and understand the process dynamics. Before being completed, each product *p* has to go through *S* process stages (machines). There are infinite capacity intermediate buffers where the products are placed while waiting their turn to be processed by the next machine or depleted by external demand if the previous operation was the last one. Somewhere during the process product splitting occurs. At this operation each product *p* can lead up to *M* sub-products.

Despite the results discussed in chapter 2 concerning yield modeling and its consequences on production decisions, the present work will consider standard base stock policies. However, it is our conviction that the base stock levels will reflect reasonably well the yield influence. We refer the reader to Chapter 5 for a discussing on future work.

The following notation is valid throughout this text:

- *P* products, indexed by p_m (the index m= 1, ..., M indicates sub-class of product *p* defined for the cold area stages);
- *S* stages, indexed by *s* (the last stage of the process is stage 1, while first stage is stage S);
- *N* time periods, indexed by *n*;
- $D_n^{p_m}$: Demand for product p_m in period *n* at stage *1* (last stage);
- $Z^{p_m s}$: Echelon base stock level for product p_m at stage *s*;
- $I_n^{p_m s}$: Inventory in time period *n* for product p_m at stage *s*;
- $E_n^{p_m s}$: Echelon inventory in time period *n* for product p_m (sum downstream all $I_n^{p_m s}$, beginning at stage *s*);
- $Y_n^{p_m s}$: Shortfall in time period *n* for product p_m at stage *s*;
- $P_n^{p_m s}$: Production amount in period *n* for product p_m at stage *s*;
- $\alpha^{p_m s}$: Random yield correction factor for product p_m at stage *s*;
- C^s: Capacity of stage *s*;
- $T^{p_m s}$: Processing time of product p_m at stage *s*;
- $U^{p_m s}$: Production limit for product p_m at stage *s*;

This nomenclature considers stage *S* as the first stage of the process and stage *1* the last one. For example, $I_n^{1_3 1}$ represents the available inventory of sub-product *3* of product *1*, at the last buffer at the beginning of time period *n*, from which random demand is satisfied or backlogged. For ease of explanation and henceforth, the reference to a product p is made generically, that is, not regarding if it is a product p or a sub-product p_m.

The issue of how to compute costs is discussed in [Glasserman and Tayur, 1995]. One may consider two alternative procedures: Cost accounting after demand takes place or cost accounting after production has been decided. In the first procedure, the production quantities are set for all products and stages at the beginning of each period. Inventory levels are updated at the end of the period according to the achieved production and the realized demand, which is assumed to occur only after production has been decided. This demand is instantly satisfied if there is sufficient inventory. If not, it is backlogged and eventually satisfied with future production runs. The second procedure assumes that demand occurs at the beginning of each time period. Next the production quantities for all products and stages are determined and, finishing the period, the inventory levels are updated according to the reached production and demand. After, costs are incurred, depending on the remaining inventory quantities. In line with the established in inventory control theory, the cost estimation procedure in this work follows the exposed on the first procedure.

Even if the model described in this chapter embodies some sources of uncertainty, one needs to consider it a relatively simple model. The demand and production yield randomness are the only two sources of uncertainty considered in the model. Production processes are under many other sources of uncertainty like machine failure, raw material properties, processing times, and so forth. Nevertheless, one needs to understand simplified versions of production processes in order to get insights and effectively manage more complex versions of them. Incorporating variable capacity and non-deterministic processing times on the present model will not be a complex mission, since its structure is easily adjustable to such scenario.

3.4.1 Variable Classes

The main model variables should be classified into three classes according to their function: state variables, control variables, and decision variables.

Echelon variables and shortfall variables, both functions of inventory variables, form together the first class of variables. Their value reproduces the process dynamics since they are updated at each simulation cycle. The echelon base stock, the production limit and the capacity variable form the control variables class. The first two variables hold the same value during all simulation, while capacity variable is time independent, being initialized at all simulation cycles. The last class consists only of the production variables, which have to be computed at each simulation cycle and reset for next cycle, because they are not explicitly time dependent.

3.4.2 State Variables Basic Recursions

The inventory, echelon inventory, and shortfall variables are the state variables that define the process dynamics. Their equations will now be presented as well as the initial conditions and a set of alternative variables.

I. INVENTORY DYNAMIC EQUATION

The inventory equations are given by:

$$I_{n+1}^{p_m s} = \begin{cases} I_n^{p_m 1} - D_n^{p_m} + \alpha^{p_m 1} P_n^{p_m 1}, & \text{last stage of the cold area} \\ I_n^{p_m s} - P_n^{p_m (s-1)} + \alpha^{p_m s} P_n^{p_m s}, & \text{remaining cold - area stages} \\ I_n^{ps} - \sum_{m=1}^{M} P_n^{p_m (s-1)} + \alpha^{ps} P_n^{ps}, & \text{last stage of the hot area} \\ I_n^{ps} - P_n^{ps} + \alpha^{ps} P_n^{ps}, & \text{otherwise} \end{cases} \tag{3.1}$$

The first line of the equation describes the inventory consumption by external demand at stage *1* (last stage) at a given time period *n*. The remaining lines of (3.1) describe the inventory evolution at a given intermediate stage *s* and time period *n*. The inventory level is consumed by an amount corresponding to the production of the downstream stage *s-1* and increased by the amount effectively produced at stage *s*. The sum on the last stage of the hot-area reflects the division of each base product in a set of *M* sub-products.

II. ECHELON INVENTORY EQUATION

The echelon inventory can be described by the following recursive equation:

$$E_n^{p_m s} = \begin{cases} I_n^{p_m 1}, & \text{last stage of the cold area} \\ I_n^{p_m s} + E_n^{p_m (s-1)}, & \text{remaining cold area stages} \\ I_n^{ps} + \sum_{m=1}^{M} E_n^{p_m (s-1)}, & \text{last stage of the hot area} \\ I_n^{ps} + E_n^{p(s-1)}, & \text{otherwise} \end{cases} \tag{3.2}$$

The sum of inventory downstream for each product *p* at a given time period *n* corresponds to the echelon inventory. At the last stage (stage 1) the echelon inventory is just the local inventory, as stated by the first line. The echelon inventory for the other stages is defined recursively by the other lines depending on which stage one is interested.

After simple manipulation of (3.2) it is possible to obtain the echelon inventory dynamic equation:

$$E_{n+1}^{p_m s} = \begin{cases} E_n^{p_m 1} + \alpha^{p_m 1} P_n^{p_m 1} - D_n^{p_m}, \quad \text{last stage of the cold area} \\ E_n^{p_m s} - \sum_{j=1}^{s-1} \left(1 - \alpha^{p_m j}\right) P_n^{p_m j} + \alpha^{p_m s} P_n^{p_m s} - D_n^{p_m}, \\ \text{remaining cold-area stages} \\ E_n^{ps} + \alpha^{ps} P_n^{ps} - \sum_{m=1}^{M} \left[\sum_{j=1}^{s-1} \left(1 - \alpha^{p_m j}\right) P_n^{p_m j} \right] - \sum_{m=1}^{M} D_n^{p_m}, \\ \text{last stage of the hot area} \\ E_n^{ps} - \sum_{j=1}^{s-1} \left(1 - \alpha^{pj}\right) P_n^{pj} + \alpha^{ps} P_n^{ps} - \sum_{m=1}^{M} D_n^{p_m}, \quad \text{otherwise} \end{cases} \tag{3.3}$$

III. SHORTFALL DYNAMIC EQUATION

The following equation describes the shortfall process dynamics. A shortfall is defined as the difference between the echelon base stock and the echelon inventory and, by definition, is always non-negative.

$$Y_n^{ps} = Z^{ps} - E_n^{ps} \tag{3.4}$$

Z^{ps} represents the echelon base stock level for product p at stage s. Considering the inventory dynamic equation (3.1), it is possible to derive a similar one for the shortfalls, which highlights the demand role in moving the echelon inventory away from the target and the production efforts trying to reduce shortfall to zero:

$$Y_{n+1}^{p_m s} = \begin{cases} Y_n^{p_m 1} - \alpha^{p_m 1} P_n^{p_m 1} + D_n^{p_m}, \quad \text{last stage of the cold area} \\ Y_n^{p_m s} + \sum_{j=1}^{s-1} \left(1 - \alpha^{p_m j}\right) P_n^{p_m j} - \alpha^{p_m s} P_n^{p_m s} + D_n^{p_m}, \\ \text{remaining cold area stages} \\ Y^{p_m s} - \alpha^{ps} P_n^{ps} + \sum_{m=1}^{M} \left[\sum_{j=1}^{s-1} \left(1 - \alpha^{p_m j}\right) P_n^{p_m j} \right] + \sum_{m=1}^{M} D_n^{p_m}, \\ \text{last stage of the hot area} \\ Y_n^{ps} + \sum_{j=1}^{s-1} \left(1 - \alpha^{pj}\right) P_n^{pj} - \alpha^{ps} P_n^{ps} + \sum_{m=1}^{M} D_n^{p_m}, \quad \text{otherwise} \end{cases} \tag{3.5}$$

IV. INITIAL CONDITIONS

At instant *n=0* the state variables will be set at their base stock levels:

$$I_0^{p1} = Z^{p1}$$

$$I_0^{ps} = Z^{ps} - Z^{p(s-1)}$$

The initial conditions for echelon inventories will be defined according to equation (3.2):

$$E_0^{ps} = Z^{ps}$$

The remaining initial variables are set equal to zero.

V. ALTERNATIVE VARIABLES

Let the following equation represent an alternative set of control variables linearly related with the multi-echelon base stock variables.

$$\Delta^{p_m s} = \begin{cases} Z^{p_m 1}, \quad \text{last stage of the cold area} \\ Z^{p_m s} - Z^{p_m (s-1)}, \quad \text{remaining cold - area stages} \\ Z^{ps} - \sum_{m=1}^{M} Z^{p_m (s-1)}, \quad \text{last stage of the hot area} \\ Z^{ps} - Z^{p(s-1)}, \quad \text{otherwise} \end{cases} \tag{3.6}$$

[Bispo, 1997] proposes this alternative set of control variables, which simplifies the perturbation propagation. In order to keep consistency, the base stock variables have to be ordered respecting $Z^{ps} \geq Z^{p(s-1)}$, which is not always simple to manage. The use of these alternative variables will simplify such propagation.

3.4.3 The Model Cost Structure

The intrinsic product value increases as it flows throughout process stages since it gets more value added at each stage. Therefore, the holding costs are estimated based not only on raw materials cost, but also on the hourly cost of resources used to perform the activities related with a given stage, while backloging costs are related with last stage's holding cost.

Let

e^s - Energy cost at stage *s*, per unit of time;

r^s - Human resources cost at stage *s*, per unit of time;

m^p - Raw material cost for product *p*; (3.7)

which allows the recursive definition of holding costs structure as:

$$h^{ps} = \begin{cases} m^p + (e^S + r^S)T^{pS}, & \text{for s = S} \\ h^{p(s+1)} + (e^s + r^s)T^{ps}, & \text{for other hot - area stages} \\ h^{p(s+1)} + (e^s + r^s)T^{p_m s}, & \text{for cold - area first stage} \\ h^{p_m(s+1)} + (e^s + r^s)T^{p_m s}, & \text{otherwise} \end{cases} \tag{3.8}$$

The penalty costs are not as simple to estimate because it is necessary to convert into numbers all the disadvantages of not satisfying the costumer, which suggests some intangible features.

3.4.4 The Performance Measures

Present research usually considers two different kinds of performance measures: operational cost based and service level based. The traditional cost based measures are related with the accounting of costs to inventories and backlogs, while service level based measures deal with the system's performance in satisfying costumer needs.

I. OPERATIONAL COST BASED MEASURES

Defining holding and backloging costs by the following notation,

h^{ps} - holding cost rate for product *p* and stage *s*;

b^{p} - backlogging (or penalty) cost rate for product *p* at stage *1*; (3.9)

and the single stage cost as:

$$J_n = \sum_{p=1}^{P} J_n^p \tag{3.10}$$

where

$$J_n^p = \sum_{s=2}^{S}\left[\left(I_n^{ps}\right)^+ h^{ps}\right] + \left(I_n^{p1}\right)^+ h^{p1} + \left(I_n^{p1}\right)^- b^p + \sum_{s=1}^{S}\left(1-\alpha^{ps}\right)\left(h^{ps}-m^p\right)P^{ps} \tag{3.11}$$

The last term of equation (3.9) expresses the parts which are defective and scrapped. Although these parts cannot continue being produced, since the raw material is glass, it can later re-enter in the furnace and used in future runs. That is why we only incur energy and labor costs for lost parts.

Hence, the finite horizon average cost is given by:

$$J_{avg} = \frac{1}{N}\sum_{n=1}^{N} E[J_n] \tag{3.12}$$

and the infinite horizon average cost by:

$$J_{\infty} = \lim_{N\to\infty} E\left[\frac{1}{N}\sum_{n=1}^{N} J_n\right] \tag{3.13}$$

In order to consider the time value of money we need to associate a discount factor $\beta \in [0;1]$. Therefore, the infinite horizon discounted cost is

$$J_{\infty,\beta} = \lim_{N\to\infty} E\left[\sum_{n=1}^{N} \beta^n J_n\right] \tag{3.14}$$

II. SERVICE LEVEL BASED MEASURES

Measures based on service level are quite important under scenarios where shortage costs estimation is often a laborious operation. The consequences of lost customers are very difficult to measure. They may choose to go elsewhere in the future, which means loss of future revenues. Whereas it is hard to place a money value on loss of future revenues and customer goodwill, it is much easier to define a service level target. This is the reason why service level measures are a widely used tool.

The most common measures are:

i) Type-1 Service Level – the proportion of periods in which all demand is met;
ii) Type-2 Service Level – the proportion of demand satisfied immediately from inventory.

Even if Type-2 service level is what one usually means by service, the Type-1 service level is the most used because of its simplicity.

Additionally, and in the context of this thesis, we are going to deal with an indicator, which expresses time delay from instant of costumer order arrival to instant of total order delivery – the lead-time.

[Bispo, 1997] introduces a result, designated as *optimality condition*, which is related with the service level. This condition states that the optimal base stock levels for the average cost measure for any production policy and any capacity sharing mode are such that the following condition is true:

$$Pr\left(D_n^p \le I^{p1}\right) = \frac{b^p}{b^p + h^{p1}} \tag{3.15}$$

3.4.5 The Derivatives of the Basic Model

The derivatives of inventory, echelon inventory, shortfall and production levels can be obtained by differentiating the respective dynamic equations. They are taken with respect to $Z^{\circ} = Z^{p_m s}$ for some *p=1, ..., P*, *s*=1, ..., *S* and *m=1, ..., M*. The same is valid for the other control variable, the production limit U°. Since, even under infinitesimal perturbations, the base-stock levels have to remain ordered, let us assume that, for the applicable strategies (MTS and DD) $0 < Z^{p_m 1} < Z^{p_m 2} < \ldots < Z^{p_m S}$ is a valid relation.

3.4.6 State Variables Derivatives

We simply detail the inventory derivatives, given that the procedure is relatively simple. The other variables follow the same method.

I. INVENTORY DERIVATIVES

The inventory derivatives are governed by the following two equations:

$$\frac{\partial I_{n+1}^{p_m s}}{\partial Z^{\circ}} = \begin{cases} \frac{\partial I_n^{p1}}{\partial Z^{\circ}} + \alpha^{p1} \frac{\partial P_n^{p1}}{\partial Z^{\circ}}, & \text{last stage of the cold area} \\ \frac{\partial I_n^{p_m s}}{\partial Z^{\circ}} - \frac{\partial P_n^{p_m (s-1)}}{\partial Z^{\circ}} + \alpha^{p_m s} \frac{\partial P_n^{p_m s}}{\partial Z^{\circ}}, & \text{remaining cold area stages} \\ \frac{\partial I_n^{ps}}{\partial Z^{\circ}} - \sum_{m=1}^{M} \frac{\partial P_n^{p_m (s-1)}}{\partial Z^{\circ}} + \alpha^{ps} \frac{\partial P_n^{ps}}{\partial Z^{\circ}}, & \text{last stage of the hot area} \\ \frac{\partial I_n^{ps}}{\partial Z^{\circ}} - \frac{\partial P_n^{ps}}{\partial Z^{\circ}} + \alpha^{ps} \frac{\partial P_n^{ps}}{\partial Z^{\circ}}, & \text{otherwise} \end{cases} \tag{3.16}$$

$$\frac{\partial I_{n+1}^{p_m s}}{\partial U^{\circ}} = \begin{cases} \frac{\partial I_n^{p1}}{\partial U^{\circ}} + \alpha^{p1} \frac{\partial P_n^{p1}}{\partial U^{\circ}}, & \text{last stage of the cold area} \\ \frac{\partial I_n^{p_m s}}{\partial U^{\circ}} - \frac{\partial P_n^{p_m (s-1)}}{\partial U^{\circ}} + \alpha^{p_m s} \frac{\partial P_n^{p_m s}}{\partial U^{\circ}}, & \text{remaining cold - area stages} \\ \frac{\partial I_n^{ps}}{\partial U^{\circ}} - \sum_{m=1}^{M} \frac{\partial P_n^{p_m (s-1)}}{\partial U^{\circ}} + \alpha^{ps} \frac{\partial P_n^{ps}}{\partial U^{\circ}}, & \text{last stage of the hot area} \\ \frac{\partial I_n^{ps}}{\partial U^{\circ}} - \frac{\partial P_n^{ps}}{\partial U^{\circ}} + \alpha^{ps} \frac{\partial P_n^{ps}}{\partial U^{\circ}}, & \text{otherwise} \end{cases} \tag{3.17}$$

II. INITIAL CONDITIONS DERIVATIVES

The initial conditions must also be differentiated, as stated in the following expressions:

$$\frac{\partial I_0^{p1}}{\partial Z^{\circ}} = 1\left\{Z^{\circ} = Z^{p1}\right\}$$

$$\frac{\partial I_0^{ps}}{\partial Z^{\circ}} = 1\left\{Z^{\circ} = Z^{ps}\right\} - 1\left\{Z^{\circ} = Z^{p(s-1)}\right\}$$

$$\frac{\partial E_0^{ps}}{\partial Z^{\circ}} = 1\left\{Z^{\circ} = Z^{ps}\right\}$$

where $1\{\bullet\}$ is the indicator function.

3.4.7 Performance Measures Derivatives

In line with equation (3.11), the operational cost derivatives are given by:

$$\frac{\partial J_n^p}{\partial Z^\circ} = \sum_{s=2}^{S} \frac{\partial \left(I_n^{ps}\right)^+}{\partial Z^\circ} h^{ps} + 1\left\{I_n^{p1} > 0\right\} \frac{\partial \left(I_n^{p1}\right)^+}{\partial Z^\circ} h^{p1} - 1\left\{I_n^{p1} < 0\right\} \frac{\partial \left(I_n^{p1}\right)^-}{\partial Z^\circ} b^p + \\ + \sum_{s=1}^{S} \left(1-\alpha^{ps}\right)\left(h^{ps} - m^p\right) \frac{\partial P_n^{ps}}{\partial Z^\circ}$$

$$\frac{\partial J_n^p}{\partial U^\circ} = \sum_{s=2}^{S} \frac{\partial \left(I_n^{ps}\right)^+}{\partial U^\circ} h^{ps} + 1\left\{I_n^{p1} > 0\right\} \frac{\partial \left(I_n^{p1}\right)^+}{\partial U^\circ} h^{p1} - 1\left\{I_n^{p1} < 0\right\} \frac{\partial \left(I_n^{p1}\right)^-}{\partial U^\circ} b^p + \\ + \sum_{s=1}^{S} \left(1-\alpha^{ps}\right)\left(h^{ps} - m^p\right) \frac{\partial P_n^{ps}}{\partial U^\circ} \qquad (3.18)$$

3.5 THE PRODUCTION DECISIONS AND THEIR DERIVATIVES

The production rules are defined to manage the dynamic capacity allocation. In other words, the production decisions are necessary to decide, in each period *n*, how the available capacity is going to be distributed among the different competing products when a shortage of capacity regarding the total production needs becomes a factor.

The shortfall equalization for every product is a way to dynamically allocate capacity. This procedure assigns capacity to products in decreasing order of their present distance to the target level, the echelon base stock Z^{ps}.

As referred in [Bispo, 1997], the intuition behind an algorithm that equalizes the shortfalls is that one should start by allocating capacity to the product with a higher difference to its target level Z^{ps}, that is, the product with the highest shortfall, until it reaches the level of the product with the second highest shortfall. The next step consists on the distribution of the available capacity in equal parts to both products until their shortfalls equalize that of the third highest shortfall. Note that, at any point, one can have different shortfalls at the end of the production decision as a consequence of capacity exhaustion or insufficient inventory for some products. However, the problem of inventory exhaustion at the first stage does not exist since raw material is always available, avoiding starvation. Since one can produce several products within a production period, one needs some rule that allows the distribution of the available capacity by the products.

3.5.1 The Production Decisions Algorithm

The next algorithm is inspired on the *Equalize Shortfall Algorithm* developed in [Bispo, 1997], with some significant modifications. The first one deals with the shortfall ordering issue. While in the above the ordering is established on a shortfall absolute value, here it will be based on a weighted shortfall, that is, the shortfall absolute value divided by the average demand. The second modification concerns the capacity allocation problem. In [Bispo, 1997], section 3.4, the production decisions are determined in order to continuously equalize each product shortfall. The next algorithm, in contrast, does not consider the equalization. For the product with the higher weighted shortfall, the production decision is the shortfall value if there is available capacity, upstream inventory and if the production limit is not exceeded. The algorithm proceeds by allocating capacity to products by decreasing order of their weighted shortfall.

■ **PRODUCTION DECISIONS ALGORITHM**

For each product p and at each time period n, the following algorithm gives the production decisions for each process stage s:

STEP 0 – Set control variable C^s, decision variable P^{ps} and their derivatives to initial values.

$$\begin{cases} P_n^{p_m s} = \dfrac{\partial P^{p_m s}}{\partial Z^\circ} = \dfrac{\partial P^{p_m s}}{\partial U^\circ} = 0, & \text{cold-area stages} \\ P_n^{ps} = \dfrac{\partial P^{ps}}{\partial Z^\circ} = \dfrac{\partial P^{ps}}{\partial U^\circ} = 0, & \text{otherwise} \end{cases} \tag{3.19}$$

$$\mathrm{C}^s = \text{capacity stage value} \tag{3.20}$$

$$\frac{\partial C^s}{\partial Z^\circ} = \frac{\partial C^s}{\partial U^\circ} = 0, \quad \text{for all stages} \tag{3.21}$$

STEP 1 – Order, for each stage *s*, the products by decreasing order of

$$\begin{cases} \dfrac{Y^{p_m s}}{\mathrm{E}\,[D^{p_m}]}, & \text{cold-area stages;} \\ \dfrac{Y^{ps}}{\mathrm{E}\left[\sum_{m=1}^{M} D^{p_m}\right]}, & \text{otherwise} \end{cases} \tag{3.22}$$

after demand is realized. Consider that $j = (1),\ldots,(P)$ expresses that ordering.

STEP 2 – The production decision and its derivative, for product corresponding to any *j* is determined as follows:

$$P^{p_m s} = \begin{cases} \min\left\{Y^{p_m s}, \dfrac{C^s}{T^{p_m s}}, I^{p(s+1)}, U^{p_m s}\right\}, & \text{for cold - area first stage} \\ \min\left\{Y^{p_m s}, \dfrac{C^s}{T^{p_m s}}, I^{p_m(s+1)}, U^{p_m s}\right\}, & \text{for other cold - area stages} \\ \min\left\{Y^{ps}, \dfrac{C^s}{T^{ps}}, U^{ps}\right\}, & \text{for hot - area first stage} \\ \min\left\{Y^{ps}, \dfrac{C^s}{T^{ps}}, I^{p(s+1)}, U^{ps}\right\}, & \text{otherwise} \end{cases} \tag{3.23}$$

$$
\frac{\partial P^{p_m s}}{\partial Z^{\circ}} =
\begin{cases}
\begin{cases}
\dfrac{\partial Y^{p_m s}}{\partial Z^{\circ}}, \text{ if } s \text{ is shortfall bound} \\
\dfrac{\partial I^{p(s+1)}}{\partial Z^{\circ}}, \text{ if } s \text{ is supply bound} \\
0, \quad \text{if } s \text{ is production bound} \\
\dfrac{\partial \left(\dfrac{C^{s}}{T^{p_m s}} \right)}{\partial Z^{\circ}}, \text{ if } s \text{ is capacity bound}
\end{cases} & \text{, first stage of the cold area} \\
\begin{cases}
\dfrac{\partial Y^{p_m s}}{\partial Z^{\circ}}, \text{ if } s \text{ is shortfall bound} \\
\dfrac{\partial I^{p_m(s+1)}}{\partial Z^{\circ}}, \text{ if } s \text{ is supply bound} \\
0, \quad \text{if } s \text{ is production bound} \\
\dfrac{\partial \left(\dfrac{C^{s}}{T^{p_m s}} \right)}{\partial Z^{\circ}}, \text{ if } s \text{ is capacity bound}
\end{cases} & \text{, remaining cold - area stages} \\
\begin{cases}
\dfrac{\partial Y^{ps}}{\partial Z^{\circ}}, \text{ if } s \text{ is shortfall bound} \\
0, \quad \text{if } s \text{ is production bound} \\
\dfrac{\partial \left(\dfrac{C^{s}}{T^{ps}} \right)}{\partial Z^{\circ}}, \text{ if } s \text{ is capacity bound}
\end{cases} & \text{, first stage of the hot area} \\
\begin{cases}
\dfrac{\partial Y^{ps}}{\partial Z^{\circ}}, \text{ if } s \text{ is shortfall bound} \\
\dfrac{\partial I^{p(s+1)}}{\partial Z^{\circ}}, \text{ if } s \text{ is supply bound} \\
0, \quad \text{if } s \text{ is production bound} \\
\dfrac{\partial \left(\dfrac{C^{s}}{T^{ps}} \right)}{\partial Z^{\circ}}, \text{ if } s \text{ is capacity bound}
\end{cases} & \text{, otherwise}
\end{cases}
\tag{3.24}
$$

$$\frac{\partial P^{p_m s}}{\partial U^{\circ}} = \begin{cases} \begin{cases} \dfrac{\partial Y^{p_m s}}{\partial U^{\circ}}, \text{ if } s \text{ is shortfall bound} \\ \dfrac{\partial I^{p(s+1)}}{\partial U^{\circ}}, \text{ if } s \text{ is supply bound} \\ 1, \quad \text{if } s \text{ is production bound} \\ \dfrac{\partial \left(\dfrac{C^{s}}{T^{p_m s}} \right)}{\partial U^{\circ}}, \text{ if } s \text{ is capacity bound} \end{cases} & \text{, first stage of the cold area} \\ \begin{cases} \dfrac{\partial Y^{p_m s}}{\partial U^{\circ}}, \text{ if } s \text{ is shortfall bound} \\ \dfrac{\partial I^{p_m(s+1)}}{\partial U^{\circ}}, \text{ if } s \text{ is supply bound} \\ 0, \quad \text{if } s \text{ is production bound} \\ \dfrac{\partial \left(\dfrac{C^{s}}{T^{p_m s}} \right)}{\partial U^{\circ}}, \text{ if } s \text{ is capacity bound} \end{cases} & \text{, remaining cold - area stages} \\ \begin{cases} \dfrac{\partial Y^{ps}}{\partial U^{\circ}}, \text{ if } s \text{ is shortfall bound} \\ 0, \quad \text{if } s \text{ is production bound} \\ \dfrac{\partial \left(\dfrac{C^{s}}{T^{ps}} \right)}{\partial U^{\circ}}, \text{ if } s \text{ is capacity bound} \end{cases} & \text{, first stage of the hot area} \\ \begin{cases} \dfrac{\partial Y^{ps}}{\partial U^{\circ}}, \text{ if } s \text{ is shortfall bound} \\ \dfrac{\partial I^{p(s+1)}}{\partial U^{\circ}}, \text{ if } s \text{ is supply bound} \\ 0, \quad \text{if } s \text{ is production bound} \\ \dfrac{\partial \left(\dfrac{C^{s}}{T^{ps}} \right)}{\partial U^{\circ}}, \text{ if } s \text{ is capacity bound} \end{cases} & \text{, otherwise} \end{cases} \tag{3.25}$$

STEP 3 – Update the available capacity as follows:

$$\begin{cases} C^s = C^s - P^{p_m s(j)}, & \text{cold-area stages} \\ C^s = C^s - \sum_{m=1}^{M} P^{p_m s(j)}, & \text{last stage of the hot area} \\ C^s = C^s - P^{ps(j)}, & \text{otherwise} \end{cases} \tag{3.26}$$

Likewise the derivatives are updated as:

$$\begin{cases} \dfrac{\partial C^s}{\partial Z^\circ} = \dfrac{\partial C^s}{\partial Z^\circ} - \dfrac{\partial P^{p_m s(j)}}{\partial Z^\circ}, & \text{cold-area stages} \\ \dfrac{\partial C^s}{\partial Z^\circ} = \dfrac{\partial C^s}{\partial Z^\circ} - \sum_{m=1}^{M} \dfrac{\partial P^{p_m s(j)}}{\partial Z^\circ}, & \text{last stage of the hot area} \\ \dfrac{\partial C^s}{\partial Z^\circ} = \dfrac{\partial C^s}{\partial Z^\circ} - \dfrac{\partial P^{ps(j)}}{\partial Z^\circ}, & \text{otherwise} \end{cases} \tag{3.27}$$

and

$$\begin{cases} \dfrac{\partial C^s}{\partial U^\circ} = \dfrac{\partial C^s}{\partial U^\circ} - \dfrac{\partial P^{p_m s(j)}}{\partial U^\circ}, & \text{cold-area stages} \\ \dfrac{\partial C^s}{\partial U^\circ} = \dfrac{\partial C^s}{\partial U^\circ} - \sum_{m=1}^{M} \dfrac{\partial P^{p_m s(j)}}{\partial U^\circ}, & \text{last stage of the hot area} \\ \dfrac{\partial C^s}{\partial U^\circ} = \dfrac{\partial C^s}{\partial U^\circ} - \dfrac{\partial P^{ps(j)}}{\partial U^\circ}, & \text{otherwise} \end{cases} \tag{3.28}$$

STEP 4 – If $C^s = 0$, STOP.

The total production for stage *s* is bound by capacity.

Otherwise go to STEP 5.

STEP 5 – If $j < k$, set $j = j + 1$ and go to *STEP 3*.
Otherwise, STOP.

The total production for stage *s* does not use up all the available capacity.

■ END OF THE PRODUCTION DECISIONS ALGORITHM

The algorithm generates the production decision and its derivative at the same time. By the end of the procedure, the parameter P_n^{ps} will contain the production decision for product p at stage s at the instant time n, while $\frac{\partial P^{ps}}{\partial Z^{\circ}}$ and $\frac{\partial P^{ps}}{\partial U^{\circ}}$ its derivatives.

3.6 INFINITESIMAL PERTURBATION ANALYSIS VALIDATION

The Infinitesimal Perturbation Analysis approach needs to be validated in the framework of this model. The validation procedure consists on showing that all variables are differentiable, providing their values, and finally to prove that expectation and derivation are permutable operators. This validation is presented in [Bispo, 1997] for multi-product, multi-stage, re-entrant flow shops, subject to random demand, using a discrete time, capacitated, production-inventory model. Since the model presented in this thesis is a particularization of the above model, the IPA approach is valid.

The IPA approach was implemented by means of the optimization procedure which uses the average total cost and its derivatives with respect to all variables to determine a new set of Z and U values if none of the stopping criteria is verified (see Figure 3.3). Examples of stopping criteria are small gradient norm and a maximum number of iterations.

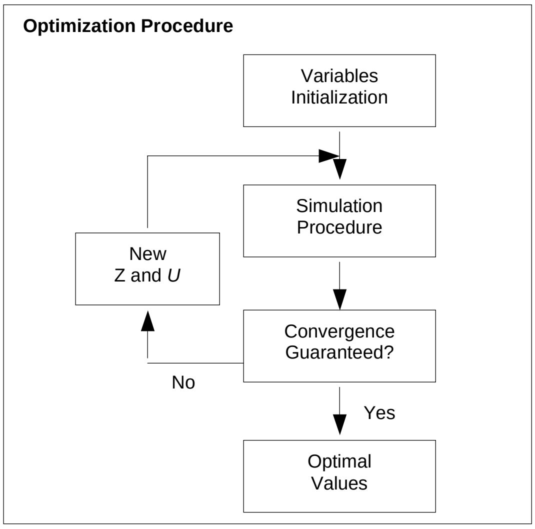

Figure 3.3 – Optimization-Simulation procedures interaction

CHAPTER 4
EXPERIMENTAL STUDY

4.1 INTRODUCTION

In this chapter we analyze some numerical experiments carried out with the simulation optimization package developed (see Appendix A for more details on the software and on the optimization procedure). Several computational experiments will be presented in order to provide some insights into the behavior of the glass manufacturing production process under different production policies.

4.2 SYSTEM AND PRODUCTS DATA

The simulation of glass manufacturing processes with a high degree of accuracy will require considering the complete set of products, studies on their demand patterns, and knowing the real cost structure of the company, just to mention some relevant issues. Obviously, such scenario will require not only a tremendous computational power but also the development of an entirely different application (with data bases), which is outside the scope of this thesis.

From the almost 900 references in the catalog, 27 products were selected, which were divided in 3 groups of 9 products each. These groups represent the three classes of demand levels – high, medium and low – and are based on a Pareto's Law categorization. The high demand total quantity is roughly 80% of the total units, while low level accounts only for 5% of total demand.

The processing times of the 27 products for different stages are according with those used on the company's information system. They were determined with the standard production rates and number of operators per stage. The wide range these values present reflect the degree of complexity and manual labor intensity of any product. Some products do not need to pass through all processing stages, which was simulated with very low processing times (0,00001 minutes/unit). The capacity of each stage, expressed in minutes, corresponds to the available working time in one production shift. The firm works 24 hours on a four-shift base, which corresponds to 480 minutes per shift.

Product split occurs at the end of the hot area, i.e., at stage 5. Stage 4 takes inventory from stage 5 to produce one of the three possible sub-products out of each main product.

We test four different production strategies: MTS, MTO, DD, and a mix of the three, called MTS/DD/MTO. This last strategy applies MTS to high demand products, DD to products of intermediate demand level, and MTO to low demand products. The first three were previously defined.

4.2.1 Cost Structure

As described in section 3.2.4, the developed cost structure accounts for three main direct costs: raw material cost, energy cost, and direct labor cost. Rather than represent real costs, the values try to illustrate the relative weight between them. With net weights varying from 0,03 Kg to 6,03 Kg, it was used a value of 2 monetary units per kilogram for the glass cost. Next table shows the values of energy cost and labor cost to all process stages. The penalty cost of each product is defined according to the respective holding cost. It is used a factor of 1,3 for high demand products, 1,4 for medium demand products, and 1,5 for products with a low level of demand.

Table 4.1– Energy and labor hourly costs

STAGES / COSTS	9	8	7	6	5	4	3	2	1
Energy Cost [m.u./hour]	0,5	3,0	1,0	3,0	0,1	3,0	3,0	3,0	0,1
Labor Cost [m.u./hour]	90	8	20	30	20	10	10	20	30

4.2.2 Demand and Yield Structures

Demand and yield are the two sources of randomness of the production system. While the first tries to replicate the uncertainty of the demand pattern, the second aims to reproduce non-deterministic production resources. The simulator uses two parameters – average demand and the inverse of the variance coefficient (ivc) – to generate a random number according to an Erlang distribution. All the experimental study was performed with an ivc =1. Under a non-perfect yield (NPY) scenario the average demand level must decrease when compared with the perfect yield (PY) scenario in order to maintain the same desirable bottleneck load (85%). Table 4.2 shows the average demand values depending on demand level and yield scenarios.

Table 4.2– Average Demand to both yield scenarios

DEMAND LEVEL [units/shift]	**PERFECT YIELD (PY) [% TOTAL]**	**NON-PERFECT YIELD (NPY) [% TOTAL]**
High	17,362 [80%]	14,886 [80%]
Medium	3,255 [15%]	2,791 [15%]
Low	1,628 [5%]	1,396 [5%]

The yield factors used on the numerical experiments are equal to those available at the company's information system, which were obtained from historical data. These rates represent an attribute of each stage, being almost product independent. The molding stage has the lowest average yield rate partially explained by a strong human labor dependency and operation degree of complexity. On the other stages, less human dependent and technically more simplified, yield is close to 100%. See Table 4.3 below for the average yield factors at each process stage.

Table 4.3 – Average yield factors for non-perfect scenario

STAGES	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Yield Factor	0,6	0,9	0,95	0,95	0,95	0,95	0,95	0,95	0,95

One should mention that the application is not limited to consider stage dependent yield factors. It is possible to define a different yield factor for all products and stages.

4.3 SIMULATION CYCLES

The decision of how many simulations cycles need to be performed must deal with two competing questions: time and accuracy.

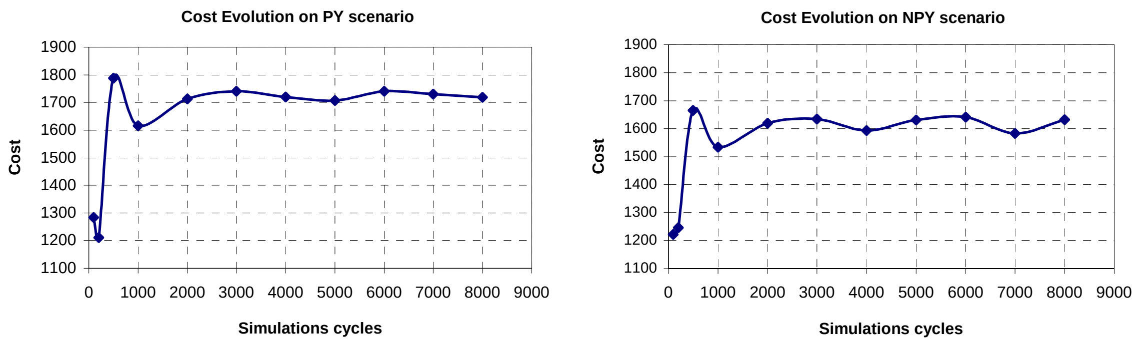

Figure 4.1– Cost vs. simulations cycles for PY and NPY scenario

The number of simulation cycles must be sufficiently high to produce the most accurate estimates, regardless of the increasing simulation time. Figure 4.1 demonstrates how cost evolves with the number of simulations.

Both in PY and NPY scenarios the cost stabilizes in the neighborhood of 2000 simulation cycles. However, and even though the oscillation is only within ± 2,0 % of the average cost, all the simulations will be performed with a length of 8000 cycles.

4.4 NUMERICAL RESULTS

An in-depth analysis of this production system would be a considerable task to complete given the number of parameters to be taken into account: average demand, demand variance, processing times, holding costs, penalty costs, number of stages and their capacity, and production policies. For this reason, the numerical study was limited to two main simulation groups with three scenarios each. The simulation groups are divided in limited production (LP) and non-limited production (NLP) approaches, which are individually formed by the following scenarios: perfect yield production (PY), random yield production (NPY) and perfect yield production with a lower machine load (PY-LL) using demand corresponding to the random yield scenario. For each one of these scenarios and production strategies, we present the optimal cost, in-house cost, and lead-times.

4.4.1 Numerical Results Confidence Interval

In all experiments, and after optimizing the values of the control variables *Z* and *U*, 25 replications were performed to verify the results' accuracy. These 25 replicas are a sample from a normal distribution with unknown mean μ and unknown variance σ^2 out of which we wish to construct a 100.(1–α) percent confidence interval for μ. Since σ is unknown, we can no longer base our interval on the fact that $[\sqrt{n}\,(X_i - \mu)]/\sigma$ has a unit normal random variable. However, by letting S^2 denote the sample variance, then $[\sqrt{n}\,(X_i - \mu)]/S$ has a *t* distribution with n-1 degrees of freedom. All the intervals presented are determined with 95% confidence and refer solely to costs.

4.4.2 Limited Production Approach (LP)

On the first group of experiments the production system was simulated assuming each production decision is bounded by some limit, *U*. *U* was taken also as a control parameter subject to optimization. The initial values of the production limit were set at their minimum stable value, that is, equal to the average demand. During the iterations, the optimization algorithm will change the *U* values according to the gradient information and ensuring the successive values of *U* to be inside their feasible region.

Scenario 1: Perfect Yield Production (PY)

Under an MTS production strategy, all the control variables are adjustable and the *Z* and *U* values are defined depending on gradient information. Any production strategy different from this one will have a higher cost, since some of the base stock values are forced to zero on the other strategies. In ascending order of cost, the production policies will then be ranked as: MTS, MTS/DD/MTO, DD and MTO.

Figure 4.2 presents the evolution of the costs during the optimization procedure and serves the purpose of illustrating the convergence properties of the overall algorithm. In terms of optimal cost the ordering of the four strategies should not be a surprise. For the MTO strategy we simply optimize the values of U and keep all Z's equal to zero. For the DD strategy, additionally to the MTO, we optimize the Z values for the hot area, keeping the Z values for the cold area all equal to zero. For the MTS strategy all values of U and Z variables are subject to optimization. The MTS/DD/MTO strategy has less variables forced to be zero than the DD strategy and more variables forced to be zero than the MTS strategy. Therefore, the strategy which allows for a smaller number of variables to be set to zero will have the lower optimal cost.

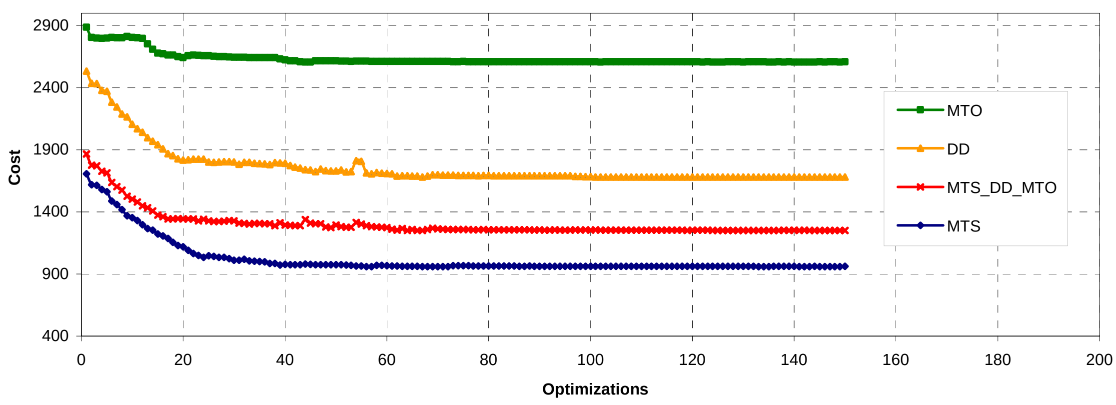

Figure 4.2 – Cost optimization evolution (LP – PY scenario)

Table 4.4 shows the minimum average total cost (Optimal Cost) reached by the optimization process, when optimization was halted, the corresponding gradient norm, the confidence interval for the costs of the 25 replications performed with Z and U placed at their optimal values, and the in-house costs. In line with what was previously exposed, the MTO production policy accounts for the largest penalty costs (given by the difference between the optimal cost and in-house cost). For the same demand profile, one must pay a higher penalty cost since there is no stock available to satisfy customer orders. On the other hand, this strategy incurs less holding costs. Understandably, the MTS production strategy achieves the lowest total cost, but paying the highest holding costs. For the process environment simulated within this experiment, having a production-to-order strategy in spite of a production-to-stock strategy could mean almost 30% less in tangible costs (in-house costs). Nevertheless, a more strategic vision that values customer service will probably implement an MTS production strategy, reducing the optimal average total cost roughly 65%. A similar analysis can be made for the remaining production policies.

See section 4.4.4 – Discussion on Optimal Convergence – for further analysis on optimal values and gradient norm.

Table 4.4 – Cost data for all production strategies (LP – PY scenario)

PRODUCTION STRATEGY	OPTIMAL COST	CONFIDENCE INTERVAL	GRADIENT NORM	IN-HOUSE COST
MTO	2606,04	± 1,79 %	15.07	431,43
DD	1679,76	± 2,27 %	13,51	490,49
MTS/DD/MTO	1249,08	± 3,08 %	11,65	560,41
MTS	957,88	± 6,07 %	8,01	608,95

One can access the system performance not also accounting for its holding and penalty costs but also measuring the system's responsiveness to customer orders. The average number (and its standard deviation, σ) of production shifts – lead-time (LT) – necessary to satisfy a customer order is presented in Table 4.5.

When it comes to the DD and MTO production policies, and since we need to have comparable values, the lead-times are counted, respectively, after 4 and 9 production shifts. These numbers represent an assumed time period during which the customer willingness to wait is significant. That is, the probability of paying the corresponding penalty costs during those periods is low. Nevertheless, if the time value of money is considered, the waiting period may represent an important cost to the company.

Table 4.5– Average lead-time and its standard deviation (LP – PY scenario)

STRATEGY	**MTO**		**DD**		**MTS/DD/MTO**		**MTS**	
Demand Level	LT	σ	LT	σ	LT	σ	LT	σ
High	2,99	3,20	2,22	1,06	1,68	2,51	1,88	2,00
Medium	2,13	1,72	1,90	0,83	1,98	0,78	1,06	1,40
Low	1,47	1,84	1,78	0,84	1,77	0,81	0,74	1,43
All Products	2,19	2,37	1,97	0,99	1,81	1,75	1,23	1,70

As expected, in terms of lead-times the MTS production policy beats DD and MTO performances, independently of the demand level. Also expected is the result of the compound strategy (MTS/DD/MTO). Since that medium and low demand level products are produced under DD and MTO policies, the system has more available capacity to satisfy high demand level orders, when compared with all products being produced under an MTS policy. This explains the improvement on high demand products' lead-time and the deterioration of the other two product groups. If we compare the MTO strategy with the DD strategy we can conclude that, a slight increase (+13,5%) in holding costs means only a minor reduction on products' lead-time. However, if we account for the penalty costs, the DD strategy achieves a 35,5% reduction in the optimal cost. The small absolute difference between the average lead-times (1,23 shifts to 2,19 shifts, i.e., 7,4 hours to 13,1 hours) of the extreme strategies could be explained by the process reliability. The presence of random yield will considerably change these values, as we will see in the next scenario.

Scenario 2: Non-Perfect Yield Production (NPY)

Introducing the production randomness (see Table 4.3 for yield factors), we are now interested in comparing the system's performance with scenario 1.

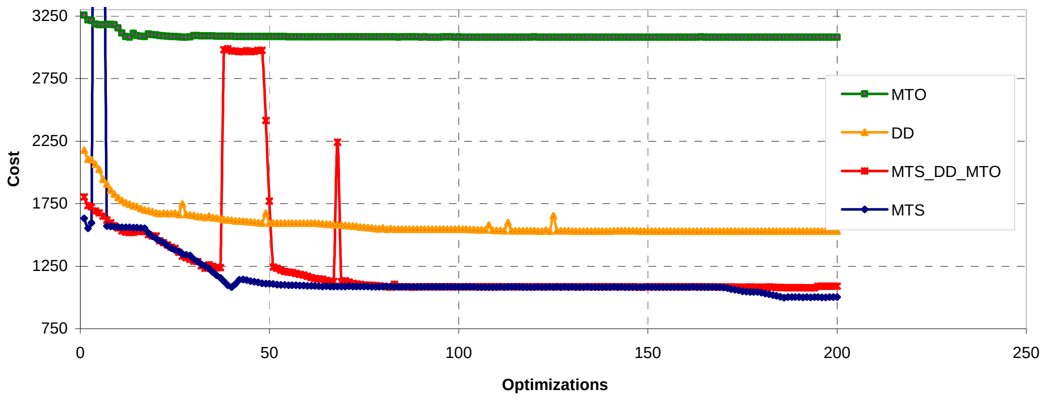

Figure 4.3– Cost optimization evolution (LP – NPY scenario)

This experiment maintains the same hierarchy for the optimal and in-house costs. One should remark that the cost values of both scenarios are not comparable in absolute terms, since the last one was simulated with lower values of average demand. Moreover, the inferior value found for the DD and MTS/DD/MTO strategies (see Table 4.6) suggests that, despite the additional cost paid for item disposal, it is not enough to compensate the holding cost decrease due to lower average inventory levels. Also in this scenario, and considering the production process implemented, the cost differences between the two extreme strategies are relevant. Although an MTS strategy could represent more 30% in holding costs, it represents 67% less if all costs are incurred.

Table 4.6 – Cost data for all production strategies (LP – NPY scenario)

PRODUCTION STRATEGY	OPTIMAL COST	CONFIDENCE INTERVAL	GRADIENT NORM	IN-HOUSE COST
MTO	3079,57	± 1,94 %	15,12	488,58
DD	1529,43	± 0,75 %	12,68	540,42
MTS/DD/MTO	1075,87	± 0,86 %	9,41	595,24
MTS	999,05	± 01,42	8,29	631,53

The degradation induced by the presence of random yield is easily identified, both in terms of costs and of average lead-times This issue is particularly significant when the system is producing to order (see Table 4.7).

Table 4.7– Average lead-time and its standard deviation (LP – NPY scenario)

STRATEGY	**MTO**		**DD**		**MTS/DD/MTO**		**MTS**	
Demand Level	LT	σ	LT	σ	LT	σ	LT	σ
High	7,21	4,64	2,71	0,98	1,24	1,83	2,73	2,68
Medium	6,66	2,59	2,67	0,96	2,61	0,89	1,17	1,56
Low	4,67	3,27	2,74	0,96	4,87	2,41	0,73	1,62
All Products	6,18	3,70	2,71	0,99	2,91	2,63	1,54	2,11

This scenario maintains the average lead-times structure presented in scenario 1. However, the differences between DD and MTO strategies must be emphasized. Even under just a tangible costs accounting, a 10,6% increase on in-house costs may represent a decrease in lead-times from 7,21 shifts (43 hours) to 2,71 shifts (16 hours). Despite the worst lead-times of medium and low level products, and an increase of 8% on optimal cost relative to MTS, an MTS/DD/MTO causes a reduction of 55% (from 2,73 shifts to 1,24 shifts) on the average lead-time of high demand products.

Recall that lead-time values are computed after the waiting period of 4 and 9 shifts, depending if we are under a DD strategy or under an MTO strategy.

Scenario 3: Perfect Yield Production – Lower Load (PYLL)

This scenario corresponds to an experiment realized with perfect yield, but with demand of scenario 2. Therefore, with the system's load reduction from 85% to 73%, we must incur less costs that those of scenario 1, and reduce lead-times of the same scenario.

Table 4.8 confirms the expectation on lowest costs and on the cost hierarchy. In line with the results obtained for the two previous scenarios, the MTS strategy represents a decrease of 67% on the optimal cost and an increase on the in-house costs of 45%. Also, the expected lead-time reduction is presented on Table 4.9.

Table 4.8 – Cost data for all production policies (LP – PYLL scenario)

PRODUCTION STRATEGY	OPTIMAL COST	CONFIDENCE INTERVAL	GRADIENT NORM	IN-HOUSE COST
MTO	2228,23	± 1,32 %	14,56	375,33
DD	1420,24	± 2,45 %	12,23	470,49
MTS/DD/MTO	908,95	± 6,02 %	11,21	523,01
MTS	741,13	± 4,66 %	3,15	545,58

Table 4.9 – Average lead-time and its standard deviation (LP – PYLL scenario)

STRATEGY	MTO		DD		MTS/DD/MTO		MTS	
Demand Level	LT	σ	LT	σ	LT	σ	LT	σ
High	2,92	3,10	1,62	1,06	1,15	1,55	1,25	1,47
Medium	2,13	1,69	1,59	0,82	1,48	0,67	1,02	1,28
Low	1,53	1,80	1,47	0,94	1,69	0,71	0,91	1,29
All Products	2,19	2,32	1,97	0,99	1,47	1,57	1,03	1,34

From these three experiments, one concludes that all the scenarios present the same cost hierarchy, both in terms of average total cost and in terms of average in-house cost. This suggests that such hierarchy should hold for any other demand and yield patterns. Additionally, as already expected, the random yield presence induces degradation of costs and lead-times.

4.4.3 Non-Limited Production Approach (NLP)

This second group of experiments emerges with several other accomplished simulations and, simultaneously, the need of try to reduce more the gradient norm. At the Non-Limited Production Approach the *U* levels are fixed at their maximum feasible values. Given that we are fixing some of the variables, the optimal values should be higher than those obtained with the LP approach. However, the results are not in agreement with such statement, suggesting that under the LP approach the optimization is halted before reaching the optimum (see section 4.4.4 for additional discussion on this topic).

All the results obtained for the three scenarios are in line with those presented for the LP approach, both in terms of costs and lead-times, degradation induced by random yield, and relative differences between costs and between lead-times. Therefore, the corresponding figures and tables are presented, in order to verify such agreement.

Scenario 1: Perfect Yield Production (PY)

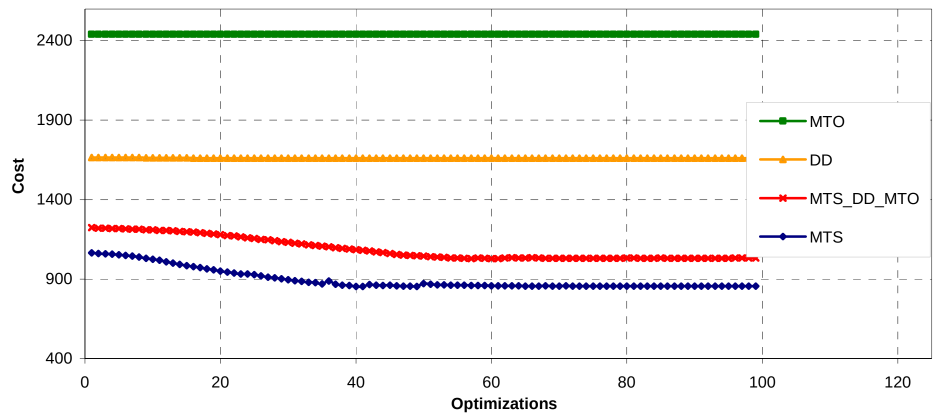

Figure 4.4– Cost optimization evolution (NLP – PY scenario)

Table 4.10 – Cost data for all production strategies (NLP – PY scenario)

PRODUCTION STRATEGY	**OPTIMAL COST**	**CONFIDENCE INTERVAL**	**GRADIENT NORM**	**IN-HOUSE COST**
MTO	2441,49	± 0,66 %	13,42	454,73
DD	1661,14	± 0,92 %	10,42	539,10
MTS/DD/MTO	1027,65	± 1,03 %	9,32	608,30
MTS	853,50	± 1,15 %	2,40	641,28

Table 4.11– Average lead-time and its standard deviation (NLP – PY scenario)

STRATEGY	**MTO**		**DD**		**MTS/DD/MTO**		**MTS**	
Demand Level	LT	σ	LT	σ	LT	σ	LT	σ
High	2,41	2,34	2,08	1,02	1,12	1,44	1,16	1,38
Medium	2,50	2,50	1,95	1,28	2,10	1,40	1,31	1,42
Low	2,12	2,53	2,09	1,29	1,81	1,43	1,54	1,44
All Products	2,35	2,37	2,04	1,15	1,68	1,67	1,34	1,46

Scenario 2: Non-Perfect Yield Production (NPY)

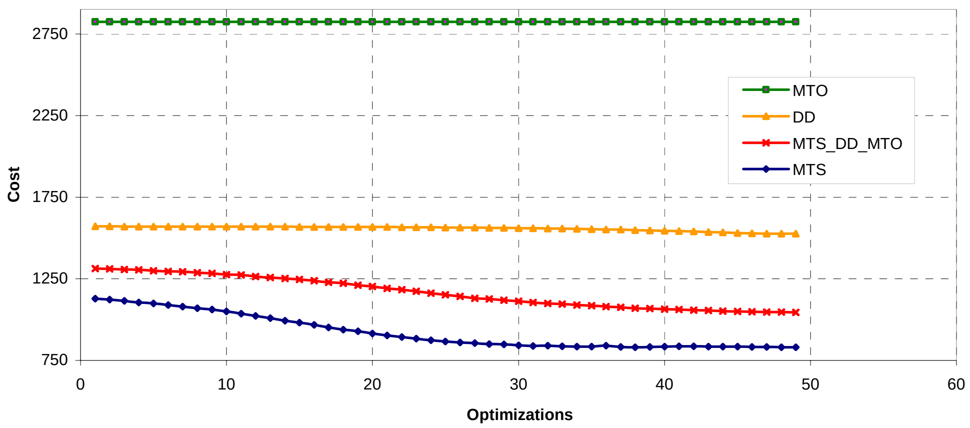

Figure 4.5 – Cost optimization evolution (NLP – NPY scenario)

Table 4.12 – Cost data for all production strategies (NLP – NPY scenario)

PRODUCTION STRATEGY	OPTIMAL COST	CONFIDENCE INTERVAL	GRADIENT NORM	IN-HOUSE COST
MTO	2822,70	± 0,53 %	12,64	489,60
DD	1525,15	± 0,81 %	10,98	545,13
MTS/DD/MTO	1044,05	± 0,87 %	9,31	626,77
MTS	829,45	± 0,90 %	2,25	647,41

Table 4.13 – Average lead-time and its standard deviation (NLP – NPY scenario)

STRATEGY	MTO		DD		MTS/DD/MTO		MTS	
Demand Level	LT	σ	LT	σ	LT	σ	LT	σ
High	6,13	3,14	2,77	1,06	1,10	1,51	1,18	1,47
Medium	6,29	3,34	2,83	1,18	2,93	1,30	1,28	1,40
Low	5,26	3,49	2,78	1,18	4,90	2,36	1,03	1,42
All Products	5,89	3,30	2,79	1,11	2,98	2,65	1,16	1,43

Scenario 3: Perfect Yield Production – Lower Load (PYLL)

Table 4.14– Cost data for all production strategies (NLP – PYLL scenario)

PRODUCTION STRATEGY	OPTIMAL COST	CONFIDENCE INTERVAL	GRADIENT NORM	IN-HOUSE COST
MTO	2043,18	± 0,39 %	12,55	374,85
DD	1310,85	± 0,46 %	10,41	467,43
MTS/DD/MTO	866,10	± 0,58 %	9,32	512,47
MTS	704,11	± 0,65 %	2,33	532,64

Table 4.15 – Average lead-time and its standard deviation (NLP – PYLL scenario)

STRATEGY	MTO		DD		MTS/DD/MTO		MTS	
Demand Level	LT	σ	LT	σ	LT	σ	LT	σ
High	2,23	2,23	1,51	0,72	1,12	1,34	1,11	1,33
Medium	2,31	2,36	1,83	0,78	1,85	0,93	1,18	1,32
Low	1,87	2,40	1,77	0,78	1,76	0,93	0,74	1,40
All Products	2,14	2,23	1,70	0,75	1,57	1,53	1,01	1,31

4.4.4 Discussion on Convergence

The results obtained with the NLP approach suggest that the optimization algorithm has some problems in reaching the optimum before the maximum number of optimizations halts the experiment. With the several other experiments performed became apparent a strong dependence on the initial step of the algorithm.

It should be remarked that, given the number of variables of the problem (306), a gradient norm of 10 represents an average derivative value of only 0,572, which is by itself a low value. Despite the non-convergence to the optimal values, it is our conviction that the qualitative behavior obtained will not change with slightly more accurate results (see Figure 4.6 and Table 4.16 for such an example).

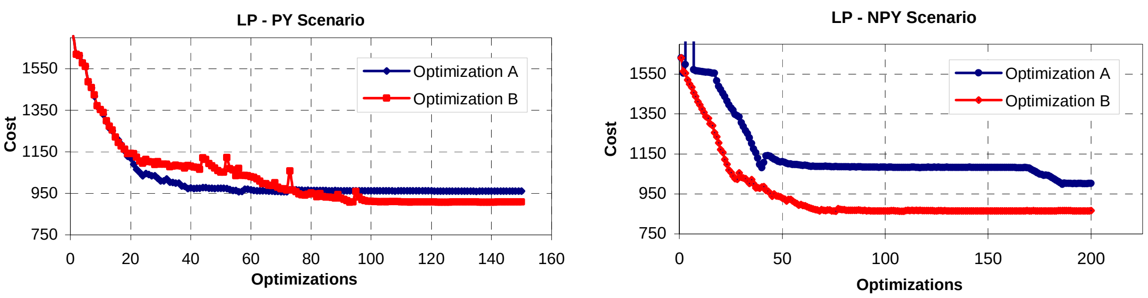

Figure 4.6 – Convergence comparison (MTS strategy)

Table 4.16– Costs data comparison (LP scenario)

OPTIMIZATION	OPTIMAL COST	CONFIDENCE INTERVAL	GRADIENT NORM	IN-HOUSE COST
A (LP – PY)	957,88	± 6,07 %	8,01	608,95
B (LP – PY)	906,79	± 5,78 %	5,54	639,12
A (LP – NPY)	999,05	± 1,42 %	8,29	631,53
B (LP – NPY)	863,79	± 0,93 %	2,88	650,45

"Optimization A" and "Optimization B" refer to different initial steps for the same problem. The gradient when the optimization is halted is lower in "Optimization B" as well as the reached average total cost.

These experiments were produced to test the optimization algorithm. The fact that different initial steps may lead to different final total costs should serve to show that the used optimization algorithm may not be the most suited (see section 5.2 for further discussion). It should be stressed that the IPA procedure and respective gradient estimation have no influence on these results.

In section 3.4.4 we presented the *optimality condition*, establishing a connection between operational costs and Type-1 service level. In order to evaluate the convergence of the optimization procedure we computed the achieved Type-1 service level for "Optimization A" and "Optimization B". Table 4.17 shows those levels.

According to equation 3.13 the optimal service level for products ranges between 57% and 60%. Table 4.17 shows that for an optimization execution which is halted with a lower gradient, the achieved service level is closer to what is established by equation 3.13.

Table 4.17 – Average Service Level Comparison (LP scenario)

PRODUCTS	**OPTIM. A (LP – PY)**	**OPTIM. B (LP – PY)**	**OPTIM. A (LP –NPY)**	**OPTIM. B (LP – NPY)**
High Demand	29,14 %	39,68 %	23,04 %	50,15 %
Medium Demand	51,75 %	55,20 %	50,25 %	56,67 %
Low Demand	68,82 %	47,66 %	67,21 %	57,33 %

According to equation 3.13 and the holding and penalty costs used, the optimal service level is the one displayed under "OPTIMAL SL" on Tables 4.18 and 4.19.

Table 4.18 – Average Service Level (LP, MTS strategy)

PRODUCTS	**OPTIMAL SL**	**AVERAGE SL (PY scenario)**	**AVERAGE SL (NPY scenario)**
High Demand	57,00 %	39,68 %	50,15 %
Medium Demand	58,00 %	55,20 %	56,67 %
Low Demand	60,00 %	47,66 %	57,33 %

Table 4.19 – Average Service Level (NLP, MTS strategy)

PRODUCTS	**OPTIMAL SL**	**AVERAGE SL (PY scenario)**	**AVERAGE SL (NPY scenario)**
High Demand	57,00 %	51,18 %	53,75%
Medium Demand	58,00 %	57,53 %	58,68%
Low Demand	60,00 %	44,90 %	64,73%

These tables present the best service levels achieved in the several experiments conducted. The difference to the optimal service level shows that the optimization algorithm was halted short of converging to the optimal average total cost. However, it should be emphasized that the final gradient is very small in some cases. Take for instance "Optimization B (LP-NPY)" of Table 4.16, where the norm of the final gradient is a little less than 3, in an optimization problem with 306 variables.

The results of Tables 4.18 and 4.19 show also that the service level seems to be more sensitive to small changes of the control variables than the average total cost. Nevertheless, the proximity to the optimal service level is quite significant in some instances. Moreover, the cases where there is a strong difference for the service level are easily explained by the fact that the final gradient is not as small as the gradient of those other cases where the difference for the service level is small.

The implementation of alternative optimization algorithms is a possible approach to try to increase the accuracy of these results (see section 5.2 for further discussion on this topic).

4.4.5 Weighted Shortfall *versus* Non-weighted Shortfall

As referred in section 3.4.1, the priority rule to allocate capacity to the different products, is based on the descending order of the shortfall weighted by the average demand. We are now interested in studying the impact of an alternative strategy on the system's performance, namely on average total cost. The comparison will be done with a non-weighted shortfall policy.

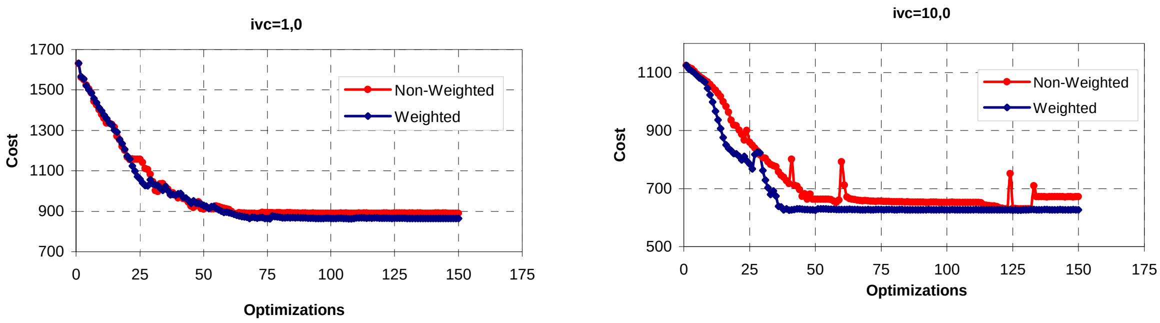

Figure 4.7 – Weighted shortfall vs. Non-weighted shortfall – ivc=1,0 and ivc=10,0 (NLP – NPY scenario)

Either with a variance coefficient of 1,0 or 1/10,0 the cost difference is less than 1%, reflecting the fact that high demand products account for 80% of total demand, as presented in Table 4.16. However, a non-weighted shortfall ordering will highly increase the lead-time of lower demand products, since it is only after several shifts that their shortfall will compete with those from high demand products (see Table 4.21). This result is in line with the discussed on the previous section, that is, a small difference on the control variables will not change significantly the average total cost, but could highly change the Type-1 service level.

Table 4.20 – Weighted shortfall vs. Non-weighted shortfall – ivc=1,0 and ivc=10,0 (LP – NPY scenario)

POLICY	OPTIMAL COST	IN-HOUSE COST	GRADIENT NORM	CONFIDENCE INTERVAL
Weighted (ivc=1)	863,79	650,45	2,88	± 0,93 %
Non-Weighted (ivc=1)	889,07	643,89	4,75	± 0,96 %
Weighted (ivc=10)	626,03	561,64	3,56	± 1,05 %
Non-Weighted (ivc=10)	629,53	572,34	4,02	± 1,34 %

Table 4.21– Average lead-time and its standard deviation comparison (Weighted shortfall vs. Non-weighted shortfall)

POLICY	**Weighted (ivc=1,0)**		**Non-Weighted (ivc=1,0)**		**Weighted (ivc=10,0)**		**Non-Weighted (ivc=10,0)**	
Demand Level	LT	σ	LT	σ	LT	σ	LT	σ
High	1,28	1,76	1,47	1,99	1,03	0,80	1,06	0,73
Medium	1,12	1,36	1,09	1,42	0,49	0,66	0,60	0,79
Low	0,99	1,37	9,70	8,72	0,41	0,67	16,18	15,60

The consequences of the results presented in this chapter in terms of defining the most appropriate strategy will be discussed in Chapter 5.

4.5 HARDWARE AND TIME REQUIREMENTS

All the experimental study was performed under personal computers with Pentium III processors at 750 MHz, and 512 MB RAM simms. Each iteration of the optimization ran for 8000 simulation periods and took 4 minutes per iteration. Finding the optimum took an average of 150 iterations. Thus, the total time required for an experiment was 600 minutes.

CHAPTER 5

CONCLUSIONS & FUTURE RESEARCH

5.1 CONCLUSIONS

This thesis proposed a framework to study the glass manufacturing production process. It considered four different production strategies – make-to-stock, make-to-order, delayed-differentiation, and a combination of these three strategies according to the demand level (MTS/DD/MTO). This thesis presented their impact on several performance measures: average total cost, in-house costs, and products' delivering time (lead-time).

The process was modeled as a discrete time, capacitated, multi-stage, multi-product, production-inventory system, with random yield, operating under multi-echelon base stock policies. The production decisions are taken in accordance with the weighted shortfall.

A simulation-based optimization was the tool used to analyze the glass production system, given the complexity of an analytical approach for those types of systems. The dynamics of the state variables, the production decisions, and their derivatives were provided by means of recursive equations. The gradient components are computed via Infinitesimal Perturbation Analysis, providing the rapid identification of good solutions. Therefore, and in the context of this thesis, the simulation is used as an optimization tool to derive the optimal parameters for the proposed production strategies.

A set of computational experiments is presented in order to get some insights about the impact of the different production strategies on the performance measures.

Recalling one of this thesis' motivations was to try to understand why, in the glass industry, the management teams usually decide for a production-to-order strategy. The numerical results of the previous chapter clearly show that a make-to-order strategy incurs less in-house costs than all the other strategies, while having the highest average total cost and the worst lead-times. Therefore, what could justify the actual common strategy? First, one needs to recognize that the average total cost reflects not only the in-house costs but also the penalty costs incurred for not immediately satisfying the customer. It is also known that these costs are hard to estimate, since they must incorporate intangible features. Additionally, if we consider that in-house costs essentially account for the value paid for raw material, energy, and labor costs, i.e., those costs with immediate impact on the companies' financial statements, the managerial decision is perfectly understandable.

It is important to stress that, despite the results presented, the MTO could represent the right strategy. Suppose that the costumer willingness to negotiate the lead-times is reasonable and that we are not under the competitors' pressure, then the referred intangible costs are negligible. Although this is not a frequent scenario, an MTO strategy would be the right one.

However, under the actual business context, where strong competition is a factor, and time and customer service level are critical issues, it sounds logical to pursue strategies different from the MTO. Despite the uncertainty associated with intangible costs estimation, management policies tending to valorize service level measures could be more profitable in the medium/long term horizon.

Moreover, given the high uncertainty induced by the random yield, an MTO strategy seems inappropriate given that lead times are higher than they would be on more reliable processes, where MTO could make more sense.

Usually, the estimation of the holding and penalty costs is a difficult task. Moreover, not only it is hard to place the real value added by a given operation, but also it is hard to measure the exact impact of backlog in terms of cost. The *optimality condition* introduced in section 3.4.3 establishes an equivalence between penalty costs and service level. Additionally, we know that the relative proportion between holding and penalty costs defines the system's performance. Therefore, setting a target service level is an easier task than to determine what should the penalty costs be.

As a final remark, one can state that the decision of produce-to-order, produce-to-stock or any other composite strategy cannot be taken independently of the business context. All the aspects must be evaluated in order to understand what are the critical issues for success, or, in other words, what are the factors most valued by the customer. The framework developed on this thesis provides means to measure the impact of a strategy change, helping management evaluating the exact trade-off involved.

5.2 FUTURE RESEARCH

This thesis provides a broad set of topics that are worthy of further investigation. Perhaps, the key one is related with the study of additional production strategies and the model development in order to incorporate more real process properties. An MTS/DD/MTO policy using the remaining free capacity to produce high demand products could be an alternative production strategy.

The introduction of random capacity, random processing times and maintenance periods on the model, in addition with the study of their impact on the system's response under different production strategies, is another possible direction for future research. Also for future study, and regarding the production decisions, it will be relevant to analyze other priority ordering rules and test different production decision algorithms. The Equalize Shortfall Algorithm is an example (see [Bispo, 1997]).

Another aspect not fully addressed in this thesis is the utilization of policies which are more adequate to deal with random yield. Some sort of *order amplification* to compensate for yield losses would be a possible approach. Additionally, the IPA technique is nothing but an excellent tool to investigate the possible benefits of alternative production rules that produce more than effectively needed to compensate for random yield. It is necessary that such rules are described by a simple set of parameters and the model guarantees the adequate smoothness properties of the cost function.

This thesis uses a discrete step version of the Fletcher and Reeves optimization algorithm. During the experimental study some difficulties were experienced in converging to the optimum. These difficulties suggest that, despite the results obtained in [Nunes et al., 1999], the above algorithm is probably not the most suitable. Then, implementing and comparing the performance of alternative optimization algorithms would be a step further to improve the quality of the generated results. The Davidon-Fletcher-Reeves algorithm or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm could be more effective alternatives. See [Bazaraa and Shetty, 1979] for a description of these and other discrete step optimization algorithms.

APPENDIX

A

THE SIMULATION & OPTIMIZATION SOFTWARE – *SIMULGLASS*

A.1 GENERAL DESCRIPTION

Despite the several process simulation software available in the market, the degree of customization one must introduce on the process model defined in chapter 3, turned out to be necessary the development of a simulation and optimization package so that the glass manufacturing process becomes represented as accurately as possible. A user-friendly interface was another important motivation on the referred tool creation.

The *Microsoft Visual Basic 6.0 – Enterprise Edition*, from *Microsoft Visual Studio*, was the development tool used to implement the software. Simplicity, design capabilities, connectivity with other software, and previous language knowledge, were the main issues considered on the software selection.

For a given production strategy selected by the user, the application, designated as *SimulGLASS – Inventory Decisions Support System* – see Figure B.1, uses the process parameters (stages' capacity and system cost structure) and products' data (demand level, processing time and yield factor) to determine the optimal values of the base-stock variables, Z, and production limit, U, for each product p and stage s, by means of an optimization procedure based on Infinitesimal Perturbation Analysis. It produces two performance measures: the total cost and products' lead-time, associated with the optimal Z's and *U's*. Although the application was designed to solve a concrete problem, minimum modifications are needed to solve any problem that requires discrete event simulation.

The application is composed of the following modules:

FORMS – Module containing all the forms used as interface with the user;

DECLARATION – Definition of all global variables used in the program;

AUXILIAR – Contains all the auxiliary procedures and functions used by the other modules;

DATA – Contains all products and system parameters used for the experimental study (see Chapter 4). This data is also available in file format;

SIMULATOR – Simulates the glass manufacturing process, using the values defined in the Data module;

OPTIMAL – Computes the optimal values of the decision variables Z and U, according to the production strategy defined by the user;

A.2 THE USER INTERFACE AND INPUT VARIABLES

The program main menu provides access to four main windows, which constitute the interface with the user. These windows are: the Product Parameters Window, the System Cost Structure Window, the Decision

Variables Window, and the Simulation and Optimization Parameters Window.

A.2.1 Product Parameters Window

As one can see in Figure B.2 (see Appendix B), the Product Parameters window has three tabs, each one for the different product groups (divided according to the demand level – High Demand Products, Medium Demand Products and Low Demand Products), on which one needs to enter the information related with the products, as displayed on Table A.1. Four combo boxes are also available to provide a rapid insertion of values for all products. The *Default Values* button provides a quick input of all default data related with the product parameters present in the DATA module.

Table A.1 – Product Parameters

PARAMETER	UNITS	DESCRIPTION
Processing Time, $T^{p_m s}$	[min./unit]	Processing time of product *p* at stage *s*
Yield Factor, $\alpha^{p_m s}$	[0 –1]	Yield factor for product *p* at stage *s*
Demand Average, $\overline{D}^{p_m}$	[units/shift]	Average demand for product *p* at stage *1*
Demand Variance Coefficient, $\sigma_D^{p_m}$	[units/shift]	Demand variance coefficient for product *p* at stage *1*

A.2.2 System Cost Structure Window

The application works with a direct cost structure, which integrates labor costs and energy costs – both proportional to working time – with raw materials costs. Hence, the user needs to input the hourly cost of labor and energy at each stage of the process, the cost of each kilogram of glass, and the glass quantity needed to produce each product. These parameters and their units are presented on Table A.2.

The *Default Values* button displayed on Figure B.3 allows a fast insertion of all default information contained in the DATA module and related with the process costs. This set of numbers will be used, as explained in section 4.2.1, to establish the products holding cost h^{ps} and penalty cost b^{ps}, used to find the total cost related with a given production strategy.

Table A.2 – System Cost Structure

PARAMETER	UNITS	DESCRIPTION
Labor Cost	[u.m./hour]	Hourly manual labor cost at each process stage s
Energy Cost	[u.m./hour]	Hourly energetic cost at each process stage s
Glass Cost	[Esc/Kg]	Cost of raw materials kilogram
Glass Quantity	[Kg]	Quantity of glass for product p

A.2.3 Decision Variables Window

The base stock levels $Z^{p_m s}$ and the production limits $U^{p_m s}$, for each product p at a given stage s, form the two groups of decision variables, which need to be initialized by the user. Despite the set of default values defined on DATA module, which can be accessed by the *Default Values* button, it is possible to start the application with any values of $Z^{p_m s}$ and $U^{p_m s}$, once the following two conditions are verified:

i) Stability condition – All base stock levels have to be ordered in accordance with $0 < Z^{p_m 1} < Z^{p_m 2} < \ldots < Z^{p_m S}$;

ii) Feasible region condition – the production limit of each product has its feasible domain defined as $\overline{D}^{p_m} < U^{p_m s} < \dfrac{C^s}{T^{p_m s}}$, where $\overline{D}^{p_m}$, C^s, and $T^{p_m s}$ represent, respectively, the average demand, stage's capacity, and processing time of a given product p at process stage s.

Another parameter that needs to be established at this time is the production policy. Each product has a combo box with the available policies – MTS, DD and MTO – for simulation. Basically, this combo box works like a switch that forces the values of the base stock levels when the policy selected is other than MTS.

As one can see at the right side of Figure B.4, there are four buttons for quick adjust of the decision variables according to the policy selected. These buttons represent four different pre-defined production strategies. *All products MTS* will simulate and optimize a scenario where all base stock levels are subjected to optimization, while *All products MTO* corresponds to a scenario with all base stock levels set to zero. The selection of *All products DD* button will run the application with all products simulated under a delayed-differentiation policy. In terms of base stock levels, this option conducts to null values at the cold-area stages and adjustable values at the hot-area process stages. Lastly, the *MTS/DD/MTO* button stands for a simulation scenario where the products with a high level of demand are produced under a make-to-stock policy, the medium level ones under a delayed differentiation policy and the low ones under a make-to-stock policy.

A final reference to the *Base Stock Z* and *Production Limit U* combo-boxes, which linked to the *Multiplying Factor* text-box, allow a quick launch of all *Z* and *U* values, guaranteeing simultaneously the Stability Condition and the Feasible Region Condition.

Table A.3 – Decision Variables

PARAMETER	DESCRIPTION
Base Stock Level, $Z^{p_m s}$ [units]	Base stock level of product *p* at process stage *s*
Production Limit, $U^{p_m s}$ [units]	Production limit of product *p* at process stage *s*
Production Policy Combo Box	Selection of one of the available policies for each product *p*
All products MTS button	Selection of Make-to-Stock policy for all products
All products MTO button	Selection of Make-to-Order policy for all products
All products DD button	Selection of Delayed-Differentiation policy for all products
MTS/DD/MTO button	Selection of a combined policy in accordance with demand level (High demand – MTS, Medium demand – DD, Low demand – MTO)

A.2.4 Simulation and Optimization Parameters Window

The Simulation and Optimization Parameters Window (see Figure B.5) exhibits all the parameters needed to control the process simulation and optimization.

The manufacturing process is simulated during a number of cycles defined by the user at the *Number of Simulations* text box. Each cycle may represent a production shift, a production day or any other time scale, depending on the time relation between products processing times and stage capacity, which the user has to define. Often, a small number of simulations do not incorporate all the possible production situations and, if there is a significant number of unusual occurrences they will greatly affect the final result. Therefore, it is heavily recommended the usage of a large number of simulations so that such effect can be, if not eliminated, strongly reduced.

The optimization has three convergence criteria that stop the application execution: a maximum number of optimizations – defined at *Nº Max. Optimizations* text box; a step, used with the cost function gradient to determine a new values for the control variables – labeled as *Minimum Step* text box; the cost difference between two consecutive optimizations – defined at *Epsilon* text box.

The optimization module has a procedure to guarantee the stability condition any time a new set of variables is determined. If the gradient information leads to a violation ($Z^{p_m s} < Z^{p_m (s-1)}$) of the above condition, the value defined at *Epsilon Delta* text box is used to correct such infringement defining $Z^{p_m s} = Epsilon\,Delta + Z^{p_m (s-1)}$.

The number in the *Step* text box is used as the initial program step, being applied after the first simulation to calculate the new set of *Z* and *U* variables. The answer for a typical value for an initial step is not obvious, depending on several factors. Hence, the best approach to find a fine value for the initial step is probably a sensitivity analysis procedure.

Two different ranking strategies of products shortfall are implemented in the code. If the *Weighted Shortfall* check box is selected the products are ranked in descending order of their shortfall divided by average demand. If it is not checked, the ranking is built only based on shortfall.

At this point one need also to define each stage capacity (using units according with those presented on Table A.4), which, as referred above, depends on the time basis of all relevant variables.

The *Policy Selection* combo box must be used to inform the program of which of the policies, already mentioned on this text, one intends to simulate.

During the program execution one can verify its evolution on the two counters available – the *Simulations Counter* and the *Optimizations Counter*. Given that these counters are processor consuming, subsequently time consuming, the user can turn them off through the *Counters On/Off* Switch. To finish, once all parameters are defined, one just needs to press the *Run* button to start the program.

Table A.4 – Simulation and Optimization Parameters

PARAMETER	DESCRIPTION
Number of Simulations	Number of production shifts
Minimum Step	Minimum step value to guarantee convergence
Epsilon	Minimum cost difference to guarantee convergence
Epsilon D	Minimum value to guarantee the Stability condition
Step	Optimization Initial Step
Nº Max. Optimizations	Maximum number of optimization iterations
Stages Capacity	Definition of each process stage capacity [min.]
Policy Selection Combo Box	Policy definition (MTS, MTO, DD or MTS/DD/MTO)
Weighted Shortfall Check box	Selection between raking or not the shortfall weighted with average demand in Production Decisions Algorithm
Simulations Counter	Number of simulations already completed
Optimizations Counter	Number of optimizations already completed
Counters Switch	Turns ON and OFF the counters
Run button	Starts the Simulation / Optimization process
Main Menu button	Returns to the application Main Menu

A.3 THE SIMULGLASS OUTPUT

The output provided by the SimulGLASS program can be divided in two main groups representing system performance measures, both related with the selected production strategy. One group, focused on the cost performance measure, has not also the historical evolution of cost during the optimization process but also the decision variables optimal values corresponding to the minimum cost scenario, amongst other less significant results. The other performance measure group contains lead-times, structured by product and by demand level group.

A.3.1 Historical Cost and Optimal *Z* and *U* Values

Before running the application one must create the folder *simuldata* at the PC root –*c:\simuldata*. The text file that contains the results – optimal.txt – is created automatically by the application. By default, if the file already exists in the folder, it is overwritten. The following results are available at the file:

i) Historical evolution of cost for all optimizations performed;

ii) Minimum cost found by the optimization process;

iii) Base stock levels Z^* and production limits U^*, for all products at each stage s, corresponding to the minimum system cost scenario – the so called *optimal values*;

iv) Cost function gradient and its norm, corresponding to the minimum process cost state;

A.3.2 The Lead-Times

The lead-times are computed with a production scenario having as base stock levels and production limits those resulting from the optimization process. These results are recorded in the file located at *c:\simuldata\leadtime.txt*, which is also an overwrite type file. One may divide the results provided in leadtime.txt file as follows:

i) Average lead-time and lead-time standard deviation for all products;

ii) Average lead-time and its standard deviation for each demand level group – High, Medium and Low demand level;

iii) An overall average lead-time and its standard deviation;

iv) The 95% confidence interval with 25 replications with Z and U at their optimal values;

v) The type-1 service level for all products;

A.4 THE SIMULATION MODULE

This module represents the core of the application since it tries to replicate the glass manufacturing process dynamics. The following seven key procedures form it: Demand Generation, Echelon Update, Shortfall Update, Production

Decisions, Inventory Update, Cost Update and Lead-Time. These procedures are nothing but the code implementation of module equations presented at Chapter 3. The flowchart presented on figure A.1 illustrates the main simulation algorithm.

A.4.1 Demand and Yield Generation Procedure

At each simulation cycle a new demand vector is determined using an Erlang Distribution, which uses two parameters – average demand and the inverse of the variance coefficient – to generate a random number between 0 and 1, which then is used to calculate the demand value for a known product *p* at a given simulation cycle *n*.

Both perfect yield and random yield can be incorporated in the production system and simulated by the application. A new vector of yield coefficients is computed at each simulation run if such feature has been selected. The random yield is generated through a uniform distribution with bounds such that the average matches the value specified by the user.

A.4.2 Echelon Update Procedure

Although this procedure is not used at the application final version, it was extremely useful during the testing period. It computes for all products both the inventory sum upstream from stage 1 of the process (see equation (3.3)) and its derivatives as stated by equations (3.16) and (3.17) but considering the echelon variable. Once the echelon inventories are computed, one can estimate the quantity needed to reset the base stock levels – the shortfall.

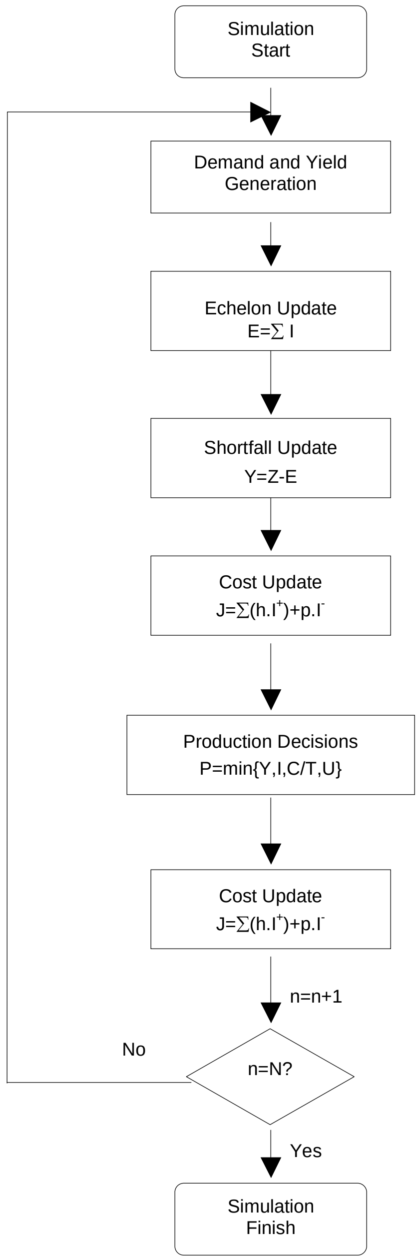

Figure A.1 – Simulation algorithm flowchart

A.4.3 Shortfall Update Procedure

During all simulation run, the shortfall procedure determines for all products and stages the difference between the base stock level and echelon inventory in line with equation (3.5). The shortfall derivatives are also computed at this stage using equations (3.16) and (3.17) but applied to the shortfall variable.

A.4.4 Production Decisions Procedure

Section 3.5 describes the production decisions algorithm, which was translate to code in this procedure. After ranking all products in descending order of their weighted shortfall (or pure shortfall, depending on the strategy selected) the production decisions are taken considering de actual shortfall, the inventory at the upstream stage, the available capacity, and the production limit. In line with equation (3.23), the minimum of these values corresponds to the quantity that will be produced, while the production derivatives were codified taking into consideration equations (3.24) and (3.25).

It was necessary to create auxiliary arrays to deal with the process transition stage – stage 4 – both for control variables and its derivatives, because at each production decision it is required to update their values. In other words, one needs to incorporate on every product the effects produced by its three sub-products.

To rank the weighted shortfall matrices, two sorting algorithms were implemented – the Selection Sort (n^2 degree) and the Quicksort ($n.log(n)$ degree) as presented in [Cormen et al., 1997]. Nevertheless, the experiments showed that even for production systems with a considerable dimension (27 products and 9 stages) an $n.log(n)$ degree sorting algorithm represents only less than 1% of processor time when compared with a n^2 sorting algorithm. Hence, the relative performance of both algorithms will become visible just for very large arrays.

A.4.5 Inventory Update Procedure

After deciding the production quantities one needs to update the inventory at each stage. Such update is made in line with equations (3.1) and

(3.16-17) corresponding in that order to the inventory variables and to the inventory derivatives.

A.4.6 Cost Update Procedure

The equations (3.11) and (3.18) describe the way the performance measure cost and its derivatives are computed every simulation cycle. Holding and penalty costs for all products are computed during the simulation run. The sum of these costs is divided by the number of cycles to get an average cost, which is the scalar returned by the Cost Update Procedure.

A.4.7 Lead-Time Procedure

After finding the optimal values for the different production policies, the application will compute the average lead-time of all orders received. If the system has sufficient inventory to satisfy all the order, the lead-time is zero. Otherwise, it starts counting the number of simulation cycles necessary to deliver the complete order. For strategies diverse from the MTS, one can define a time lag during which the costumer is willing to wait for the order. After that, the clock starts to count the waiting period. This procedure will also compute the type-1 service level, according to the selected policy.

A.5 THE OPTIMIZATION MODULE

In this thesis context, and considering that the IPA algorithm needs the calculation of the performance measure function gradient, one must choose an algorithm that requires evaluation of the function derivative. Algorithms using the derivative are somewhat more powerful that those using only the function value, but not always enough so as to compensate for the additional calculations of the derivatives. There are two major families of algorithms for multidimensional minimization with the calculation of first order derivatives. The first family goes under the name *Multidimensional Search Using Derivatives*, which includes the *Steepest Descent* method and the Newton method. The second family, *Multidimensional Search Using Conjugate Directions*, includes the *Davidon-Fletcher-Powell* method, the conjugate gradient method of *Fletcher-Reeves* and the *Zangwill* method. The work of [Nunes et al., 1999] highlights the better performance of *Fletcher-Reeves* method in the context of this thesis, when compared with the other optimization algorithms.

A.5.1 The Fletcher and Reeves Algorithm

As described in [Bazaraa and Shetty, 1979] the conjugate gradient method of Fletcher and Reeves deflects the direction of steepest descent by adding to it a positive multiple of the direction used in the last step.

INITIALIZATION STEP

Choose a terminating scalar $\varepsilon > 0$ and an initial point x_1, which is a value for the decision variables (Z, U). Choose the initial step size δ_1. Let $d_1 = - \nabla J(x_1)$, $i=k=1$. Let $count=0$.

STEP 1

If $\nabla J(x_i) < \varepsilon$, STOP. Otherwise, let $x_{i+1} = x_i + \delta_k . d_i$.

If $J(x_{i+1}) < J(x_i)$ go to *STEP 2*, otherwise go to *STEP 3*.

STEP 2

The iteration is a *success* and a new direction is constructed:

$$d_{i+1} = \nabla J(x_{i+1}) + \alpha_i . d_i \text{ where } \alpha_i = \frac{\left\|\nabla J\left(x_{i+1}\right)\right\|^2}{\left\|\nabla J\left(x_i\right)\right\|^2}.$$

Also, $\delta_{k+1} = 1,1 . \delta_k$. Let $i=i+1$, $k=k+1$, count=0, and go to *STEP 1*.

STEP 3

The iteration is a *failure*. Let $count=count+1$.

If *count=K* go to *STEP 4*. Otherwise let $\delta_{k+1}=\dfrac{count}{count+1}\delta_k$ with *count=1, 2, ..., K-1*. Let *k=k+1* and go to *STEP 1*.

STEP 4

Because the maximum number of successive failures is reached, take a *Spacer Step*. Use as direction the gradient of the present iteration, i.e., $d_i = -\nabla J(x_i)$. Reset $\delta_{k+1} = \delta_{k-K}$, that is, the value it had after the last success. By successively reducing the step size, let x_{i+1}, be the first value that leads to a success for x_i, using always direction d_i. Let *p* be the number of iterations within *STEP 4*. Let *i=i+1*, *k=k+p*, and go to *STEP 2*.

A.6 TESTING THE SOFTWARE

Several tests were made to the application to guarantee its stability, to understand its behavior under abnormal conditions, and essentially, to verify the code conformity with module equations and decision algorithms. Among those tests one can refer the following:

- *i)* Derivative Definition;
- *ii)* Cuts along gradient direction function;
- *iii)* Control routines and error messages;

As an example and because of its relevance, the test referred in *i)* is based on the Taylor series development of the cost function *J* (not considering second order terms) and consists on perturbing by an infinitesimal amount ε one of the variables and running a simulation. The derivative of the cost function in order to the perturbed variable must verify the following relation:

$$\frac{\partial J}{dZ} = \frac{J_{perturbed} - J_{nominal}}{\varepsilon} \qquad \text{(A.1)}$$

where *Z* can represent any system control variable.

APPENDIX

B

SIMULGLASS USER-INTERFACE WINDOWS

Figure B.1 – The *SimulGLASS* Enter Window

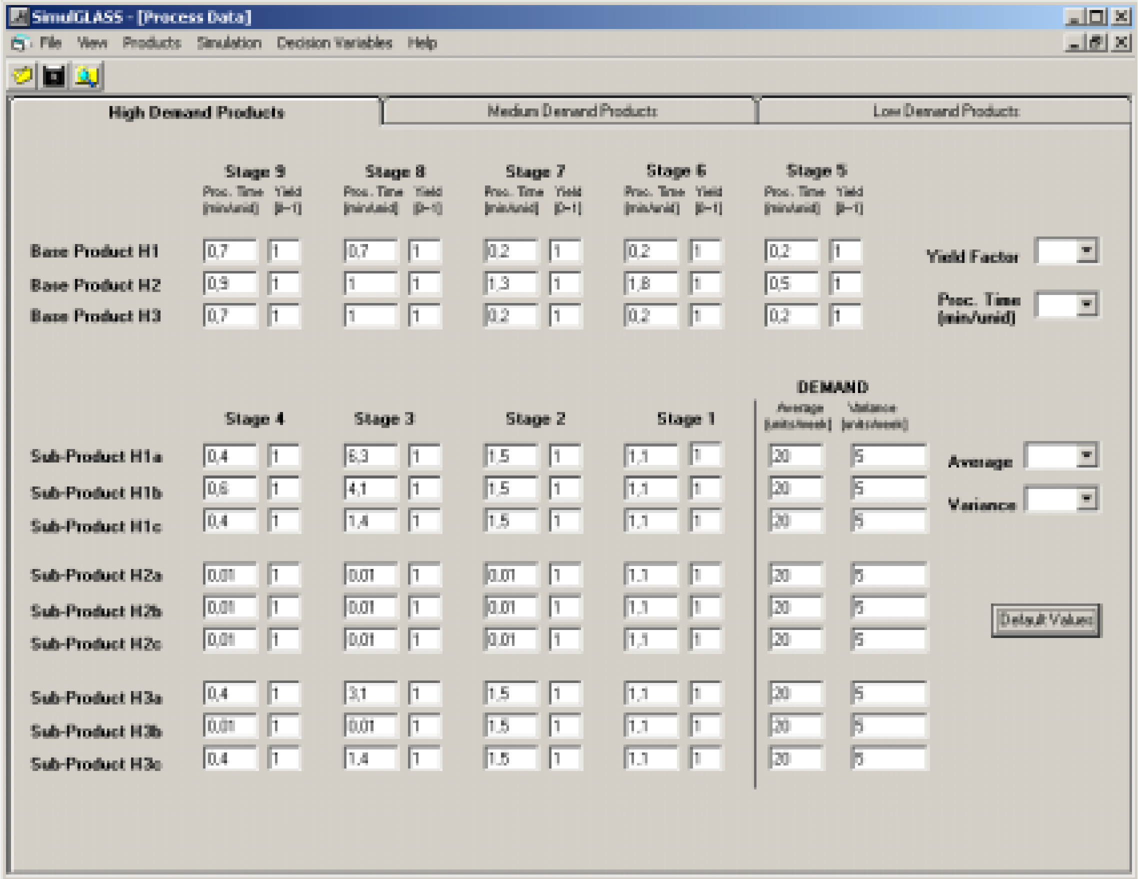

Figure B.2 – Products Parameters window

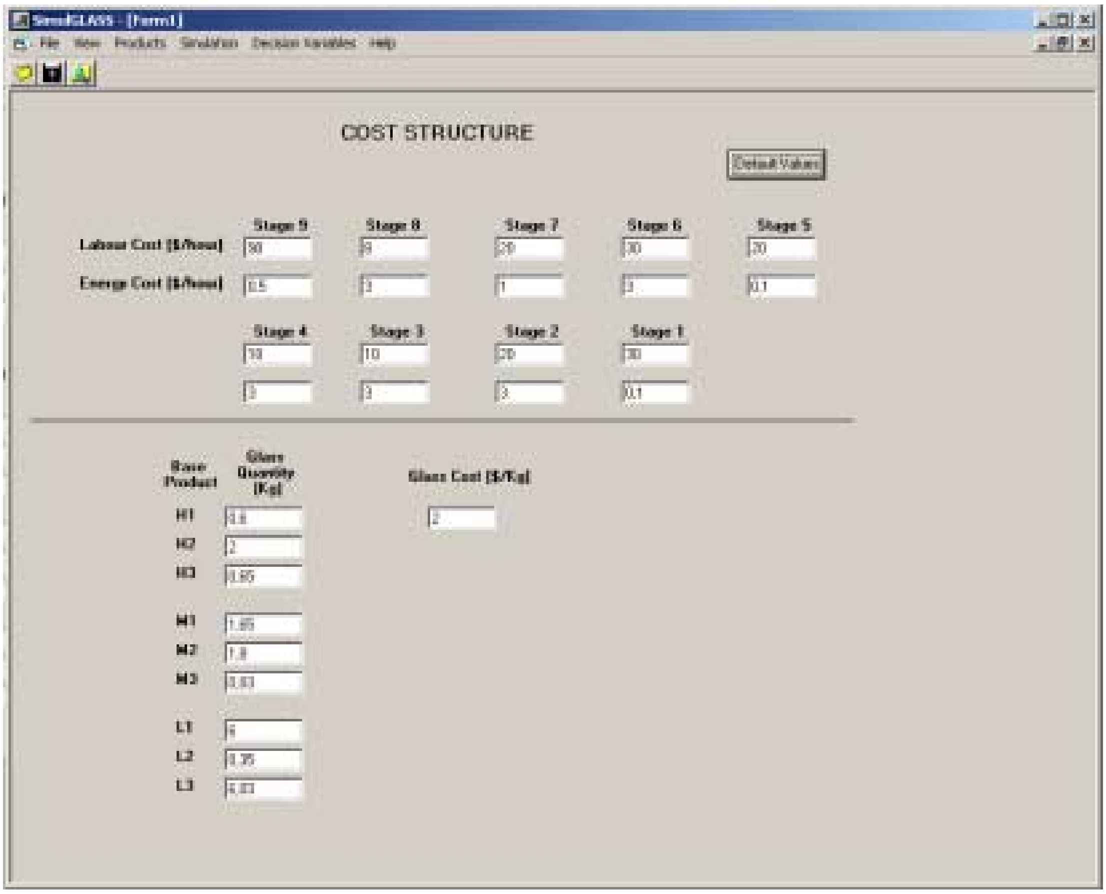

Figure B.3 – System Cost Structure Window

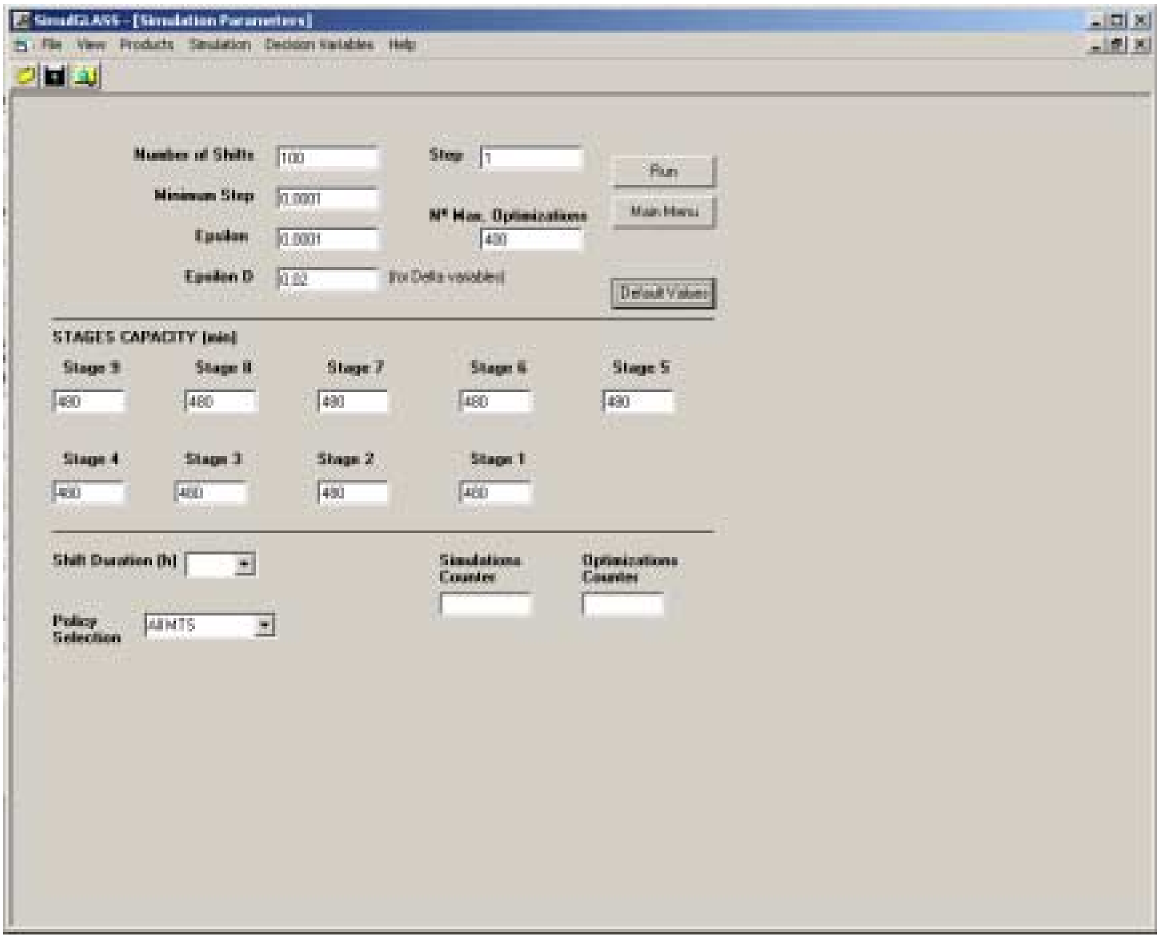

Figure B.4 – Decision Variables Window

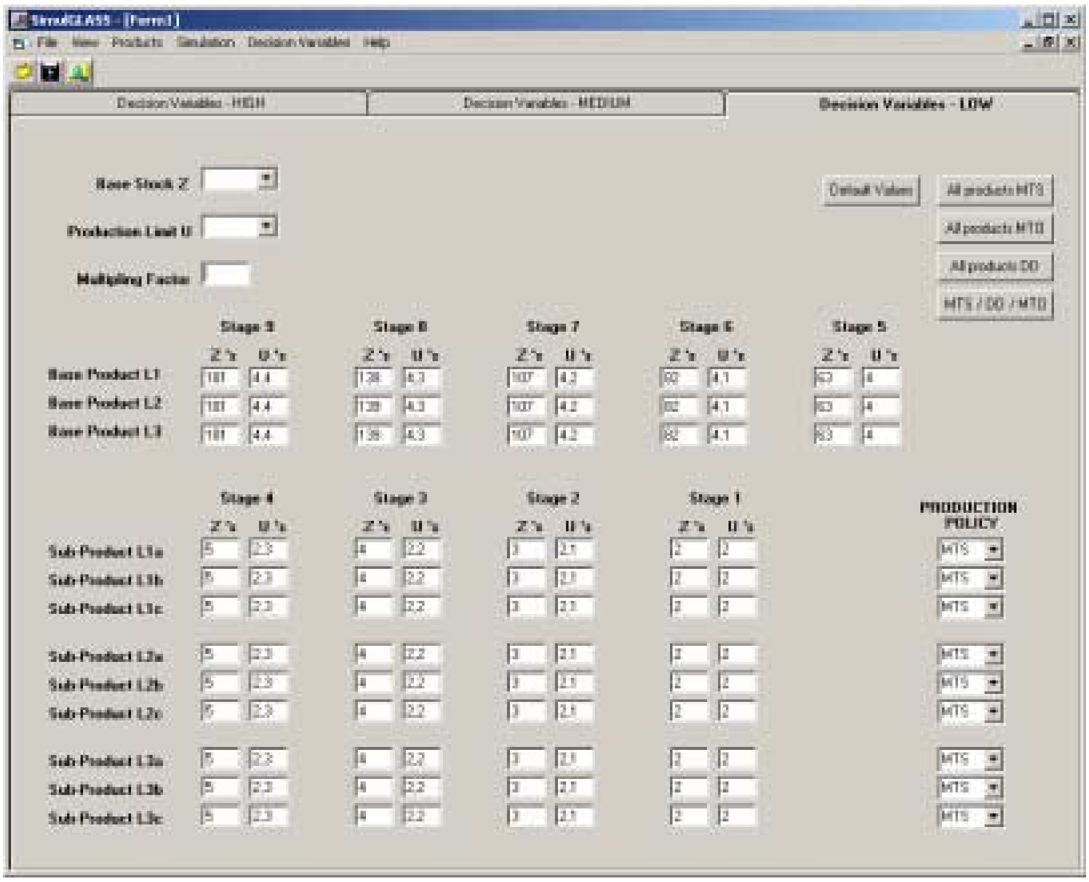

Figure B.5 – Simulation and Optimization Parameters Window

APPENDIX C

THE SIMULGLASS PACKAGE

The simulation software package – SimulGLASS – is enclosed with this thesis. Appendix A gives some insights in how to use the software. For installation, it is only necessary to put the cd-rom on the drive, click on the *setup.exe* file and follow the installation instructions. Before running the application, it is also necessary to create the following folder at the hard drive root: c:\simuldata\.

A complete version of this thesis is also available on the cd-rom (see the *MasterThesis.pdf* file).

BIBLIOGRAPHY

[Akella and Kumar, 1986] Akella, R.and Kumar, P. R. (1986). *Optimal Control of Production Rate in a Failure Prone Manufacturing System*. IEEE Transactions on Automatic Control, AC-31(2) : 116-126.

[Bai and Gershwin, 1994] Bai, S. X. and Gershwin, S. B. (1994). *Scheduling Manufacturing Systems with Work-In-Process Inventory Control: Multiple-part-type Systems*. International Journal of Production Research, 32(2): 365-385.

[Bai and Gershwin, 1995] Bai, S. X. and Gershwin, S. B. (1995). *Scheduling Manufacturing Systems with Work-In-Process Inventory Control: Single-part-type Systems*. IIE Transactions, 27(5): 599-617.

[Bai and Gershwin, 1996] Bai, S. X. and Gershwin, S. B. (1996). *Scheduling Manufacturing Systems with Work-In-Process Inventory Control: Reentrant Systems*. OR Sepctrum, 18(4): 187-195.

[Bazaara and Shetty, 1979] Bazaara, M. S. and Shetty, C. M., (1979). *Nonlinear Programming: Theory and Algorithms*. John Wiley and Sons, New York.

[Bispo, 1997] Bispo, C. F. (1997). *Re-Entrant Flow Lines*. Ph.D. Thesis, Graduate School of Industrial Administration and the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

[Cassandras and Ho, 1983] Cassandras, C. and Ho, Y. C. (1983). *A New Approach to the Analysis of Discrete Event Dynamic Systems*. Automatica, 19(2): 149-167.

[Clark and Scarf, 1960] Clark, A. and Scarf, H. (1960). *Optimal Policies for a Multi-Echelon Inventory Problem*. Management Science, 6: 474-490.

[Cormen et al., 1997] Cormen, Thomas H., Leiserson, Charles E. and Rivest, Ronald L. (1997) *Introduction to Algorithms*. The MIT Press, Cambridge, Massachusetts.

[Desrochers, 1990] Desrochers, A. A. (1990). *Modeling and Control of Automated Manufacturing Systems*. IEEE Computer Society Press, Washington, D.C.

[Gerchak et al., 1988] Gerchak, Y., Vickson, R. G., and Parlar, M. (1988). *Periodic Review Production Models with Variable Yield and Uncertain Demand*. IEE Transactions, 20(2): 144-150.

[Glasserman and Tayur, 1995] Glasserman, P. and Tayur, S., (1995). *Sensitivity Analysis for Base-stock Levels in Multiechelon Production-inventory Systems*. Management Science, (41(2): 263-281.

[Graves et al., 1993] Graves, S. C., Rinnooy Kan, A. H. G. and Zipkin, P. H. (1993). *Logistics of Production and Inventory*, volume 4, chapter 1, 4, 6, 7 and 11. Handbooks in Operations Research and Management Science. Elsevier Science Publishers B. V.. Amsterdam, The Netherlands.

[Graves, 1996] Graves, S. C. (1996). *A Multiechelon Inventory Model with Fixed Replenishment Intervals*. Management Science, 42(1): 1-18.

[Grosfeld-Nir and Gerchak, 1996] Grosfel-Nir, A. and Gerchak, Y., (1996). *Production to order with random yields: single-stage multiple lot-sizing*. IIE Transactions, 28: 669-676.

[Henig and Gerchak, 1990] Henig, M. and Gerchak, Y. (1990). *The Structure of Periodic Review Policies in the presence of Random Yield*. Operations Research, 38(4): 634-643.

[Ho et al., 1979] Ho, Y. C., Eyler, M. A., and Chien, T. T. (1979). *A Gradient Technique for General Buffer Storage Design in a Production Line*. International Journal of Production Research, 17(6): 557-580.

[Ho, 1988] Ho, Y. C. (1988). *Perturbation Analysis Explained*. IEEE Transactions on Automatic Control, 33(8): 761-769.

[Johnson and Jackman, 1989] Johnson, M. E. and Jackman, J. (1989). *Infinitesimal Perturbation Analysis: A Toll for Simulation*. Journal of the Operational Research society, 40(3): 243-254.

[Karmakar and Lin, 1986] Karmakar, U and Lin S. (1986). *Production Planning With Uncertain Yields and Demands*. Working Paper, William E. Simon Graduate School of Business Administration, University of Rochester, N.Y..

[Kapuscinski and Tayur, 1999] Kapuscinski, R. and Tayur, S., (1996). *A Capacitated Production-Inventory Model with Periodic Demand*. Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University.

[Kimemia, 1982] Kimemia, J. G. (1982). *Hierarchical Control of Production in Flexible Manufacturing Systems*. PhD thesis, Mass. Inst. Technology, Cambridge, MA.

[Kimemia and Gershwin, 1983] Kimemia, J. G. and Gershwin, S. B. (1983). *An Algorithm for the Computer Control of a Flexible Manufacturing System*. IIE Transactions, 15(4): 353-362.

[New and Mapes, 1984] New, C. and Mapes, J., (1984). *MRP With High Uncertain yield Losses*. Journal of Operations Management, 4(4): 315-330.

[Nunes et al., 1999] Nunes, C., Sofia, Isabel (1999). *Controlo de Inventário para Múltiplos Produtos*. Trabalho Final de Curso da Licenciatura em Engenharia Electrotécnica e Computadores, Instituto Superior Técnico.

[Rao, 1996] Rao, Singiresu R., (1996). *Engineering Optimization – Theory and Praactice*, chapter 1, 2, 5, 6 and 12, third edition. John Wiley and Sons, New York.

[Ross, 1990] Ross, Sheldon M., (1990). *Introduction to Probability and Statistics for Engineers and Scientists*, chapter 2, 3, 4, 8 and 12. John Wiley and Sons, New York.

[Sharifnia, 1988] Sharifnia, A. (1988). *Production Control of a Manufacturing System with Multiple Machine States*. IEEE Transactions on Automatic Control, AC-33(7): 60-625.

[Silver et al., 1998] Silver, E. A., Pyke, D. F., and Peterson R., (1998). *Inventory Management and Production Planning and Scheduling*. John Wiley and Sons, New York.

[Suri and Zazanis, 1988] Suri, R. and Zazanis, M. (1988). *Perturbation analysis Gives Strongly Consistent Sensitivity Estimates for the M/D/1 Queue*. Management Science, 34(1): 39-64.

[Tayur, 1996] Tayur, S. (1996). *Recent Developments in Single Product, Discrete-time, Capacitated Production-Inventory Systems*. Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University.

[Yano and Lee, 1995] Yano, C. A. and Lee, H. L., (1995). *Lot Sizing with Random yields: A Review*. Operations Research, 43(2): 311-314.