OPTIMIZED MARKET INTRODUCTION OF LARGE CAPITAL PRODUCTS (LCP) WITH LONG DEVELOPMENT AND LEARNING CYCLES

by

ANTJE LEMBCKE

Dipl. Ing. (FH) University for Applied Science Weihenstephan, Germany, 1998
MBA University of Central Florida, 2004

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2010

Major Professor: Linda Malone

© 2010 Antje Lembcke

ABSTRACT

Any product sold is expected to be reliable and available when the customer wants to operate it. Companies that produce large capital products (LCP), such as rockets, satellites, or large gas turbines to generate electrical energy, tend to shy away from extending their testing and validation method above the requirements by law, mainly due to the very high costs of each additional test and the uncertain return on investment.

This research shows that today's state of the art validation methods for LCP, required by law, or suggested in literature, and adapted by these industries, are not capable of capturing all significant failure modes (or even enough failure modes), with the consequence that the subsequently sold commercial product will still experience failures with significant effects on product reliability, and subsequently on the companies' bottom line earnings projections.

The research determines the type of data (significant variables) necessary to correlate a company's validation policy to product failures after commercialization, and predicts the financial impact of the current validation policy on the company's profitability. An optimized validation plan and testing policy is suggested, and its impact on a company's profitability is demonstrated through simulation.

A generic methodology is derived and its viability is illustrated using a specific product and a dynamic model developed with data available to the researcher. The generic method can be applied by any company to develop its own model for optimizing product reliability prior to market introduction.

ACKNOWLEDGMENTS

Foremost I would like to express my sincere gratitude to my advisor and committee chair, Dr. Linda Malone for her valuable guidance and support throughout the process of writing this thesis. Very special thanks goes to my committee member and mentor Dr. Hans-Juergen Kiesow who encouraged me all the way through my graduate study and for his invaluable advice on my research. I would like to cordially thank my committee member Dr. Tim Kotnour for all his insightful suggestions during our discussions and express my warmest appreciation to my other committee member Dr. Wilhelm Thompson, and to Dr. Konrad Vogler who devoted their time to review my work.

I would like to thank Dr. Olaf Koenig for his advice in gas turbine testing and my colleague Dr. Marco Brunelli, whose very useful knowledge of statistics brought new ideas to this project. I would like to thank Joy Tatlonghari for kindly helping me go through all required administrative procedures at UCF. Finally I would like to thank Siemens AG for allowing me to access their data and last but not least my boss, Vinod Philip, for the continuous support he offered throughout this venture.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

LCP	Large capital products
LS	Least Squares Method
ML	Maximum Likelyhood Method
MTTF	Mean Time to Failure
NCC	Non-Conformance Cost
NPD	New Product Development
OEM	Original Equipment Manufacturer
OR	Operations Research
PDF	Probability Density Function
R&D	Research and Development

CHAPTER ONE: INTRODUCTION AND RESEARCH SCOPE

Practical Need and Motivation

Any product sold is expected to be reliable and available when the customer wants to operate it. There are many different definitions for the terms reliability and availability. In this document, availability is defined as the probability that a product is operating at a given time interval when used as specified as given in Equation 1:

$$Availability = \frac{\text{uptime}}{\text{uptime} + \text{downtime}} \qquad \textit{Equation 1}$$

Every interruption of product usage, be it planned or unplanned affects availability. Reliability is defined as the ratio of the actual system uptime and the planned system uptime. Planned system uptime is total possible system uptime time minus planned downtime for maintenance (see Equation 2). Every unplanned interruption due to product failure reduces reliability.

$$Reliability = \frac{\text{actual uptime}}{(\text{total possible uptime} - \text{planned downtime})} \qquad \textit{Equation 2}$$

High reliability and availability is important for the customer of any consumer products such as telephones and cars, but also for products such as rockets, satellites, or large gas turbines to generate electrical energy. These products are called large capital products (LCP), because companies buying them have to make significant upfront investments in the order of hundreds of millions of dollars to build/acquire them, but a company’s return on investment does not occur until years later.

LCPs are defined in this document as having long development times (five to eight years), long learning cycles (on average eight years), and low yearly production rates (generally less than 100 products a year within one product class). A learning cycle of a LCP encompasses the continuous collection of experiences during concept studies, design and manufacturing, instrumentation and component validation, final product testing and validation, and some years collecting experience while the product is operated by a customer at a customer site. If failures occur, there has to be a redesign, manufacturing of the new part, implementation of the new part and operation again to prove the redesign is correcting the error. These two feedback loops (implementation of the design and the error loop) complete the learning cycle. The gathered and gained knowledge during this learning cycle is used to improve future product developments and upgrades. The fact that the development time, and the learning cycle for one product is so long and the product is so expensive has significant impact on decisions made on the number of units tested before release, the testing duration, and when to serially release the product, all of which require substantial financial commitment and each of these decisions can hugely affect profit and loss potential.

LCP development is following a common new product development process (NPD). Typically, after full lab testing of the product's constituents has been completed, late stage design system testing is conducted with one prototype product due to high costs. After several validation runs, and reiterative fixing of occurring issues during validation, the product's design is considered mature enough for serial release and commercialization (Bobrow, 1997). Commercialization of a product in this document means that the product specification is released for serial production and products of this type can be offered to the customer. However, different customers are using these products with highly varying operating conditions. Manufacturing uncertainties also add to variations of these products. Together these issues can randomly cause a product to fail, despite the product being tested in one prototype testing phase (Kiesow, 2008). According to Swift roughly three quarters of product failures are established in the product design process phase. Swift also states that about 60% of the failures (caused in the design process) are detected only after first start-up and during service (1997). It is apparent that many companies in all industry sectors and for all product types, are struggling to develop a product that is really error free after commercialization. As an example, the official government website www.Recalls.gov, which shows recent recalls for products such as cars, motorcycles and leisure boats in the United States due to safety issues, states that there were thirty safety recalls on these three products in the month of March 2008 alone in the US.

Cars, motorcycles and the typical leisure boat are inexpensive to qualify, compared to LCP, and their producers can more easily afford to test several prototypes to detect design and manufacturing flaws. LCP producers however, which are the subjects of this research, experience testing costs that are substantially higher. Also with the high prototype testing cost in mind, it is common understanding that the testing of one prototype is sufficient to introduce the product into the market with acceptable risk. This research intends to investigate and propose a process that enables OEM's to evaluate alternative late stage testing methods and schedules for LCPs to reduce producers' risk of non-conformance cost linked to correction of errors after product release. The process is seeking to optimize a companies' product market introduction strategy by evaluating the cost impact of errors and for the example in this research the impact on reputation caused by product errors at the customer site after product release. There will be an attempt to factor these costs into a different testing approach. Industries producing LCPs such as gas turbines, airplanes, submarines, satellites and the like can benefit from an application of such a process or policy. This process helps to estimate and simulate the more comprehensive picture rather than only a part of it. The results of this research may also be of interest for insurers issuing insurance policies for large capital goods.

Research Scope

This research intends to develop and build a generic dynamic model (considering interdependencies and feedback loops) that supports to assess and optimize product testing policies of companies that build large capital products (LCPs; *German*: Anlageinvestitionsgüter).

The generic model will be derived by first developing an industry specific model, using LCP data available to illustrate the deficiencies in detail in today's testing strategies. Then an appropriate optimization method will be applied to the model, with the goal of providing guidance for a better testing strategy to detect as much as possible product deficiencies before serial release of the product and to improve the decision-making process.

Failure and cost data of any product are confidential and therefore difficult to get. In fact, hardly any company likes to admit that errors occurred in their product, and likes even less to report about them. The actual dynamic model used will therefore be partly product and perhaps company specific, created with data that is available to this researcher but has been normalized for privacy. However, each company would be able to generate their own specific in-house model based on the generic process developed in this research. The scope of the research is depicted in Figure 1. The flow chart describes schematically the design and implementation process of a product that, for example, Company A is using for their R&D and commercial product development. The process steps from prototype testing to field operation, as depicted in Figure 1, marks the scope of this dissertation.

This study will not address questions as to the right timing of market introduction, pricing, correct product mix, opening of new markets, or similar topics. All calculations are made under the assumption that there is an average market for the product and the product demand would be typical for the average market.

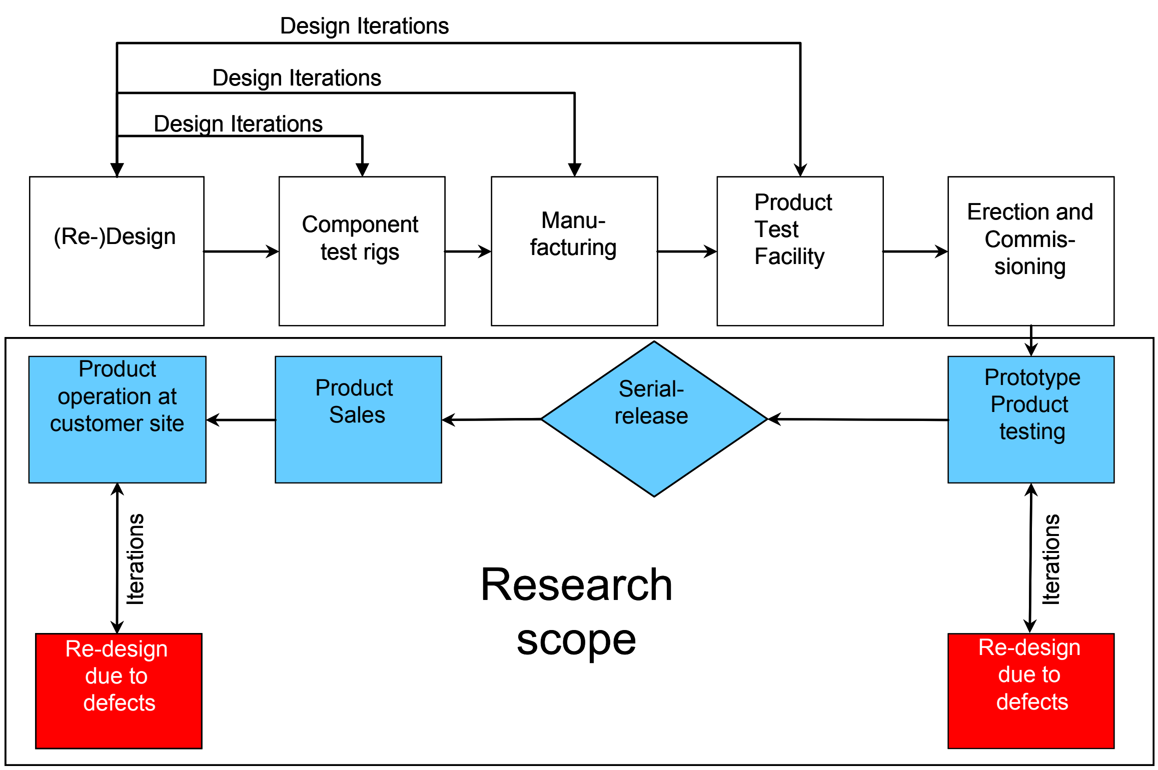


Figure 1: New Product Design Process and Defect Occurrence Opportunities

<u>Hypothesis One</u>

<u>Hypothesis one:</u> For original equipment manufacturers (OEMs), which are defined in this context as suppliers who design and manufacture LCPs, it is not economically appropriate to

build only one prototype product, test it, correct issues as they arise during testing (also called reliability growth testing), and then release the design for serial manufacturing.

Figure 2 illustrates a typical product development and testing process with a simple cost and profit calculation (Payback), assuming that all errors will be detected and corrected prior to serial release and subsequent sales. The payback calculation method was preferred over a Net Present Value calculation method because of its simplicity. Return on investment method (ROI) could have been another viable technique. A payback period is defined as the time for an asset to pay for itself through the profit it makes over the years. Future earnings and costs are not discounted to the present in this method. On the x-axis in Figure 2 for Product Class 1 and Figure 3 for Product Class 2 one can see the years from start of development, to testing and commercialization and a six year sales period of the product. The numbers on top of the squared blue boxes show the amount of products sold over the years. The pink line shows the cost associated with the development and testing and the income due to profit from product sales and maintenance contracts with the customer.

However, it is believed that a single prototype is not likely to capture all failure modes that are to be expected due to

- Design features which do not fulfill the design intent;
- Manufacturing variations and material properties' variation;
- Different customer operation modes and uncertainty;
- Varying environmental conditions that the product faces at customer sites.

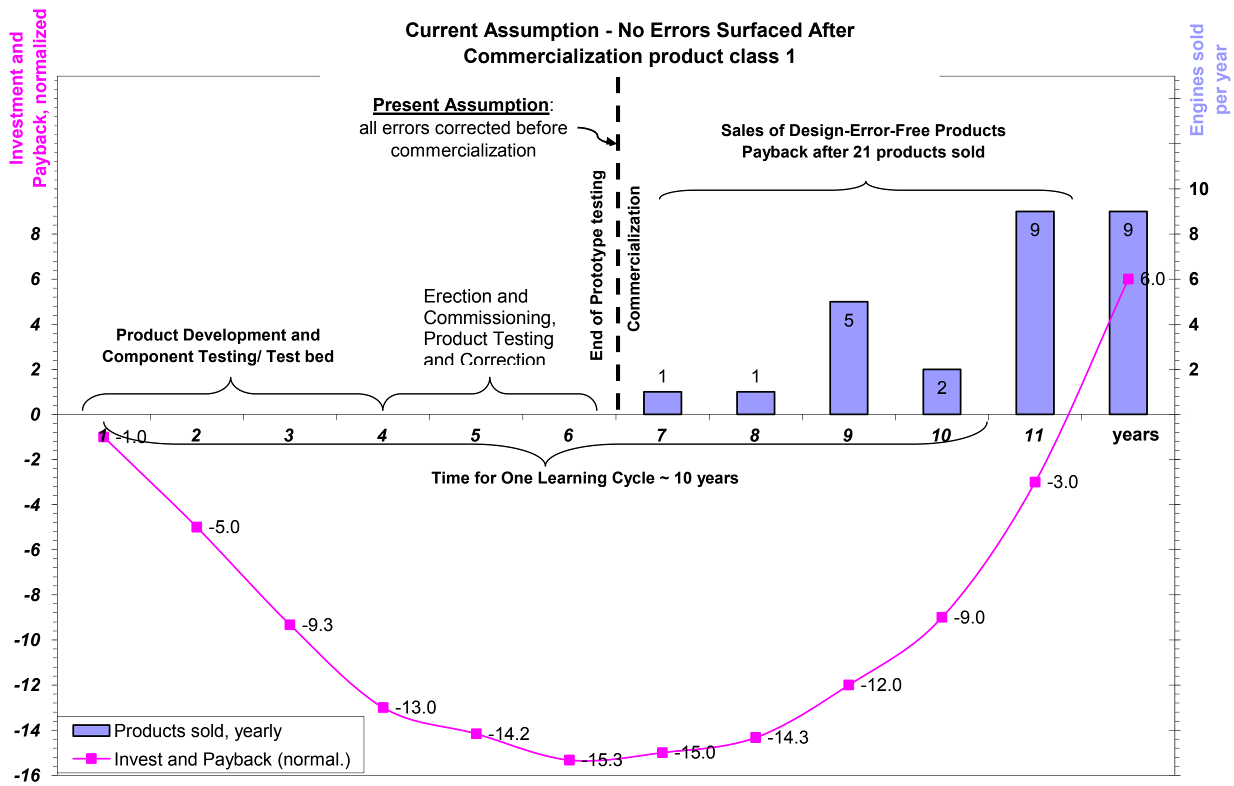


Figure 2: Payback Assumption of a Large Capital Product, Product Class 1

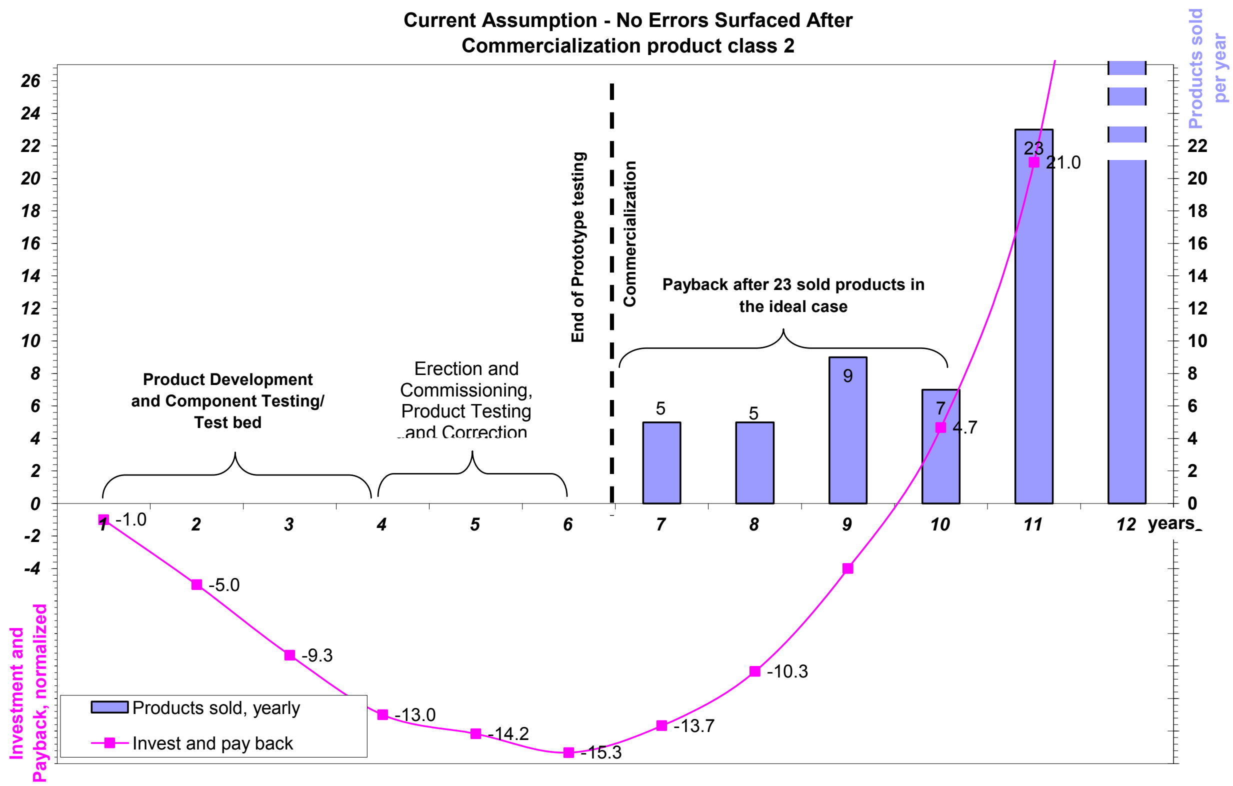


Figure 3: Payback Assumption of a Large Capital Product, Product Class 2

Non-conformance-cost due to failures in design after commercialization will multiply in sold products and thus strongly impact companies' bottom lines.

Hypothesis Two

The second hypothesis is that a generic process can be derived and a model can be built for LCPs, that

- adjusts to respective industries
- supports a company's policies on decision making on an economically optimal number of prototypes

- supports a company's decision making on length of testing
- takes into account non-conformance cost due to late product errors
- considers time-to-market impacts,

and thus enables the company to optimize its testing and validation strategy.

A representative development and implementation time line for LCP are reflected in Figure 2. For the example in this research, design time ranges around four years, including the time needed to build component test hardware. Two years are required to build and erect the first prototype, perform start up and loading tests, and conduct longer durability runs, while corrective actions are implemented and documented.

After a commercialization review all further products are built and sold to the finally released specification. The general assumption is that negligible non-conformance costs occur after commercialization. Future sales are assumed to incur no or negligible (ideally less than 5% of revenue) non-conformance cost in their life cycle. In this specific example and under these assumptions, which are highly dependent on sales volume, break-even would occur after 21 sold products, and four years after commercialization, as is illustrated in Figure 2, considering a typical margin for this industry of 14%

If failures occur after product introduction when more of these products have been sold, the effort to repair these products will have a huge monetary impact on the performance of a firm.

CHAPTER TWO: LITERATURE BASIS FOR THE PROBLEM

Introduction

The two most related literature streams for prototype-testing methods with respect to economical advantages are

- The field of new product development (NPD) including research and development (R&D).
- Strategic marketing, particularly product introduction strategies.

The NPD and R&D research is relevant because it gives insight into the front end of a product development project and may involve information about the quality of testing methods prior to product release. The focus on marketing and strategy related research is of interest because literature in this area may cover introduction strategies and their financial impact, as well as key success factors for a product launch.

Typical LCP and Product Selection for Illustration

Typical examples for products that have long learning cycles, a long product development time, and extremely high development costs are airplanes, ships, submarines, satellites or large land-based gas turbines for the power generation industry. All the above goods could benefit from an improved testing policy.

Due to lack of access to other industries' data (highly confidential) this thesis focuses on using data from operating gas turbines from the power generation industry, and intends to develop a generic model to improve decision making on late stage testing policies which is illustrated by a specific example from the power generation industry.

Power generation producers or Original Equipment Manufacturers (OEMs) are developing, selling and maintaining power plants that convert fossil, green or nuclear energy into electrical energy. Fossil energy is commonly defined as energy derived by processing coal, oil or gas; green or renewable energy encompasses wind, solar and hydro power among others. In the late 1960s, BBC Switzerland, GE in the USA and Mitsubishi in Japan built the first combined cycle power plants with a net efficiency of 43% (Lechner et al, 2002). Today's modern combined cycle power plant can reach an efficiency of 60% of primer energy by using a gas turbine cycle and steam cycle, also called bottoming cycle. The bottoming cycle is utilizing the hot exhaust gas from the gas turbine to produce steam for a Heat Recovery Steam Generator's (HRSG) steam cycle. A gas turbine operated without a steam turbine still has an efficiency of up to 40% and an aero-derivative gas turbine of about 45%. In some countries, even the bottoming cycle's exhaust gas energy is used for district heating, which leads to a utilization of primer energy of about 85%. The rising attractiveness of gas turbine plants as well as combined cycle power plants can be explained by the plant's high efficiency and lowest investment cost per kilo watt output, the discovery of large sources of natural gas, and that burning gas is also by far cleaner than the combustion of brown and mineral coal with respect to emissions such as NO_x, CO_2 and SO_2. Additionally, such a plant is more rapidly built compared to a coal power plant, its payback period is short, and it can be operated with very few operators. Also, if the coal is first gasified, even coal can be fired in a gas turbine, and the carbon may be captured for improved emissions, which has direct impact on climate change. In addition, a gas turbine plant is very flexible in accommodating peak, medium and base load electricity demands (Lechner et al., 2002). "The trend was also accelerated through a change in power plant customer structure – besides the classical utility companies, Independent Power Producers (IPP) financed and operated gas turbine and combined cycle power plants."(Lechner et al., 2002). Half of the new built plants today are gas turbines and combined cycle plants. It is prognosticated that 29% of power generation will be coming from natural gas by 2010 (Lechner et al., 2002).

The gas turbine is the most complex component in a combined cycle or power plant. Due to market demand on increased efficiency and power and a pressure to get to the market faster with better products, gas turbine technology arrived at a stage where only novel cutting-edge knowledge and methods could accomplish the targets of efficiency, life cycle cost, first-time cost, emissions and time to market. To reach these targets, it is critical that all components in a gas turbine are harmonized with each other. The high firing temperature and pressure ratio, the lowest possible cooling and leakage air and optimized turbine and compressor efficiencies are of utmost importance to meet process efficiency, emissions and life requirements of the gas turbine. The firing temperature, for example, in a gas turbine exceeds by several hundred degrees Celsius the melting point of the high temperature super alloys used in the hot gas path of a gas turbine. Therefore, highly advanced cooling and coating technology is developed to prevent the meltdown of these components in a turbine (Lechner et al., 2002). The main OEMs are Alstom, General Electric, Mitsubishi Heavy Industries, and Siemens (Gibson, Bhatnagar, 2006). These companies are each investing hundreds of millions of dollars into product development every year to increase efficiency, reduce emissions, increase flexibility, and reduce first time and life cycle cost for gas turbines alone. GE announced in a press release on October 23, 2007 that they were investing one billion in 2007 on a broad array of technologies, two of which were integrated coal gasification and fuel efficiency. This heavy investment into technology is necessary to stay competitive, but is only possible when the OEMs generate a sufficiently high return on their investments. One way to improve the return is to minimize non-conformance cost. Non-conformance costs are unanticipated expenses to mend failures after product release, such as repair and redesign cost, as well as payments to the customer for their incurred cost due to failure.

Small defects in a gas turbine can lead to a shutdown of the power plant with high cost to the operating company and to the OEM due to contractual obligations, called liquidated damages. Just the opening of a cover of a gas turbine to replace a malfunctioning part, for example, costs several hundred thousand dollars, excluding any repair work, hence reduction of defects before market release could save a large amount of money on the after market side.

The viability of the process will be illustrated via data from land-based gas turbines as an example for LCPs. However, all actual values of company data used will be normalized when appearing in this document.

New Product Development Challenges of LCP

Over 30 papers and books were examined in the field of new product development (NPD), and research and development (R&D) in view of testing and validation methods that might be specifically applicable for LCP.

The definition for new product development is described by Wheelwright and Clark in their book "revolutionizing product development", as using existing technologies to create a commercial product whereas R&D means focusing on basic research, the know-how and know-why of new materials and technologies which get into a commercial product later (Wheelwright, Clark, 1992). Other literature on new product development concentrate on the importance of understanding customer and market needs rather than on the testing and validation requirements of LCP (Belliveau, Griffin, Somermeyer, 2002; Hari, Kasser, Weiss, 2007). It is proposed to use tools such as a quality function deployment to understand customer needs and focus on the importance of customer perception of product value (Hari, Kasser, Weiss, 2007). Further literature suggests managing risks of malfunctioning parts by applying appropriate risk mitigation tools such as Failure Mode and Effect Analysis (FMEA) or Fault Tree Analysis (FTA), and applying proper project management methods (Miller, Swaddling, 2002; Sachon, Cornell, 2004). In comparison to the process in Figure 1 described earlier, the generic product development process depicted in Figure 4, has not even foreseen a real product testing and validation phase. Wheelwright merely mentioned "tested design" in the phase titled: "Development" without further itemization on the "how". Wheelwright implied that serial release occurs during the phase he described as "preproduction phase", but without detailing the steps (1992).

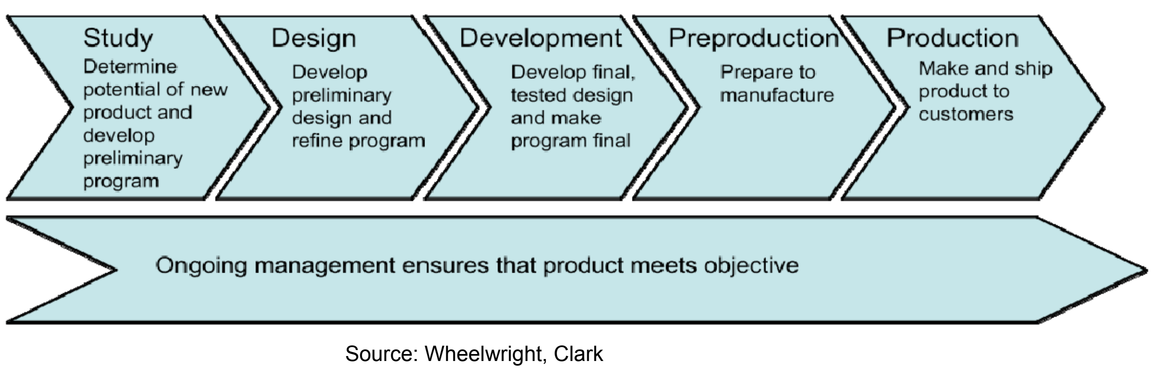


Figure 4: Generic New Product Development Process

Concurrent engineering approaches are suggested to improve quality of the product and establish its manufacturability early in the process. It also is said to increase development speed and reduce rework and cost (Componation, Utley, Armacost, 1999).

In general, new product development processes have three distinctive phases (Belliveau, Griffin, Somermeyer, 2002):

- fuzzy front end,
- new product development phase and
- commercialization phase

Product development is a known and predictable process and very similar from product to product (Cooper, Kleinschmidt, 1986). Cooper summarized that 70 – 85% of a development process is the same, typically including product planning, lab tests, preliminary design, large scale testing, manufacturing scale-up, support training, market launch preparation and commercialization. However, no papers or books concentrated further on final-product-testing or late stage testing concepts for LCP. Not any after market considerations, such as service processes and life cycle cost, which are of great importance for products with long learning cycles and long lifetimes have been discussed.

Although there is a huge variation in NPD process steps according to Barclay, which can range anywhere from four to 27 steps, he stated that the main contributors to product success were not the number of steps, but investment in people, their skills and competencies, working in teams with good leaders, and operating according to procedures and guidelines (2002). Barclay based this assessment on a NPD assessment tool and methodology that he developed and used to compare successful products with unsuccessful ones based on product development steps ranging from application of quality tools (e.g. QFD or FMEA) to team leadership and formal NPD procedures. The importance of testing methods or prototype testing was implied through the emphasis on the "importance on the progression from development stage to full production and the success of these steps that ultimately determines the attainment of a successful development stage."

Balachandra researched over 60 papers on success factors of new product development (1997). He found many success factors from different authors, some of which were even contradicting each other due to varying contexts, and different definitions of success. He categorized the major potential success factors into market, technology, environment, and organization. All researched papers agree that a strong market is of high importance for the new product; also a good market analysis is important, and an early market analysis is better than a late one. All success factors were categorizing the product as failed or not failed. The particular area of this research, which is the importance of an economically sound prototype testing method, was not addressed in any paper researched by that author. Cooper described how companies are utilizing the product development process and the impact that each process step has on the final product. He mentioned as an "important factor for project success" the pre-commercialization business analysis and trial production. He also focuses on test market and trial sell, which is applicable for mass products but not for the large capital products, which is the object of this research (1986).

None of the authors investigated or elaborated on how a company should handle testing of a newly developed LCP on a system level. The research papers on new product development that were found focused mainly on development of far less expensive mass market products, where due to their relatively low cost many products can be tested and their survival statistically evaluated in standard reliability analyses.

Marketing

The debate in the literature on market introduction focuses on:

- Strategies for appropriate market segmentation;
- Marketing strategies and their impact on segments in the market place;
- Product selection and selection of their distribution channels of mass market products (Urban, Hauser, 1993; Bowersox, Stank, Daugherty, 1999)

In general these authors emphasize the importance of a technical success ("consumer accepts the product"), next to a financial one, but do not embark into the area of how to improve technical success. Bowersox, however, implies specifically that failure rates of 50 to 67 percent of all U.S. product introductions are typical, caused by improper product development and product launch (1999). Both, Cooper and Bowersox brush on the importance of proper product development process execution, but do not touch on prototype testing methods in their research (1999; 1986). Bowersox explains that new product launch is "a part of the overall new product development process" and states that product launch has received only limited research so far. He, however, focuses then on supply chain and distribution for mass market products rather than on prototype testing and introduction methods.

Burgelman, in his article about technology ventures, distinguishes between technology and product development and acknowledges the difficulty of testing, especially when the product contains some new technologies (2007). However, the main focus of his article is on the importance of a methodic and systematic strategy making process for high technology ventures, rather then on exploring testing methods of high-tech LCP.

Messica focuses on the financial aspects of time-to-market of a product (2002). He stated that it is not always good to be as fast as possible to get to market. Although faster time-to-market is crucial, the “window of opportunity poses an additional factor” which might be different from the fastest time-to-market. He also point out that competition might lead to compression of the development period and increase the risk of a premature product launch. He analyzed time-to-market as a function of the expected return and compared development cost with opportunity cost for new products. He derived a mathematical model as a supporting tool to decide when to terminate a development project or when to continue. No specific research on the length or amount of prototype testing was conducted, but helpful insight into the modeling aspect of time-to-market was given, which is be part of this research.

Cooper was often quoted in the research literature for his extensive investigations on product success factors. Cooper investigated what key success factors contributed to a successful product by studying 203 launched products (1987). He found in his research that new products are generally high-risk endeavors. Key attributes of successful products are:

- Having a clear product advantage,
- Having a thorough understanding of customer needs,
- Having performed a thorough market research.

He also stated that proficient R&D management and technology activities are important. He then remarked that major failure reasons were, among others, product defects, and demanded better concept testing. He stated that success is determined partly through product design and product launch, and that product quality was a major determinant of profitability. He did not further explore prototype testing concepts but mentioned the importance of them.

Lynn stated in his research that clear goals and a functioning NPD process were critical success factors for product success without further detailing important NPD steps such as prototype testing (1999).

Ofek studied aspects of success of next-generation products (2003). Among other factors he clearly stated the reputation advantage. As an example, he claimed that new technology in a product often bears uncertainties for customers. He also stated that a firm with a highly valued current product is more likely to sell its new product, and the customer has a more favorable attitude towards the new product if it is an extension of the older one.

According to Ofek, "there have been attempts to quantify the value of reputations" (2003). Charles Fombrun, professor at Stern School of Business and author of many books to this topic, conducted, according to Sisk "a study of 35 companies in which he found a strong relationship between a company's market value and its reputation" (2001).

According to Groenland the relationship quotient (RQ) that Charles Fombrun developed to better define reputation, and Groenland further qualified through a study he conducted in the Netherlands, structures reputation into six areas, which are: emotional appeal, products and services, financial performance, vision and leadership, workplace environment, and social responsibility (2002).

One aspect that destroyed customer's trust into a company was, according to Groenland's study, "delivery of substandard products"(2002).

In summary it can be said that literature with focus on marketing and on marketing strategy is helpful for confirming the need for this particular area of research on testing of LCP, but does not provide further insight into the "how to" improve prototype testing.

Current Approaches for Validating LCP

There are different recommendations and rules for validation, depending on the type of LCP. Some examples below show the variance in government regulations to OEM's own procedures for certifying or approving of LCPs.

Aircraft Engine Certification

An aircraft engine type certification is achieved when a new designed engine type was tested and approved to airworthiness according to standards of Federal Aviation Regulation (FAR), Part 33 "Airworthiness Standards: Aircraft engines."

Official engine certification tests are to be conducted in accordance to FAR Part 33, Federal Aviation Administration (FAA) Order 8129.2A "production approval and surveillance procedures," and several other supporting regulations. The certification of air worthiness requires engine calibrations, endurance and operation tests, followed by a teardown inspection (specialization tests as vibration measurement, detonation, rotor integrity, rotor blade containment, icing and ingestions tests). The test authorization identifies the engine to be tested, the specific tests, including schedule of runs, test limits and the test equipment, instrumentation and facilities (rigs, cells) and the inspections to be conducted, covering engine and test equipment.

Design changes to an engine desire a retesting of engines with all recent design changes, including development, endurance, vibration and fatigue testing, depending on the kind of changes. For example minor design changes are defined as slight variations in clearances, reasonable increase in radius of fillets, increase in thickness, where the design permits it, change to equivalent or improved material, in minor parts, improvement in heat treatments, small changes in design in non-critical parts, and improvement in manufacturing. Major changes are defined as changes in compressor pressure ratio, or super charger gear ratio, increase in temperature limit or speed limit; change in material of highly stressed parts, rotating or non-rotating parts etc.

The FAR regulation specifies the components to be tested, and how. For instance, it describes that an engine has to survive an endurance test over 150 hours according to a test procedure, specified in the regulation of the FAR. The regulation however does not request more than one test engine of the same type to be tested for passing of the certificate. In fact, the regulation allows the use of separate engines of identical design and construction in so-called block tests for the vibration, calibration, detonation, endurance and operation tests. The regulation also allows that numerical analysis may substitute some of the tests. For instance blade containment for both rotor and turbine, and rotor unbalance tests and must be demonstrated. However, the test may be replaced by analysis based on rig testing, component testing or service experiences.

Space Shuttle Main Engine Testing –RS-68

According to Wood, the space shuttle's (RS-68) main engine was "developed and certified using eight new engines, and four rebuilt engines" with 183 different tests. Wood states that "five engines were run beyond the maximum mission duty cycle endurance factor limit and three engines were run to over three times the maximum flight duty cycle" (duty cycle requirements are to survive 8 Starts and 1200 seconds each) and 105% power to validate reuse potential (2002). Prior to the certification testing, the developing company was performing component qualification testing to supplement engine level testing and to decrease the risk of engine level testing. The company researched prior developments and claims that, based on their data, the main cost for the development of LCP came from so called fail-fix issues, meaning the final product is tested, and the upcoming errors are being fixed and the product gets retested. The different approach for testing RS-68 reduced the fail-fix issues from an average of 73% of total cost to much less according to Wood (2002)

Gas Turbine Testing –Approaches Across Industry

Fluck describes how Alstom is validating one large upgrade (2005). Validation starts with component validation through extensive rig testing, followed by testing of the entire system using two engines. Both engines were equipped with various sensors to measure engine behavior continuously. Fluck explains further, that after these tests the component's behavior in an engine was monitored over a time span of 10000 operating hours within a two year window. Fluck claims that Alstom's validation concept reflects "today's market expectations for proven equipment". Fluck does not explain in his paper why Alstom tested exactly two engines for this particular validation, nor explained the reason for the length of testing.

Variables and Policies for Testing

Policies in the Gas Turbine Arena

This research is concentrating on investigating variables that drive design errors after commercialization creating companies' non-conformance costs and policies that Gas Turbine producing companies have attributed to direct development and testing. The previous chapter, section "Current approaches for validating LCP" describes in testing methods for the major OEMs. Table 1 summarizes component testing facilities for gas turbines from major OEMs.

Table 1: Major OEM Testing Facilities for Advanced Gas Turbines.

Manufacturer	Location	Remarks
Alstom	Birr, Switzerland	In-house test facility with generator, connected to grid
Siemens Energy	Berlin, Germany	In-house test facility with water brake
	Cottam UK	Demonstration site
	Unit 4 of E.ON's Irsching station Germany	H-class* demonstration site
Mitsubishi Heavy Industries	Takasage, Japan	In-house verification plant with generator
General Electrics	Baglan Bay Power Station, Uk	H-class* validation site
		In-house engine test facility

Source: Bechtel Power Corp

* arbitrarily defined technology level used by this industry

One can see from Table 1 that three major testing methods are being used:

- Testing of the product in an engine test facility not connected to the grid (the generated power will be terminated with a waterbrake).
- Testing of the product in a validation plant that is connected to the grid, owned by the OEM, and its components are equipped with specific measurement devices for validation purpose. Since most of these sensors for diagnostics and monitoring are life limiting for components, the product will be completely torn down and rebuilt with new parts for the customer after validation is completed. The tested components will be investigated with non-destructive and destructive evaluation methods.
- Testing of the product in a validation plant that is connected to the grid and is owned by the OEM. After testing, the product will be handed over as is to the customer with an appropriate service plan.

The major advantage of a test facility not connected to the grid is that it enables off-frequency tests, which are important to better understand the operation, frequency response and stall margins of the compressor, and reveal the true operational range of the engine. The build up of such a test facility takes about one year.

The disadvantage is the short testing run time of only a couple of hours at a time under full speed, and full load conditions.

The advantage of an OEM owned validation plant, connected to the grid that is torn down after testing is that the produced energy can be fed into the grid and allows therefore long endurance tests, which are important for understanding wear behavior and other effects that occur only after a certain run time. However, such a plant is restricted to the operational range of the electricity grid and cannot be operated in off-frequency conditions. The insurance may demand a premium price on the prototype plant, depending on the “unprovenness” of the new technologies applied. The build up of such a prototype validation plant takes about two years. The components are typically heavily instrumented for online monitoring of the engine. Due to the holes that host the sensors and other diagnostic instruments, the component’s life time is often drastically reduced and must be replaced prior to handing over to the customer, requiring tearing the plant down and rebuilding with new components. This takes typically another year. Relevant components of this test engine are non-destructive evaluated (NDE) and cut up for metallurgical investigation.

Test engines at customer sites, which are operated by the OEM and then handed over to the customer are common practice. There is no special set up phase, and no special measurement equipment that limits life on components. However, non-life limiting equipment (infrared cameras attached to existing ports, spray on sensors that don’t require rotor drilling) is used for a better insight in component behavior.

Cost of LCP and Test Facilities

The following paragraph gives some reference values for a better understanding of the financial dimensions a company has to cope with, when deciding to invest into testing facilities or a customer into a new LCP.

For a combined cycle power plant, the typically required capital investment, usually measured as cost per KW (kilowatt), was $ 800/KW in 1990 and dropped to $450 / KW in 1998 mainly due to the liberalization of the local energy markets and overcapacity (Huppmann, Groetschel, Mueller, 1998). A typical power plant in simple cycle operation, with one gas turbine on one power train costs about $350/KW (Rafai, 2003). That price has dropped due to overcapacity and deregulation and is more likely to be around $250/KW today. A test facility with only one gas turbine connected to a waterbrake would cost about $200/KW.

According to this cost structure, the cost for a 150 MW prototype test facility with no connection to the grid including erection and commissioning, water brake, turbine hardware and measurement equipment ranges around 30 to 40 Million. A 150 MW prototype plant connected to the grid using the value of $250/KW cost about $45 Million (including a 20% cost increase for "new developed hardware" cost).

Testing operation modes that can be performed in a validation plant are peak mode (many starts of limited time, for example several hours operating time), base load mode (long operation time over several 1000 hours, without turning the engine off) and intermediate load mode, which is a mix of medium long operation times with occasional shut downs and starts of the turbine. A testing facility can perform many short term tests (peak) over a couple of hours.

Personnel and maintenance cost to run the test facility / validation plant are estimated to be about $5 million per year, including insurance premium. Insurance premium for a testing facility or validation plant is estimated to be about 1 to 2% of the erection, procurement and commissioning cost.

Fuel prices are volatile. For this study a value of $17/MWh is assumed, which equals $5 per 1 Million Btu (with a conversion factor of 1 KWh to every 3412 Btu (British thermal units)). $5 per one million Btu is the average short term energy outlook gas price according to Henry Hub's natural gas prices (source: Short Term Energy Outlook, August 2010, Reuters News Service). For 1000 operating hours a 150 MW plant operator has to pay \$ 2.5 million for fuel based on \$17/MWh gas price.

The total fuel cost of a test facility is significantly lower than of a validation plant due to the lower operating time.

In total the cost for a 150 MW validation plant that is erected, commissioned and operated over a two year test period with 8000 hour testing time would be \$70 Million. The cost for a 150 MW test facility with 1000 hour test time in a two year time frame would be \$49 Million.

Approach to Derive the Generic Dynamic Model

To illustrate how to derive the generic model to investigate the number of prototypes needed for testing to optimize profitability, operation and market data will be collected from two new product classes. To understand the severity of the failures, non-conformance costs of each failure mode will be assessed from available data and the financial impact calculated. Defects can develop from design errors or from uncertainties. Uncertainties can derive from manufacturing uncertainties in dimensions, tolerances, and material deviations, environmental influences and operation differences.

With these data, an industry specific dynamic model will be developed that takes into account:

- the impact of the product development process,
- the time delay between design release and failure occurrence,
- the time to market introduction,

- the reliability improvement methods
- other potentially influencing factors.

Sanders describes in detail how to build a dynamic model (1980). He proposes to focus first on model conceptualizing, a most basic step necessary to help ensure that the later model has the right focus and time horizon according to the problem statement, and that the researcher has the right understanding "mental model" of the operation in the real world. The subsequent steps he proposed after conceptualization are "formulation" of the part of reality of interest, the testing stage that helps to reveal weaknesses of the model and to determine its limits and quality, and lastly, the implementation phase, to transfer the insights gained to a user community for application.

The Need to Study the Policies

Typically, industries have a prototype-testing and market introduction policy of some sort which will be integrated into their new product development process. These policies are rather vague in requesting specific tests and are driven by requirements of insurance companies. Companies' intent is to develop useful and safe products for their customers while maximizing their income. Product errors are costly in several dimensions: It creates non-conformance cost (NCC), it may dissuade future customers from buying the product, and insurance rates may increase.

Table 2 describes typical NCC ranges for companies and their evaluation.

Table 2: Non-Conformance-Cost Ranges (Kiesow, 2008)

Range of NCCs [in percent sales]	Evaluation
25%-40%	Bad
15% - 24%	Below standard
5% - 10%	Typical
1%	Ideal

Insurance companies with their experience across the industry typically have specific definitions on what constitutes mature technology. If a technology is considered as “new”, the insurance rate for the plant deploying this technology will be quite high. Insurance cost for test products are typically about 1% of the plant value, ranging from 500k to 1000K USD. A conservative insurance company considers a technology as “proven” when the technology has been deployed in one to seven products (depending on the part), for at least 8000 hours, which approximates on average about one to two years of operating time, including maintenance intervals.

A company’s policy for example may have the following requirements on testing of “new” components in a product class:

- Testing of each “new” component in appropriate test rigs.
- Extensive testing of the components integrated in a product at a company owned test facility

- One prototype test product testing at a customer site.

It is believed that it is sufficient to test with one prototype. Further testing is considered to be too expensive for too little gain, and would potentially delay market introduction. However, this research will investigate product failures and costs after product serial release (non-conformance costs) across the product fleet and will develop a model that helps derive policies for an economically better late stage prototype-testing concept. The study will produce a generic process to build a specific dynamic model for any company to aid them in the decision making on late stage testing. It gives a company a strategy for assessing its testing policies.

The Need to Use Systems Dynamics

System Dynamics "Pros"

The nature of this study requires looking at a system which includes aspects such as time delays, soft factors that are typical gained through expert knowledge, feedback loops and other causal dependencies that are not closely linked in time and space but their mathematical relationships need to be understood and mapped. System Dynamics (SD) is seen as a better fit to this type of real world problem than pure operations research (OR) (Forrester, 1994). Forrester points out that most of the OR tools are of linear nature, such as linear programming, queuing theory, regression analysis, and scheduling algorithms, and static and cannot depict the dynamic and non-linear processes of the real world as it is required for this research. SD's main advantage lies in its unique capability to deal with non-linearity, high complexity and feedback loops.

SD has "a long tradition of using formal models to study research and development processes", according to Repenning (2000), and especially problems with respect to resource allocation among projects. According to Sterman, SD can improve the understanding of the relationship of companies' internal policies and processes and how they relate with their "customers, competitors, and suppliers, and then use that understanding to design high leverage policies for success" (2000). The numerical simulation technique of system dynamics provides an opportunity to create a holistic view of a system. It supports the representation of time delays and other non-linear relationships between variables that are mathematically difficult to calculate and allow building empirical data tables into the model.

Jay Forrester founded the field of systems dynamics (SD) in 1956. He described SD as a method that depicts how structure and information flow determine the behavior of a system and how feedback loops can change this system over time. SD provides insight on why a system behaves a certain way and helps the observer to redesign the structure to improve behavior. The use of computer simulation combined with the underlying theory makes this field very applicable for a real world problem. It allows modeling a problem and simulating the outcome which can be used to derive more desirable policies.

Frank Severance declared that systems are "a set of cause-effect relationships that can be decomposed into subsystems and applied over a restricted application domain."(2001).

Repenning described applying system dynamics to a real world problem, ideally combining the disciplines of operations research and economics and their use of formal mathematical models, along with the insight of psychology, sociology and anthropology (2003). In his view SD is a valid method to those "whose primary interest lies in understanding real world phenomena", (Repenning, 2003).

Sterman explained system dynamics as “a perspective set of conceptual tools that enable us to understand the structure and dynamics of complex systems” (2000). SD is “grounded in the theory of nonlinear dynamics and feedback control developed in mathematics, physics and engineering” (Sterman, 2000). SD is used as an approach to build realistic models using computer simulation techniques to simulate complex systems and to design more effective policies for organizations (Sterman, 2000). The key of systems dynamics is discovering and modeling the feedback processes into the system and aligning them with stock (levels) and flow (rates that cause the levels to change) structures. SD also enables one to model delays of activities, and other nonlinear behavior in the system, which is mathematically difficult, if not impossible, to solve other than through a numerical approach (Sterman, 2000).

Forrester compares SD, systems thinking and soft operations research (Soft OR) with respect to understanding and improving a system (1994). While Soft OR and systems thinking help in understanding the system, they both do not require model creation and simulation, and therefore rely on intuition for evaluation of the results.

SD is a unique field that enables one to map hard mathematical logic with non-rational hard to quantify elements of a problem (e.g. in the form of empirical data tables), such as reputation, word-of-mouth or time effects such as delay, that quickly increase the complexity of the underlying mathematical equations (Bossel, 1992; Repenning 2004). System dynamics models are systems of nonlinear ordinary differential equations. However, their analytic solutions often cannot easily be found and the behavior of the models must therefore be computed numerically (Sterman, 2000). For further research on computational solutions on nonlinear differential equations Sterman pointed to Atkinson, Burden and Maron, who performed research in the theoretical field of numerical integration for differential equations (Atkinson 1985; Burden, Faires 1989; Maron 1987).

As stated in the literature review above, there are a number of reasons why system dynamics would be an appropriate method for this research problem. However, one has to be aware of some issues as described below that have to be addressed when using this method.

System Dynamics "Cons"

The SD method requires one to describe the system, but this step is not straightforward. According to Forrester, it is generally very difficult to convert a real-life model into a simulation model, especially when one is new to the concept (1994). The process of "taking various bits of information about the real world and turning them into a coherent and unifying theory can be most challenging" (Forrester, 1994). Another difficulty being faced is of a mathematical nature. Solving differential equations of a non-linear nature in closed form can become impossible. To circumvent this problem, system dynamics solves its differential equations numerically through simulation. This could lead to certain integration errors, due to finite time steps and a resulting approximation. When using higher order models, they are to be carefully assessed whether they are compatible e.g. with discontinuous events in the model, as it is the case in this research.

Typically a first order model (Euler) is used initially, when large errors in variables are expected, for example when using historical data.

Summary

This literature research of the areas of marketing and new product development shows that neither marketing related nor product development related research focuses, addresses or investigates the final product validation and market introduction strategy of original equipment manufacturers that produce goods with

- long development and validation times (six to ten years),
- very high development and validation costs, and
- low production rate.

It does not sufficiently address the sensitive area where the product is a success in the sense that it fulfills customer needs, but has not quite experienced all or at least most of the possible defects in the testing phase since typically only one prototype product is used for testing. The literature does address the need for investigation in this area, but does not give solutions on how to better handle this situation by also considering the economic impact.

CHAPTER THREE: CHALLENGES AND GENERIC PROCESS

Challenges of Studying this Problem

Data Availability, Accessibility and Reliance on Past Data

Product data have been retrieved from products sold over several years (these data are also called field data). Each product has had different operating conditions, running times (time the product is operating), starts (count of how often the product has been started), and service intervals (planned and unplanned). Time to failures are ungrouped (recorded at the time they occurred) as well as grouped (failure time estimated, and recorded during a scheduled service interval). During the life time of each product, manufacturing modifications of parts have been implemented; also failures may have gone unreported or been misinterpreted. Due to the fact that time of occurrence of failures happened up to 15 years back in time, it is difficult to verify the data collected. This should not influence the conceptual model that can be used by any industry, but may have an impact on the optimization results of the specific example used in this research.

Proof of Concept

This research develops a generic model (relevant for all OEMs with similar development times, cost and complexity) for determining the policy of optimum prototype testing using data from the gas turbine industry to illustrate the steps necessary.

Part of the model building blocks will be non-tangibles, such as reputation or the loss of it, and its impact on revenue. The values used for this type of variable will be based on expert opinion. The validation of the model is based on past data. The time to validate the model on current data will take decades, and is therefore not feasible for this research. However, the model building will be done stepwise and the simulation results of each new step will be compared to expert opinion (engineers in related industry with 10 to 30 years experience) and real world data as part of the validation process.

The Generic Process

Approach to Building the Systems Dynamics Model

Because a purely theoretical analysis is very difficult, if not impossible, an empirical analysis will be relied on to derive actual impact and implications. This research will utilize the system dynamics method to simulate the dependencies and causal relationships of different testing approaches on product errors, non-conforming costs, and their impact on business decisions (policies) of the manufacturer (OEM).

The process to derive a dynamic model that can be applied by any OEM is iterative. First, a specific model will be developed and verified using data from a specific industry. A general process will then be derived from this specific example. Following steps have been performed:

1. Collecting failure data, rates and related cost data on different product classes of a large capital product to understand error occurrence during prototype testing and in subsequently sold products after design freeze.
2. Analyzing past failure data and deriving reference modes (e.g. bottom line impact over years) with respect to late stage testing methods from data or hypothesized in case no historical data can be retrieved.
3. Determining which elements should be part of the quantitative dynamic model and why, using expert knowledge during the mapping process.

4. Deriving cause and effect loops with the main variables that define the reference graphs
5. Linking variables logically and mathematically through sets of equations into a dynamic model that can be used to predict the effect of varying testing approaches.
6. Improving the model gradually (scenarios) by introducing more variables where necessary e.g.
    - Introducing different market introduction scenarios
    - Adding economic effects from delaying market introduction.
7. Comparing the model's simulation results after each improvement with the reference mode and discuss with industry experts for verification and validation.
8. Optimizing the model for profit by varying the input variables with respect to testing
    - Testing length.
    - Number of test products.
9. Assessing current testing policies with the optimized variables.

This research explains the steps necessary to assess the need for an improved testing policy (Hypothesis One), and illustrates a generic process to develop a company specific dynamic model that can be used by any OEM to optimize their testing policy. The main steps are depicted in the graph below.

<table>
<tr>
<td>Hypothesis One – Assessing company's need for a better testing policy

1. Collecting failure and cost data

2. Analyzing failure and cost data to derive failure rates and non-conformance cost after product release</td>
<td rowspan="1">Hypothesis Two – Developing the dynamic model

3. Determining variables for dynamic model

4. Deriving cause and effect loops

5. Linking variables through meaningful mathematical equations</td>
<td rowspan="2">Hypothesis Two – Assessing the policy with optimization results

8. Optimizing the model for testing length and number of test objects

9. Assessing company's current testing policy with optimized variables</td>
</tr>
<tr>
<td colspan="2">6. Improving the model gradually by introducing different scenarios

7. Comparing model results with expected results and reference modes</td>
</tr>
</table>

Figure 5: Major Process Steps of Hypothesis One and Two

Verifying and Validating the Model

For verification, the model was simulated for various scenarios, and a sensitivity study was conducted. The underlying equations with which the parameters are linked were numerically solved through simulation. The model output was subsequently compared to reference modes from past data and discussed with subject matter experts, to verify the plausibility of the model.

For the sensitivity study, the parameters were varied depicting a variety of scenarios with some small deviations, which showed either the magnitude of the impact of a single relatively small change in the inputs or the opposite, evaluating the robustness of the outcome for a variety of changes in the assumptions.

After verification of the model, the optimization was performed; an appropriate specific policy was proposed and verified with industry experts in the field.

Studying the Testing Policies

Policies are definite courses of actions for the sake of expediency (Webster's college dictionary). Companies define policies to advocate important rules to their employees to improve and to stream line their behavior. Typically federal laws and governmental regulations are the backbone of company policies, as well as economic reasons, as it is the case for example for travel policies. In the realm of validation and testing, expert experience - own or industry – also regulates these policies.

The target of this research is to challenge OEM's current testing policies and derive a model that supports the assessment of a company's current testing policy with respect to quantity of prototype products and testing duration that also takes into account long term impacts on a company's bottom line, including loss or gain of reputation.

The contribution of this research is:

A failure mode analysis method

A specific dynamic model of a testing environment

Evaluation of testing policies and the capability to manipulate testing variables in the model to optimize testing policy

A generic process to build such a dynamic model for any OEM allowing to evaluate and improve their testing policy

CHAPTER FOUR: BUILDING THE CONCEPTUAL MODEL

The goal of this research is to define a process and build a realistic, sufficiently sophisticated and complex dynamic model that supports a company's decision making policy on testing of large capital products.

Because a purely theoretical analysis does not provide a clear conclusion, an empirical analysis will be relied on to derive actual impact and implications. This research utilizes the system dynamics methodology to simulate impact of different testing approaches on product errors and non-conformance costs to support the evaluation of manufacturers' (OEM) testing policies.

The process to define a dynamic model is iterative. The concept to build a model is described in the following paragraphs. The verification of the conceptual model with actual data can be seen in Chapter Five: Building the Simulation Model and Chapter Six, Subchapter "Conclusions for the Specific Model."

Collecting and Analyzing Failure Modes

Before starting to collect data, one has to determine which data are needed for the model.

These are

- Failure related data, such as failure modes and failure times (mean time to failure)
- Cost related data, such as product development cost, testing and operation cost, failure costs, total product cost and
- Income related data, such as profit margin, revenues, and number of products sold.

Those data should be collected for at least two product classes of a large capital product. These data are the basis to understand and later model error distributions expected after prototype testing and design freeze in subsequently sold products.

The first step, and a most difficult and time-consuming one, is to collect these data. Typically there is a large time lag of many years between the releases of large capital product classes and data for more than one product class may therefore be difficult to come by. At least failure data of two product classes should be used for the analysis to get an understanding of the product class failure and cost variation. Failures appearing for the same part with the same failure mode, but in different products belong to the same failure class. Failure times of all failures that occurred in a product after serial release have to be collected and documented per failure class, so are the cost to correct the failure.

Data Analysis

Failure distribution:

During data analysis, failure distributions are generated for each failure class using reliability theory for time dependent failure models. Each failure class of each product class should be assigned a distribution, such as Weibull, or Lognormal and tested for is appropriateness using standard tests. Then one averages the failure probabilities of all failure classes per each time intervals, for example every 500 hours. The target is to derive one "virtual" failure distribution that captures the average probability of failures over time to be used in the simulation model.

Chapter Five, "Reliability Testing Method" describes how this process was performed for a specific product.

Cost data:

The summarized cost of all failure classes in a product class needs to be calculated and will also be used in the simulation model (non-conformance cost). This is not to be confused with the summarized cost of all failures in a product class. In the former case, the cost of each failure class will be added together, independent of the frequency of failures in each failure class.

All product development, product capital (hardware and assembly) and product incommissioning and operation (testing) cost need to be calculated.

Income data:

All income, which is a function of product sales need to be estimated or measured so that it can be used in the model.

The independent variables (which are changing factors in the simulation) and controlled variables (parameters which are kept constant during the simulation) and the dependent variables need to be determined. Parameters for this model are those variables that are anticipated to not change significantly during the simulation time frame (e.g. profit margin), or determined to be kept constant to better compare simulation results (such as gas prices, or inflation).

Variables (independent, dependent or controlled) selected during this first step may change or be added during the model refining process from building the simplest model to the final one, if it increases the quality of model results.

Independent variables are those that the company has the choice to determine in its testing policy. In the specific example it was duration of testing and number of test products.

The dependent variable that is used to compare different testing approaches is profit. Other dependent variables as intermediates to calculate profit need to be determined as well (such as all cost items, income, time to market, non-conformance cost)

Creation of the (Simple) Baseline Model

All variables determined during the former phases need to be added to the model. A causal diagram (tree diagram) is a good way to visualize cause and effect relationships. A generic tree diagram is shown in Table 3.

All variables (controlled, independent and dependent) must be linked to each other according to their causal relationships using mathematical equations or relationships diagrams.

For example to calculate profit from the yearly income and cost in the model through simulation, the following Equation 3 below was used.

$$\text{Profit} = \int_0^{t_1} (\text{Income} - \text{Total Cost})\,dt \qquad \text{Equation 3}$$

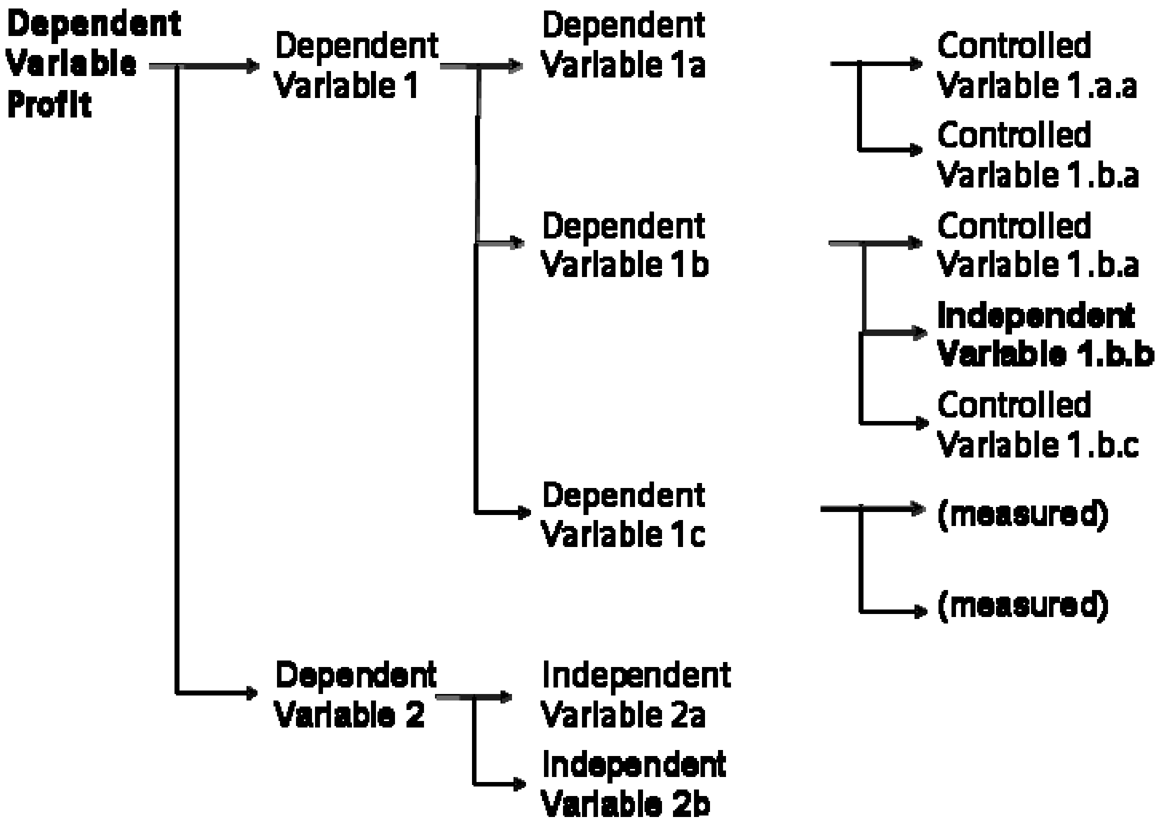
The value of the dependent variable "probability to detect failure" depends on testing length and number of prototype products. The underlying probability distribution was generated by analyzing failure classes and time to failures, as described in the Chapter Four, Subchapter "Data Analysis."

Subsequently all variables must be linked according to their causal relationships to each other. The relationships and the underlying equations are product and process specific.

An example for equations used in the specific example to link variables according to their causal relationships can be seen in the equations for the specific model in Chapter Five.

The decision on what to model in detail, and what estimates to use are decided by the modeler and her team of experts (three to five engineers with 10 to 30 years experience in related industry). The decision will depend on the difficulty to get more details, and the impact of this variable on the dependent variable (Profit). This decision making process is indicated in Table 3 below with the word “measured” in parenthesis. For example fuel cost can be estimated as two million dollars for a one year test period for one product and be modeled as a controlled variable, or it can be a dependent variable, a function based on gas price, product size, and operating hours. Development costs for example can be estimated as ten million dollars per year, or they can be calculated as a function of engineering time, number of resources, and cost of resources per hour and so on. The former (fuel cost) is dependent on testing length, one of the independent variables, and therefore important enough to show the dependence of fuel cost on testing length. The latter (development costs) is independent of different testing policies, and therefore is at a lower level of detail.

Table 3: Generic Causal Relationship Diagram



An example for a baseline model can be seen in Chapter Five, Figure 28.

For verification of the baseline model, data from either product class can be used (but not the averaged data over all product classes). The simulation result on profit over the years (profit curve) should match the measured or estimated results. The measured (or estimated) profit curve is called the "reference model" to which the simulated profit curve is compared. If the model is correct, these curves should be the same, and thus verify the baseline model.

Optimization of Testing Approach

Once the model is developed and verified, it should be simulated and optimized for profit by varying the independent parameters. Optimization is often part of a software package. In the specific example it was done using Vensim software.

The difference from baseline (original testing approach) to the optimized test settings is expressed in the gain in profit. A model used to optimize the test settings for the specific example is shown in Chapter Five, Figure 31. The model and the results should be opened up for a critical review by industry experts. In the specific example in Chapter Five, the model was revised after a discussion with industry experts to incorporate in the model the cost of exchanging all "risky" parts (that had failed in some products, but not yet in all products), during a planned downtime.

Modeling of Feedback Loops

Feedback loops are used to capture reactions from events as they feed back to the event and either reinforce or balance this event. As an example, a single loop is described below and shown in Figure 6. The profit-error loop has five entities until it circles back to it starting point. This loop is called a reinforcing loop, because it reinforces a certain behavior as is explained in the following: If the probability of detecting errors during testing is low, then there will be more product errors in the field (this inverse correlation is symbolized with a little minus sign next to the error). More product errors in the field will lead to higher non-conformance cost (the direct correlation is symbolized with a plus sign). Note that a plus sign does not mean more of something, but merely indicates that if the predecessor is <u>increasing</u>, the successor is <u>increasing</u>, if the predecessor is decreasing, the successor is decreasing, too. The higher the non-conformance cost, the lower is the profit hence a minus sign. Note that a minus sign does not mean less of something, but merely indicates that if the predecessor is decreasing, the successor is increasing or the other way around. The lower the profit the lower will be funding that is available for future testing. Lesser testing will lead to even less errors detected during testing (plus sign).

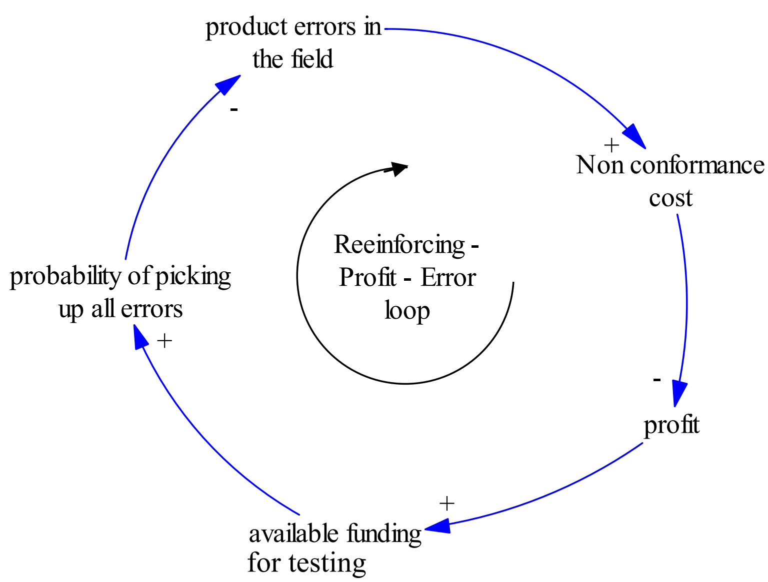


Figure 6: Reinforcing or "Positive" Causal Loop

The selection of the feedback loops that are important to the result of the model is different from industry to industry and perhaps even from company to company of the same industry. After each modeled feedback loop, the model should be simulated and optimized for profit. The results and the causal links that were added to the model should be discussed for plausibility with a team of experts.

Three feedback loops have been modeled in the specific example and can be seen in Chapter Five, Scenario 2 - Integration of Feedback Loops.

Confidence Bounds and Sensitivity Study

Confidence bounds need to be determined and a sensitivity study should be conducted to better understand the limits of the model results and the degree of influence of the variables on the dependent variable profit.

Confidence Bounds:

The constants (controlled variables) used in the final model should be varied by a small amount to determine the confidence bounds with respect to profit. A Monte Carlo simulation technique can be used, which automatically simulates the changes of the constants at random according to a selected distribution. If the 95% confidence bound on profit using the optimized testing settings does not overlap with the confidence bound using the base line test settings, then the difference in profit from the two test settings is statistically significant.

Sensitivity Analysis:

The sensitivity analysis is conducted by changing one variable at a time, conducting the simulation and documenting the difference in profit for each run. The difference can be normalized to change of profit in percent and a Pareto chart can be established to visualize the largest to the smallest contributors to a change in profit. Chapter Six, Conclusions for the Specific Model shows a sensitivity study conducted for this specific example.

A simulation model and its underlying mathematical equation were developed following the concept described in this chapter. The model was simulated using data from an LCP to which this researcher had access. The specific simulation model and its results are described in Chapter

Five: “Building the Simulation Model” and Subchapter Six: “Conclusions for the Specific Model”.

CHAPTER FIVE: BUILDING THE SIMULATION MODEL

Approach to Test Hypothesis One

Analysis of Product Class 1 and 2

Hypothesis one: For original equipment manufacturers (OEMs), which are defined in this context as suppliers who design and manufacture LCPs, it is not economically appropriate to build only one prototype product, test it, correct issues as they arise during testing (also called reliability growth testing), and then release the design for serial manufacturing.

Historical failure times and correlated cost data from two product classes, which fit the definition of a LCP have been collected, for confidentiality reasons named here Product Classes 1, and 2. All data used in this research have been normalized to honor the confidentiality of the company. All cost numbers are normalized with the cost for one product.

At first the data analyses were done by using calendar data to failure occurrence. For example, product A was incommissioned on April 1st, failure A occurred on August 1st. Time to failure was 122 days. However, due to the different product operating modes, those calendar dates were not useful. The product may not have operated for five full months. There may have been a planned down time, and potentially, an unscheduled one as well, reducing the true operating time from 122 days to perhaps 90 days. A much better approach was to count true operating hours to failure which was considered as being adequate for this type of product. For this research either exact failure times were collected, where available, for example when the failure caused the product to stop operating, or the failure time was used, when the failure was found during a scheduled or unscheduled product service interval.

Two different methods were used to provide evidence of the validity of Hypothesis one:

- Reliability testing method
- Visual analysis.

Reliability Testing Method

Typical reliability methods that would be usually employed are severely constrained for LCPs by the time it would take to complete enough tests to failure, and extremely high testing costs. For example, fuel costs alone are estimated to be between 100k and 250k per month, other costs such as hardware cost, personnel cost and so on adds to the soaring testing costs in addition to the base cost of the unit itself. The recommended testing time for the gas turbine industry is between 8000 hours to up to 16000 hours to validate a component or system. That is approximating two to two and a half calendar years of testing time for 8000 hours of testing, because one has to plan for operation time interruptions for maintenance and parts assessments. Long design and manufacturing times of critical parts limit the testing time of redesigned parts further. These issues restrict the overall achievement of statistically significant sample sizes which could produce statistically significant results. However, a company must make a decision to proceed to the next phase, which is serial release of a product. Therefore it must derive as much information as possible from the small set of sample data. The following paragraph describes a method to derive information from a small data set.

Standard Reliability Testing

A life time of a part or product can be described in three distinctive phases, the early-failure-period, also called burn-in-time or infant-mortality, the useful-life-period with constant failure rates, where some failures occur at random time intervals, and the final period with increasing failure rates, called the wear-out-period or aging. This characteristic curve is also known as bathtub curve (Lewis, 1994; Rausand, 2004). A picture of a bathtub curve can be seen in Figure 7 with lambda (λ) being the failure rate at a time or start (t)

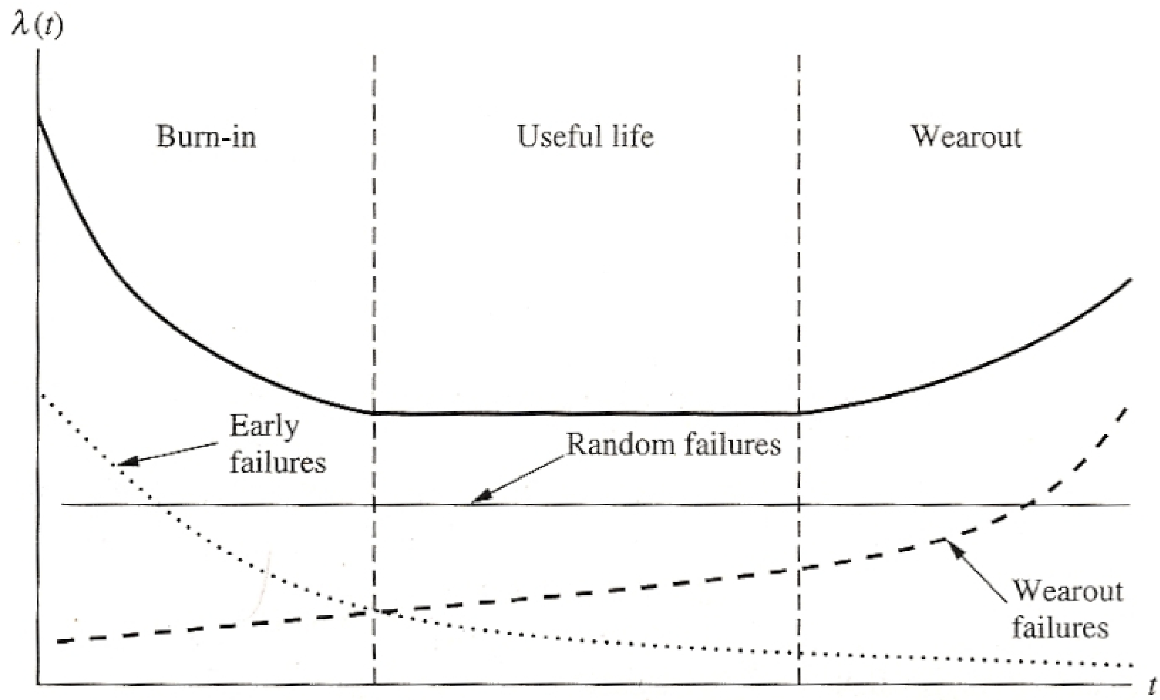


Figure 7: Bathtub Curve (Source: Charles Ebeling, 1997)

To use any parametric statistical method, one has to select an adequate life distribution model for the components to be tested. The Weibull distribution was chosen, because it is a good fit to the products' failure data. The first approach was to statistically show that the failures that were occurring after product introduction are statistically different from failures occurring before product introduction. This turned out to be an infeasible approach mainly because of the lack of a significant sample size. A proper test could have been to test the population proportion of failure categories of the first product equal to one (the total proportion of failure categories in all the other subsequent sold products). Company policy that requires one prototype test, expects that the majority of the failure categories occur in the first product (prototype) before occurring in subsequent products as they are serially introduced. A sufficiently large sample size for using the standard normal z statistic would be 30 or more samples. Small sample tests are also available, using the binomial distribution, but in this research, only 2 samples were available (2 product classes), which does not even provide enough samples to perform a non-parametric test, such as the binomial (Mendenhall, 1995).

Another way of showing that not all failures did occur in the first product is to create Weibull plots, and show estimated failure probabilities. If the probability of a failure within a given test time is low, say less then 40 %, it would mean that one needs more products tested to have a higher likelihood to experience the given failures. That would be another way of saying that not all errors occur in product 1, which would support hypothesis one.

The following paragraph gives an outline of the Weibull analyses and how this researcher has interpreted the results of the failures in Product Class 1 and 2.

Weibull Probability Density Function PDF f(t):

$$f(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}} \qquad \textit{Equation 4}$$

The shape of the Weibull plot indicates when the majority of failures are presumably occurring, according to the three phases of the bath tub curve described earlier. The total area under this curve equals 100%. t stands for operation time or starts. The area under the curve from left to a time t (or to a number of starts t) represents the probability that the product will have failed by this time.

θ is the scale parameter, which is indicating the measurement aspect of failure occurrence (here: time or starts), and is influencing the spread of the curve of the probability density function (PDF). When t equals theta, 63.2 percent of all failures will have occurred regardless of the shape of the curve. Theta is therefore also called the *characteristic life*. β is the shape parameter, providing the plot with a shape according to one of the three phases of the bathtub curve. $\beta < 1$ indicates early failures with a decreasing failure rate over time (or starts), a $\beta \approx 1$ is a sign of a constant failure rate, and $\beta > 1$ specifies wear out with an increasing failure rate over time (or starts). t is the measurement unit, here: operating time, measured in hours, or number of starts, depending on the type of failure the Weibull curve is modeling.

The following equations are describing important parameters to better understand the particular failure distributions.

The mean or mean-time-to-failure is according to Equation 5 a measure of central tendency of the failure distribution where $\Gamma(x)$ is the Gamma distribution.

$$\mu = \theta \ \Gamma(1 + 1/\beta) \qquad \textit{Equation 5}$$

MTTF (mean-time-to-failure) is also called expected value and is not to be confused with another measure of central tendency, the median, which divides the distribution in two halves, with 50% of the failures occurring before this point and 50% after.

The hazard rate function is time or start specific and describes the type of failure occurrence such as infant mortality, random failures or aging. A picture of a hazard function can be seen in Figure 7 (bathtub curve).

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}, \theta, \beta > 0;\ t \geq 0 \qquad \textit{Equation 6}$$

The reliability function depicts the probability of survival at a certain time (or start) t. Therefore the plot for this function is also called the survival plot.

$$R(t) = e^{-\left(\frac{t}{\theta}\right)^{\beta}} \qquad \textit{Equation 7}$$

The literature recommendation for Weibull plots with significant parameters β and θ, is to have ideally at least six failure data points (Smith, 2007). Smith does not recommend relying on results from Weibull plots that have been generated with less than four failure points. As a simple quality test, he recommends to compare maximum likelihood (ML) parameter estimates (for β and θ) with least squares (LS) parameter estimates. It is stated that if the Weibull parameter estimates are in a similar range, the estimates should be reasonably good. Smith indicates that it could be a sign of random failures, if the parameter estimates are largely differing from each other. One of the differences between the LS method and the ML method is the ranking method of the failures. LS follows a relative ranking concept (Bernards's approximation), and does not take into account the actual hours contributed by censored items. Due to the way the parameters are computed, higher values of time are favored. Smith recommends using LS for uncensored data. If the data set contains both uncensored (failed) and censored (survived) data, he recommends ML.

For the purpose of this study, the two methods are used as one measure to assess the quality of parameter estimates. Based on Smith's recommendation only failure categories with a minimum of four or more failure occurrences will be used, and of these, only the ones with similar parameter estimates using least square and maximum likelihood methods will be used for further analysis. As a second quality test, the correlation between the points and the fitted line needs to be larger than 0.9 (using the least squares method). As a third quality test, the results (shape and scale parameter estimates) have been discussed with an engineering expert to verify that they are in line with experience.

The results of the quality test are presented below in Table 4.

Table 4: Weibull Analysis - Remaining Failure Categories

Product Class	Failure categories	Failure categories passing quality test	Remark
Class 1	26 failure categories total 20 failed quality test - 19 had < 4 failure occurrences - F24 failed correlation requirement > 0.9	Six, namely: F5, F15, F17, F18, F20, F26	All time based
Class 2	41 failure categories total 37 did not pass quality test - 36 had < 4 failure occurrences - F1 failed comparison test between ML and LS method	Four, namely F15, F20, F40 F9	→ Time based → Starts based

From the 26 failure categories in class 1, six passed the quality test. Of the remaining 20, 19 had too few failure occurrences (less than four) for analysis. Failure 24 did not pass the correlation requirement of larger than 90%.

From the 41 failure categories in class 2, four passed the quality test. Of the remaining failure categories, numbers 36 and 37 had less than four failure occurrences per failure mode. Failure 1 (F1) did not pass the ML / LS comparison test. Although the correlation coefficient was larger than 90% the scale and shape parameters were too different (shape was 0.7 in the ML and 1.4 in LS method, scale was 32000 months versus 7300 months respectively).

Table 5 depicts the probabilities with which the six failure categories of Product Class 1 and four failure categories of Product Class 2 (which passed the test of similar parameter estimates for the LS and ML methods) would have been detected according to the probabilities from the Weibull analysis using

- The operating time of the respective first products (P1) of each class,
- The respective products with the least hours and lastly,
- An average length necessary for an approximately two year testing cycle (hour based and for the one failure – start based).

The experience shows that a test year has about 4000 hours of run time and averages 100 starts (these numbers are experience based and caused by down time for testing and evaluation of parts; typical run time per year for this specific product when operated by a customer in base load operation is 8000 hours / 430 starts). However, product run times at customer sites vary largely. For example, the first product P1 of Product Class 1 experienced in a 30 month calendar time frame only 2000 operating hours, but had more than 500 starts.

Based on the results of the Weibull distribution, one can expect to detect on average only 29% of all failure categories in class 1 after 1951 hours, 47% of all failure categories after 5479 hours and 54% after 7668 hours. The first product (Prototype) happened also to be the engine with the lowest operating hours.

A similar analysis for Product Class 2 has been done with the following result: The product with the lowest operating hours had experienced on average 19% of the failures. The first product (P1) in this class would have experienced on average 29% of all failures, and if a product test had run for 7668 hours, statistically 38% of the failure categories should have emerged in that product.

Table 5: Weibull Failure Probabilities for Product Class 1 and 2

Class 1 ML-method		Failure categories						
		F5	F15	F17	F18	F20	F26	Avg
Weibull Probability that failure occurs within:	1951 hrs*	16%	36%	25%	27%	54%	13%	29%
	2000 hrs	20%	36%	26%	28%	28%	28%	27%
	4000 hrs	33%	48%	34%	37%	37%	37%	38%
	6000 hrs	44%	55%	40%	43%	43%	43%	45%
	8000 hrs	53%	61%	45%	47%	47%	47%	50%

Class 2 ML-method		Failure categories				
		F9**	F15	F20	F40	Avg
Weibull Probability that failure occurs within:	2000 hrs 50 starts	38%	1%	1%	24%	16%
	2400 hrs 60 Starts*	42%	3%	2%	27%	19%
	4000 hrs 100 starts	53%	7%	4%	29%	23%
	6000 hrs 150 starts	63%	18%	9%	33%	31%
	8000hrs 200 St	70%	35%	16%	36%	39%

* Product with least operating hours in time 30 month time frame

** Start based failure mechanism, all others hourly based

The following figures are depicting the Weibull plots for the ten "quality approved" failure categories. Figure 8 and Figure 9 show the data of failure 5 for the comparison of the Maximum

Likelihood (ML) and Least Squares (LS) methods as part of the data quality test. The data of these figures demonstrates that both shape parameters (LS and ML) of failure 5 are below one, which indicates a good similarity of the section in the bathtub curve – infant mortality. Also the Anderson Darling tests have very similar values, 28 for both. Anderson Darling statistics is a way to measure the goodness of fit (how far the measured points are from the fitted line). Here it is used as a measure of “relative” goodness of fit. Similar Anderson Darling values indicate a similar goodness of fit of ML and LS curves. The correlation of the points to the fitted line of 0.91 according to the LS method is sufficient. Although there were only four failures that support the shape and scale parameter, the resulting plot was accepted by expert engineers as a good representation of the failure.

All failure categories have been evaluated according to this process.

For a better readability, only the plots for Maximum Likelihood method have been attached in the following, instead of also adding all the LS plots. All failure categories have passed the ML and LS comparison-test as described above and depicted in Figure 8 (LS) and Figure 9 (ML).

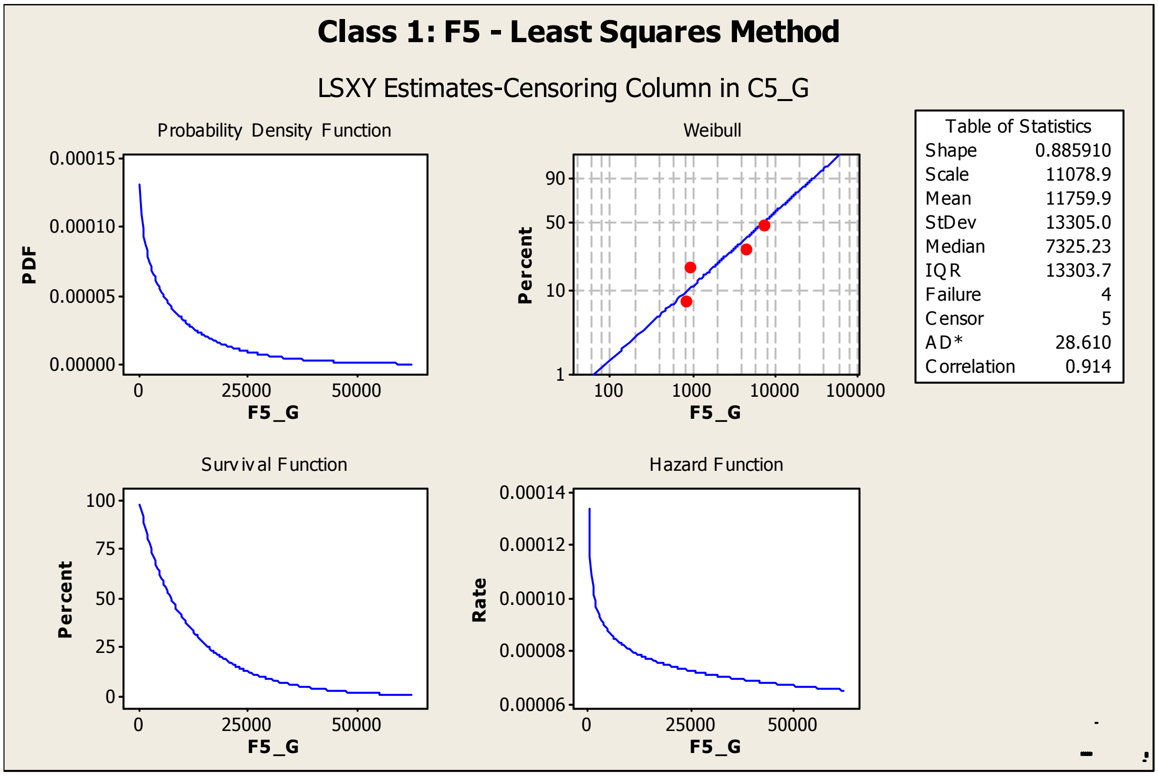


Figure 8: Least Squares Method on Failure 5, Class 1

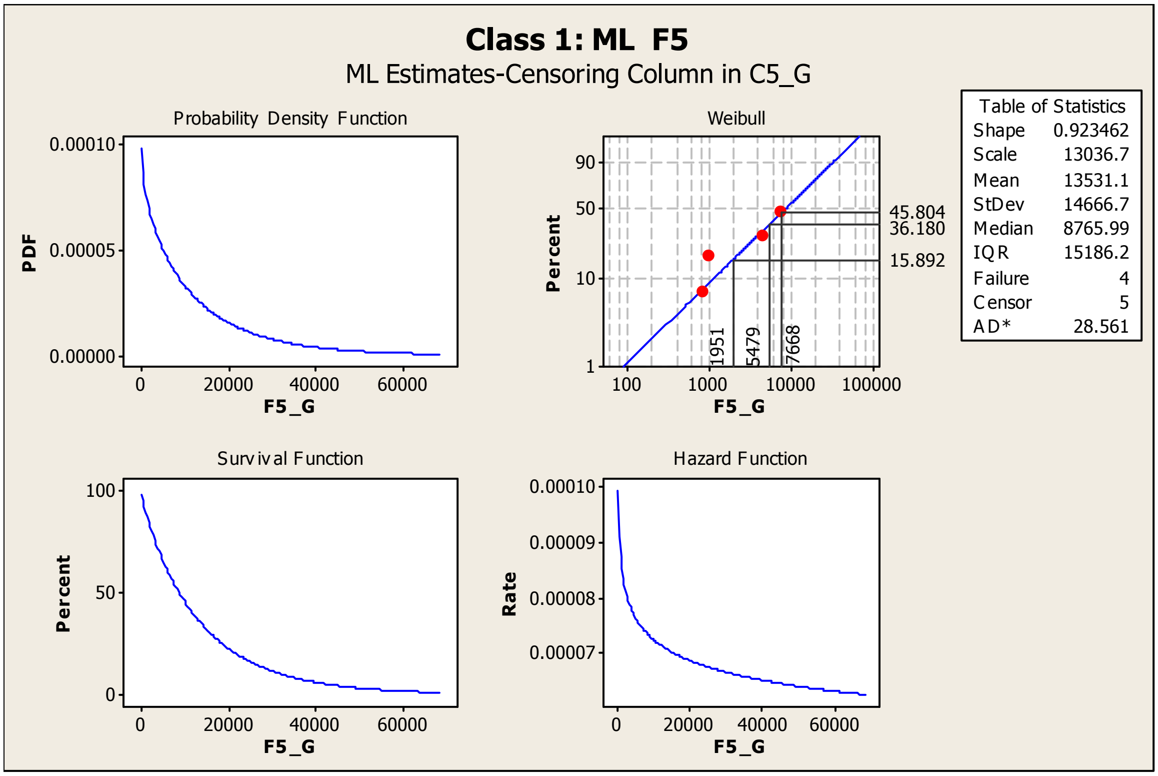


Figure 9: Maximum Likelihood Method on Failure 5, Class 1

The following plots are all similar in pattern giving an overview of the essential results of the Weibull analysis for each failure category. On the upper left quadrant of Figure 10, depicting failure 9 (F9) of Product Class 2, one can see the probability density function, found from Equation 4. The "L" shaped form, which represents the front end of the bath tube curve, suggests early failure occurrence. The Hazard Rate curve from Equation 6 on the lower right quadrant shows that the failure rate decreases over time. The shape parameter with 0.66 is smaller than one, indicating the same behavior. The scale parameter provides insight into the life of the product. After 152 product starts, there is a 37 % chance of survival. That is also depicted in the Survival Function Plot from Equation 7 in the lower left side of Figure 10. The mean time to failure is 204 from Equation 5. The Median means that after 87 starts there is 50% chance of survival. The curve was generated with 10 product data points. Of the ten products, six failed on failure 9, and four did not indicate a failure in the given start frames. The Weibull plot on the top right corner has the probabilities of failure for 60, 149 and 200 starts respectively. The corresponding failure probabilities are depicted in Table 5. All Weibull Plots were generated using the statistical software package MINITAB.

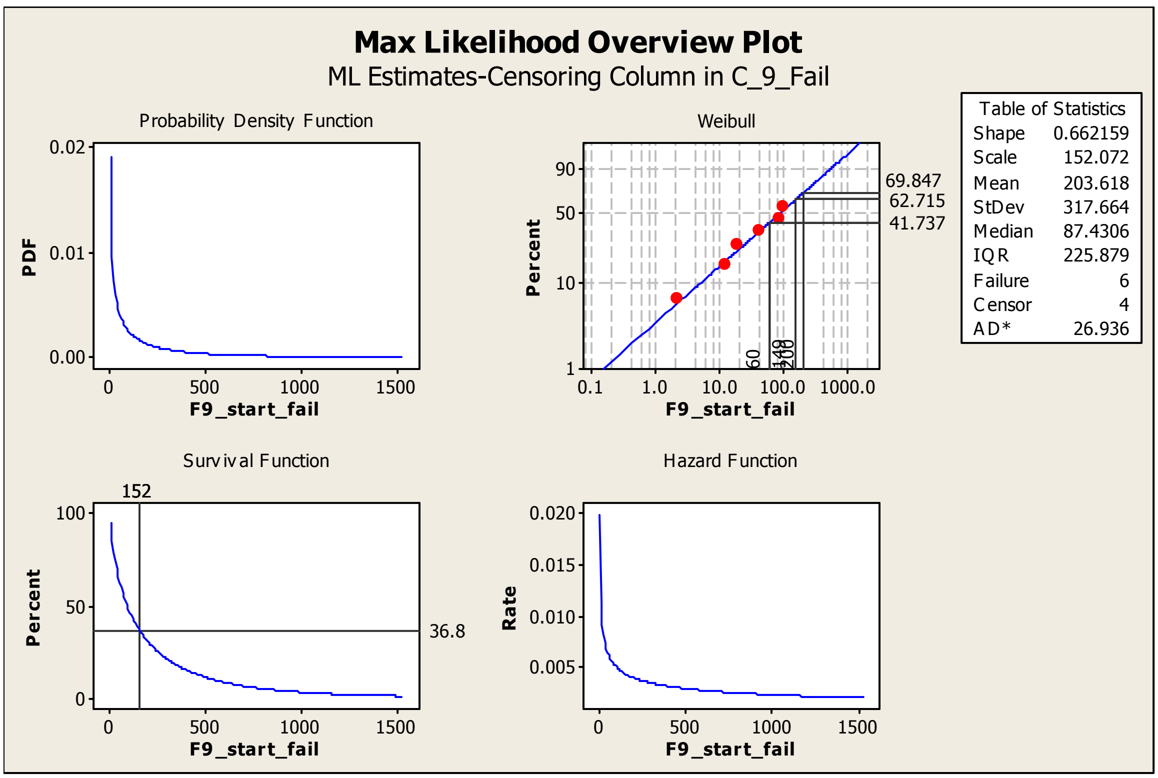


Figure 10: Weibull Plot – ML, Failure 9 Product Class 2

Figure 11 below portrays failure 15 (F15). The shape parameter is 2.6 and indicates a wear out failure. After 11084 hours, there is a chance of 63.2% that the product will have failed. Eleven products have been investigated, of which six failed, and five survived.

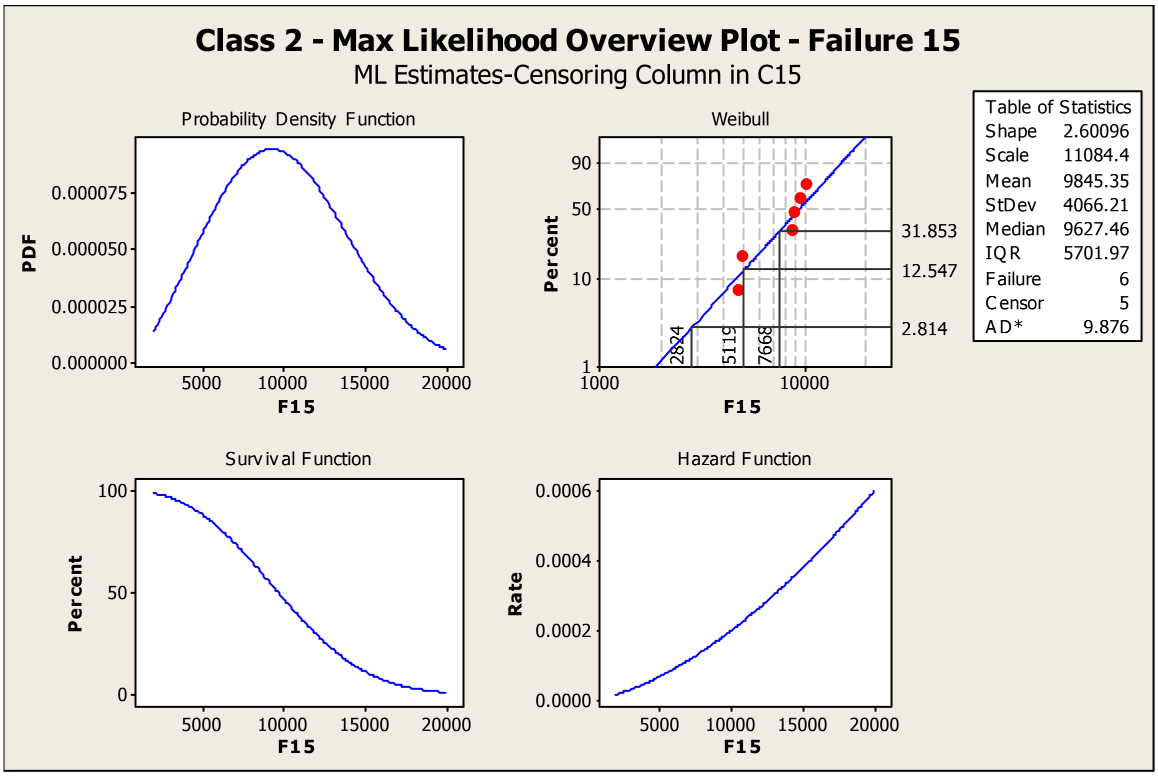


Figure 11: Class 2, ML - F15

Figure 12 depicts failure 20, with a shape parameter of 2.2, indicating a wear-out type failure with Mean Time To Failure of 15,600 hours. Failure types such as in Figure 11 and Figure 12 are especially difficult to detect with only one test product and test times as short as 2800 hours.

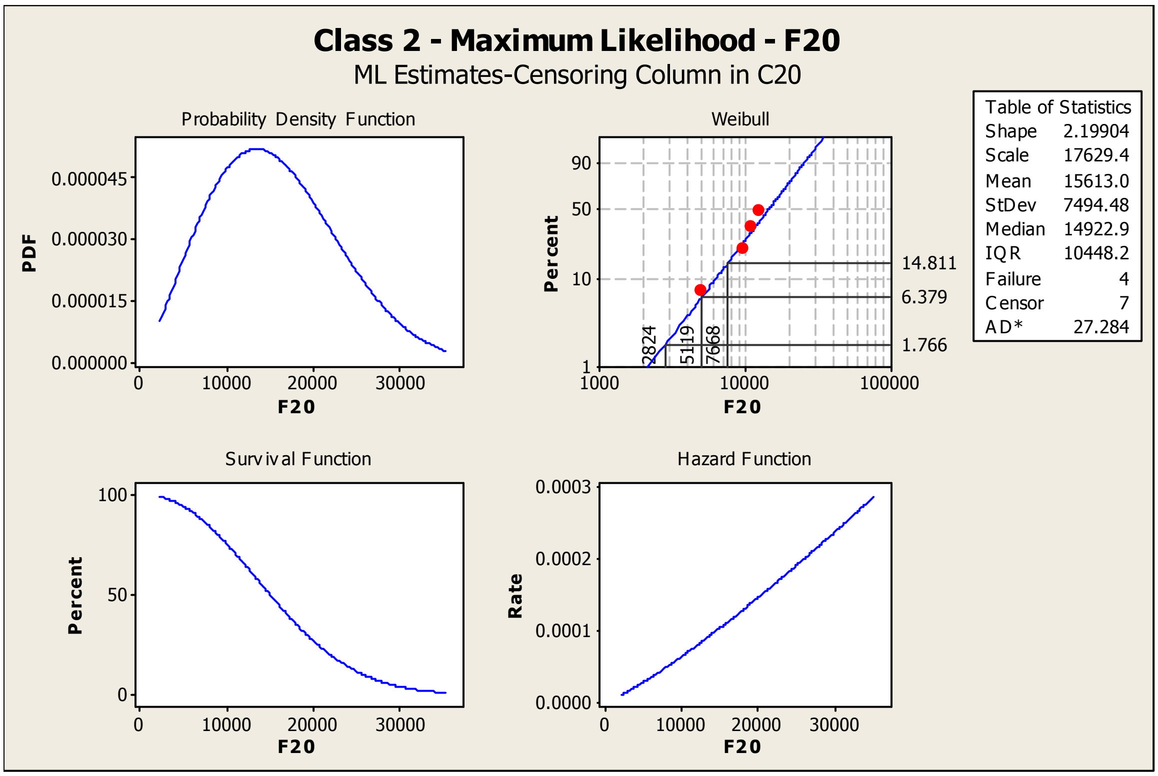


Figure 12: Class 2, ML - F20

Figure 13 depicts failure category 40. The shape parameter of 0.34 indicates an infant mortality type failure category, which is easier to detect with one prototype.

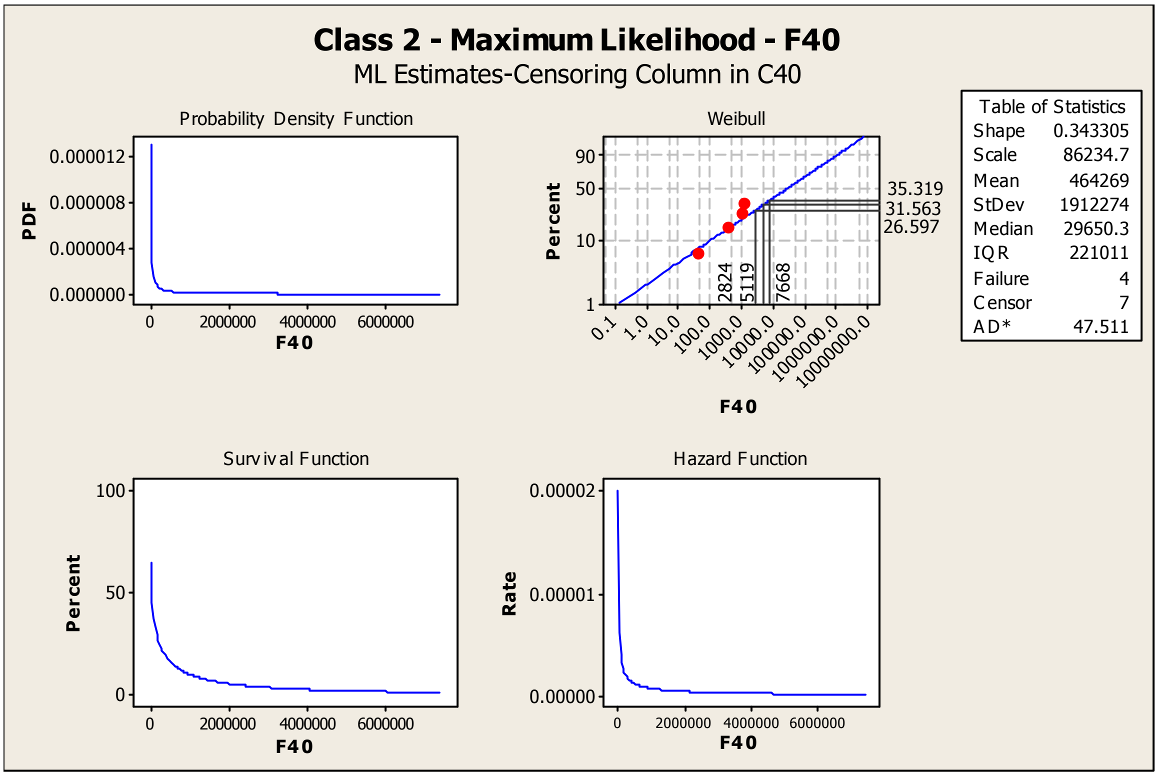


Figure 13: Class 2, ML - F40

The following six Weibull plots are with respect to Product Class 1.

Failure five in Figure 14 has a shape parameter of 0.9, indicating infant mortality type failure. There is a 63.2 % probability that the product will have failed after 13000 hours, indicated by the scale parameter.

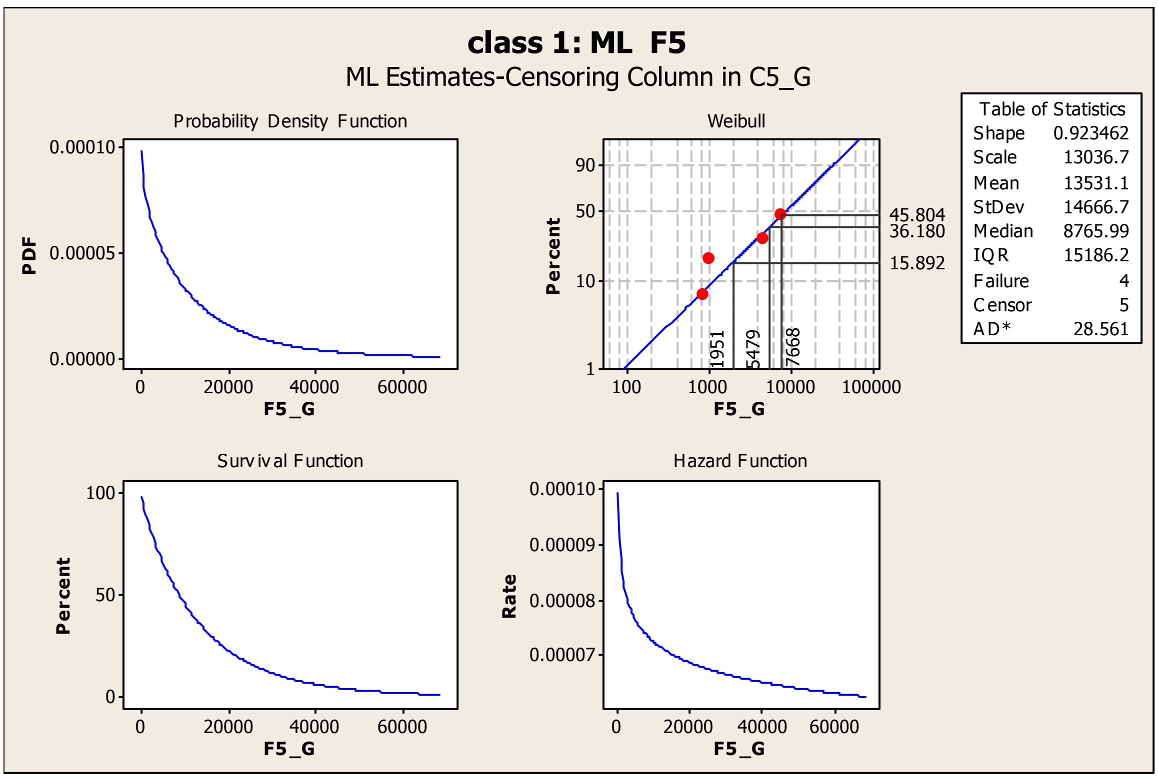


Figure 14: Class 1, ML - F5

Failure 15 in Figure 15 has a shape parameter of 0.5, indicating infant mortality type failure. There is a 63.2 % probability that the product will have failed after 9200 hours, indicated by the scale parameter.

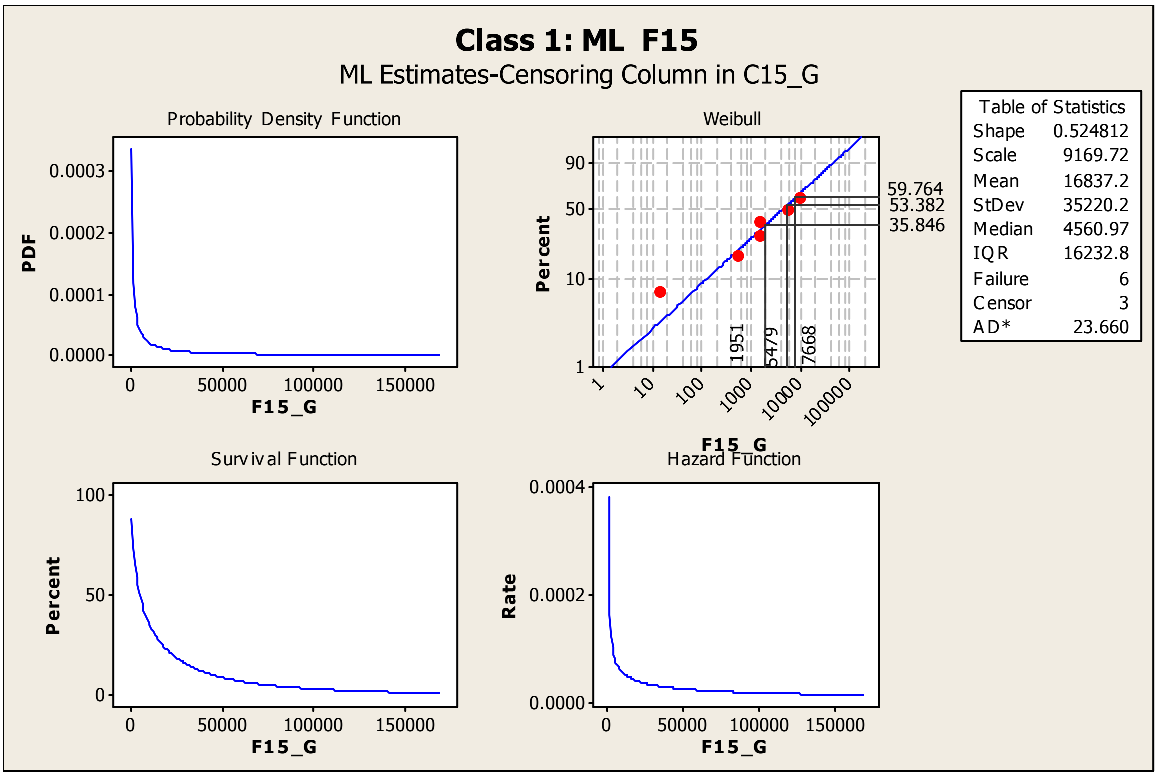


Figure 15: Class 1, ML - F15

Failure 17 in Figure 16 has also a shape parameter of 0.5, indicating infant mortality type failure. There is a 63.2 % probability that the product will have failed after 22400 hours, indicated by the scale parameter.

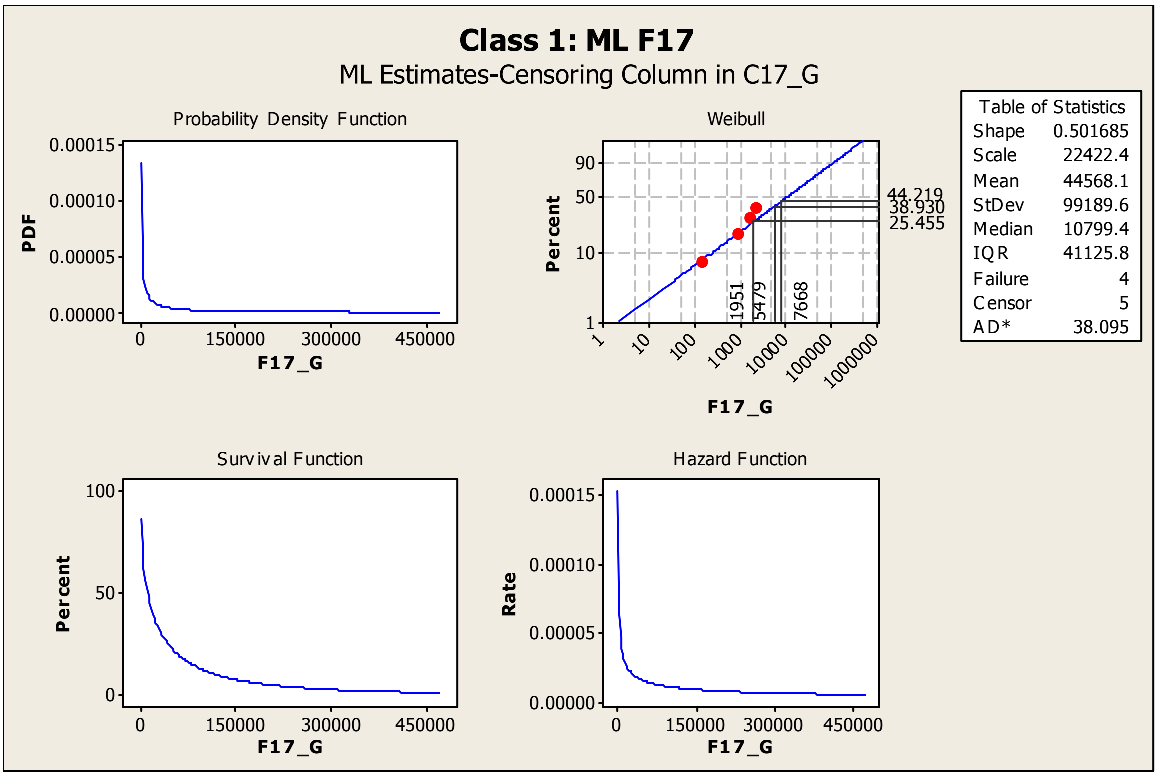


Figure 16: Class 1, ML - F17

Failure 18 in Figure 17 has also a shape parameter of 0.5, indicating infant mortality type failure. There is a 63.2 % probability that the product will have failed after 19400 hours, indicated by the scale parameter.

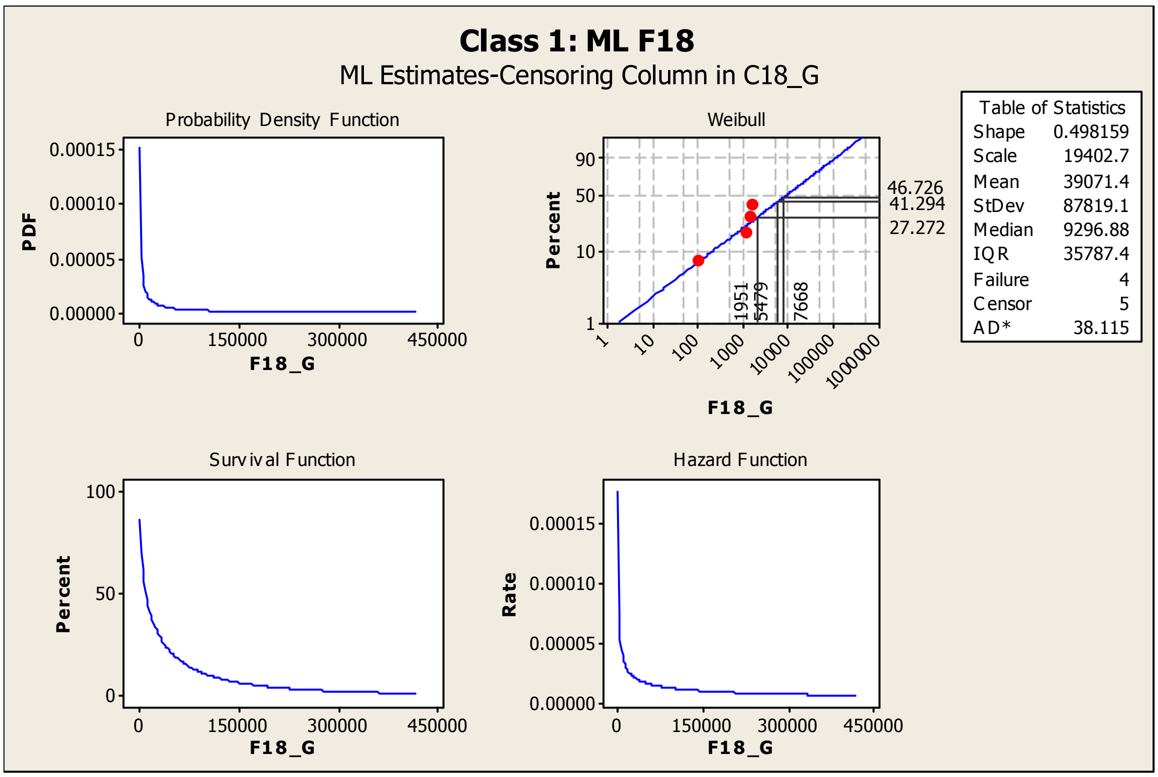


Figure 17: Class 1, ML - F18

Failure 20 in Figure 18 has a shape parameter of 0.7, indicating infant mortality type failure. There is a 63.2 % probability that the product will have failed after 2800 hours, indicated by the scale parameter.

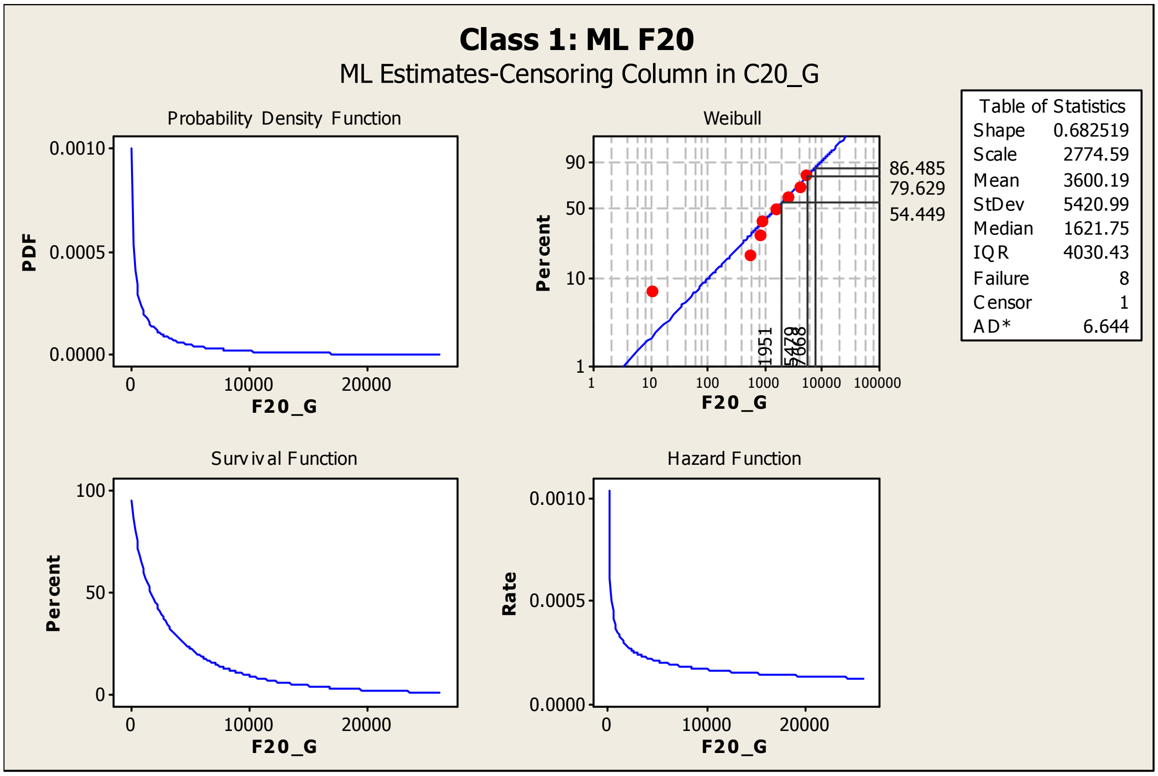


Figure 18: Class 1, ML - F20

Failure 26 in Figure 19 has a shape parameter of one, indicating random type failures. There is a 63.2 % probability that the product will have failed after 13,200 hours, indicated by the scale parameter.

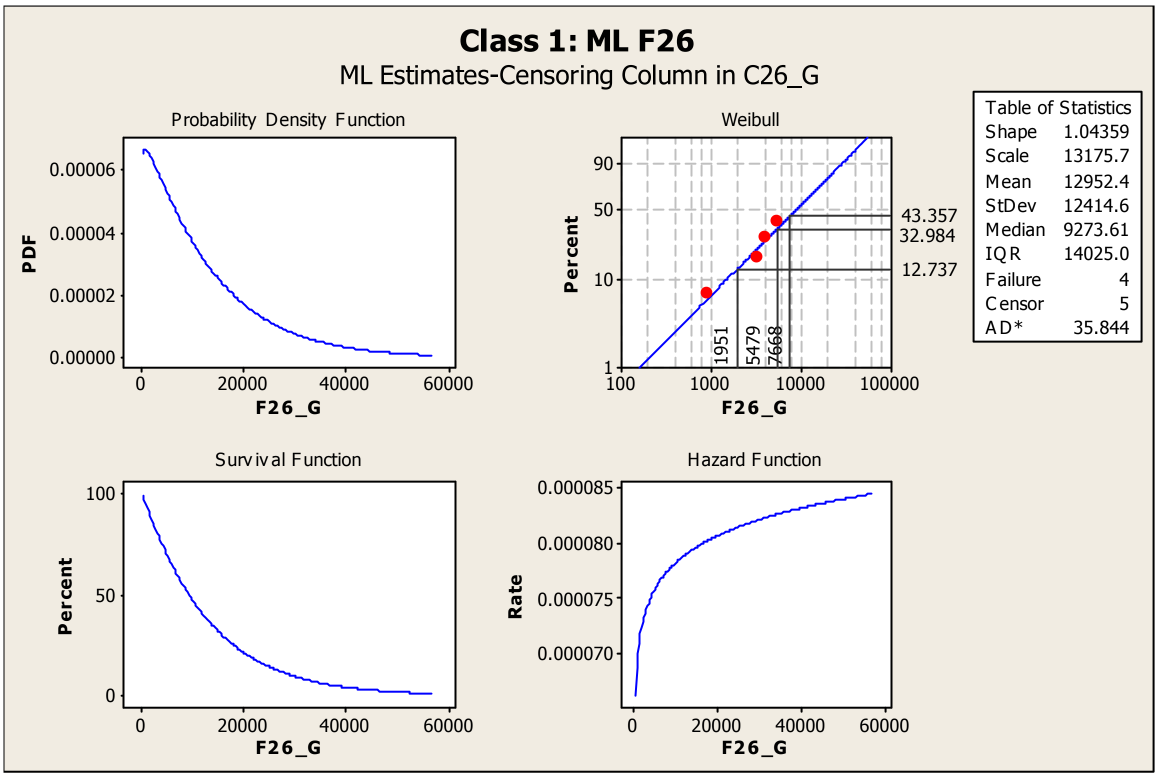


Figure 19: Class 1, ML - F26

In summary it can be said that, looking at the cases above, on average only one third of all failure categories would have been captured in Product Class 2 and 50% would have been captured in Product Class 1 with one prototype test in a time frame of about two test years or 8000 hours.

This illustrates the necessity of changing the current testing methods and warrants further investigation of the impact of the large percentage of later failures on the profitability of the product.

Visual Analysis

Another way to present these data with respect to hypothesis one is through a graphical display of the distribution of failure categories across the product fleet.

If all products would have begun their operational life at the same calendar time, as it is typical for an experiment, then it would be easy to see which product has a particular new failure class first. Although that exact approach is not possible when selling products, Figure 20 shows a similar approach. The products are numbered in the order they were sold, however, their failure occurrence time is counted from the time this product started its operation to failure. Because of this adjustment one is now able to determine which product experienced a particular failure class first ("relative" first), independent of when the product was sold.

Results of the Visual Analysis

Figure 20 shows that within Product Class 1 only 54% of all failures classes occurred first in Product 1, 8% of all failures classes first in Product 2, no failures happened in Product Three. 4% of failures classes happened first each in Products Four and Five, 19% of failures classes happened first in Product 6, and 12% of new failures classes incurred first in Product 7. In all subsequent products, no new failure category surfaced during the product's first 30 months of operation life. The total product count of Product class 1 was 24 products. All failure categories took place within the first seven subsequently sold products. However, the third product sold, Product 3, did not have any new failure category occurring first. Only roughly half of all failure categories detected did surface in Product 1 first, the other half occurred distributed over the six subsequently sold products after Product 1 (namely in Products 2, 4, 5, 6 and 7).

Product class 1's first seven products were sold within three years; it took five and a half years before discovering all of the major failures counting from the sale of its first product.

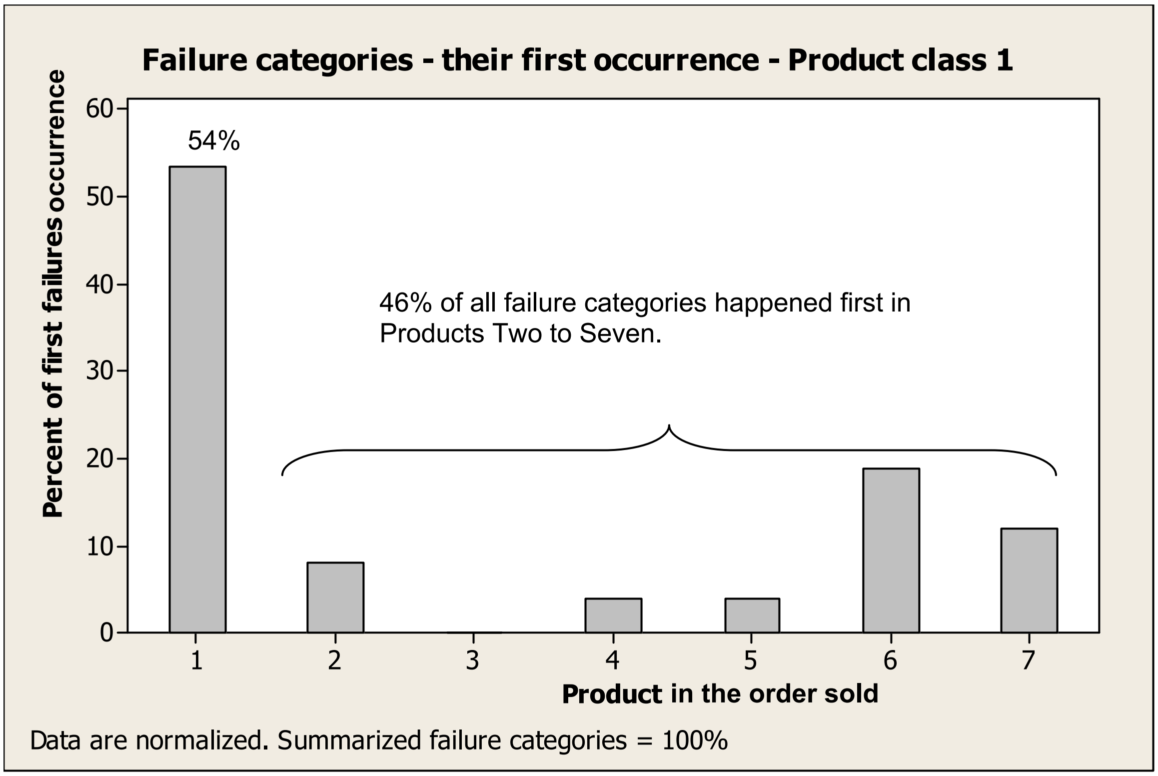


Figure 20: Class 1 - Distribution of First Failure Categories (Relative to Product Operation)

Product class 2 has a much higher product count (roughly 10 times more products sold than in Product Class 1). A similar distribution pattern of failures is demonstrated, compared to class 1; however, it has a wider distribution of first failure occurrences as shown in Figure 21.

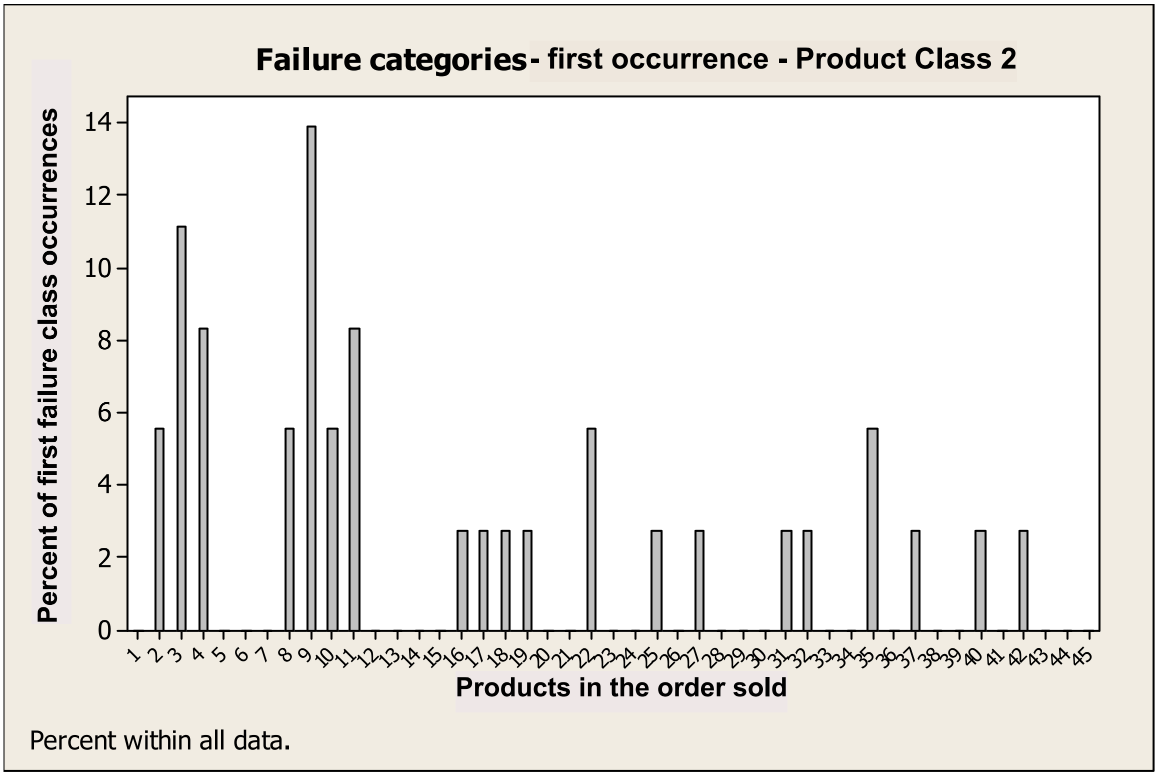


Figure 21: Class 2 - Distribution of First Failure Categories (Relative to Product Operation)

In Product Class 2, new failures classes are distributed over the first 42 sold products.

A cumulative graph (Figure 22) following the sales order shows that 50% of "relative first" failure classes are distributed over the first 10 products sold, 80% of relative first failure classes spread over the first 27 products sold. The meaning of "relative" first failure is that a failure class occurred first measured from the time this product was starting operation, not on an absolute time line of all products. For example product A starts operation on March 1st, and a failure occurred on June 1st, three months after start of operation. Product B, the second product sold of this Product Class, starts operation on April 15th, and the same failure (from the same failure class) that occurred in product A, occurred in Product B on June 15th, which is on a relative time scale only two months after start of operation of Product B. The failure class, although first seen in Product A on June 1st, happened ”relative first” in Product B, because it happened after two months of operation, versus three months of operation in Product A.

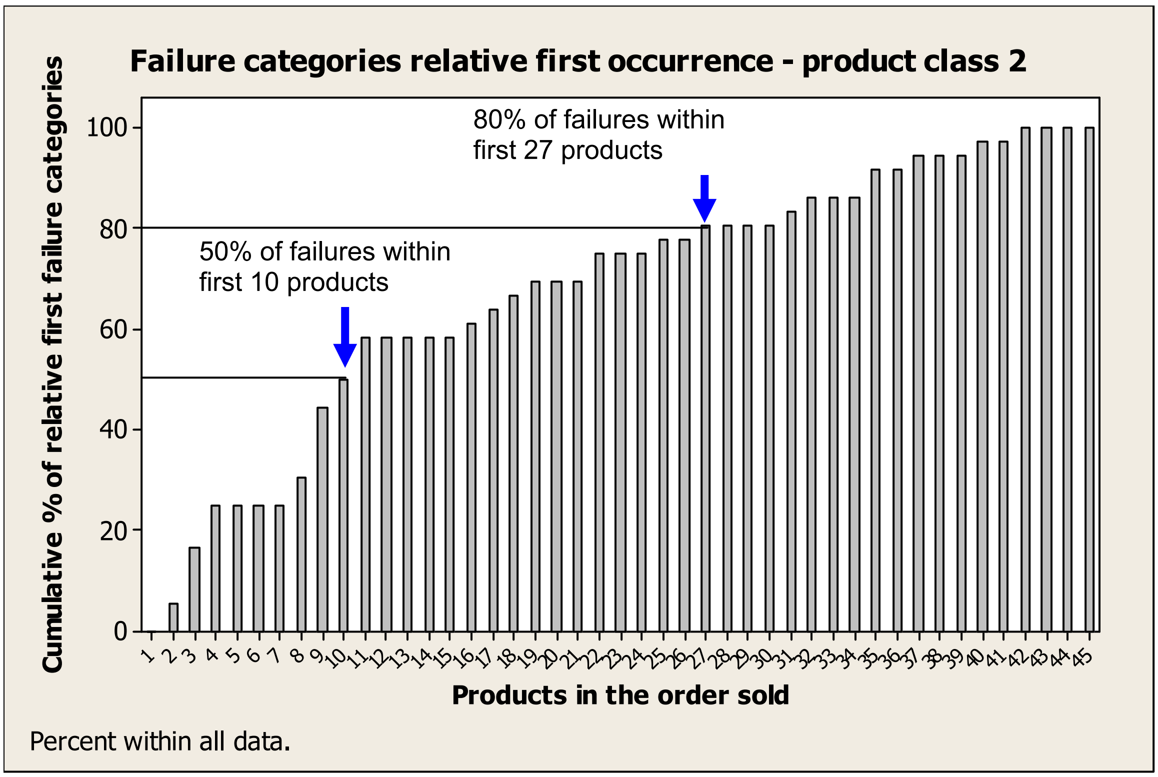


Figure 22: Cumulative – Relative First Occurrence of Failure - Product Class 2

Even though no first failure class occurred first in Product 1 (P1) on a relative time scale, P1 was not error free. Figure 24 gives a picture of the distribution of failure classes based on product sales order. P1 experienced 8% of all failure categories, P1 and P2 experienced 19%, the first three products experienced more than 40% of all failures, the first six products covered more than 50%, and the first 18 covered more than 80% of all failure categories. A similar graph for class 1 can be seen in Figure 23, showing that 50% of the failure categories occurred in product 1, and 100% of all failure categories showed up within 7 subsequent products.

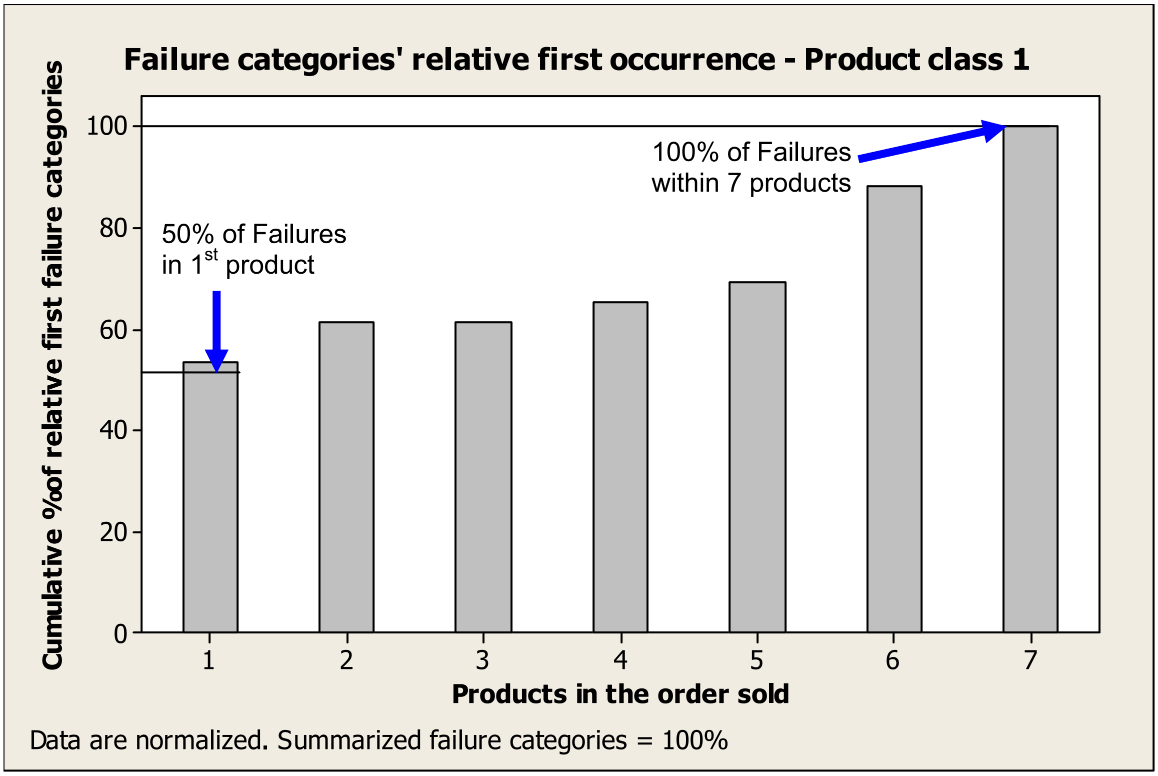


Figure 23: Cumulative – Relative First Occurrence of Failure - Product Class 1

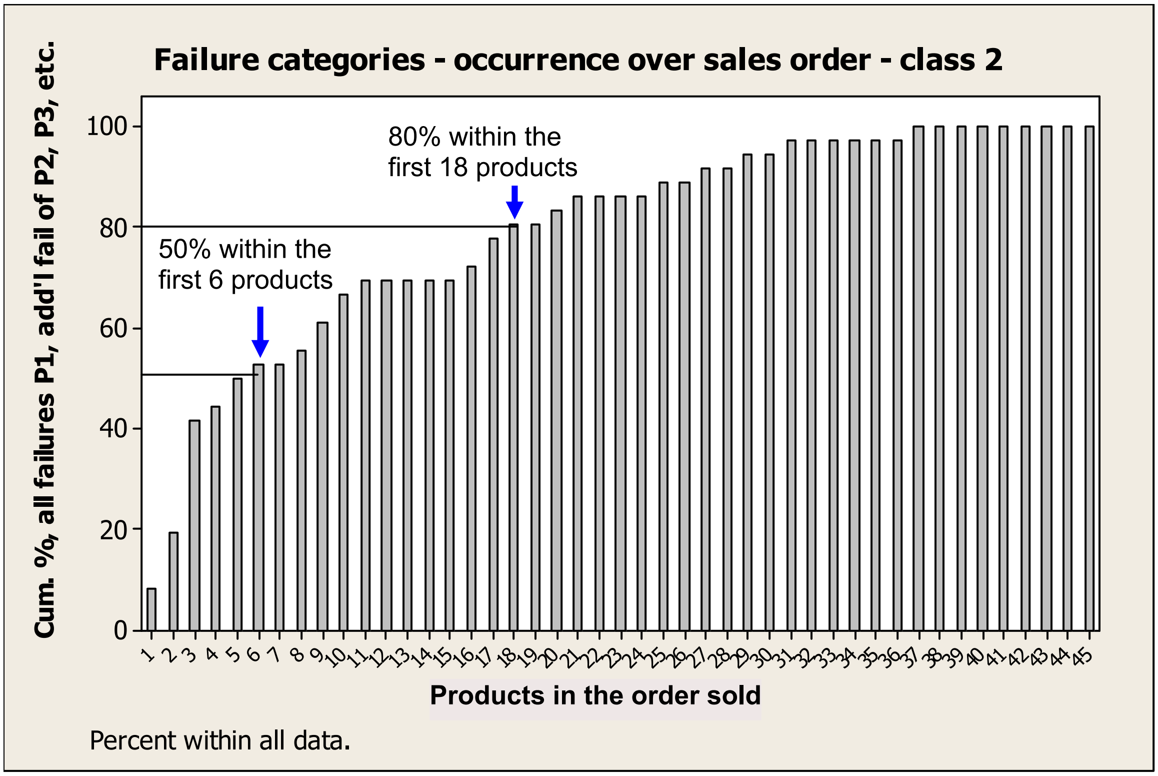


Figure 24: Cumulative Failure Categories (not Relative to Product Incommissioning Time)

A company also wants to know how many products have been sold before all failure categories have surfaced in at least one product, after commercialization. This is illustrated in Figure 24 for Product Class 2, showing failure categories that appeared in product P1 (in percent), plus additional failure categories that appeared in P2, plus additional ones that appeared in P3 and so on until all failure categories have been surfaced.

One can see in Figure 24 that the first six products that were sold accumulated 50% of all failure categories. The first 18 products accumulated 80% of all failures. The more products have been sold within a given time frame, the more distributed are the failures over all these products.

This is illustrated using the sales data from Table 6: Sales of New Product Class 1 and 2 Over Years 19 products from Product Class 2 were sold within the first two and a half years. If all failure categories equal 100%, then 80% of these errors have been found in the first 18 products after two and a half years (as is shown in Figure 24). On the other hand Product Class 1 had a slow start in selling products, only seven products were sold within the first two and a half years. 100% of the error categories that surfaced after testing appeared in the first two products.

The time frame of 30 months of operation time was chosen as maximum industry testing time for a particular product from the time the first product leaves manufacturing (ex works). This time encompasses five major service intervals and enough data to capture most of the "early failures" that a prototype test can capture.

Table 6: Sales of New Product Class 1 and 2 Over Years

Year / Product	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6 and ff
Class 2	5	5	9	7	23	49+
Class 1	1	1	5	2	9	6

Product-non-conformance-costs are estimated from the cost it takes to locate the failed parts at site, which, depending on the type of failure, includes cover lifts, man power cost to exchange the failed part, and the cost of the new part. For the purpose of the later model, typical costs for field work and hardware were averaged. For example a part failure in the compressor requires a cover lift, an appropriate work crew and a set of new parts.

Normalized, average cost data for different types of inspection work on site and hardware cost can be found in Table 7. Cost with respect to redesign is not taken into account, because the cost of redesign is the same, regardless of whether the failure occurs at a product test site or six sold products later down the line after serial release. It will not influence the cost model. The below data are normalized to disguise confidential cost data. Relative cost numbers are sufficient for the cost model.

Table 7: Normalized Cost Data for Typical Service Work Packages, Product Specific

Scope	Average total cost, normalized	Description
Major Combustor inspection	~ 0.04	Visual inspection of combustion parts and first row turbine vanes and blades and last row turbine blades and vanes. No cover lift.
Hot Gas Path outage inspection	~ 0.07	Same as combustor outage inspection. In addition turbine cylinder removal and inspection of all blades and vanes.
Major inspection with destack on site	~ 0.17	Same as hot gas path outage inspection. In addition removal of inlet, compressor, combustor and turbine cylinders. Visual and NDE inspection of inlet, compressor, combustor, turbine, rotor and exhaust.
Material Cost	0.013 – 0.2	Depending on which parts are to be replaced

In Chapter 1, hypothesis one, Figure 2, one can see a graph with the assumption that no product errors occur after product implementation. Figure 25 and Figure 26 show a revised version of this graph with the non-conformance-cost-data that did occur after product commercialization of Product Class 1 and 2. The x-axis (for each graph) shows that it takes six years for product development, design to testing and commercialization, and another six years (from year seven to 12) of product sales after serial release.

The pink curve illustrates the cost associated with each year, starting with investing into research and development, hardware and testing cost and, starting from year seven, non-conformance cost due to errors in customer owned products and income from product sale and maintenance contracts. The numbers on the squared blue rectangles are the number of sold products in that year.

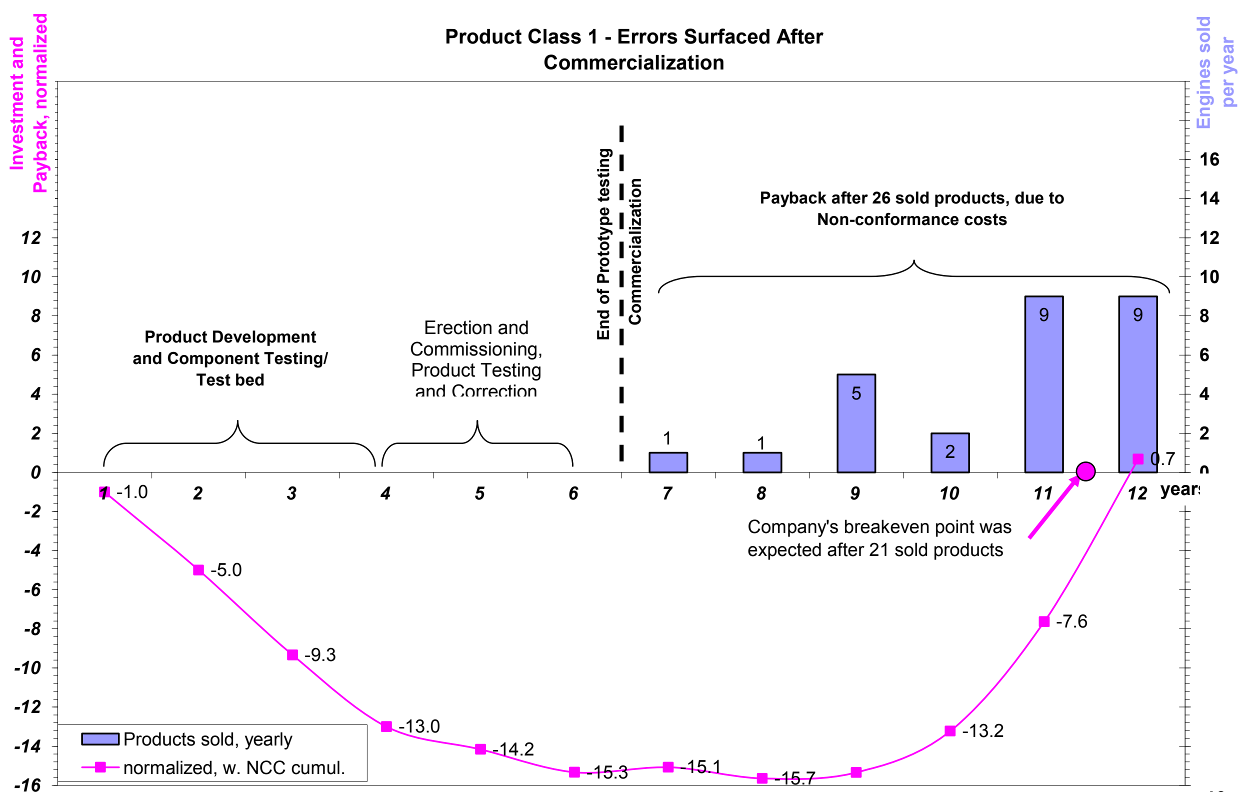


Figure 25: Cost Impact of Errors that Occurred After Commercialization- Class 1

One can clearly see that the errors after commercialization were costly enough that the company had to sell nine additionally products to break-even (from the baseline in Figure 2 for Product Class 1) and seven additionally products from the baseline in Figure 3 for Product Class 2 respectively.

Table 8 illustrates a breakdown of cost and income for Figure 26, Product Class 2. Costs that remained unchanged from the base line cost assumption in Figure 2 and Figure 3 respectively were product development, component testing cost, test hardware cost, system level test costs and the related costs to redesigning faulty parts (independent of time of occurrence). The income to the company consists of profit made by selling the product and a yearly service-fee related income for maintaining the product. Cost that did increase when comparing Figure 2 with Figure 20 and Figure 25 and negatively impacted the break-even point, was failure cost of flawed parts in customer products after commercialization that had to be repaired or replaced with a better designed one. A summary of these costs that are often associated with the correction of errors in the product can be found in Table 7. Other error related market specific costs that apply depending on the region in which the product is operated, are not listed and not accounted for in this calculation. An example for such costs is liquidated damages, requiring the manufacturer to compensate the customer for the down time of the product (lost business opportunity).

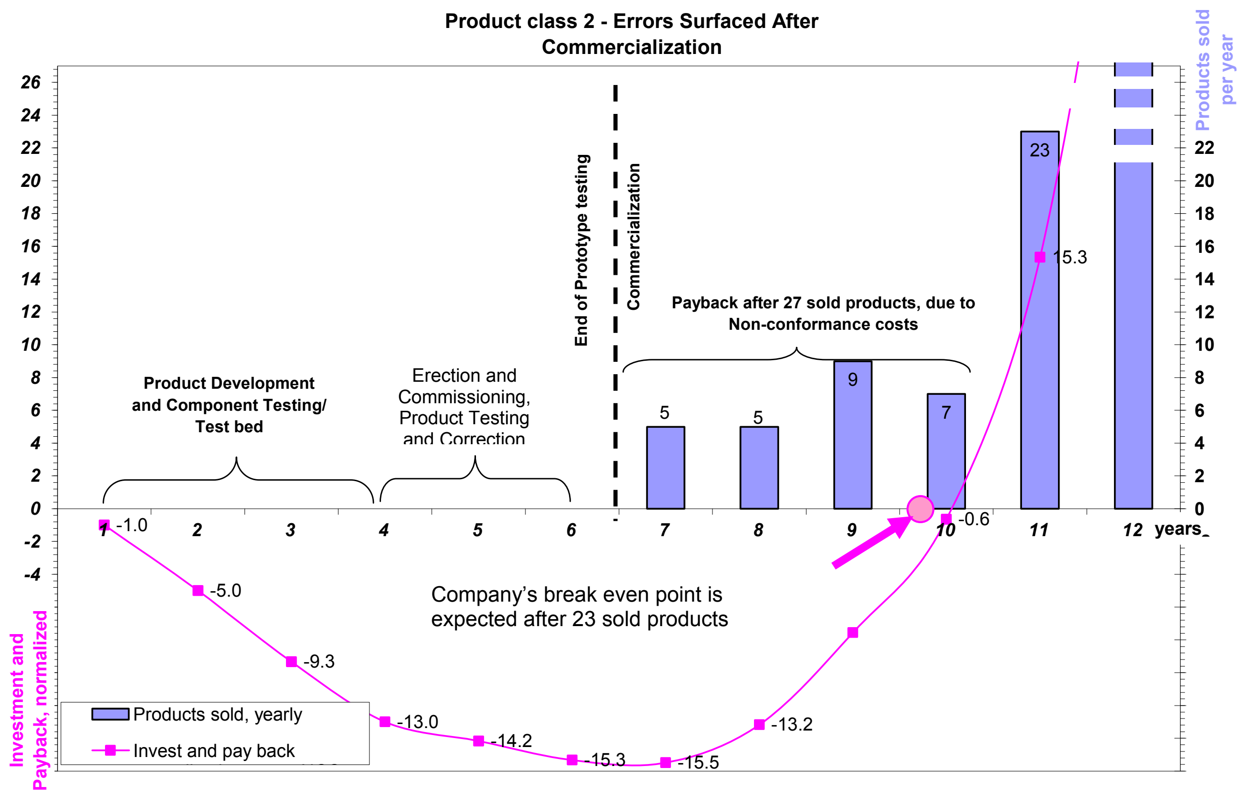


Figure 26: Cost Impact of Errors that Occurred After Commercialization- Class-2

Table 8: Payback Calculation Table, Class 2, Normalized

Design development, testing, serial release Class 2		Year	Products sold, yearly	Ideal Invest and pay back	Invest and pay back; Cumulative	Total cost per year	Product development cost	Component tests cost	testing cost and HW	NCC cost	Maintenance income
Pre-commercialization	development &	1	0	-1.0	-1.0	1.0	1.0				0
	component testing	2	0	-5.0	-5.0	4.0	2.7	1.0	0.3		0
	end of year 4:	3	0	-9.3	-9.3	4.3	1.7	1.3	1.3		0
	first product ex work	4	0	-13.0	-13.0	3.7	1.3	1.0	1.3		0
	Testing, service support & re-engineering	5	0	-14.2	-14.2	1.2			1.2		0
	End of year 6: first engine sold	6	0	-15.3	-15.3	1.2			1.2		0
					Commercialization						
After commercialization	NCC year 1 after commercialization	7	5	-13.7	-15.5	1.8				1.8	1.7
	NCC and income year 2	8	5	-10.3	-13.2	1.0				1.0	3.3
	NCC and income year 3	9	9	-4.0	-7.6	0.7				0.7	6.3
	NCC and income year 4	10	7	4.7	-0.6	1.8				1.8	8.7
					From here on downward, company makes profit despite NCC cost						
	NCC and income year 5	11	23	21.0	15.3	0.3				0.3	16.3
	NCC and income year 6	12	49	53.7	47.3	0.7				0.7	32.7

Table 8 shows the differences in payback between the current assumption with no NCC (greenish column, called "ideal invest and pay back") and the real data from Product Class 2 with non conformance cost after product serial release (column labeled investment and payback to the right of the green column).

The table also shows the other costs such as component development cost over the years, the testing cost in HW cost and fuel, and the income by year. The income is defined as profit made with the sale of the product and, on a yearly basis, through a maintenance contract with the customer. Income does occur only after commercialization at end of year six.

The observed <u>failure</u> <u>cost</u> behavior is demonstrated in Figure 27 for both Product Class 1 and Product Class 2. In the graph, the violet and blue columns are the costs that each first product (P1) of each class incurred. The striped violet and striped blue columns are costs that would have occurred if none of the errors had been repeated in subsequent products (the "ideal" test prior to commercialization) respectively. The values of the striped bars are the sum of each failure category cost without counting a multiplication of failures in other products, distributed over two (assumed) testing years.

The ratio of the actual cost to the "ideal" cost for classes 1 and 2 are:

$$class1: \frac{CPall}{CPX} = \frac{5.33}{1.69} = 3.15 \quad \textit{Equation 8}$$

$$class2: \frac{CPall}{CPX} = \frac{6.32}{2.38} = 2.65 \quad \textit{Equation 9}$$

CP_{all} are costs that impacted all sold products over their first 30 months lifetime.

CPX are the costs of all single error categories, regardless of the product in which they occurred, and without counting multiples of the same error classes. In Product Class 1 this ideal cost (CPX) is 1.7 (normalized with 1 = cost of one product), in Product Class 2 this cost are 2.4. Ideally, all error categories only appear once, and then get corrected. The actual error correction cost of the two classes was 5.3 (normalized), and 6.4 (normalized). In the ideal case, it would have been 1.7 (normalized) and 2.4 (normalized) respectively.

The ratio of 2.65 for class 2 and 3.15 for class 1 respectively from Equation 8 and 9 means that the actual costs are 2.65 and 3.15 times higher than the ideal case, if one were to assume that all error categories only appear once, and then get corrected.

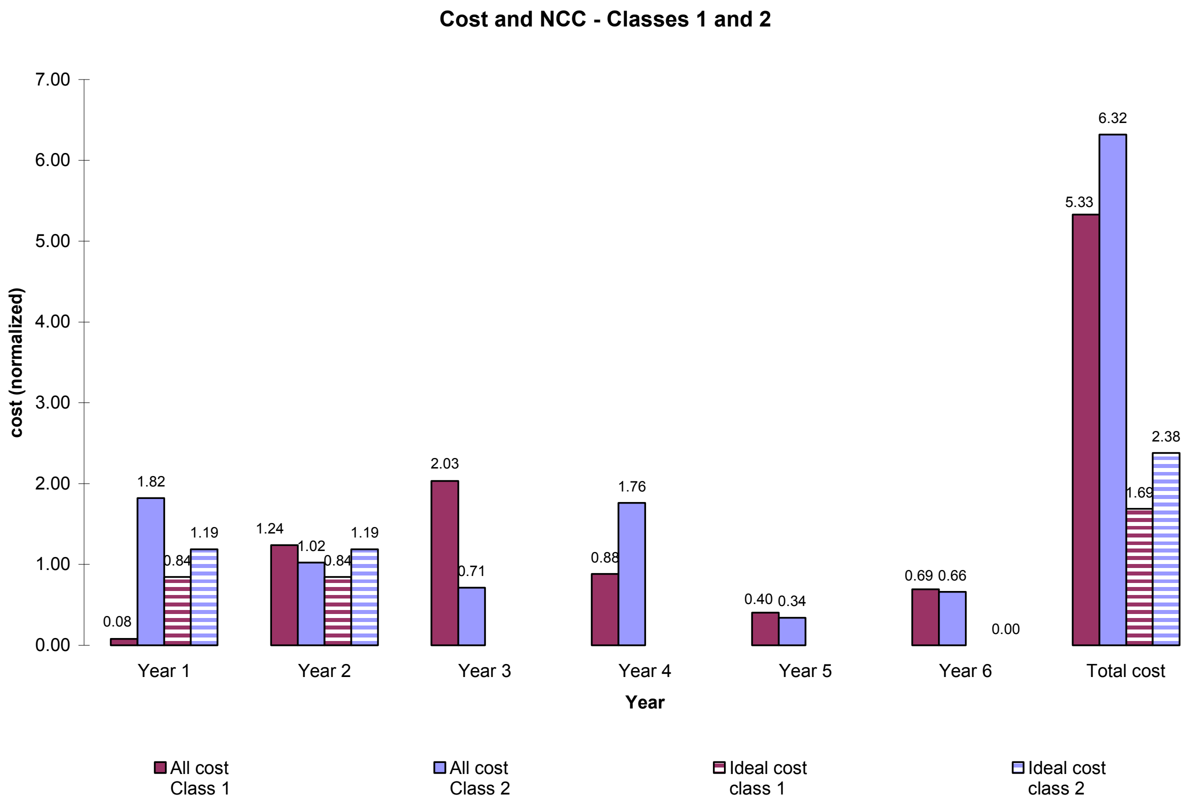


Figure 27: Cost Behavior - Ideal and Real for Classes 1 and 2

Data Quality

It is important to address data quality and the difficulties in gathering data. Data used for this research was first documented on paper rather than in an electronic database. Those data were transferred to an electronic system in early 2000. Engineers involved at that time have left the company or retired, and that made it harder to verify data correctness.

This researcher's approach to gathering data was to focus on collecting product data for expensive field issues with significant cost and customer impact. Those cases were the best documented ones. Although that may not give a complete picture of failures and failure cost of products after serial release, it is assumed to capture about 80% of the total cost, which was considered to be reasonable by the company and to be sufficient for this investigation.

Approach to Testing Hypotheses Two

Problem Boundary

Hypothesis Two: The second hypothesis is that a generic process can be derived and a dynamic cost model can be built for LCPs, adjustable to the respective industry and supporting companies' policies on decision making with respect to an economically optimal number of prototypes and testing length. This assumption will be tested with real product data from two product classes, Product Class 1 and Product Class 2 using systems dynamics simulation methodology.

The validity of the following assumptions will be investigated through simulation in the model:

- One prototype product test will only lead to detection of a fraction of the typical failure modes
- Products of the same type will experience different failure modes.
- Parallel product tests have higher initial cost, but costs associated with failures after product introduction would be reduced (and break even points would occur sooner for a company).
- Fewer failures after product release may increase the company's reputation and potentially result in more products sold.
- Later product introduction causes the company a loss of income due to later sales of products; thus return of investment will be delayed.

Scenario 1: Base Line Model

The approach used in this research to validate the final dynamic model is to first build a model that reflects the reference data (expressed in the cost-benefit curve) from Product Class 1 and 2 when simulated, and then increase the model's complexity to reflect more selected causal relationships between the independent variables and the events they cause (such as influencing customer buying behavior or the time when the new product gets to market).

The variables for this specific model that will be used to simulate various testing scenarios that are held constant are:

- Profit margin (for this specific product set to 14%; percentage of revenue of sold operation and maintenance (O&M) contracts)
- Product development cost (engineering and component testing; normalized to one product and fixed – for values see Table 11 below)
- Fuel price (for scenario 1 values see Table 11)
- Test product cost (for one set of hardware – for values see Table 11 below)
- Cost to operate and maintain (O&M) one test product (cost of personnel and resources other then fuel, needed to operate this specific product – for values see Table 11)
- Non-conformance cost (measured from Product Class 2 – for values see Table 11)
- Impact of time to market (loss or gain of sales due to later or earlier time to market, due to longer or reduced test phase in years; base line is six years of development and testing – values see Table 11 below)

The testing variables that will be varied to determine an optimal testing policy are:

- Test length per year (1 interval = 500 test hours; 4000 test hours per year possible)
- Number of years of testing
- Number of test products
- Number of products sold over time (sales forecast of products per year – for Product Class 2 sales see Table 11).

The dependent variables are:

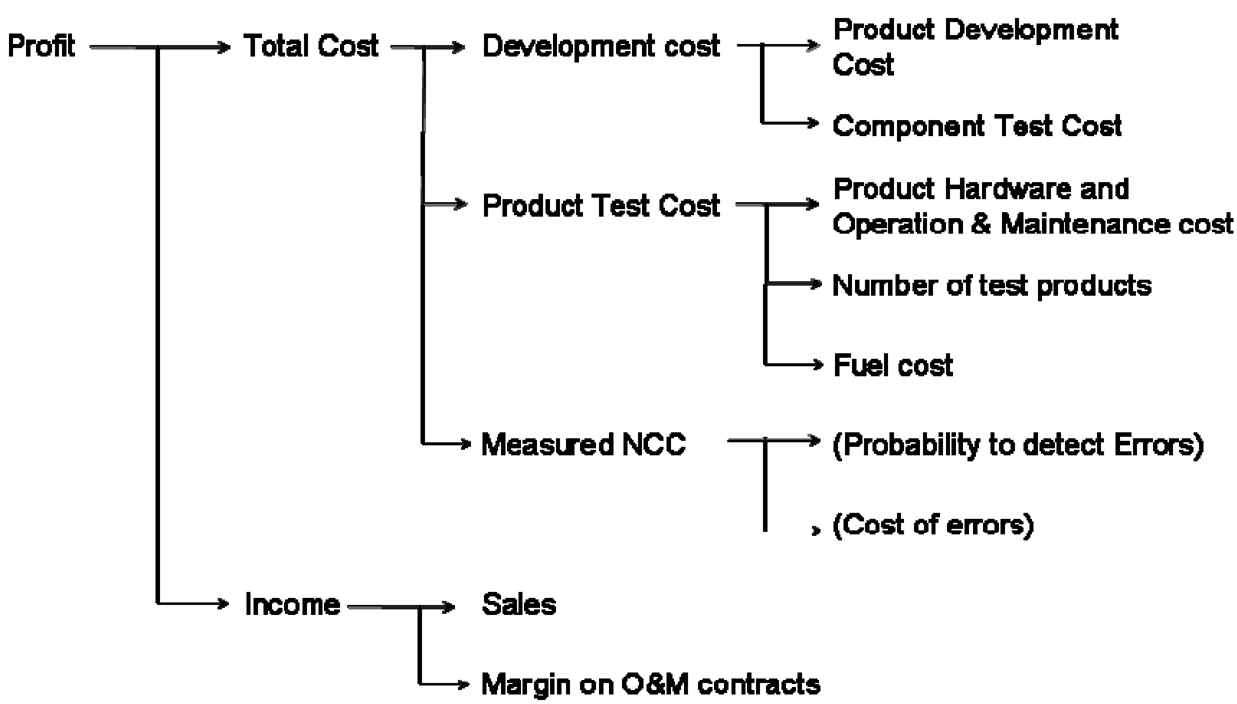
- Profit
- Income (from operation and maintenance contracts)
- Total costs
- Time to market (depending on testing time)
- Buying behavior (depending on errors in products after commercialization)
- Affordability of test products (depending on profit made)
- Non-conformance cost (depending on testing intensity).

The dependent variables will be calculated according to the equations below.

The total costs are dependent on developing costs, testing hardware cost, O&M costs and non-conformance costs. The non-conformance costs are dependent on the probability to detect errors during testing. Table 9 displays these relationships. The left hand side illustrates the dependent variables; the right hand side shows the independent variables. Values for the independent variables have been acquired from Product Classes 1 and 2 and are shown in Table 11. The probabilities to detect errors were calculated using the Weibull distributions, as described in

Hypothesis 1. Those distributions will be used in the enhanced model later. For this first model, the measured related failure costs (hardware replacement cost and the cost to repair and reinstall) per year have been used according to Table 11. The “probability to detect errors’ and the ‘cost of errors’ are parenthesized in Table 9 to indicate that they are not yet modeled. The probability to detect errors at given operation times (Weibull distributions) and cost of error will be replacing Table 10 values for non-conformance cost in the later scenarios.

Table 9: Causal Structure of the Product Development and Sales Process



The causal model was developed and verified using data from Product Class 2. Since it is irrelevant which product class data are used to verify the model, it was arbitrarily decided to be data from Product Class 2. The model is called scenario 1 baseline model, and was used as the basis for extension to a dynamic model through a step-wise integration of more causes that influence testing results and testing process. The baseline model can be seen in Figure 28.

The first rectangle in this picture (representing stock, similar to a container) called “Profit” integrates income and cost over 12 years. The second rectangle called “Cumulative Units” integrates the units sold over the years and is used to calculate operation and maintenance (O&M) income from contracts sold per unit. Single arrows in this diagram represent causal influences from one link to the link to which the arrow points. The causality is built into the model through equations. The specific equations used for this model are described later in the text. The values of each parameter used in this model over the years are shown in Table 11.

For the extended dynamic model the averaged data sets from Product Class 1 and Product Class 2 have been used as values for the variables.

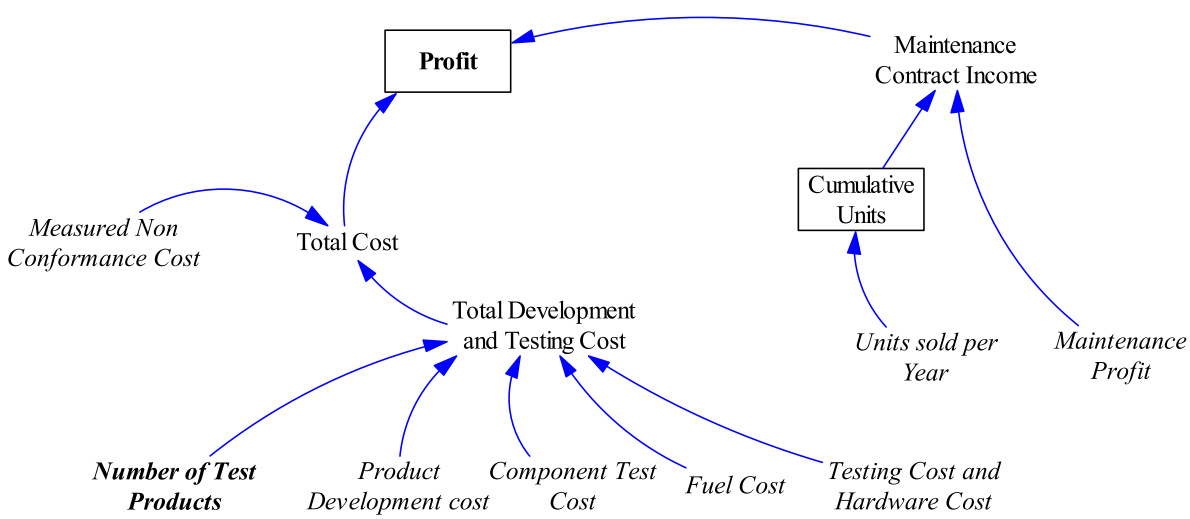


Figure 28: Causal Model - Baseline-Class 2

The baseline model in Figure 28 does not have any feedback loops yet (hence it is called a static model). This is a reflection of the initial mental model of the testing policy decision makers. As one can see in Figure 28, the investments (total development and testing costs) consist of product development cost, cost for component test rigs, test product cost, O&M cost for the test product(s) and fuel cost. A typical time line for product development and component and system testing can be found in Table 10.

Table 10: Example for a New Product Development Time Line

Description	Time line* (ideally: six years)
Development and parallel component testing (on test rigs)	48 months
Product manufacturing and shipping	12-24 months
Erection and commissioning	6 months
System testing (product testing), redesign and retest of failed parts in parallel, if necessary	10-24 months
Decommissioning of test product, structural and destructive testing, and redesign of parts, if necessary	14-24 months

* Development time lines overlap using the concurrent engineering approach;

As a next step to verify the correctness of this model, it was simulated over 12 years. The values used for the simulation are given in Table 11.

The simulation was conducted using the software VENSIM, a visual modeling tool for conceptualization, simulation and optimization of dynamic systems. The equations used in the model in Figure 28 can be seen in Equation 10 to Equation 15.

Table 11 has values of the variables (same values that have been used for the cost and benefit graphs in Figure 3 and Figure 26).

$$\text{Profit} = \int_{0}^{t1} (\text{Maintenance Contract income} - \text{Total Cost})\,dt \qquad \textit{Equation 10}$$

with t_1 equals 12 years and time step equals one year;

$$\text{Maintenance Contract income} = \text{Maintenance profit} * \text{cumulative units} \qquad \textit{Equation 11}$$

$$\text{Cumulative units} = \sum_{n=1}^{12\text{ years}} (\text{Units sold})_n \qquad \textit{Equation 12}$$

$$\text{Total Cost} = \text{Non Conformance Cost} + \text{Total Development and Testing Cost} \qquad \textit{Equation 13}$$

$$\text{NonConformance Cost} = \text{ set to 0 for the baseline} \qquad \textit{Equation 14}$$

$$\begin{aligned}&\text{Total development \& testing cost} = \\ &(\text{Product development cost} + \text{Testing cost \& hardware cost} + \text{Fuel cost} + \\ &+ \text{Component test cost}) * \text{Number of test products}\end{aligned} \qquad \textit{Equation 15}$$

Table 11 lists the values (normalized to the product cost) used in this simulation:

- Sales per year (starting from year seven)
- Component test cost (occurred in years two to four)
- Fuel cost (occurred in the years five and six)
- Operation and maintenance (occurred in years five and six)
- Test product hardware cost (in the years two to four)
- Product development cost (occurred in the years one to four)
- Non-conformance cost (distributed over the years 7 to 12).

The costs are specific for this particular product. The bulk of the product development costs in year two to four come from component tests that have been conducted prior to product testing. Fuel cost and operation and maintenance cost are cost related to testing. The cost for the prototype test product is distributed over three years for ordering, manufacturing and assembly of the different parts of the test product.

Table 11: Data Table for Simulation of Scenario 1 (Class 2)

Name of variable:	Product class 2 data - costs and benefits normalized to the cost of one product - units are measured per year											
Year	1	2	3	4	5	6	7	8	9	10	11	12
Units sold per year	0	0	0	0	0	0	5	5	9	7	23	49
Component test cost	0.00	1.00	1.33	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fuel cost	0.00	0.00	0.00	0.00	0.83	0.83	0.00	0.00	0.00	0.00	0.00	0.00
Operation and maintenance	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.00
Test product hardware cost	0.00	0.33	1.33	1.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Product development cost	1.00	2.67	1.67	1.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Non conformance cost (NCC)	0.00	0.00	0.00	0.00	0.00	0.00	1.82	1.02	0.71	1.76	0.34	0.66

After conducting the simulations (one with the measured NCC per year, the other one with non-conformance cost assumed to be zero) the cost-profit graphs in Figure 29 and Figure 30 were created. They display in blue the cumulative cost and profit distribution over 12 years since the start of the product development program of Product Class 2. Years one to six represents the development and testing period, years seven to 12 are the years of product sales to the customer, plus income through O&M contracts and non-conformance costs.

The purpose of these first two simulations was not to improve the testing procedure yet, but to show that the model is correctly simulating profit development according to the baseline. The first run simulates a 12 year development and sales program under the assumption that no non-conformance cost (NCC) occur. The result is depicted in Figure 29. The second simulation run included NCC, and can be seen in Figure 30. Both graphs are mirroring exactly the cost curves (in pink) in Figure 3 and Figure 26. The curves have been successfully reproduced and thus verify the validity of this model by simulating correctly the known results.

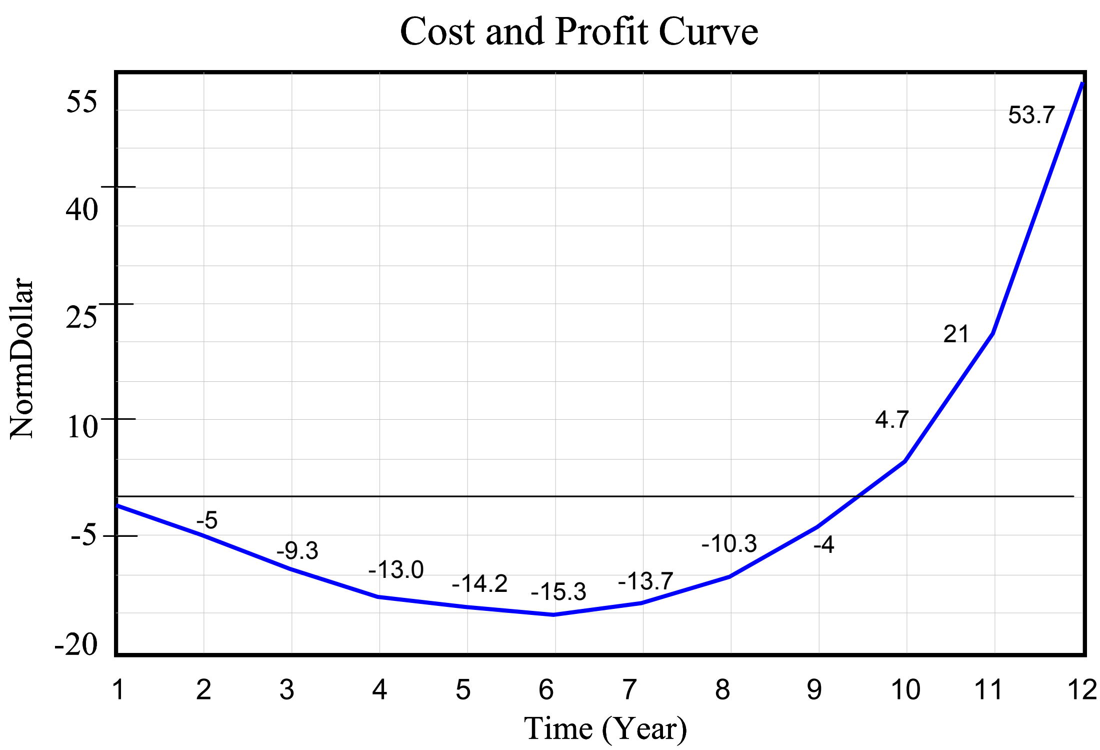


Figure 29: Payback Simulated from the Baseline Model Product Class 2 – no NCC

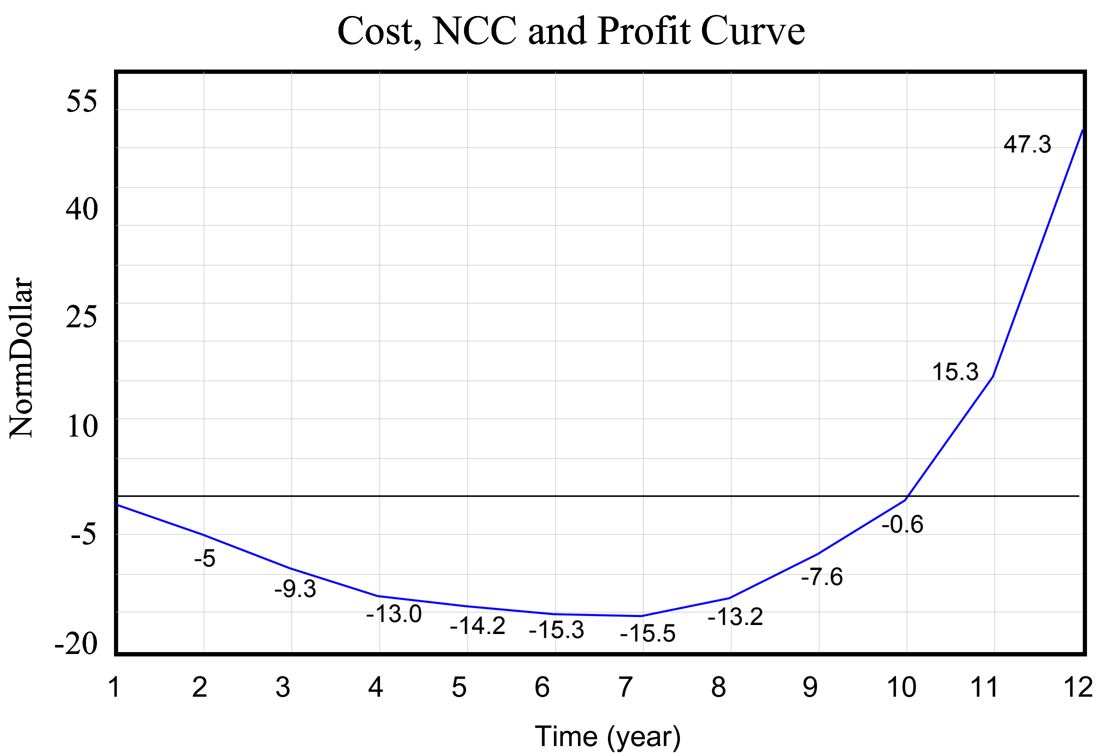


Figure 30: Payback Simulated from the Baseline Model Product Class 2 – with NCC

The main target of developing simulation models in this research is to provide a model for decision making to derive optimal testing policies. In all following simulations, the change in profit is the indicator whether a simulated testing procedure with varied years of testing and varied numbers of test products have improved the company's bottom line. It also shows in a bigger picture through feedback loops, how testing decisions made today will influence future sales, and future testing decisions, and ultimately future profit.

Scenario 2: Integration of Profit Loops and Optimization of Testing

After successful development of the base line model, the model was further extended to be capable of optimizing input such as testing length and number of test products to optimize profit. The main purposes for the scenario 2 simulation model is to find the best settings of the number of test products, the number of test intervals per year and the number of test years that optimize profit and to discuss the results with experts for verification. It shall also illustrate how cost (non-conformance cost and testing cost) is impacting the decision making process on future testing. For that reason, two feedback loops have been modeled in addition, the NCC - Profit loop and the Cost-Profit loop.

The enhanced model including the feedback loops in green and orange can be seen in Figure 31.

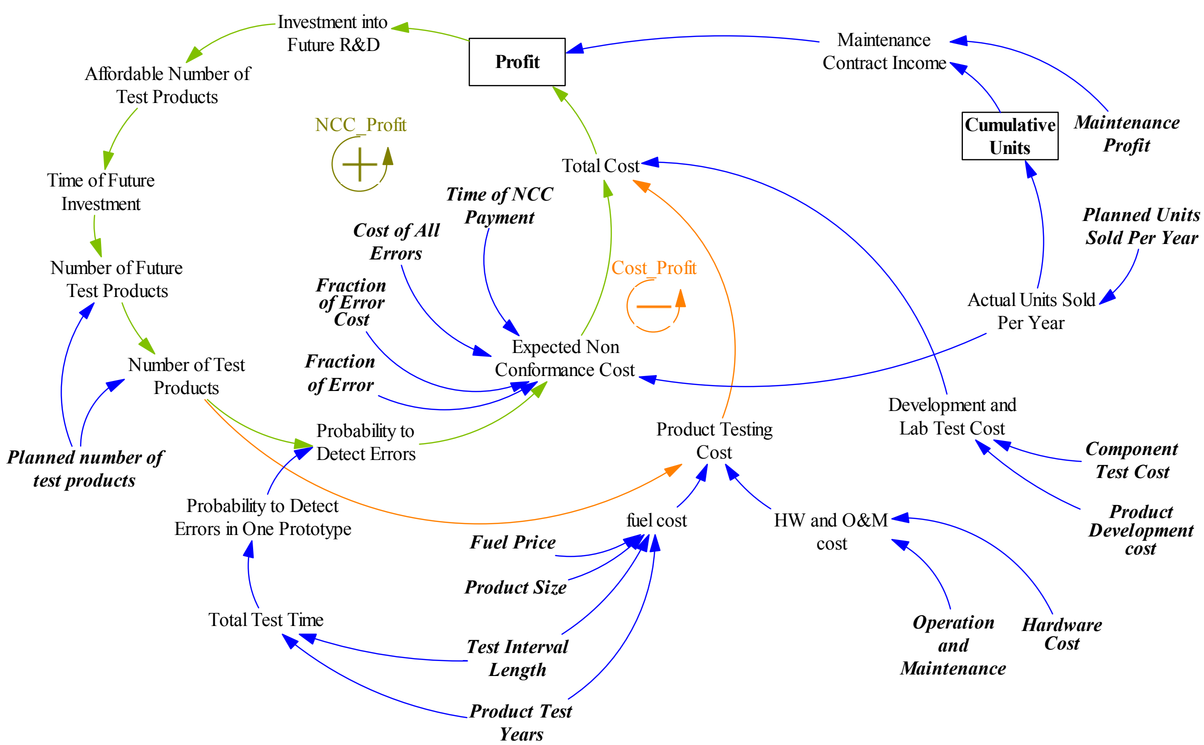


Figure 31: Scenario 2 - Optimization of Profit and Feedback Loops

The three variables with respect to testing that will be varied to see their influence on profit are

- Number of test products,
- Test length per year (in intervals of 500 hours)
- Number of years of testing.

Figure 32 illustrates the causal relationships of the three independent variables (in bold) with profit.

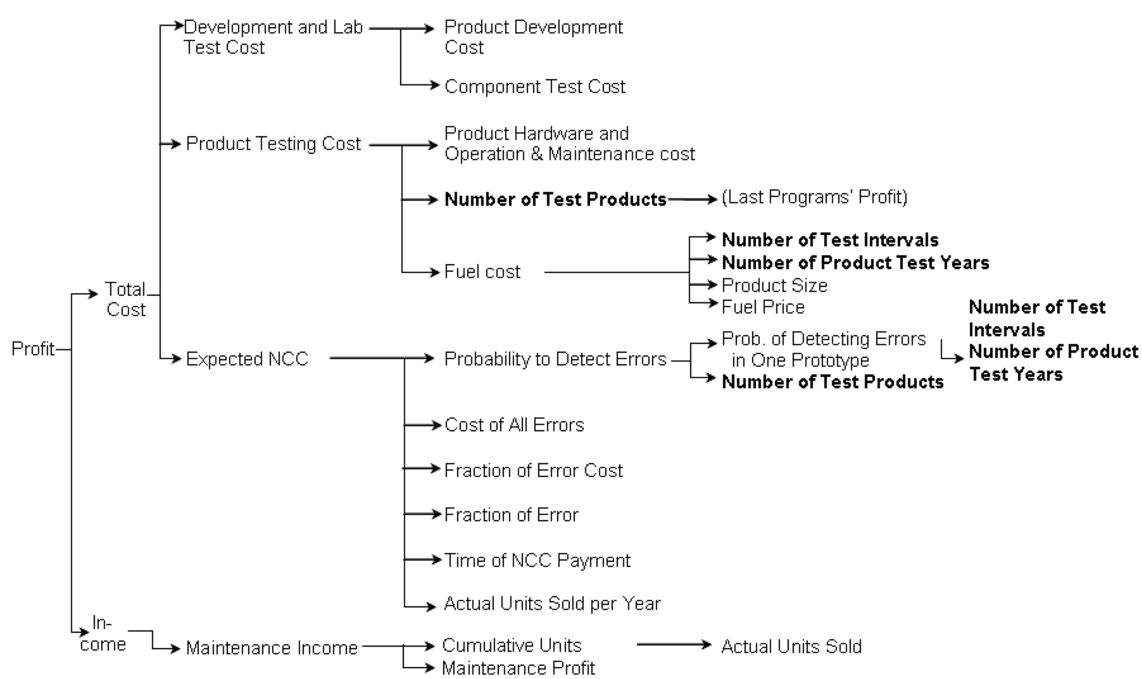


Figure 32: Causal Relationship of Profit with Independent Variables - Scenario 2

As an enhancement to scenario 1, the following causal relationships have been defined or refined:

- Fuel cost is a function of product size (set to 150 Mega Watt (MW) for this specific example), fuel price (an average natural gas price of 17 USD / MW hour, normalized in the model), testing hours (number of intervals and testing years)
- Probability to detect failures in one prototype is derived from average Weibull distributions of Product Class 2 (see Table 11).
- Cost of all errors (repair cost and replacement hardware cost for each failure class) is derived from average failure cost of Product Class 2, normalized to one product (the normalized value is 3).

The equations linking profit and income in scenario 2 can be seen in equations 10 and 11 from scenario 1. The new equations used in scenario 2 to calculate cost can be seen in Equations 16 to 19.

$$\text{Total Cost} = \text{Development and Lab Testing Cost} + \text{Product Testing Cost} + \text{Expected NCC}$$

Equation 16

$$\text{Product Testing cost} = (\text{HW and O \& M cost} + \text{fuel cost}) * \text{Total number of test products}$$

Equation 17

$$\text{Fuel cost} = \text{Fuel price} * \text{product size} * \text{number of test intervals} * \text{Product test years}$$

Equation 18

In contrast to scenario 1, fuel price is now linked to product size (which is measured in Mega Watt (MW)). One test interval has 500 test hours.

$$\text{Expected Non-Conformance Cost} = (1 - \text{probability to detect errors in all tests}) * \text{cost of all errors} * \text{Units sold per year}$$

Equation 19

Table 12: Weibull Probability Distribution of Failures

Hours	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500	6000
Avg. Prob. Of Failure Class 2	0.08	0.11	0.14	0.16	0.18	0.20	0.22	0.23	0.25	0.27	0.29	0.31
Avg Prob of Failure Class 1	0.14	0.20	0.24	0.27	0.30	0.33	0.35	0.38	0.39	0.41	0.43	0.45
Prob of Failure Avg Both Classes	0.11	0.16	0.19	0.22	0.24	0.26	0.28	0.30	0.32	0.34	0.36	0.38
Hours	6500	7000	7500	8000	8500	9000	9500	10000	10500	11000	11500	12000
Avg. Prob. Of Failure Class 2	0.33	0.35	0.37	0.39	0.41	0.44	0.46	0.48	0.50	0.52	0.54	0.56
Avg Prob of Failure Class 1	0.46	0.48	0.49	0.50	0.51	0.52	0.54	0.55	0.56	0.56	0.57	0.58
Prob of Failure Avg Both Classes	0.39	0.41	0.43	0.45	0.46	0.48	0.50	0.51	0.53	0.54	0.56	0.57

To determine the probability to detect errors (in one test), the average Weibull probabilities, in intervals of 500 hours, have been taken from Product Class 2 failure data. The data can be seen in *Table* 12, in the line with the title "avg. probabilities of failure class 2". The average Weibull probabilities have been derived by averaging the probability of each failure at 500 hours, at 1000 hours at 1500 hours and so on for Product Class 2.

The thus derived failure probabilities related to test hours were used to model the average probability of detection of failures as a function of testing length (testing years * number of testing intervals per year * 500 hours) and number of test products as can be seen in Equation 20.

$$\text{Probability to detect errors} = \text{f(testing length)} * \text{number of test products} \qquad \textit{Equation 20}$$

A testing year has been divided into eight intervals, with 500 test hours each in order to vary testing length. A calendar year has 24 hours * 365 days = 8760 hours. A commonly used time in industry for one year of product operation is 8000 hours. The difference to a calendar year comes from time subtracted for planned maintenance. A company would like to have as much time as possible for testing, however, a test year for this specific product reaches maximal 4000 hours per year, due to much more frequent down times for adjustments and part evaluations. The number of intervals has therefore been set to eight per year. This number will be varied as a constraint in one of the later scenarios.

The scenario 2 model was iterated to optimize for profit by varying years of testing, and number of test products, the results can be seen in Table 13.

The base line settings were:

Number of test products = 1;   Product test years = 2;   Number of test intervals= 8

After iterating the model, the maximum payoff was found at the following settings:

Number of test products = 1; Product test years = 3; Number of test intervals = 8

Given the sales forecast from Product Class 2, there would be a slight reduction of non-conformance cost if one conducted a three year long test with 4000 testing hours per each year and one test product. The small profit gain compared to the base line testing approach comes from avoided non-conformance cost due to longer testing (failure detection increased from 39% to 56 %, according to Table 12, Product Class 2 probability data). The higher test costs are more than off-set by NCC avoidance, as can be seen by comparing the profit of the baseline 49.7 with the profit of the optimization 50.2 (*Table* 13).

Table 13: Comparison of Profit Base Line of Product Class 2 and Profit after Optimization

Time Years	Optimization Profit (cumulative)	Baseline Profit (cumulative)
1	-1.00	-1.00
2	-5.00	-5.00
3	-9.33	-9.33
4	-13.00	-13.00
5	-14.35	-14.01
6	-17.25	-17.17
7	-17.13	-17.65
8	-13.80	-14.32
9	-7.46	-7.98
10	1.20	0.68
11	17.54	17.02
12	50.2	49.7
Number of Test Products	1	1
Years of Testing	3	2
Interval Length	8	8

The small cost advantage of 1% has been calculated from Equation 21 below.

$$\frac{\text{Profit optimized settings}}{\text{Profit base line settings}} = \frac{50.2}{49.7} = 1.01 \approx 1\% \qquad \textit{Equation 21}$$

A discussion of the results with experts in this field lead to the conclusion that an additional non-conformance cost factor was missing in the model.

So far, only failure costs that directly interrupted operation have been taken into account. However, typically over regularly scheduled downtimes, parts that had actually failed in other products within one product class will also have to be exchanged in the yet intact products, but only for a fraction of what it would have cost in the case of a break down. To calculate this fraction of cost as part of the non-conformance cost, only the replacement cost (hardware and some field work cost), no cost related to a break down (such as cover lifts, stand still times) have been estimated for each failure class and averaged over Product Class 1 and Product Class 2. The cost of such repairs is on average 40% of the failure cost of that of a break down. This factor was added to the model (named: fraction of error cost). Each failure caused on average three out of ten products to break down, which means 70% of the products will have no failure occurrence, but will need the parts replaced at a convenient time, as a risk reduction measure (named fraction of error). The model was updated with two additional variables to reflect these factors and additional costs.

- Fraction of error
- Fraction of error cost.

The refined Equation 22 for expected non-conformance cost is now:

$$
\begin{aligned}
&\text{expected non-conformance cost} = \\
&(1-\text{Probability to Detect Errors})*\text{Cost of All Errors}*\text{Fraction of Error}*\text{Actual Units Sold Per Year}+ \\
&(1-\text{Probability to Detect Errors})*\text{Cost of All Errors}*\text{Fraction of Error Cost}*(1-\text{Fraction of Error})* \\
&*\text{Actual Units Sold Per Year}
\end{aligned}
$$

Equation 22

The values for 'fraction of error' and 'fraction of error cost' based on Product Class 1 and 2 data are

- Fraction of error = 0.3;
- Fraction of error cost = 0.4.

In addition the probability to detect failures was averaged over both product classes. The averaged values can be seen in *Table* 12 in the line "probability of failure – averaged over both classes". The enhanced model is consistently using the averaged values of both product classes for probability of detection of failures and for the fraction of errors and error cost.

The enhanced model was reiterated to optimize for profit, by varying number of test products and number of product test years. The number of test intervals was kept constant. The base line settings for the variables were:

Number of test products = 1;   Product test years = 2; Number of test intervals = 8

The maximum payoff was found at the following settings:

Number of test products = 2; Product test years = 3; Number of test intervals = 8

The difference in profit over years between baseline and an optimized test plan can be seen in

Table 14. The best result with respect to profit can be achieved when investing in two test products and testing them over a three year period with 4000 hours operation time per year. The cost advantage over the base line of one test product and two test years with 4000 hours per year is 9% higher then the expected profit when following the base line testing approach as shown in Equation 23.

$$\frac{48.3}{44.3} = 1.09 \approx 9\% \qquad \textit{Equation 23}$$

Table 14: Testing Optimization Results – Enhanced Model

Time Years	Optimization	Base Line
1	-1.00	-1.0
2	-5.33	-5.0
3	-11.00	-9.3
4	-16.00	-13.0
5	-18.37	-14.0
6	-20.73	-19.8
7	-19.07	-23.0
8	-15.73	-19.7
9	-9.40	-13.3
10	-0.73	-4.7
11	15.60	11.7
12	48.3	44.3
Number of Test Product	2	1
Years of Testing	2.5	2
Interval Length	8	8

The improved testing scheme would improve a company's bottom line, despite the initial higher testing cost, due to a reduction in non-conformance cost after product commercialization. The longer testing time from 8000 hours to 10000 hours (8 intervals x 2.5 years = 20 intervals) increased the probability of detecting failures from 45% to 51%, according to Table 12. Two test products double the probability to detect failures from 45% to 100%. The cost for the longer testing time and an additional set of hardware and operating cost is more than offset by the disappearance of failures in customer products that caused non-conformance cost.

Integration of Profit Feedback Loops in Scenario 2

Each company has its own process to decide when to release a new or upgraded product class to the market. In this specific example, it is assumed that a company releases a new large capital product (LCP) class every 12 years. To do so, the company must have earned enough money during the sales period (after commercialization, from year seven on) to finance a new product development program including its testing approach. Figure 33 illustrates an income and cost distribution of two consecutive new development programs. From year one to year six, a company invests in a new project development program, but will not receive any income from this new development program. From year seven to year 12 and further, the company will earn money with that product and, potentially, may have to pay NCC. To bring a new product to market every 12 years after release of the previous one, a company has to start investing in the newest development program by year 12 and in testing at the latest by year 15 (see Figure 33). Assuming a company invests 5% of its revenue into R&D (innovation factor), which is a typical investment strategy for highly innovative companies, a company has six years time to accrue the required R&D funds until the new product development program starts (see Equations 24 to 26 below)

$$\text{Revenue} = \frac{\text{Profit}}{\text{Profit Margin}^{*}} \qquad \textit{Equation 24}$$

* Profit Margin = 14% for specific example

$$\text{Profit} = \int_0^{15\text{ years}} (\text{Income} - \text{Total Cost})dt \qquad \textit{Equation 25}$$

$$\text{R \& D funds} = \text{Revenue} * \text{Innovation Factor}^* \qquad \textit{Equation 26}$$

*Innovation factor = 5% (industry specific)

If non-conformance costs occurred during that timeframe, the company has to use part of its R&D money to correct these errors. Only the remainder can then be used for new product development.

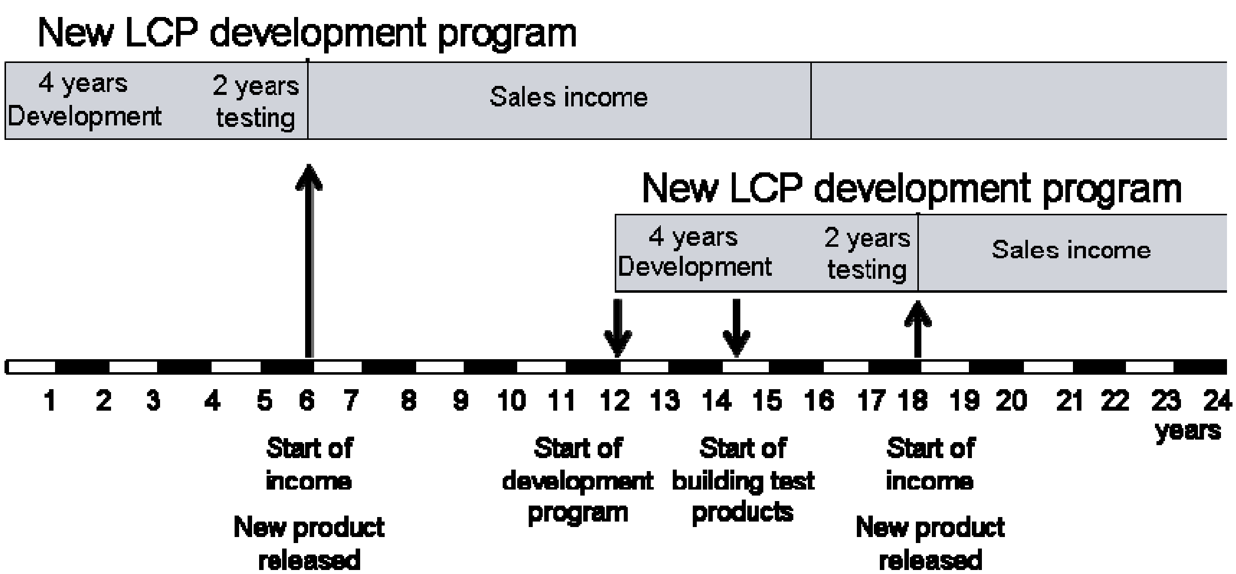


Figure 33: Cost and Income Timeline

Depending on sales, the amount of non-conformance cost and the profit a company makes, the company may perform the following steps: From year 12 to 15 the company will be investing into a new development program. Next, depending on R&D funds available, a company must decide about the testing policy. If there are enough R&D funds available at year 15, the company has the option to proceed with the economically best solution (provided they know which one that is). If there are not enough funds available, it is most likely that the company will allocate only one test product. This option, checking to see if there is or is not enough R&D money available, is modeled through the combination of the NCC-profit loop and the cost-profit loop.

These two closely related feedback loops (cause and effect loops) were integrated into the model (making it a dynamic model) to evaluate affordability of a more intense testing method. One loop is called "non-conformance cost – profit loop", shown in green in Figure 31, the other loop is called "cost – profit loop", supplementing the green loop, the supplemental part is shown in orange in the same figure. These two loops illustrate the connection from today's earnings to future testing policies.

The NCC-Profit loop (in green) together with the cost-profit loop (partly orange) calculates the affordability of test products. The NCC-profit loop is a reinforcing loop. A reinforcing loop means that when tracing the effect of a small change in one variable around the loop, the change is reinforcing the direction of the original change once it arrives at its origin. The variables involved in the NCC-profit loop are

- Number of test products
- Probability to detect errors
- Expected non conformance cost
- Total Cost
- Profit

- Investment into future R&D
- Affordable number of test products
- Number of future test products

The loop is a reinforcing loop, because increasing the number of test products from one to two doubles the failure detection probability during testing. The more failures that are detected during testing, the lower are the non-conformance cost after product release, and the larger will be the profit. Larger profit may allow a company to invest in more test products if they knew more economical test procedures.

The ‘cost-profit loop’ marked in orange, is a balancing loop. A balancing loop means that when tracing the effect of a small change in one variable around the loop, the change is causing a damping to the original change once it arrives back at its origin.

The variables involved in the ‘cost-profit loop’ are

- Profit
- Investment into future R&D
- Affordable number of test products
- (Time of future investment) – is used to set time of future investment for the simulation
- Number of future test products (number of test products that are affordable)
- Number of test products (number of test products that are planned or, if not affordable, the closest number that is affordable)
- Product testing cost

- Total cost

The loop is balancing because increasing the numbers of test products will increase the total cost (due to higher hardware and fuel cost). The higher product testing cost will lower profit. Reduced profit may negatively influence the decision on investing into more prototype test products in the future (even if more is more economical). Less investment into test products increases profit due to the lower investment and results in more R&D money. More R&D money may result in deciding to have more test products (if economically reasonable to reduce NCC). However, this investment will lower profit due to increased testing cost and so on.

Both loops are strongly interrelated, but weak, meaning that the cause and effect relationships between the variables in the loops are also influenced by external factors not integrated in this model. External factors that influence indirectly the “testing decision” may for example be that a company chooses to use the available R&D funds to invest into a second product development program instead of investing more into the first one.

The variable “affordable number of test products” is modeled as a constraint, constraining the “number of test products” to the “affordable number of test products” in the simulation.

The profit is highly dependent on the number of products sold per year. For this specific product, Table 15 gives lead numbers for different sales scenarios. These different sales forecasts have been used in the model to simulate potential R&D funds availability for future R&D projects. The simulation results of the model with respect to “affordability of testing” are shown in Table 16.

Table 15: Average Sales per Year for a Specific Product

Sales forecast per year	Baseline Product class 2	Slow	Conservative	Average	Boom
Count	1st year 5, then see Table 11	1st year 5, then 10	1st year 5, then 20	1st year 5, then 30	1st year 5, then 60

The testing settings that were used in Table 16 are described in Table 17. One can see that "number of test products", "number of test years" and "number of intervals per year" have been varied between Baseline and Optimized. Optimized is defined as the best setting of the three variables that maximizes profit. Baseline represents the settings of today's testing approach.

Table 16: Affordability of Testing in Subsequent Development Program, Based on Profit

Sales forecast	**class 2**	**class 2**	**weak**	**weak**	**conservative**	**conservative**
Setting	Baseline	optimized	Baseline	optimized	Baseline	optimized
Affordable number of test products*	9	9	3	4	9	10

* in year 15

Table 17: Setting of Input Variables for a Weak Sales Forecast

Settings	Baseline	Optimized
Number of Years of Testing	2	2.5
Number of Intervals (of 500 hours each)	8	8
Number of test products	1	2

Table 16 shows that even with the most conservative settings – a weak sales forecast assuming selling only 10 products per year from year seven to year 15 (after assuming selling five products in year seven, which is the same for every forecast), using the baseline setting with one prototype test, and 4000 hours testing per year over two years, the company would still generate enough R&D money (from Equations 24 to 26) to finance a new product development program with three test products and operation and maintenance for two years with 4000 hours per year testing time.

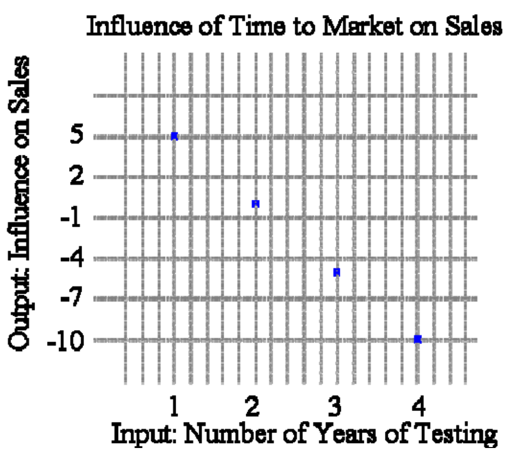
The revenue stream is large enough to generate enough R&D funds to cover the cost of one new product development program with as much as nine test products when using sales forecast of Product Class 2.

A conservative sales forecast with 20 products per year would generate enough R&D funds to finance as much as ten test products. Note that baseline in this document always means one test product, testing over 4000 hours per year over two years, representing the current testing approach. The optimal setting is generated with the optimizer of the software Vensim, by optimizing profit through the variation of numbers of test products, numbers of test years and the number of intervals per year. For example for a weak sales forecast the optimum setting would be testing with two test products for two and a half years with 4000 test hours per year using sales forecast of class 2.

However, the number of test products and the number of testing years will influence the time when a product reaches the market and generates revenue. The effect of time to market on profit will feed back into the decision on the number of testing products and the number of years to test. Scenario 3 below will attempt to capture the time to market aspect with the next enhancement of the model.

Scenario 3: Time to Market Aspect

Time to market is an important aspect when reflecting on a testing policy. It is important for the decision maker to understand the influence on profit when bringing the product to market earlier and therefore generating revenue earlier, versus reducing non-conformance cost by increasing the intensity of testing. The impact of faster time to market was calculated as follows: It was assumed that each additional year of testing (from the baseline of two years of testing) will reduce sales by five products. Being one year earlier to market (one year of testing instead of the baseline two years of testing) results in five more products sold. Five products were chosen, because it equals the sales of Product Class 2 in each of its first two years of sale. Figure 34 below illustrates the relationship between years of testing and sales.



Input	Output
1	5
2	0
3	-5
4	-10

Figure 34: Influence of Years of Testing on Sales

According to Figure 34, one year of testing results in five additional products sold. Two years of testing results in no change to the original forecast. Three years of testing means five less products sold. Four years of testing means 10 less products sold.

Table 18 shows how specific settings (number of years of testing, number of test products and number of intervals per year) influence profit under allowance of time to market. Settings one to six have fixed interval settings of eight (equaling 4000 test hours per year) while varying number of product test years and number of test products. Setting 6 “optimized” shows that the most economical solution under time to market aspects is to employ three test products and test them for one year, resulting in profit of 53.7. Results for other, less economical settings can be seen in Table 18 in settings 1 to 5.

Table 18: Integrated Time to Market Aspect and its Influence on Profit

Settings using Sales Forecast Class 2	**1**	**2**	**3**	**4**	**5**	**Optimized**
Number of Intervals*	8	8	8	8	8	**8**
Number of Product test years	2	1	1	2	3	**1**
Number of test products	1	2	1	2	2	**3**
Profit in year 12	44.3	50.1	46.5	47.2	37.6	**53.7**
Product sales time to market effect	0	5	5	0	-5	**5**
Probability of Errors after Release	0.6	0.4	0.7	0.1	0.0	**0.1**
Settings using Sales Forecast Class 2	**7**	**8**	**9**	**10**	**11**	**constraint Optimized**
Number of Intervals*	3	3	3	3	3	**3**
Number of Product test years	2	1	2	3	2	**1**
Number of test products	1	1	2	2	3	**3**
Profit in year 12	42.0	*43.9*	42.4	37.1	42.8	**46.0**
Product sales time to market effect	0	5	0	-5	0	**5**
Probability of Errors after Release	0.7	0.8	0.5	0.3	0.2	**0.4**

* 1 interval = 500 test hours

Table 18 also shows the influence on profit when the testing intervals are constrained to three per year (equals 1500 hours of testing per year). Setting "constraint optimized" in Table 18 shows that the most economical setting for "number of product test years" and "number of test products" is the same as for the previous one with interval settings to 8 (4000 hours per year), but the profit is only 46 rather than 53.7. Due to the shorter testing time, more errors remained undetected after product release, and thus reduced profit. According to *Table* 12, 19% of failures would have been detected after 1500 hours of testing compared to 30% after 4000 hours. Three test products would increase the probability of detection to 57% (3x19%) in the former case and to 90% in the latter respectively.

The constraint of three intervals was chosen, because it was the minimum test time performed per year for this specific example.

One can also see in Table 18 that reducing test time by one year off-sets the higher non-conformance cost generated due to less testing. For example setting 2, one year of testing with two test products, results in a profit of 50.1 compared to setting 4, two years of testing and two test products, results in a profit of only 47.2.

This enhanced model is portrayed in Figure 35 below. The time to market aspect is illustrated on the bottom right of the model, where one can see the new links “influence on time to market” and “sales reduction or addition”. These links contain the translation of test years into additions or reductions on sales forecast and feed into the link “actual units sold per year”, according to the diagram shown in Figure 34. When running the simulation, the reduced or increased number of sales will influence profit according to Equations 10 to 15 explained in scenario 1.

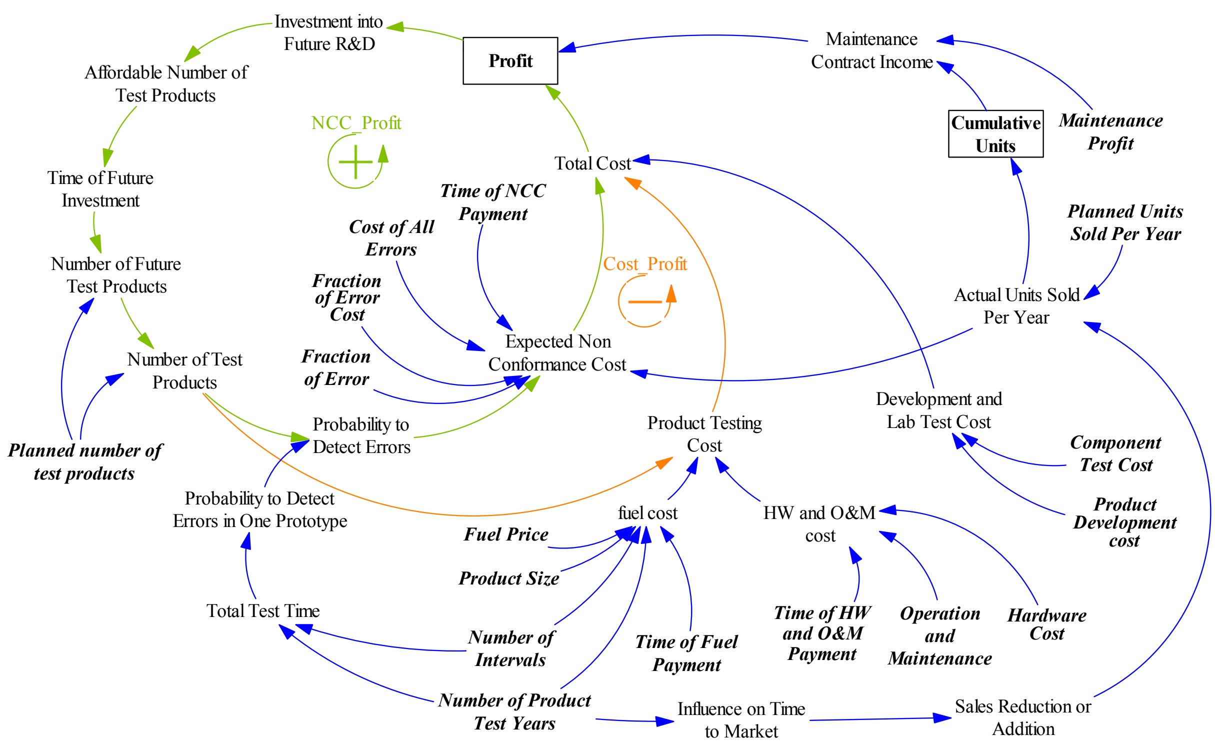


Figure 35: Scenario 3 - Integration of Time to Market Aspect

However, apart from the time to market aspect, another important influence, non-conformance cost, needs to be considered. Errors in a product will influence customer buying behavior. News of failed products travels fast, when the circle of buyers is limited, as it is with LCP buyers. Scenario 4 below attempts to capture the influence of product errors on buying behavior.

Scenario 4: Influence of Non-Conformance Costs on Buying Behavior

Scenario 4 is shown in Figure 37. The enhanced model illustrates the loop “product errors - buying behavior” in pink on the upper right corner of the model. The loop is reinforcing, because undetected errors after testing will impact the products in the field. Once the errors surface, buyers may delay their buying decision until they perceive the teething problems of the product as satisfactorily corrected. The delay in buying products will influence the number of products sold and ultimately influence profit. Since the lower profit may influence negatively the decision on a higher testing intensity, the error rate on future products will stay higher and reinforces a hesitant buying behavior at the time of product introduction. The model attempts to reflect this behavior with the two links "influence on buyer behavior" and "buyers hesitation" with the Equations 27 and 28 that are only impacting the second year after serial release.

Figure 35 shows the relationship of errors in the field (from zero to one, where one equals 100%) to sales. An error fraction between 0 and 0.24 would not affect buyers' behavior. Errors between 0.25 and 0.34 would delay purchase of one product. Errors between 0.35 and 0.44 would cause a sales reduction of two products. Errors between 0.45 and 0.64 would cause a sales reduction of three and so on. These values have been discussed with industry experts and were considered very conservative. Further discussion to this topic can be seen in the last paragraph on page 129 and in the sensitivity study in Subchapter Six: Conclusions for the Specific Model.

It is assumed in this model that the years after the second year of product release are not impacted any more by field errors and the errors that surfaced in the first year after product release are corrected the following year and are not influencing buying behavior any further. This is a conservative approach. Chapter Seven discusses this as one point for further research.

$$\text{Influence on buying behavior} = 1 - \text{probability to detect errors in the test} \qquad \textit{Equation 27}$$

$$\text{Buyers' hesitation} = \text{figure 36} \qquad \textit{Equation 28}$$

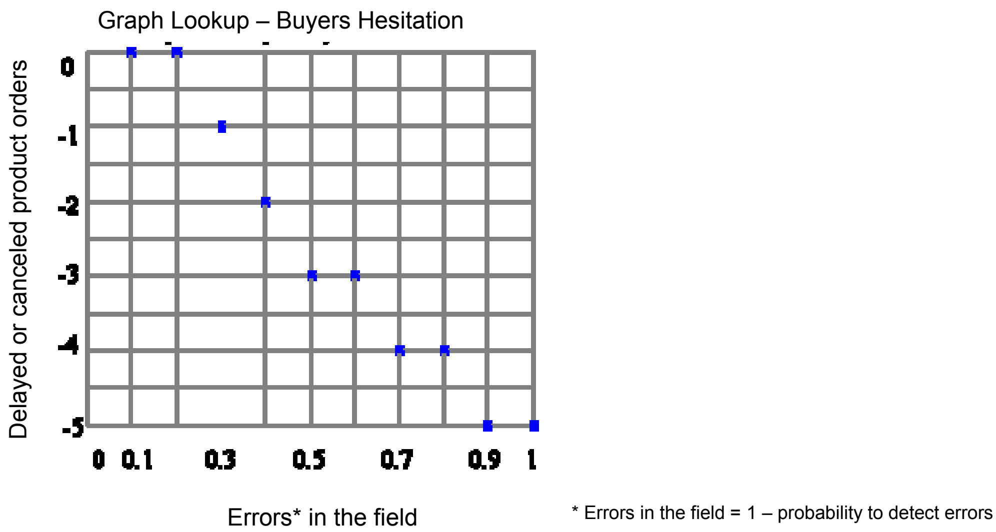


Figure 36: Relationship of Errors in the Field to Delayed Orders

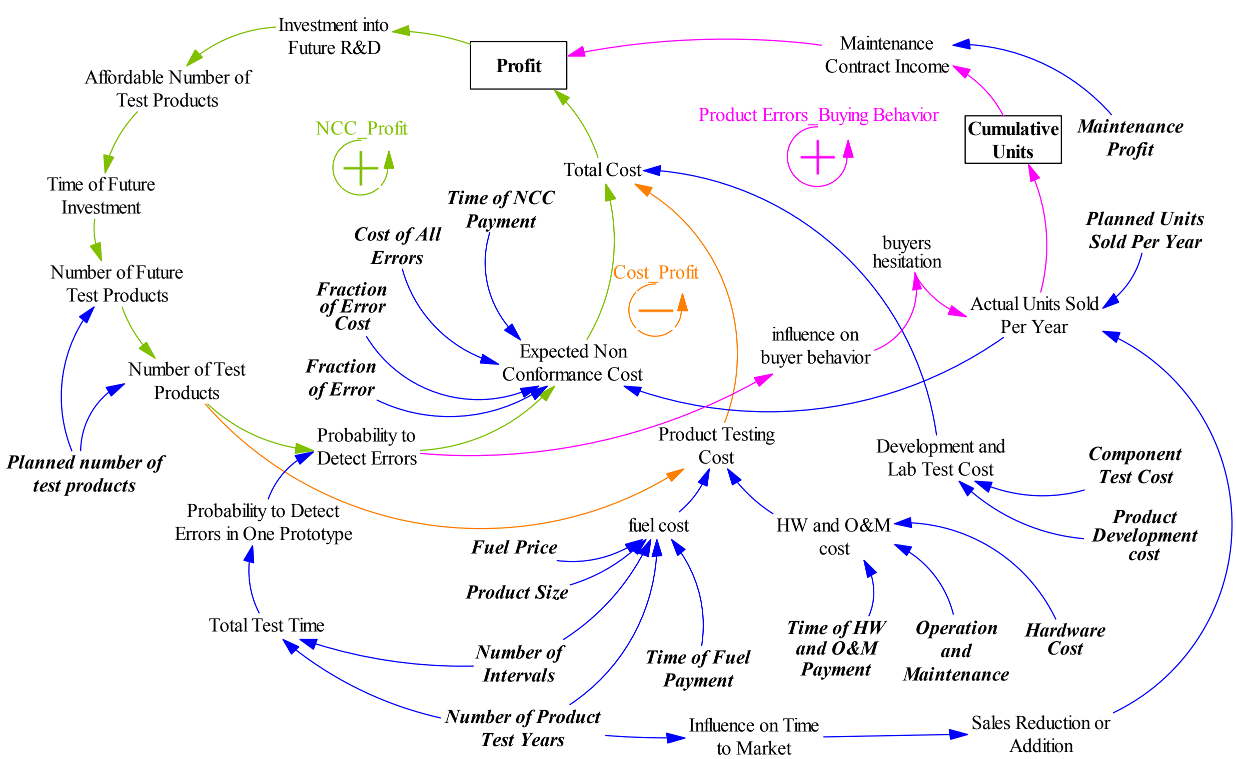


Figure 37: Scenario 4 - Impact of Product Errors on Buying Behavior

The impact of reluctant buying behavior due to perceived product immaturity can be put into perspective by re-simulating the model with the same settings as was used to show the influence of time to market on profit and compare the difference in profits.

Table 19is a reiteration of Table 18 of scenario 3 with the results of both the simulations. This table shows in addition to Table 18:

- New profit, including influence of errors on buying behavior
- Influence of reluctant buying behavior (showing reduced number of sales).

One clearly realizes in setting one, two and three that sales were reduced by three, two and four units respectively due to customer reaction to errors in the field. Settings four, five and setting called “optimized” show no impact on profit, because the majority of errors (more than 90%) was caught during testing and the remainder of errors was small enough not to impact buying behavior.

Settings seven to twelve had the number of intervals constrained to three (1500 hours testing time per year). The simulation of the model revealed for settings seven, nine and ten a reduction of product sales by four, three and one product respectively, reducing the profit by 0.2, 1.4 and 0.8 respectively.

In summary it can be said that errors in the field were reducing profit because future buyers canceled or delayed their orders until they perceived these errors as resolved. However, if the company has optimized their testing approach, not enough field errors surfaced to influence buyers’ behavior. If the company decided on any of the other settings for a testing approach for any reason, the impact was not big enough to further revise a testing approach due to customer buying behavior.

The impact of errors on buying behavior was estimated for this specific product to show that this aspect can be captured in the model. A much more thorough market analysis on customer buying behavior would be necessary to verify the impact of this aspect on the model.

Table 19: Impact on Profit for Different Settings (Testing Time, Test Products), incl. Impact of Time to Market and Product Errors on Buying Behavior

Settings using Sales Forecast Class 2	Baseline 1	2	3	4	5	6 Optimized
Number of Intervals*	8	8	8	8	8	**8**
Number of Product test years	2	1	1	2	3	**1**
Number of test products	1	2	1	2	2	**3**
Profit in year 12	44.3	50.1	46.5	47.2	37.6	**53.7**
Product sales time to market effect	0	5	5	0	-5	**5**
Profit in year 12 with reduction in sales	42.2	49.1	45.1	47.2	37.6	**53.7**
Product sales reduced by	-3	-2	-4	0	0	**0**
Probability of Errors after Release	0.6	0.4	0.7	0.1	0.0	**0.1**
Settings using Sales Forecast Class 2	**7**	**8**	**9**	**10**	**11**	**constraint Optimized**
Number of Intervals*	3	3	3	3	3	**3**
Number of Product test years	2	1	2	3	2	**1**
Number of test products	1	1	2	2	3	**3**
Profit in year 12	42.0	*43.9*	42.4	37.1	42.8	**46.0**
Product sales time to market effect	0	5	0	-5	0	**5**
Profit in year 12 with reduction in sales	40.5	*42.9*	40.7	36.0	42.8	**44.1**
Product sales reduced by	-4	-4	-3	-1	0	**-2**
Probability of Errors after Release	0.7	0.8	0.5	0.3	0.2	**0.4**

* 1 interval = 500 test hours

Verification of Results

The enhanced model in scenario 4 can be used to evaluate a company's relevant testing policies. The models in scenario one to four each builds onto its predecessor, capturing successively more of the real world issues with respect to testing. The following verification steps were performed while building the models up to scenario 4. Scenario 4 is the final model to be used for this specific company to understand and potentially modify its testing policy.

- Each scenario was verified using values from Product Class 2 and checked to see if the reproduced profit distributions were mirroring those of the first cost graph with NCC of Product Class 2.
- All causal relationships established in the model were verified with experts in this field by printing and discussing the respective graphs or tables, for example
    - Figure 34, "Influence of Years of Testing on Sales"
    - Figure 36, "Relationship of Errors in the Field to Delayed Orders",
- The table of results of each simulation run for a new scenario was compared to the simulation table of the previous scenario with the same settings. Values that differed from each other were investigated to determine if the difference can be explained by the enhancement, or if a modeling error occurred that needed to be corrected. Typical model issues were:
    - Integer numbers (units) became fractions.
    - The timing of an impact was wrong, e.g. the cost of errors and NCC should only impact the products sold in the first two years after commercialization.
    - Wrong use of functions, e.g. the step function that models the sales behavior

- All results of the individual scenarios were discussed with experts in this field and modified as required e.g.
    - Table 13"Comparison of Profit Baseline and Profit after Optimization" was reviewed and the model was updated with a better model of NCC, resulting in Table 14 "Testing Optimization Results – Enhanced Model"
    - All equations have been reviewed, and found to be correct.

Summary of Simulation Results

The second hypothesis was that a generic process can be derived and a cost model can be built for LCPs, which is adjustable to the respective industry and that supports a company's decision making on an economically optimal number of prototypes and testing length. This assumption was tested with real product data from two product classes (Product Class 1 and 2) using a systems dynamics simulation methodology.

A dynamic model for simulation was built and successively enhanced from scenario 1 to scenario 4. Product data was used and simulated for the purpose of optimizing testing policies of this specific company. It was shown that one can iterate different testing approaches by varying testing times and the number of prototype tests and thus could improve (by increasing profits) the current testing policy of this specific company. Additionally it was illustrated through simulation how product time to market influences the company's bottom line and how the amount of product errors influences buying behavior of customers, and ultimately company's profit.

CHAPTER SIX: CONCLUSIONS

Conclusions for the Specific Model

From the results of hypothesis one, it can be concluded that based on the analysis of the Weibull distributions of selected failure categories in Product Classes 1 and 2 it is very unlikely that all failure categories will surface during one prototype test with a testing time between 2000 hours and 8000 hours over a calendar time of two years, and therefore it is expected that a company will face significant cost after commercialization. The average probability of failure detection lies between 27% and 50% in Product Class 1 and between 16% and 39% in Product Class 2 for the timeframe of 2000 and 8000 hrs respectively. This means that statistically only every third to fifth failure category is detectable in one test with a length of 2000 hours and only every second or third error will be detected in a test length of 8000 hours. This trend is confirmed by the visual data analysis of Product Class 1 and 2, demonstrating that in Product Class 1 only 54% of all first failures occurred in P1 while the remaining 46% were distributed over six following products.

In Product Class 2, the first 50% of failures were distributed over the first six products sold, P1 having only about ten percent of all first failures. According to the two product classes investigated, a company has to sell seven to nine more high priced, LCP with all errors resolved, to make up for the non-conformance costs occurring after serial release.

From the results of hypothesis two it can be concluded that a process can be derived and a cost model can be built for LCPs of a specific industry that support a company's policy on decision making with respect to an economically optimal number of prototypes and testing length.

The model was developed in scenarios of increasing complexity, and each scenario was verified with data, where available, and expert knowledge. The model in scenario one was set up to vary the number of test products, test length and sales forecast. Test length is composed of the number of intervals with 500 test hours per interval per year, and the number of test years.

The baseline was the current testing approach, using one test product, testing it between 2000 and 8000 hours, and using the actual sales of Product Class 2.

Profit was used to show if improvements can be reached by varying the above described variables.

The model was enhanced in scenario 2 to allow estimating the amount of future R&D funds available for next generation product. This aspect was important with respect to affordability of more than one test product, if economically needed to improve profit. The result was that even with the baseline setting and a weak sales forecast, a company selling this specific product should generate enough R&D funds to afford the economically best testing approach in their next R&D project.

The model was enhanced to integrate the aspects of time to market as shown in scenario 3.

Each year that a product can be released to the market earlier, the company starts generating profit on that product earlier through operation and maintenance contracts. If this is at the cost of testing time, the company will face errors in customer products and resulting non-conformance cost. Introducing more test products reduces the amount of errors in the field, but adds more initial cost (R&D cost). The optimum (with respect to profit generation) for this specific example was to test three products over one year each for 4000 hours (one year) compared to one test product and 8000 hours (two years) of testing.

The last enhancement was to model the influence of product errors after product introduction on future customer's buying behavior, modeled in scenario four. This aspect does not influence profit for the optimal settings in this specific example, because only an error fraction of 0.1 or smaller remains to be detected in customer products. However, it does influence profit in today's settings, because the failure fraction in customer products is big enough to reduce sales forecasts by two to four products in the second year after product release.

The baseline for the model calibration came from data of two product classes to which this researcher had access. The data of Product Class 2 was used to calibrate the model in scenario 1. Both data sets were then combined (averaged) to calibrate the enhanced models in scenario two, three and four.

The models are deterministic, because no data uncertainties have been accounted for in the model.

The following constants used in the model will be varied in scenario four to determine if there was a statistical difference between the two models and also to determine the sensitivity of the model to small changes in the parameter settings and to see the confidence interval of the model with respect to profit. A Monte Carlo simulation technique, (changes of all parameters together) was used to determine the confidence interval with respect to profit.

The two testing settings that were compared with each other were:

- Three test products, with 4000 test hours per year, one year of testing (**Profit was 53.7**, setting is called “optimized setting”)
- One test product, with 4000 test hours per year and two years of testing (**Profit was 42.2**, setting is called “baseline setting”)

The values of the constants and the small variation of constants can be seen in parenthesis behind the listed parameter below. Since it is unknown which shape the uncertainties have, a random uniform distribution was assumed:

- Maintenance profit (0.3 ±5%);
- Cost of all errors (3 ±5%);
- Fraction of error (0.4 ±5%);
- Fraction of error cost (0.3 ±5%);

The confidence bounds of both sets of settings (optimized and baseline) can be seen in Figure 38.

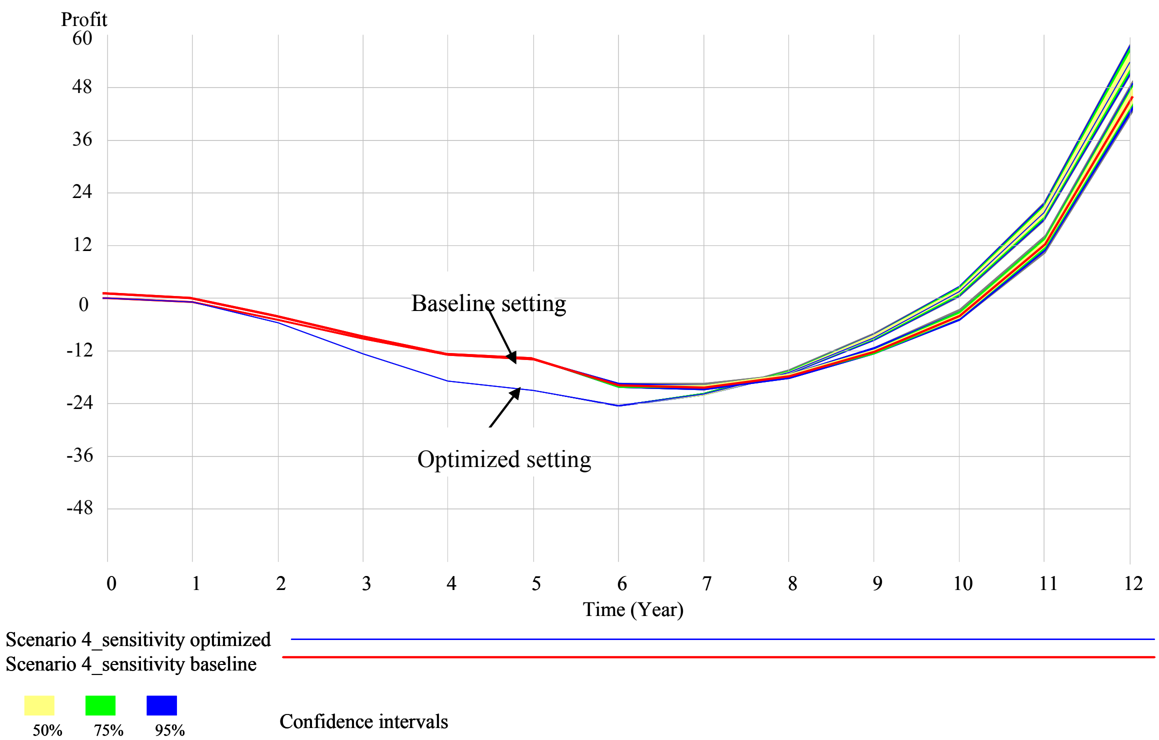


Figure 38: Sensitivity Analysis of Baseline and Optimized Testing with Respect to Profit

The red solid line is the profit development over time for the baseline testing setting; the blue solid line is the profit development with the optimized setting. The lines around these curves are the confident bounds for the output variable PROFIT. The 95% confident intervals (in blue) hugging each profit development curve do not overlap in year 12, indicating that the difference between the two profits is statistically significant. The variation (95% confidence bound) around both profit values is in absolute value about ± 3. These values have been taken from the graph in Figure 39, which is the enlarged version of Figure 38 showing profit only for the years 11 and 12.

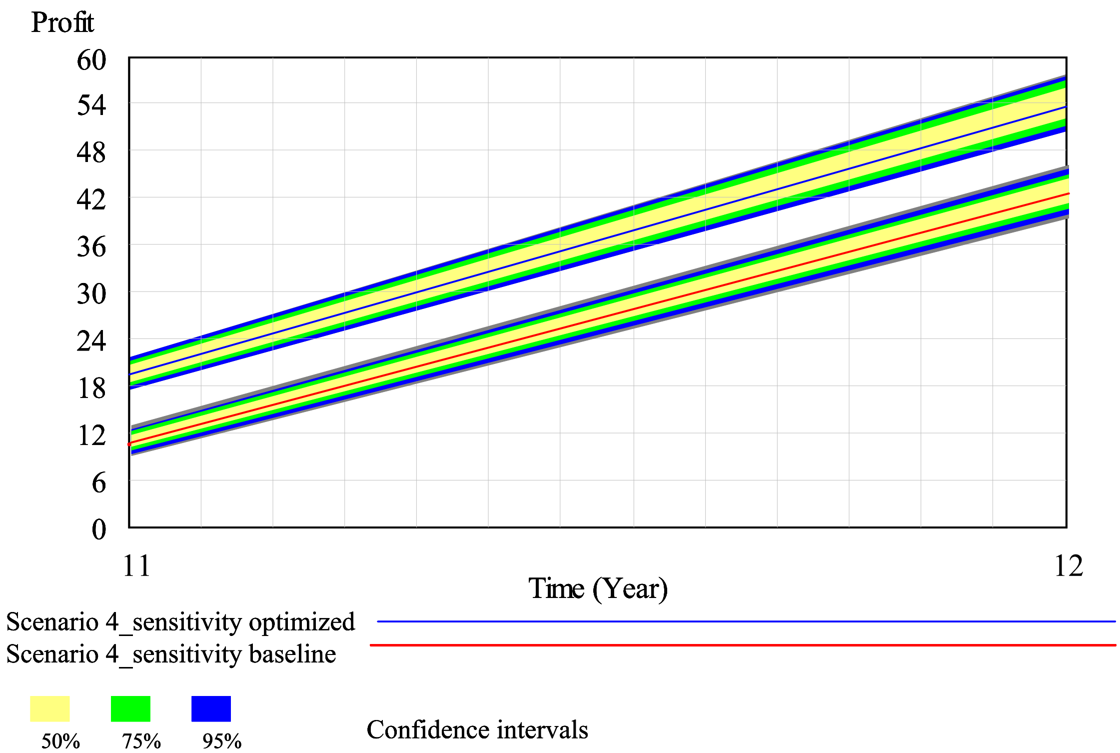


Figure 39: Confidence Intervals of Profits (Optimized and Baseline Setting)

A sensitivity analysis on parameters has been conducted to understand which parameters (those that were fixed during simulation of the model) would influence profit most if they would change over time, and must therefore be closely watched, when using the model over the years.

All following parameters have been varied by a small amount, one at a time, to understand their influence on profit if they would change over time.

These parameters have been varied by plus and minus five percent

- Maintenance profit
- Cost of all errors,
- Fraction of error,

- Fraction of error cost.

These parameters have been varied according to the Tables 20, 21 and 22 below:

- Probabilities to detect errors
- Influence on buying behavior
- Influence on time to market.

Table 20: Data Table for Buyers Hesitation

Buyers hesitation										
Error rate after product release Original	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Reduction in Sales	0	0	-1	-2	-3	-3	-4	-4	-5	-5
Error rate after product release variation down	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Reduction in Sales	0	0	0	-4	-4	-4	-4	-4	-4	-4
Error rate after product release	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Reduction in Sales variation up	0	0	0	-5	-5	-5	-5	-5	-5	-5

Table 21: Data Table for Probability to Detect Errors

Probability to detect errors in one prototype												
Intervals (500 hours)	1	2	3	4	5	6	7	8	9	10	11	12
Original Probabilities of error detection	0.112	0.156	0.189	0.217	0.241	0.264	0.284	0.304	0.323	0.341	0.359	0.377
Intervals (500 hours)	13	14	15	16	17	18	19	20	21	22	23	24
Original Probabilities of error detection	0.394	0.412	0.429	0.446	0.463	0.480	0.496	0.512	0.528	0.543	0.558	0.572
Variation +5%	0.117	0.164	0.199	0.228	0.253	0.277	0.299	0.319	0.339	0.358	0.377	0.396
	0.414	0.432	0.451	0.468	0.486	0.504	0.521	0.538	0.554	0.570	0.586	0.601
Variation -5%	0.106	0.148	0.180	0.206	0.229	0.250	0.270	0.289	0.307	0.324	0.341	0.358
	0.375	0.391	0.408	0.424	0.440	0.456	0.471	0.486	0.501	0.516	0.530	0.543

Table 22: Data Table for Time to Market

Influence of time to Market														
Years of testing	**0**	**1**	**1.5**	**1.8**	**2**	**2.3**	**2.5**	**2.8**	**3**	**3.3**	**3.5**	**3.8**	**4**	**4.5**
Reduction or addition to sales original	5	5	2	1	0	-1	-2	-3	-5	-6	-7	-8	-10	-10
Reduction or addition to sale variation up	6	6	3	2	0	-2	-3	-4	-6	-7	-8	-9	-10	-10
Reduction or addition to sale variation down	4	4	1	0	0	0	-1	-2	-4	-5	-6	-7	-9	-9

The sensitivity results in form of a data table can be seen in *Table* 23. The tables show the results of the sensitivity study for the optimized setting on the left hand side, and for the baseline setting on the right hand side. The table shows for each of the two settings the parameter names on the left hand side, then next to it the profit when the parameter was changed by plus five percent or according to the Tables 20 to 22, and in the next column the change of profit in percent. Then follows in the next column the profit when the parameter was changed by minus five percent (or according to the Tables 20 to 22), followed by the change in profit in percent in the last column. The change of profit in percent was calculated for the optimized setting and for the baseline setting as shown in the equations below:

$$\text{change in \% (Profit)} = \frac{\text{profit (optimized setting) with changed constant}}{\text{profit (optimized setting)}} \qquad \textit{Equation 29}$$

$$\text{change in \% (Profit)} = \frac{\text{profit (base line) with changed constant}}{\text{profit (base line)}} \qquad \textit{Equation 30}$$

Profit with an optimized test setting had the value of 53.66 (normalized, with 1 = cost of a product), where the optimized test setting was one year of testing using three test products. Profit with baseline testing setting had the value of 42.22 (normalized, with 1 – cost of a product), where the baseline test setting was two years of testing with one test product.

Table 23: Sensitivity of Profit with Respect to Small Change in Parameters

Parameters and uncertainties	Optimized settings				Baseline Settings			
	Profit +5%	**Change**	**Profit -5%**	**Change**	**Profit +5%**	**Change**	**Profit -5%**	**Change**
Maintenance profit (0.3 ±5%);	57.61	-7.4%	49.71	7.4%	45.42	-7.6%	39.02	7.6%
Cost of all errors (3 ±5%);	53.54	0.2%	53.77	-0.2%	41.88	0.8%	42.56	-0.8%
Fraction of error (0.3 ±5%);	53.62	0.1%	53.69	-0.1%	42.12	0.2%	42.33	-0.3%
Fraction of error cost (0.4 ±5%);	53.37	0.5%	53.47	0.4%	42.06	0.4%	42.39	-0.4%
Probabilities to detect errors during testing +/- 5%	54.85	-2.2%	52.47	2.2%	42.49	-0.6%	41.95	0.6%
Influence on buying behavior if Errors > 0.4 --> -5 if Errors >0.4 --> -4	53.66	0.0%	53.66	0.0%	40.82	3.3%	41.52	1.7%
Influence on time to market; x -1 if > two years testing x+1 if < two years of testing	55.5	-3.4%	51.81	3.4%	42.22	0.0%	42.22	0.0%
Probabilities to detect errors during testing +/- 5%	54.85	-2.2%	52.47	2.2%	42.49	-0.6%	41.95	0.6%

Another representation of the data can be seen in the Pareto graphs (Figure 40 and Figure 41) below.

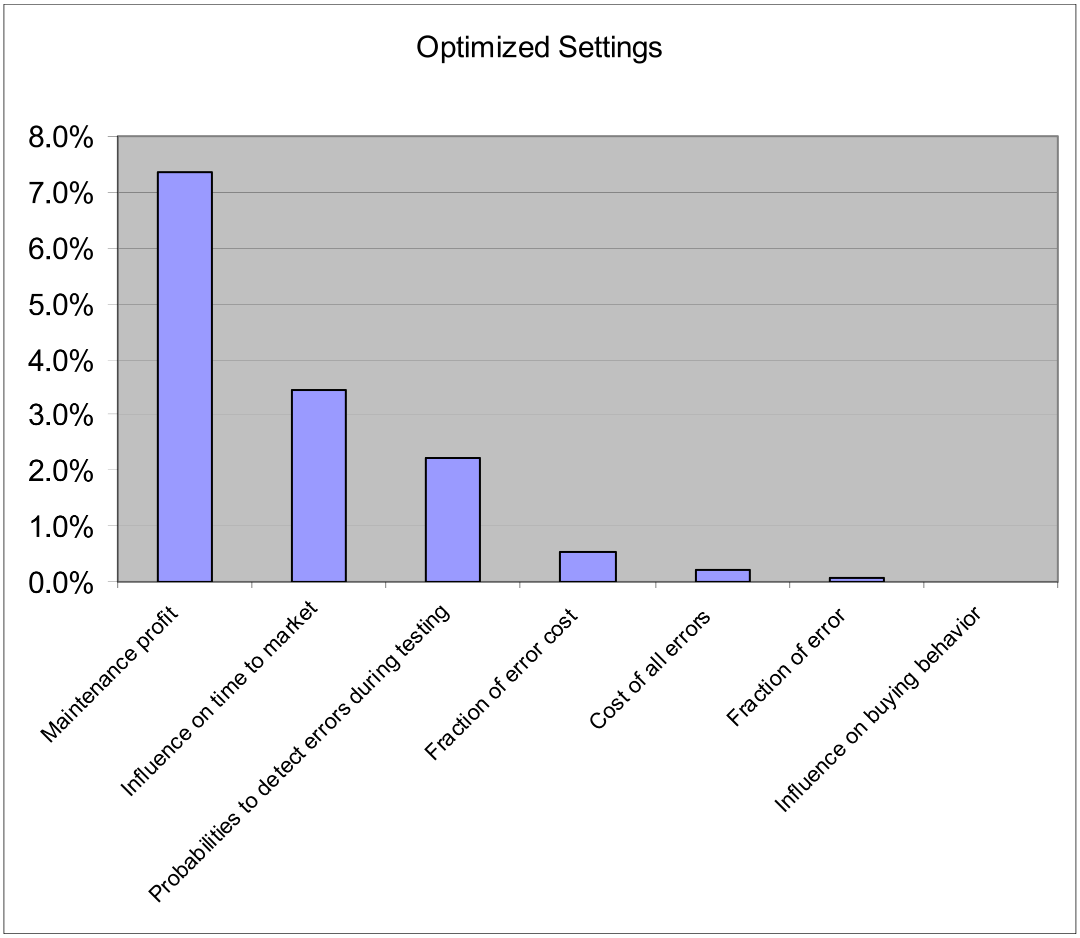


Figure 40: Pareto Chart with Sensitivities of Optimized Test Settings

One can see in the figure above that "maintenance profit" has the biggest influence on profit. If this value changes by five percent, then profit changes by 7.5 percent. The second most important factor is the influence of time to market on profit. If the assumption on time to market changes according to Table 22, then profit will change by roughly 3.5 percent. If the "probabilities to detect errors" changes by five percent, then profit will change by two percent.

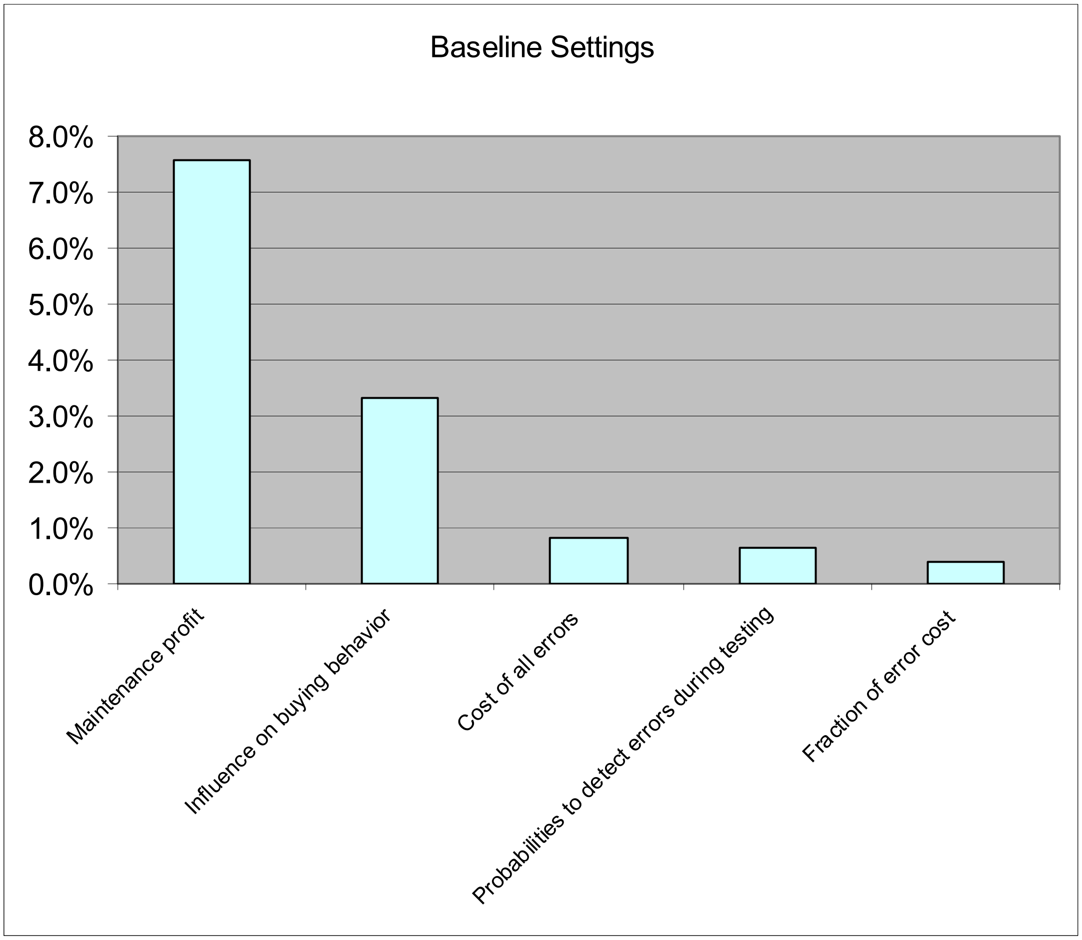


Figure 41: Pareto Chart with Sensitivities for Baseline Test Settings

For the baseline test settings maintenance profit also has the highest influence on profit. There is no influence of time to market, because the two year testing duration in the baseline testing approach is zero. The buying behavior impact is larger for the test settings than it is the case for the optimized test settings, because baseline testing approach only detects 45% of errors. If the assumed influence on buying behavior changes from three to five, then profit changes by three percent. This parameter has zero influence when using the improved testing approach, because less then ten percent of errors will remain in customer products.

In summary it can be said that the specific dynamic model is capable of supporting a company's decision on a policy on testing. In this specific example a company could use this model to revisit its existing testing policy and consider incorporating more test products prior to serial release of their next product class. The advantage would be faster time to market and a reduced error rate when the product is serially released with higher customer satisfaction. This would positively affect the company's bottom line.

Conclusions for the Generic Model

The conceptual model is the foundation for the creation of a company specific dynamic simulation model that supports a company's quantitative evaluation of their current testing approach with respect to testing length and number of prototype products. The generic process to build a dynamic model is illustrated in Chapter Four "Building the Conceptual Model."

It describes the theory of building company specific models and points out specific areas where a company for example has to define their own feedback loops appropriate for their specific product class.

Regardless of the industry, there is a need to collect

- Error-related data (time to failure and definition of error classes)
- Cost – related data (development cost, testing cost, hardware cost, cost of errors)
- Income – related data (profit margin, sales).

Once these data are collected for at least two product classes the analysis part can start, resulting in the distributions of probability of errors. These distributions are used to determine one average distribution of probabilities of failure occurrences for the model.

During the conceptual stage, a company has to conceptualize what feedback loops are important to be modeled. Data have to be collected that define the relationships of feedback loops with the rest of the model.

A common step for all industry is then the optimization of testing settings for profit. The optimization can use the model to compare different scenarios with each other, for example changing variables such as different buying behavior, different sales forecast and so on.

This concept has been applied to build a company specific simulation model that can be used to evaluate the company's current testing policy, as illustrated in Chapter Five and Subchapter Six: " Conclusions for the Specific Model."

It can be concluded that the generic model enables a company to derive a useful specific model. However, industry specific knowledge with respect to the product class under investigation and customer knowledge with respect to buying behavior is of great importance to developing a meaningful dynamic model that can support a company's testing policy. The specific dynamic model derived by following the generic model development process can be further extended to company specific needs.

CHAPTER SEVEN: FURTHER RESEARCH RECOMMENDATION

The generic model and the specific model (used to validate the generic model) have some areas into which further research would be useful.

Net Present Value Consideration and Financing

Simulation of a model that spans fifteen years and more could benefit from a time-value of money approach. Modeling the time value aspect of money would allow for more accurate investment decisions. Also aspects such as financing instead of cash payment could influence investment decisions and would be of value in a dynamic model.

New Technologies Impact

In the model that was developed, it is assumed that the next generation product is very similar to the previous product class with respect to degree of innovation. All cost information and error information was from product classes with technology enhancement rather then breakthrough new technology.

Giving the user of the model the option to select from different degrees of innovation would enhance the model and increase is usefulness. Especially the probability to detect errors and the cost of errors needed to be adapted to different scenarios of innovation. It would also necessitate changing some parameters used, for example the development cost needed to be revisited, or as a result of a more innovative product, the margins on income.

Customer Buying Behavior

The model uses a simple relationship between product errors in the first year of product release and customer buying behavior. This part of the model could be more refined through customer interviews and market research. For example product errors may be spread out over several years after product release. The time to correct errors and the time to refurbish the product with the improved part and cost involved could be investigated and implemented into the model.

Competitor Reaction

So far, the reaction of competitors to time to market aspects and product errors after product release has not been accounted for. It would be a useful enhancement of the model to integrate the relationship between these two.

Learning Cycle

Extend the model to get insight into how to reduce the learning cycle time. Define the influencing factors that build the learning cycle time. It may reveal potentials to reduce this cycle time, with potentially huge financial benefits.

LIST OF REFERENCES

Electronic Sources

Your Online Resource for Recalls. March 2008. from www.recalls.gov

Non-Electronic Sources

Balachandra, R., & Friar, J. H. (1997). Factors for success in R&D projects and new product innovation: A contextual framework. *IEEE, Transactions on Engineering Management, 44*(No.3), 276-287.

Barclay, I. (2002). Organisational factors for success in new product development. *IEE Proceedings –Science, Measurement and Technology, 149(2),*105-112.

Belliveau, P., Griffin, A., Somermeyer, S (2002). *The PDMA Toolbook for New Product Development*. New York: John Wiley & Sons. Inc.

Bowersox, D. J., Stank, T. P., & Daugherty, P. J. (1999). Lean launch: Managing product introduction risk through response-based logistics. *Journal of Product Innovation Management 16*, 557-568.

Bobrow, E. (1997). *New product development.* New York: Alpha Books

Burgelman, R. A., & Siegel, R. E. (2007). Defining the minimum winning game in high-technology ventures.*California Management Review 49*(3), 6-26.

Componation, P., Utley, D., & Armacost, B. (1999). Prioritizing components of concurrent engineering programs to support new product development. *John Wiley & Sons, Inc. CCC,* 168-175.

Cooper, R. G., & Kleinschmidt, E. J. (1986). An investigation into the new product process: Steps deficiencies, and impact. *Journal of new product innovation management, 3*, 71-85.

Cooper, R. G. (1979). Identifying industrial new product success: Project NewProd. *Industrial Marketing Management, 8*(2), 124-135.

Cooper, R. G., & Kleinschmidt, E. J. (1987). New products: What separates winners from losers?. *Journal of Product innovation management, 4*, 169-184.

Fluck, M. (2005). Gas turbines – a flexible workhorse. *Energize, April 2005, p.46-47*

Freeman, C. (1982). *The economics of industrial innovation*. Cambridge, MA: MIT Press, 1982

Forrester, J (1994). Systems Dynamics, Systems Thinking, and Soft OR. *System Dynamics Review, 10(2),* 1-13

Gibson, C., & Bhatnagar, T. (2006). *Austerlitz to borodina* (Capital Goods, Global Power Technology, 18 Oct. 2006HSBC Global Research, Industrials.

Groenland, E. (2002). *Qualitative research to validate the RQ-demensions.* Corporate Reputation Review, 4(4), 308-315.

Huppmann, H., Groetschel U., Mueller, G. (1998). *Gas Turbines - Continuous Trend Towards higher Performance -The Insurance Aspects.* IMIA conference Interlaken, September 1998

Kiesow, H.-J. (2008). *Interview April 2008*. Orlando.

Kirkwood, C (1998). New Product Dynamics – Illustrative System Dynamics Models, *Arizona State University, College of Business*

Lechner, C., Seume, J., Adam, G., Balling, L., Bals, H., Bauer, A., et al. (2002). *Stationaere Gasturbinen*. Berlin: Springer.

Lewis, E. E. (1994).*Introduction to reliability engineering*. New York: J. Wiley and Sons

Lynn, G. S., Abel, K. D., Valentine, W. S., & Wright, R. C. (1999). Key factors in increasing speed to market and improving new product success rates. *Industrial Marketing Management, 28*(4), 319-326.

Maidique, M. A., Zirger, B. J. (Nov. 1984). A study of success and failure in product innovation: The case of the U.S. electronics industry. *IEEE, Transactions on Engineering Management, EM-31*(4), 192-203.

Mansfield, E. (Nov-Dec 1981). How Economists See R&D. *Harvard Business Review, 59*(9), 98-106.

Mendenhall, W.; Sincich, T. (1994). *Statistics for Engineering and the Sciences.* Upper Saddle River, New Jersey, Prentice Hall

Messica, A., & Mehrez, A. (2002). Time-to-market, window of opportunity, and salvageability of a new product development. *Managerial and decision economics, 23*, 371-378.

Nevens. (May-June 1990). Commercialization technology: What the best companies do. *Harvard Business Review*, 154-163.

Ofek, E., & Sarvary, M. (2003). R&D, marketing, and the success of next-generation products. *Marketing Science, 22*(3), 355-370.

Porter, M. (1979). The structure within industries and companies' performance. *The Review of Economics and Statistics MIT Press, 61*(2), 214-227.

Rausand, M; Hoyland, A. (2004). Systems Reliability Theory – Models, Statistical Methods, and Applications. Second edition.Hoboken, New Jersey. Wiley and Sons Inc.

Repenning, N (2003). Selling system dynamics to (other) social scientists. *Sys. Dyn. Rev,19(4),*303-327

Repenning, N (2000). A dynamic model of resource allocation in multi-project research and development systems. *Systems dynamics Review, 16,* p173-211

Rogers, E (2003). *Diffusion of Innovations 5th edition.* New York: Free Press, A Division of Simon & Schuster, Inc.

Severance, F. (2001). *System Modeling and Simulation: An Introduction*. West Sussex: J. Wiley & Sons Ltd.

Sisk, M. (2001). The Value of reputation: As Sectors Converge, Financial Reputations Take Center Stage. *Financial Service Marketing, 3 (6): 23, September 2001*.

Smith, D. (2007). *Reliability Maintainability and Risk.* Burlington, MA: Elsevier

Sterman, J. (2000). Business Dynamics Systems: Thinking and modeling for a complex world. Boston: Irwin McGraw-Hill

Swift, K. G., Raines, M., & Booker, J. D. (1997). Design capability and costs of failure. *Proceedings Instn Mech Engrs,211(Part B),* 409-423.

Sachon, M., & Cornell P. (2004). Managing technology development for safety-critical systems.*IEEE Transactions on engineering management, 51*(4), 451-462.

Urban, G. L., & Hauser, J. R. (1993). *Design and marketing of new products* (2nd ed.). New Jersey: Prentice-Hall, Inc.

Wheelwright, S. C., & Clark, K. B. (1992). *Revolutionizing product development*. New York: The Free Press.

Wood, B. (2002), Propulsion for the 21st Century – RS-68. AIAA 2002-4324.

Yoon, E., & Lilien, G. L. (1985). New industrial product performance: The effect of market transactions and strategy. *Journal of Product Innovation Management, 3*, 134-144.