MASTER

Development and application of a decision model for synchronizing condition-based maintenance at ASML

Timmermans, B.L.R.

Award date:
2012

Eindhoven, November 2012

Development and application of a decision model for synchronizing condition-based maintenance at ASML

by
Bas Timmermans

BSc Industrial Engineering – TUE 2010
Student identity number 071211

in partial fulfillment of the requirement for the degree of

Master of Science
in Operations Management and Logistics

Supervisors:
dr. Hao Peng, TU/e, OPAC
prof.dr.ir. Geert-Jan van Houtum, TU/e, OPAC
ir. Gert Streutker, ASML, CS Marketing

TUE. School of Industrial Engineering.
Series Master Theses Operations Management and Logistics

Subject headings: proactive maintenance, condition based maintenance, maintenance planning, parameter monitoring, degradation modeling, renewal theory, downtime, cost reduction

Abstract

This master thesis is concerned with opportunistic maintenance on a high tech, complex, capital intensive system that is subject to random failures, deterioration, and high downtime cost. One of the system components is subject to continuous condition monitoring, which implies that the monitored component's condition, status or "health" is known. Once the monitored condition exceeds a predetermined critical limit, a scheduled down (SD) of the system is arranged as quickly as possible. During this SD preventive maintenance (PM) is performed on the monitored component. Besides the condition-based maintenance (CBM) of the monitored components, the other system components are subject to failure-based maintenance (FBM), and periodic maintenance (PerM).

For the monitored component a control limit for opportunistic maintenance is introduced. The setup of this limit implies that once it is crossed, the monitored component becomes eligible for opportunistic preventive maintenance (OPM). OPM is PM that is combined with the maintenance of another system component, which can be either FBM or PerM. Each time maintenance is performed, the system is down, which leads to high downtime cost. By introducing the OPM control limit, and performing CBM simultaneously with either FBM or PerM, the downtime of the system can be reduced. This is called *synchronizing* CBM with both FBM and PerM.

Using degradation modeling and Renewal theory, a mathematical optimization model is constructed that attempts to find the OPM control limit that minimizes the expected total maintenance cost. A case study is performed that describes the application of this model in the current maintenance operations of ASML. It is shown how ASML benefits from using the condition data for synchronizing CBM, and how the results can be implemented.

keywords and phrases proactive maintenance, condition based maintenance, maintenance planning, parameter monitoring, degradation modeling, renewal theory, downtime, cost reduction

Preface and acknowledgements

This report is the result of my master thesis project in completion of the Operations Management and Logistics program at Eindhoven University of Technology. This report also marks the end of my internship at the Customer Service department at ASML. I want to use this opportunity to express my gratitude to many people for their help and support.

First off, I would like to thank my first supervisor Hao Peng, for guiding me through the process, her constructive feedback and critical reviews, and especially for her confidence in me. Also, I would like to thank my second supervisor Geert-Jan van Houtum for introducing me to ASML and ProSelo, for his critical view on the project, and the useful feedback.

At ASML, I would like to thank Gert Streutker, my supervisor at ASML, for giving me the opportunity and the freedom to conduct my research at ASML. Gert's enthusiasm, experience, anecdotes, and insights have been very valuable to me and my project. I also want to thank David Sigtermans for his time and guidance. He always helped me out when I was in doubt or asked for another data set. Furthermore, I thank Fred Hallebeek for sharing his insights on ASML's preventive maintenance practice, and Joost Beke for almost instantly providing me with the downtime reports. Lastly, I thank all colleagues and fellow interns at ASML for the pleasant working atmosphere and the great time I have had.

I would like to thank my dear family and friends, for their support and interest during my entire study and this project in particular. Special thanks to my mother, brother, sister, and girlfriend for their unconditional love, support and understanding. Finally, I want to dedicate this thesis to my father, who unexpectedly passed away during my studies. He is deeply missed.

Bas Timmermans

Leveroy, November 2012

Executive summary

This report is the result of a master thesis project conducted at ASML customer service. ASML is the world's leading provider of lithography systems for the semiconductor industry. A lithography system can be characterized as a highly technical, advanced, and complex capital good with an extremely high commercial price. The opportunity cost in case of lost production when the lithography system is down is enormous, and therefore customers require a high availability. In order to maximize the availability of the systems, ASML attempts to resolve system downtime as quickly as possible. With technology developing at an unprecedented rate, and each new generation of lithography systems becoming more complex, a pro-active maintenance strategy is required to prevent systems from failing unexpectedly, and minimize unscheduled downtime.

Currently however, most maintenance on the lithography systems is failure-based. Random failures of critical system components result in unscheduled downs (USDs), during which the system is restored as quickly as possible through failure-based maintenance (FBM) of the faulty component. Additionally, ASML prescribes standard schedules for periodic maintenance (PerM). By performing maintenance pro-actively, that is prior to failure, the downtime of the system as well as the total cost of maintenance can be reduced. Also, at ASML, parameter data is continuously monitored through sensors embedded in the systems. This parameter monitoring data can be utilized to model the condition of a component. Local initiatives at ASML have resulted in various condition models which are used for scheduling preventive maintenance (PM) based on monitored condition of a component (i.e., condition-based maintenance or CBM). Altogether, the current maintenance situation can be depicted as shown in the figure below.

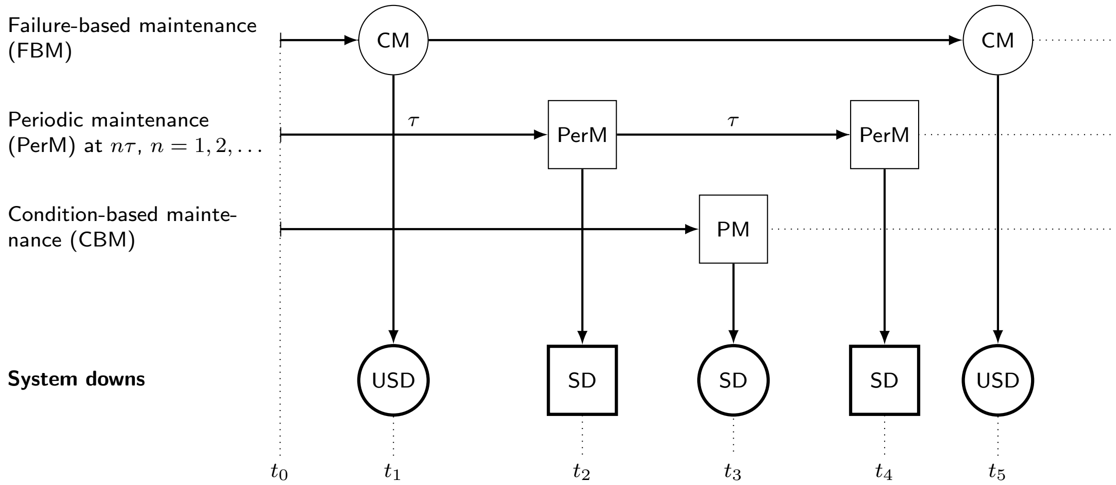


Fig. 1: *Schematic overview of the current ASML maintenance operations*

Accordingly, we proposed to *synchronize* condition-based maintenance with both failure-based and periodic maintenance. Under this synchronized CBM policy, the unscheduled and scheduled downs of the system — due to respectively random failures of critical system components and periodic maintenance — are considered as *opportunities* for performing PM on a monitored system component. For example, in the figure above, by performing the PM of the monitored component opportunistically, during the scheduled downtime that has been incurred by PerM at time t_2, no additional system downtime is required, and maintenance (setup) cost can be saved. Based on this proposal, the research question was formulated as:

> *"To what extent can (synchronizing of) condition-based maintenance reduce the system downtime as well as the total maintenance cost?"*

To this end, a mathematical decision model was developed for (synchronizing) CBM of a monitored component of a single system with both FBM and PerM, and to evaluate the performance of such a policy. This decision model requires a degradation model of the monitored component as input. A degradation model describes the deterioration of the monitored component's condition over time. In this project two general types of degradation models were considered: the linear random coefficient model that is suitable for modeling unit-to-unit variation and the stochastic process formulation that is appropriate for time-dependent variation in degradation data. In order to fit the degradation model to the available degradation data, most-likelihood estimators were provided to estimate the model parameters. To select the degradation model that best fits the data, a graphical method has been developed. Subsequently, an (extended) evaluation was developed to determine the expected total maintenance cost (ETMC) and the corresponding expected downtime (EDT) of synchronizing CBM with FBM (and PerM). Through optimization of this model, an optimal OPM control limit is obtained that minimizes the ETMC. This control limit defines the optimal opportunistic maintenance decision making strategy.

In a ASML case study, the performance of the synchronized policy for two monitored components of the XT/AT lithography system (A and B), was compared against the performance of a default CBM policy, as well as the current ASML policy. Based on this comparison, the benefits of CBM were assessed, as well as the extra benefits of synchronizing CBM. The general results are briefly summarized below. The results show the immense benefits that can

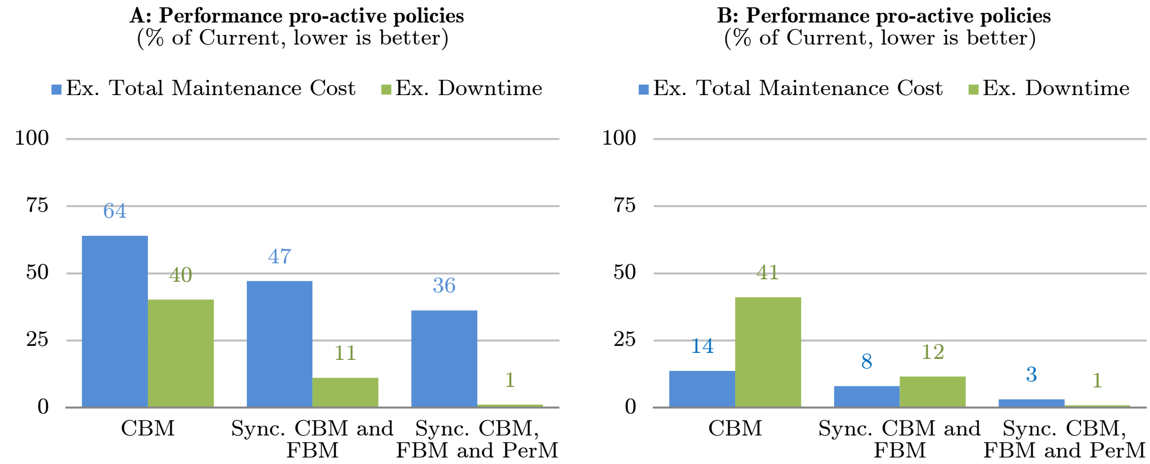


Fig. 2: *Performance of the reference and pro-active maintenance maintenance policies in the case of component A and B (expressed as a percentage of the current performance).*

be achieved be performing maintenance pro-actively, based on the condition of a component. Furthermore, the results indicate the power of synchronizing this condition-based maintenance, especially in terms of the expected downtime which can be almost completely eliminated through opportunistic maintenance.

Based on the results of the case study, and the subsequent sensitivity analysis, a description was given in order to adjust the current maintenance decision making, for the incorporation of opportunistic maintenance, so that the CBM of the monitored component can be synchronized with FBM (and PerM). Also, an implementation plan has been provided, that shows what steps must be undertaken in order to apply the synchronized policy to other monitored components.

The developed evaluation model for (synchronizing) CBM can be used by the ASML decision maker to (i) optimize and evaluate the benefits of synchronizing CBM for a monitored component, and thus as a selection method for “high potentials”; (ii) determine the optimal OPM control limit for synchronizing CBM, so that the ETMC are minimized, and (iii) as a method to improve the default warning limit for scheduling PM.

Contents

0. Introduction

This master thesis focuses on the development of a evaluation model for synchronizing condition-based maintenance, with application at ASML. In the next chapter ASML, their lithography systems, and the semiconductor industry, are introduced to the unfamiliar reader. Then, a thorough analysis is given of the current ASML maintenance operations. Based on this analysis, possibilities for improvement are identified, and a research design is formulated in chapter 2. In the research design, the main research is formulated, and a detailed overview of the project approach is given.

To improve the readability of this thesis, a list of acronyms, notation, and a description of all important definitions, is available for reference on pages 57, 58, and 60, respectively.

1. Company description

This chapter briefly introduces the unfamiliar reader to ASML and the semiconductor industry. First, a company profile is given in section 1.1, followed by an introduction of the ASML lithography systems in section 1.2. The main part of this chapter entails the analysis of the current ASML maintenance operations (section 1.3).

1.1. Company profile

ASML is the world's leading provider of lithography systems for the semiconductor industry [2]. ASML provides semiconductor manufacturers an integrated portfolio, including design, development, integration, marketing and servicing, with highly-advanced imaging tools (lithography systems) that make the smallest, most advanced and cost-effectively produced electronic chips. These chips, or integrated circuits, are used in most computers, TV and hi-fi systems, mobile phones, MP3 players, cameras, banking cards and automotive control systems. In fact, they are at the heart of most electronic equipment used in the home, the office and on the move [3]. ASML's customers include major players in the semiconductor industry, such as Samsung Electronics and Intel.

ASML is a multi-national company, located in more than 60 countries, spread over three continents: America, Europe and Asia. As of December 31, 2010 ASML employs over 9,100 FTE. Manufacturing sites and R&D facilities are positioned in Connecticut, California and the Netherlands. Technology development centers and training facilities are located in Japan, Korea, the Netherlands, Taiwan and the United States. ASML's headquarter, main manufacturing facilities, applications laboratory, assembly clean rooms, and R&D facilities are located in a single 45,000m2-site in Veldhoven, the Netherlands. Other major operating subsidiaries are located in the US and Asia; their head offices are located in Tempe (Arizona) and Hong Kong (Republic of China), respectively. Although the ASML organization is extensive, it is characterized by close cooperation internally, as well as with customers and suppliers all over the world. The organizational structure bears out ASML's business strategy, which is based on technology leadership, customer focus and operational excellence (cf. [54]). The business strategy is supported by continuous innovation, (e.g. with 500 million EURO spent on R&D efforts, ASML is a major investor in innovation) that focuses on the four strategic business segments.

With every generation, the complexity of producing integrated circuits with more functionality increases. Semiconductor manufacturers need partners that provide technology and complete process solutions. ASML is committed to providing customers with leading edge technology that is production-ready at the earliest possible date. ASML technology is supported by process solutions, enabling customers to gain and sustain a competitive edge in the marketplace [4]. Successful development of new products is highly dependent on a variety of key factors, including successful management of research and development programs, rapid product development and design (i.e. timely introduction of new systems). Accordingly ASML's competitiveness increasingly depends upon the ability to develop new and enhanced semiconductor equipment that is competitively priced and introduced on a timely basis.

1.2. Lithography systems

A lithography system may be thought of as a very large (the size of a sea container) high-tech copy machine or photo camera. The latter records light as a digital image by means of a sensor. Likewise, in photo lithography machines, patterns are optically imaged onto a silicon wafer that is covered with a film of light-sensitive material, a photo resist. This procedure is repeated numerous times, after which the actual electronic circuits are created on the silicon.

Fig. 1.1: *The ASML EUV lithography system*

Lithography systems is ASML's the largest business segment and has played a central role in enabling the growth of the semiconductor business. Other ASML business segments are Brion Technologies, Customized imaging solutions, and Optics. ASML has developed various lithography systems including dry lithography, immersion lithography, and EUV. As of 2011, the high-end Twinscan NXT:1950i system produces features down to 32 nanometers at up to 200 wafers per hour, using a water immersion lens [5]. ASML's latest lithography machine uses extreme ultraviolet light with a wavelength of only 13.5 nm and can produce features smaller than 22 nm. The total EUV lithography system equals the size of a cargo container, and is depicted in Fig. 1.1.

Within lithography systems, three main markets exist: Asia, Europe, and United States. Based on 2011 installed base data, the Asian market accounts for roughly three quarters of the net system sales. The US and European markets are ranked second and third in terms of net system sales, respectively.

1.3. Current maintenance operations

In this section the current ASML maintenance operations are analyzed in detail. A summary of this analysis is available in the final subsection (1.3.8).

1.3.1. The need for pro-active maintenance

A lithography system can be characterized as a highly technical, advanced, and complex capital good with an extremely high commercial price. The ASML lithography systems are embedded in the core production processes of the ASML customers (chip makers). The opportunity cost in case of lost production when the lithography system is down is enormous. In fact, it could potentially could run into millions of euros per day. For this reason, it is important that the availability of the systems is high, and that system downs are minimized and resolved as quickly as possible [26]. With technology developing at an unprecedented rate, and each new generation of lithography systems becoming ever more complex, a *pro-active maintenance* strategy is required to reduce both downtime and maintenance cost, and fulfill customer availability requirements in the future. Later on in this section the immense potential of pro-active maintenance is demonstrated.

The present maintenance strategy at ASML however, takes on a rather *reactive* character. That is, most maintenance is failure-based. Only a small portion is classified as preventive maintenance (PM), namely condition-based and periodic maintenance.

1.3.2. Failure-based maintenance

Failure-based maintenance (FBM), the dominant ASML maintenance approach, is often characterized as firefighting and the maintenance engineers as firefighters. The analogy is that maintenance is initiated upon a failure of a critical system component, and the objective is to restore the system as quickly as possible. FBM results in random failure behavior of critical system components and thus random system downs. A system down following a random failure of a critical system component is an *unscheduled down* or USD. A serious drawback of this reactive approach is the inability to plan for maintenance resources, due to the randomness of system downtime. This includes planning of spare parts, tools, and labor.

Once the system suddenly breaks down, a series of steps must be undertaken to restore the system as quickly as possible. These steps are better known as the USD process, and generally consist of the following actions:

- Customer interface;
- Fault diagnosis;
- Ordering parts & tools (if no stock is available);
- Repair of the failty component;
- Metrology of the repaired component; and
- System stabilization.

The USD process is depicted in Fig. 1.2. Please note that the precise number of steps may differ per USD. For example, when a spare part is available on stock no transhipment is required, and thus no spare part delay is incurred. Each of the aforementioned process steps are now described in detail.

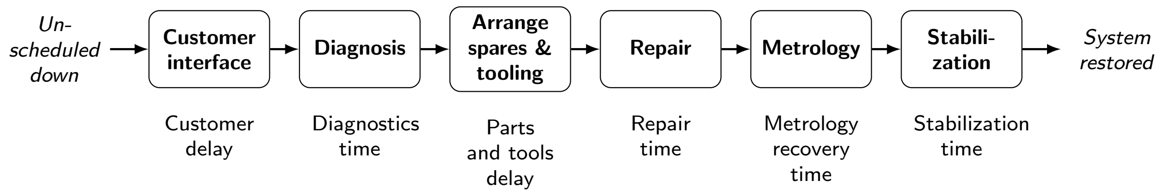


Fig. 1.2: *USD maintenance process*

1.3.3. USD maintenance process

Upon the failure of a critical system component, some time expires until the failure is noted by the customer. This is the failure response time. Once the failure is noted, the system is shut down and the production stops. The customer contacts and informs the nearest ASML customer service center ("office"). The customer interface delay also includes the failure response time. Next, the system failure is diagnosed either remotely by customer service office or on site by a field engineer. The ASML lithography machines are embedded with numerous sensors that monitor thousands of system parameters. An engineer can diagnose the failed system through running models that analyze these parameter values, and can detect any deviations or peaks that could be the possible cause of failure. Once the failure has been successfully diagnosed, the required spare parts and tooling are ordered. Usually, these are not directly available, but must be shipped from the nearest warehouse with a specific lead-time. In the worst case, an emergency shipment is required, which ships a spare part via plane as fast as possible (approximately 24 hours). The time that expires between failure diagnosis and the arrival of the required spare part (and tooling) is known as the part & tool delay. After the arrival of both the required spare parts and tools, the defect part or component is repaired. This required a specific amount of time, which is the time to repair. Once the faulty component has been repaired, metrology is performed which ensures that the repaired component is properly calibrated and that the system reconfigured accordingly. In final step, the system is stabilized so that it is ready for production. Usually this implies that the system's internal temperature and pressure is restored (i.e., C&T stabilization). The total time that this unscheduled maintenance process consumes is called the unscheduled downtime of the system.

A failure of a critical system component that results in a system down, is related to a root error code. The resulting downtime of each root error is carefully reported by ASML field engineers. Hence, for every root error a specification of the resulting downtime is available. Using this data, we can construct an average *downtime composition* for each component failure. For example, a failure of component X results on average in a total downtime of 100 hours, of which 5 is related to the customer interface, 10 to the fault diagnosis, etcetera. In Fig. 1.4, the average downtime composition for all components of the ASML AT/XT Twinscan system during 2008-2012 is given. From this figure it follows that typically, the diagnosis and the repair

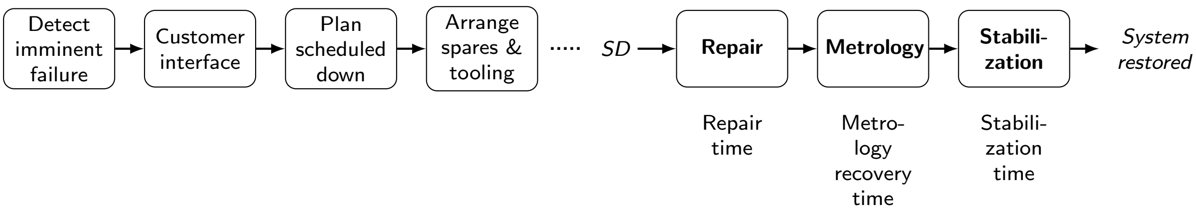


Fig. 1.3: *SD maintenance process*

consume a high portion of the total downtime. Note that there is an "other" delay. This is the downtime that is related to miscellaneous or uncommon actions, e.g., a laptop crash, or software incompatibility. Interestingly, on average, roughly 75% of the downtime that is not dedicated to the repair of the faulty component. Even more important is that about 50 to 60% of the downtime is consumed by the so called *setup* of the repair (i.e., customer interface, diagnostics, and waiting for spares and tooling). Therefore, if one could predict an imminent system failure, the critical component could be repaired preventively, and a large portion of the downtime could be eliminated. That is, by taking maintenance action pro-actively, i.e., prior to the actual failure, the downtime can significantly reduced.

1.3.4. Reducing downtime by pro-active maintenance

By pro-actively planning for preventive maintenance of a component that is about to fail, several delays in the downtime process can be eliminated. This can be explained as follows. Suppose that an imminent failure of a critical system component is detected (e.g., through condition monitoring which is discussed later in this section). Then the customer can be contacted to arrange a suitable date to shut down the system for a preventive repair of the soon-to-fail component. If a scheduled down (SD) of the system is planned for, then the required spare parts, tools, and maintenance engineers can be requested in advance, so that they are available at the time of the SD. Later, if the system is shut down at the planned date and all resources are available, a preventive repair of the component can be immediately carried out. Then, the repaired component is calibrated followed by stabilization of the system. This SD maintenance process that is enabled by a pro-active maintenance strategy such as condition-based maintenance (CBM), is depicted in Fig. 1.3.

If a failure is detected in advance, several delays in the maintenance process are eliminated, and since the imminent failure is detected in advance, the diagnosis step becomes obsolete. The only downtime that remains is the time repair, and the time that is required for metrology and stabilization of the system. In short, the downtime savings from pro-active maintenance thus amount to:

$$\textit{Downtime savings} = \textit{Customer delay} + \textit{Diagnostics time} + \textit{Parts and tools delay} \tag{1.1}$$

Obviously, the effectiveness of such a pro-active strategy depends on the accuracy of the prognostic method for predicting/detecting an imminent failure. When the predicted failure time is shorter than the actual failure time, some useful life is wasted as pro-active maintenance is performed too early. On the contrary, if the predicted failure time is larger than the actual remaining useful lifetime, then there still exists an additional unexpected downtime. In sum, for reducing the downtime (close to the theoretical maximum) an accurate prognostic is required.

Consider again the example of the expected downtime composition of a AT/XT system component in Fig. 1.4. Then using eq. (1.1) it follows that the downtime can be reduced by almost 60%. In other words, more than half of the downtime can be eliminated with a pro-active maintenance strategy. As already mentioned, these savings can only be achieved in if the prognostic methods for detecting imminent component failures are accurate, such that planning SD and arranging the neccesary resources is possible.

[redacted]

Fig. 1.4: *Average downtime composition of AT/XT components during 2008-2012*

In the next section we will describe how ASML attempts to perform pro-active maintenance based on the condition of components (i.e., condition-based maintenance), with the help of condition models and parameter monitoring.

1.3.5. Pro-active maintenance

All ASML lithography machines are embedded with numerous sensors that continuously measure various machine parameters (e.g., pressures, temperatures, etc.). The parameter data is stored locally, and transferred daily to the central file archive in Veldhoven. This data is first processed and subsequently used for all kinds of analysis, including condition modeling (see Fig.

1.5). A condition model indirectly estimates the true state or *condition* of a component, from a (set of) parameter(s). This means that the actual "health" or condition of a system component can be determined from the parameter monitoring data through its condition model. This monitoring process is summarized in Fig. 1.5.

A condition model can be used to plan pro-active maintenance. This is done by setting a warning limit. The setup of this warning limit implies that once the monitored condition of a component exceeds this warning limit, a notification is triggered and preventive maintenance must be performed as soon as possible. The warning limit must be set so that there is enough time left before the actual failure occurs, to schedule a preventive maintenance (PM) action. A component is supposed to have failed if its condition crosses the critical failure limit. The minimum time that is required to arrange a scheduled down for preventive maintenance is 48 hours (this is known as the lead time of a scheduled down). This includes the delay of the parameter monitoring data and the minimum lead time time of spares and tools, as described below.

- The parameter data is transferred once per day (i.e., 24 hours) to the central file archive, where its processed and "fed" to the condition models. So the monitored condition has a delay of 24 hours.This includes a 24 hour delay of the monitored condition.
- If a spare part is not on stock (in the nearby customer warehouse), then an emergency transshipment is required from the central warehouse with a lead time of 24 hours.

Both the warning limit and the critical failure limit is set by ASML engineers either by (i) design, or (ii) by trial and error. In the former case, the failure limit is determined on the basis of the physical properties of the component (e.g., maximum crack size is X millimeter). In the latter case, the warning limit is estimated, by comparing the actual failure data of the component with its monitored status. The actual failure time can then be linked to the monitoring status, and subsequently the critical failure limit can be obtained. Consequently, the warning limit for scheduling a preventive maintenance action is set such that there is enough time available to schedule a PM action (i.e., 48 hours or more). Because the time of failure is random, the warning limit is set well below the critical failure limit, so that it is ensured that the actual failure occurs after 48 hours, and sufficient time is available to perform a PM action during a scheduled down of the system. If the warning limit is set too close to the critical failure limit, then one runs the risk that the actual failure occurs before a PM can be planned and executed, and failure-based maintenance must be performed. Obviously, this would render the pro-active maintenance policy useless.

As previously mentioned, the effectiveness of a pro-active maintenance strategy is dependent on accuracy of the prognostic methodology. In other words, it is crucial that a condition model and the corresponding failure and warning limits, describes the true "health" of the underlying component. The performance of a condition model and its limits, is assessed in terms of model precision, which can be computed using the entries of the confusion matrix [24], as shown below in Table B.1. Precision is a measure that shows the fractions of correct notifications generated by model. Furthermore, model *coverage* is defined as the fraction of part failures that are explained by the model. Ideally, a model should generated no false alarms and cover all failures. However, far from ideal models can still yield considerable cost savings. For example, a model with a high coverage rate may still be useful for condition-based maintenance regardless of its precision. Suppose that out of 10 triggers, 9 are false positives, and 1 is a true positive (i.e., 10% precision). However, if this one true positive can avoid an unscheduled downtime and save

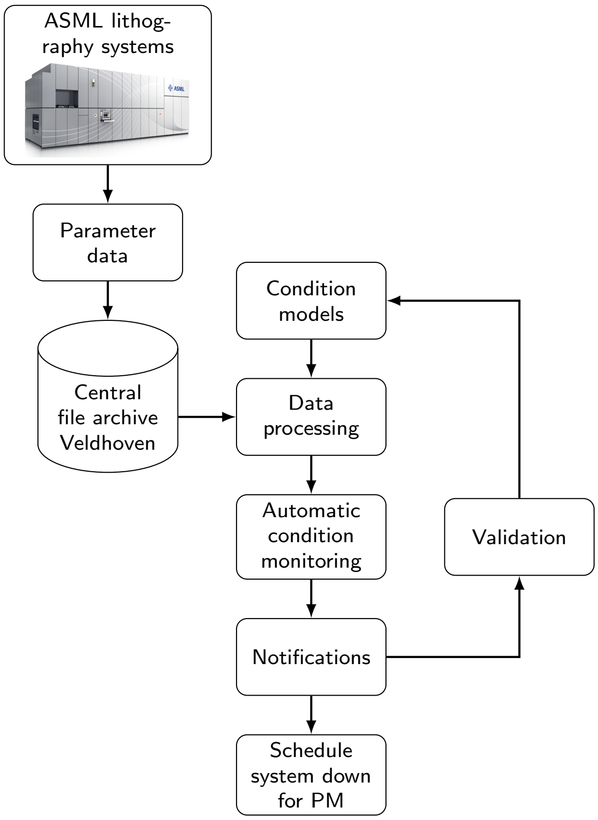


Fig. 1.5: *Condition-based maintenance process*

millions of euros, and those 9 false positives require only a small inspection cost, the model can save a significant amount of maintenance costs, despite its poor precision. Furthermore, the failure limit (and subsequently the warning limit) can be adjusted so that the predictive performance of the condition model is maximized.

Ideally, all maintenance on the ASML lithography systems should be condition-based. Currently however, only a handful of accurate condition models are available for CBM. In order to support the development of a pro-active maintenance strategy, a central ASML infrastructure is being developed that standardizes the development of condition models and supports their deployment (under the heading of the Be-Warned project lead by Gert Streutker).

1.3.6. Periodic maintenance

For some system components ASML, prescribes standard periodic maintenance (PerM) schedules. This periodic maintenance schedule includes a set of preventive actions that should be performed after fixed periods of time (i.e., weekly, monthly, quarterly, yearly, or three-yearly). That is, every period a set of preventive maintenance actions should be carried out, in order to enhance the reliability of the system, so that the number of unscheduled system downs reduces. It is the customer's responsibility to adhere to this prescribed PM schedule. The actions include mostly minor adjustments, calibrations, cleaning, measuring, and inspections. For example: inspection of critical wafer stage items, cleaning of the wafer table, and adjustment of both lens gas pressure and flow.

1.3.7. Total cost of maintenance

Thus far we have solely focused on (reducing) the downtime of a system. However, ASML's primary objective is to minimize the total cost of maintenance, of which downtime is a major factor. Needless to say, system downtime comes at a huge cost. This is particularly because a lithography machine is usually the core of the customer's production process. Downtime of the lithography system thus implies that the entire production process is shut down, and huge potential sales are lost. However, for ASML the cost of downtime is not visible on the balance sheet. Still, system downtime hurts ASML because of the key performance indicators in the customer contracts. If a specified level of uptime or availability is not satisfied, ASML is penalized. This cost is accounted for by the downtime penalty cost. This cost penalizes the (indirect) results of downtime for ASML. Besides downtime penalty cost, a system failure "hurts" ASML, because maintenance engineers (labor) and spare parts must be provided according to the warranty contracts to restore the system as quickly as possible. This implies that a system down results in engineer labor cost, spare part unit cost. In fact, if a spare part is not available in the regional customer warehouse and the system suddenly fails, then a very costly emergency shipment from the central warehouse is required.

Table 1.1: *ASML total maintenance cost*

Cost/Downtime		Customer delay	Diagnostics time	Parts delay	Tools delay	Repair time	Metrology recovery	C&T stabilization	Other delay
Per unit downtime	Penalty	✓	✓	✓	✓	✓	✓	✓	✓
	Labor		✓	✓	✓	✓	✓	✓	✓
Per system down		Cost of spares, tooling, and (emergency) shipping							

In summary, each hour of system downtime corresponds with penalty and labor cost, and each system down corresponds with (expected) spare part & tooling, and (emergency) shipment cost. Since the downtime composition for each component failure is known (cf. section 1.3.4), the expected labor and downtime penalty cost that contribute to a system down can be determined[1]. For ASML customers the primary focus is to minimize system downtime, whereas for ASML itself the total maintenance cost must be considered that takes into account the customer downtime cost indirectly through a downtime penalty cost. The total cost of maintenance is summarized in Table 1.1. In this table the downtime composition of is related to both labor and downtime penalty cost. In addition, a system downtime is related to an expected shipment and part cost. Note that the customer delay is not related to labor cost. This is because the maintenance engineers are sent to the customer site, directly after the customer has contacted the nearest ASML field office.

1.3.8. Summary

The most important conclusions and observations from the analysis of the current maintenance operations are now listed.

[1] This is illustrated in greater detail in the case study in chapter 4

- With the increasing complexity of lithography systems, pro-active maintenance becomes a key driver in reducing downtime and satisfying customer availability requirements in the future. However, the current ASML maintenance operations are not optimized for a pro-active maintenance strategy. In fact, most maintenance is failure-based.
- Failure-based maintenance (FBM) is a reactive approach that leads to long unscheduled system downs. Most of this downtime is consumed by the time waiting for parts, tooling, and engineers. Only small fraction is dedicated to the actual repair of the system.
- By performing maintenance pro-actively, system downs can be scheduled in advance in agreement with customers. This way, several steps in the maintenance process can be skipped, and the the downtime can be reduced with roughly 60% on average.
- The effectiveness of a pro-active maintenance strategy is dependent on the ability to detect imminent or predict system component failures.
- ASML is increasingly developing and deploying condition models of components for condition-based maintenance (CBM). These models can assess the status of a component from parameter monitoring data. By determining the critical failure limit, and subsequently setting a warning limit for scheduling PM before the actual failure occurs, failures can be resolved pro-actively.
- Unfortunately, only a handful condition models are available that are only used locally. However, a large effort is undertaken to develop a central infrastructure that supports the development and deployment of condition models for condition-based maintenance.
- In addition to FBM and CBM, periodic maintenance schedules are available for all lithography systems. In order to ensure system availability, customers are advised to follow up on these schedules.
- For ASML customers, system downtime relates to extremely high opportunity costs due to lost sales. For ASML the reduction of downtime, as well as the minimization of the total cost of maintenance is important, which includes downtime penalty cost, and the part, shipment, and labor costs.

2. Research design

The purpose of this chapter is to describe the project's research design. First, two improvement possibilities are described (2.1). Accordingly, the research question and sub questions are formulated in section 2.2. This is followed by a summary of the project's literature review (2.3), a definition of the project's scope (2.4), and deliverables (2.5). Finally, in section 2.6 the research approach is described and an outline of the report.

2.1. Improvement possibilities

The analysis in section 1.3 clearly shows the huge potential of pro-active maintenance. In accordance with this analysis, two improvement possibilities are now described in turn.

2.1.1. Be pro-active through condition-based maintenance

As described section 1.3, because of the increasing complexity of lithography systems, ASML requires an appropriate maintenance strategy to ensure system availability to meet customer requirements. The current maintenance approach is rather reactive. It is characterized as firefighting, and consists of mostly failure-based maintenance (FBM). Due to random failures of critical system components, large unscheduled system downtime occurs, that impact system availability. Downtime is related to high opportunity costs of lost sales, emergency shipment cost, and labor (overtime) cost. The scheduled downtime is only 40% of the unscheduled time time, a reduction of 60% (cf. section 1.3.4).

Hence, in an attempt to reduce the system downtime as well as the total cost of maintenance, preventive maintenance can be planned pro-actively using condition models. Because these condition models are able to estimate the true health of system components, it is possible to initiate preventive maintenance pro-actively, i.e., based on the condition of the monitored condition of a component (cf. 1.3.5). Because currently only a handful of accurate condition models are available for deployment, the realization of a fully pro-active maintenance strategy of the lithography systems is not yet feasible.

2.1.2. Synchronizing condition-based maintenance

ASML can employ three types of maintenance on their lithography systems: most components are subject to failure-based maintenance (FBM), a couple of maintenance tasks are carried out periodically (PerM), and a handful of monitored components are eligible for condition-based maintenance (CBM). Recall that FBM results in random downs or unscheduled downs (USDs), that periodic maintenance requires periodic downs of the system (i.e., scheduled downs or SDs). This can be summarized schematically as shown in Fig. 2.1.

So, the USDs due to FBM occur randomly over time, the PerM activities are carried out at fixed times $n\tau$, $n = 1, 2, \dots$, where τ is the periodic maintenance interval. Only the downs of the CBM process can be adjusted, since the condition of a component is known. Using this property, it is possible to save downtime (costs). For example, in Fig. 2.1 the system is down for FBM or PerM at times t_1, t_2 before CBM is performed at time t_3. If the monitored component would have been replaced in parallel with either FBM at time t_1 or PerM at time t_2, than an SD of the system could have been saved at time t_3 (as shown in Fig. 2.2). This called opportunistic maintenance of multi-component systems. The *synchronized* CBM process considers the opportunities that come from both FBM and PerM to simultaneously perform

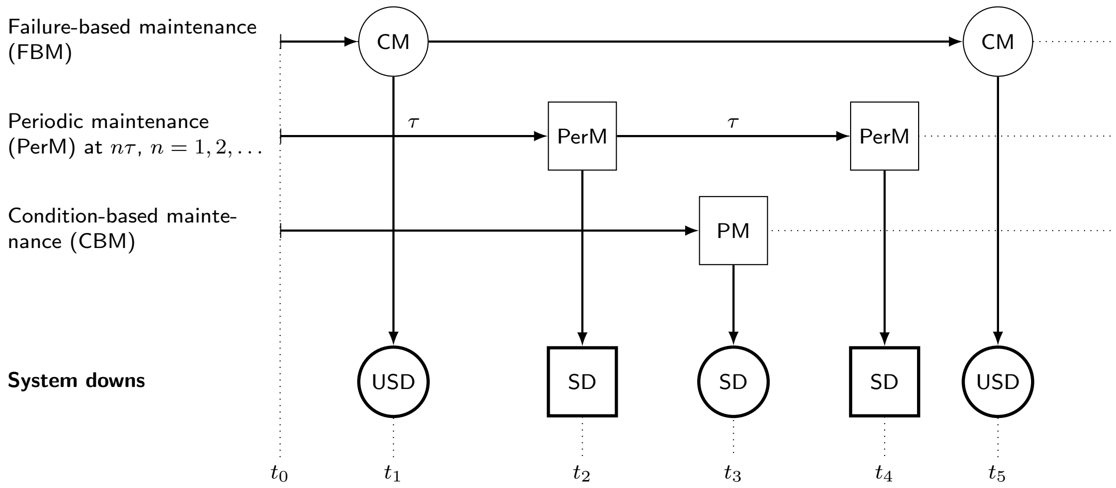


Fig. 2.1: *Schematic overview of the current ASML maintenance operations*

PM on the monitored component, so that downtime as well as maintenance cost can be saved due to economies of scale.

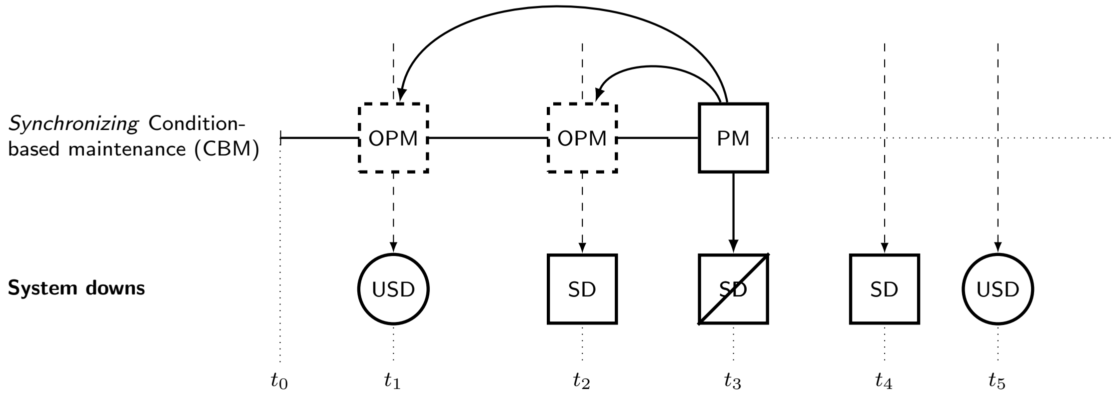


Fig. 2.2: *Improved ASML maintenance operations through synchronizing*

Accordingly, a possibility for improvement is by considering both the unscheduled and scheduled system downs due to resp. FBM and PerM, as *opportunities* for preventive maintenance of monitored system components. This is known as opportunistic maintenance or opportunistic preventive maintenance (OPM) in literature. A clear advantage of this strategy is that no additional system downs are required for PM of the monitored component. Traditionally, a disadvantage of opportunistic maintenance is, that it is often not known in advance which actions are taken, so that no planning and no work preparation are possible [12]. However, by setting control limits for opportunistic maintenance of a monitored component, one could prepare for the arrival of an opportunity, as opposed to opportunistic maintenance of FBM with preventive maintenance of non-monitored components.

To summarize the current maintenance operations can be improved by (i) applying condition-based maintenance (CBM) on monitored components, and subsequently (ii) synchronizing CBM with both FBM and PerM. Through opportunistic preventive maintenance of the monitored component, downtime as well as maintenance cost can be saved. We believe several benefits are apparent from this improved strategy:

- Since parameter monitoring equipment is already available at ASML, as well as condition models, no significant investment cost are required to implement this improved strategy.
- Even though failure-based maintenance is a reactive and costly policy, it is still the dominant ASML policy. The proposed strategy however, considers the unpleasant FBM policy as a source of opportunities for opportunistic preventive maintenance of other monitored system components.

2.2. Research question

According to the analysis in section 1.3 and the proposed improvements in the preceding section, we formulate the main research question of this project as:

To what extent can (synchronizing of) condition-based maintenance reduce the system downtime as well as the total maintenance cost?

The research questions can be split up in the following sub-questions:

1. How can a maintenance policy for synchronizing condition-based maintenance be evaluated and modeled?
   (a) How can the opportunities (due to both FBM and PerM) for simultaneous maintenance of the monitored component be modeled?
   (b) How can a monitored component's "health" or condition be described?
   (c) How can the opportunistic maintenance decision for the monitored component be modeled?
   (d) What are the relevant maintenance cost parameters to be included?
   (e) What is the reference (or as-is) policy and how can it be modeled?
   (f) What are the key performance indicators (KPIs)?
2. What are the benefits of the (synchronized) condition-based maintenance policy?
   (a) How does the (synchronized) pro-active policy perform in terms of the KPI?
   (b) How does the performance of the (synchronized) pro-active policy compare to the reference policy (i.e., no CBM)?
   (c) How can the model be properly deployed so that it can be used in the ASML maintenance operations?

2.3. Literature

A summary of the literature review that has been conducted for this project (cf. [52]) is available in Appendix A.

2.4. Scope

In accordance with the research questions, the project scope is defined as follows.

(+) *Degradation modeling*: Appropriate methodology is reviewed to fit a degradation model to degradation data from ASML condition models.

(+) *Evaluation and optimization methodology*: Evaluation models are developed to determine the performance of the (synchronized) CBM policies. In addition, the optimization of these model is discussed, such that the performance of the models can be maximized.

(±) *AT/XT lithography machine family*: This project predominantly focuses on improving the current maintenance operations of the AT/XT product family. The analysis and case study of the project are performed on AT/XT data. This implies that the newer product families (NXT and NXE) are out of scope. However, it is likely that the main conclusions and recommendations for the AT/XT system are also useful for the improving the maintenance operations of the newer product families.

(−) *Spare parts planning*: The evaluation and optimization of the pro-active policy does not include the planning of spare parts. Pro-active planning for spare parts is a topic for future research.

2.5. Deliverables

The main product of this research is a mathematical evaluation model the determines the performance of synchronizing condition-based maintenance of one monitored component. This model can be used for synchronizing condition-based maintenance, failure-based maintenance, and periodic maintenance, such that the total expected maintenance cost of the system is minimized. A case study is presented for the AT/XT lithography system, that demonstrates the benefits of synchronizing over non-synchronizing and the current ASML policy. By-products of this case study are: (i) mathematical degradation models for the AT/XT monitored parts; and (ii) a ASML cost specifications to translate the downtime composition, unit cost, and emergency shipment cost to the model's input cost parameters. In addition, an implementation plan is presented that describes how the results of this project can be deployed in the existing ASML maintenance operations. Finally, a set of MATLAB classes is provided that can supports the evaluation, optimization and implementation of the pro-active maintenance decision model.

2.6. Project approach

The project approach includes the next research steps that are taken in order to answer the research questions. The project remaining steps are now listed in chronological order. A complete schematic overview is depicted in Fig. 2.3.

1. *Modeling* (Chapter 3): Based on the results of the literature study, the analysis of the current maintenance operations, and in accordance with the improvement proposal, an evaluation model is developed for synchronizing CBM with both random system failures and PerM. In addition, the current ASML policy is evaluated, and serves as a reference policy to which the results of the proposed policy can be compared to. The modeling process includes three steps:
   (a) *Degradation modeling* (section 3.1): Degradation models are provided for the description of the deterioration of a monitored component over time. In addition, parameter estimation procedures are described to fit a degradation model to the available degradation data, as well as a graphical method to determine the goodness of fit and to select the degradation model that best fits the data.
   (b) *Evaluation models* (section 3.3 and 3.4): It is necessary to develop a mathematical evaluation model for synchronizing condition-based maintenance with both failure-based and periodic maintenance. This evaluation model will take the degradation

model of the monitored component as input, as well as various cost parameters, and characteristics of the FBM and PerM process, to determine performance of a given synchronized maintenance policy in terms of the expected total cost of maintenance, as well as the corresponding expected downtime.

(c) *Optimization* (Section 3.5): After the evaluation model has been developed, the model must be optimized to maximize its performance, for which useful heuristics are given.

2. *Case study* (Chapter 4): The evaluation model is then applied to two monitored components of the ASML AT/XT lithography system. In this case study, the performance of (synchronizing) CBM is evaluated and compared to the current ASML policy. Furthermore, the effect of important model parameters is investigated by a sensitivity analysis.
3. *Implementation* (Chapter 5): Based on the results of the case study, an implementation plan is presented for the deployment of the optimization policy in the ASML maintenance operations.
4. *Conclusions and recommendations* (Chapter 6): The thesis concludes with the main insights, implications, limitations, and directions for future research.

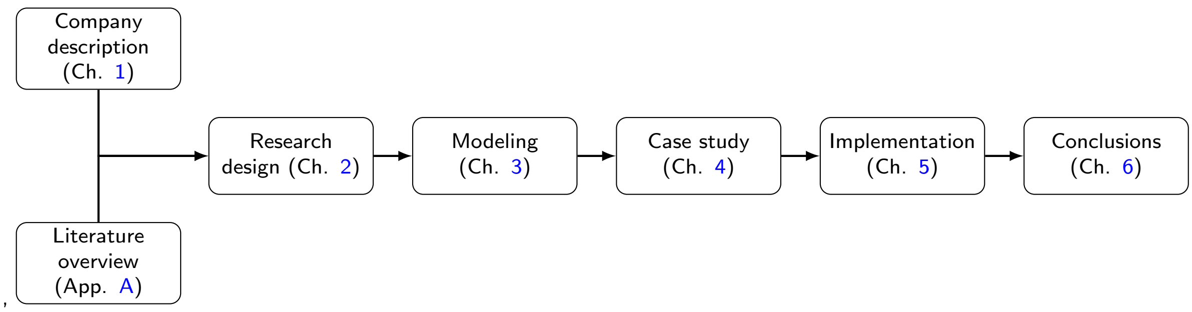


Fig. 2.3: *Schematic overview of project approach*

3. Modeling

The purpose of this chapter is to derive the benefits of condition-based maintenance (CBM) of a monitoring component, and the added value of synchronizing this CBM with random system failures and periodic maintenance. We start of by modeling the monitored component's degradation over time (section 3.1). Then, in section 3.2 this degradation model is used to evaluate the total expected maintenance cost of a pro-active condition-based maintenance policy. This cost is compared to the current ASML policy (i.e., failure-based maintenance of this monitored component). In section 3.3, a decision model is described for *synchronizing* CBM of the monitored component with random system failures and unscheduled downs. Subsequently, this preliminary decision model is extended with periodic maintenance (PerM) and scheduled downs, in section 3.4. In section 3.5 the optimization of both decision models is discussed, and finally the chapter is briefly summarized.

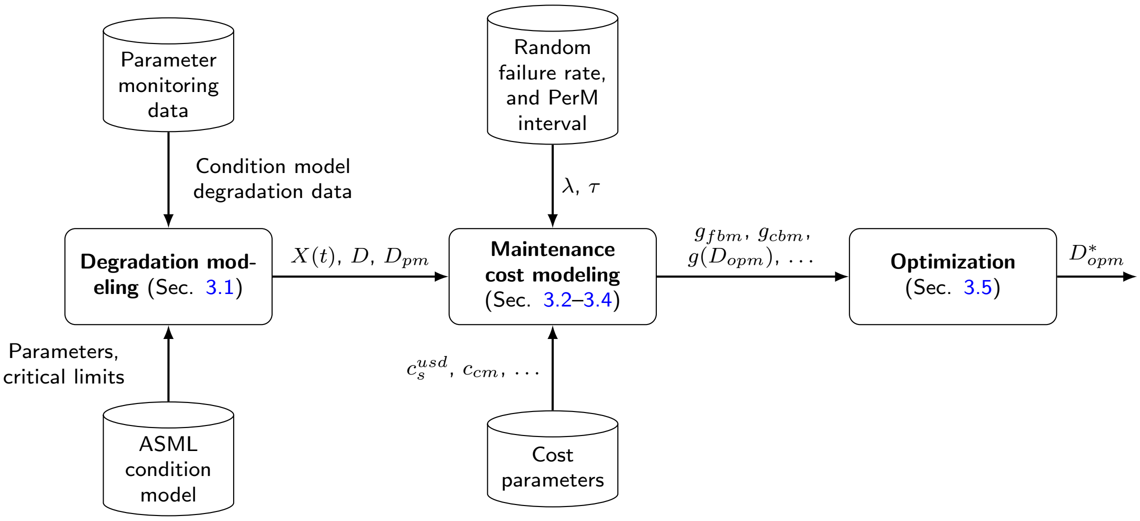


Fig. 3.1: *Modeling process*

The entire modeling process is depicted in Fig. 3.1. The model input and output parameters are discussed throughout this chapter. An overview for all model parameters is available on page 58.

3.1. Degradation modeling

As previously mentioned in section 1.3.5, at ASML a handful of condition models are available. A condition models estimates the true "health", condition, or status of a system component. The condition is remotely assessed from a set of system parameters, that are continuously monitored by sensors embedded in the system. This implies that the deterioration of monitored components can be assessed using the condition models. We require a *degradation* model to describe the deterioration of the monitored component's condition as a mathematical form. By setting a critical failure limit for the deterioration level, it is possible to deduce a component's lifetime distribution function which can subsequently be used for optimizing maintenance decision making.

In practice, degradation models are widely used to assess the lifetime information of highly reliable products if there exist quality characteristics whose degradation over time can be related to reliability [45]. The performance of a degradation model is largely dependent on the appropriateness of the model describing an item's degradation path. For example, the lifetime of a motor can be related to bearing wear-out, and the health of a laser can be assessed by monitoring the optical power.

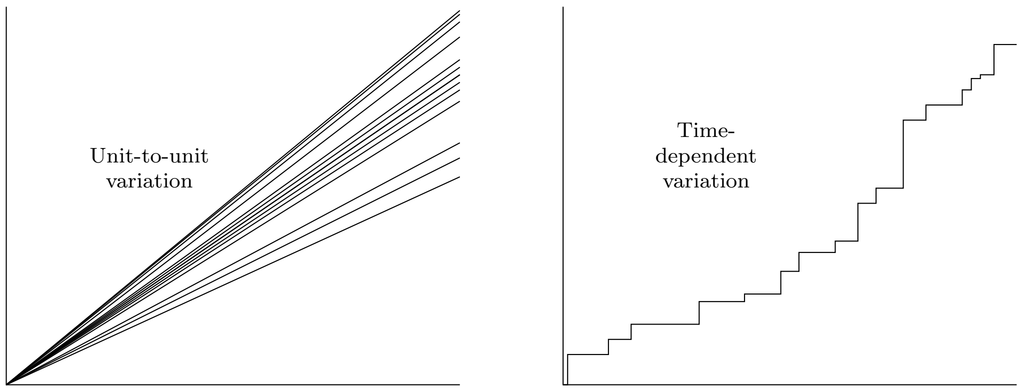


Fig. 3.2: *Linear random coefficient model (a) versus the stochastic process formulation (b). The former implies a linear deterioration with a degradation rate that varies randomly from unit to unit, whereas the latter implies that degradation is a gradual process with random increments over time*

In this thesis, we limit the scope to two common and relatively simple formulations of degradation modeling: (i) the *random coefficient* and (ii) the *stochastic process* formulations. The former implies a mean degradation path that varies from unit to unit, which is suitable for modeling unit-to-unit variation. The second approach describes the variations in degradation data by time-dependent stochastic processes. For the ASML condition models that are discussed in the case study later on (see chapter 4), these two formulations prove to be appropriate. However, it is likely that for other monitored components, other formulations are required. For example, a so called mixed-effect formulation [45], that takes into account both unit-to-unit and time-dependent variation. The difference between the two approaches is summarized in Fig. 3.2.

In this section we will discuss one instance of both formulations: a linear random coefficient model with a Weibull distributed degradation rate (section 3.1.1); and a stochastic *Gamma process* (section 3.1.2). In section 3.1.3 the degradation model parameter estimation as well as the model selection. It is shown how one can (graphically) determine which estimated degradation model fits the degradation data best. This section is concluded with a brief summary.

3.1.1. Linear random coefficient model

The linear random coefficient formulation with a Weibull degradation rate is suitable for modeling degradation processes in which deterioration is supposed to take place linearly over time with a Weibull distributed degradation rate. This model is particularly appropriate for incorporating unit-to-unit variation, implying that the degradation rate varies from unit to unit. Furthermore, the linear model has a closed-form expression for the lifetime of a part (as shown below). Basic literatures of the random coefficient model are available in [34, 33, 37].

Suppose that a part's degradation model can be described by the following linear random

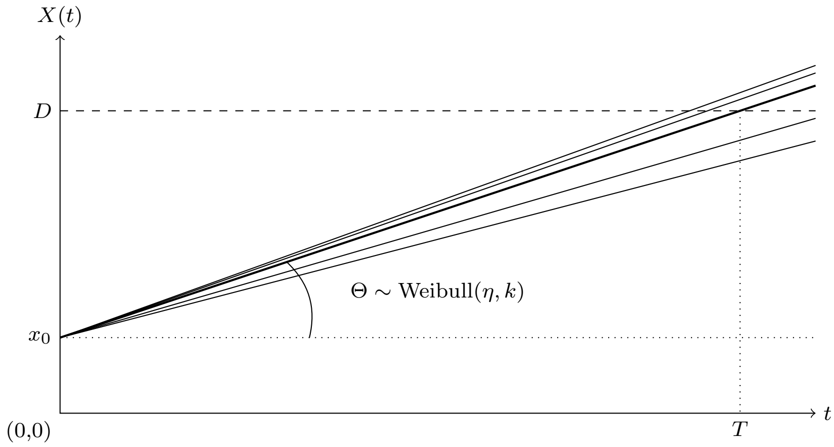


Fig. 3.3: *Linear random coefficient model*

coefficient model, as depicted in Fig. 3.3:

$$X(t) = x_0 + \Theta t, \quad t \geq 0 \tag{3.1}$$

where x_0 is a constant that denotes the initial amount of degradation (usually set to 0), and Θ represents the unit-to-unit variation of the degradation rate. Θ has a Weibull distribution function

$$F_\Theta(t) = 1 - e^{\left(\frac{t}{\eta}\right)^k}, \quad t \geq 0 \tag{3.2}$$

where $\eta > 0$ is the scale parameters, and $k > 0$ the shape parameter. Then it follows that $X(t) - x_0$, $t > 0$, follows a Weibull distribution with scale parameter ηt, and shape parameter k.

Now, let $D > x_0$ denote the critical failure level (also see Fig. 3.3). The part works until the process reaches D. Hence its lifetime T is then obtained by setting

$$T = \frac{D - x_0}{\Theta}, \quad \Theta > 0$$

and has the cumulative distribution function (cdf)

$$F_T(t; D, x_0) = e^{-\left(\frac{D-x_0}{\eta t}\right)^k}, \quad t > 0 \tag{3.3}$$

and the probability density function (pdf)

$$f_T(t; D, x_0) = \frac{k\eta}{D - x_0} \left(\frac{D - x_0}{\eta t}\right)^{k+1} e^{-\left(\frac{D-x_0}{\eta t}\right)^k}, \quad t > 0 \tag{3.4}$$

The distribution of T is known as the Inverse Weibull distribution and is also known as the complementary Weibull distribution or the reciprocal Weibull distribution. Furthermore, using the result of Gusma et al. [17], the nth moment of T is

$$E[T^n; D, x_0] = \left(\frac{D - x_0}{\eta}\right)^n \Gamma(1 - n/k) \tag{3.5}$$

The full derivation of distribution and density functions of T is available for reference in Appendix C.

3.1.2. Stochastic Gamma process

The Gamma stochastic process is a natural model for degradation processes in which deterioration is supposed to take place gradually over time in a sequence of tiny increments [29]. The choice of the homogeneous Gamma process in applications is motivated by the fact that it has non-decreasing trajectories, independent increments and is homogeneous in time. In the class of stochastic processes, the Gamma process has the advantage that the probability density function of the working period has an explicit expression. An extensive introduction to the Gamma process is available in [49].

Now, we describe the degradation of a component as a homogeneous Gamma process noted as $X(t)$, with the initial degradation level of x_0. Typically a Gamma process models the degradation of a component at time t [49]. The Gamma process is depicted in Fig. 3.4.

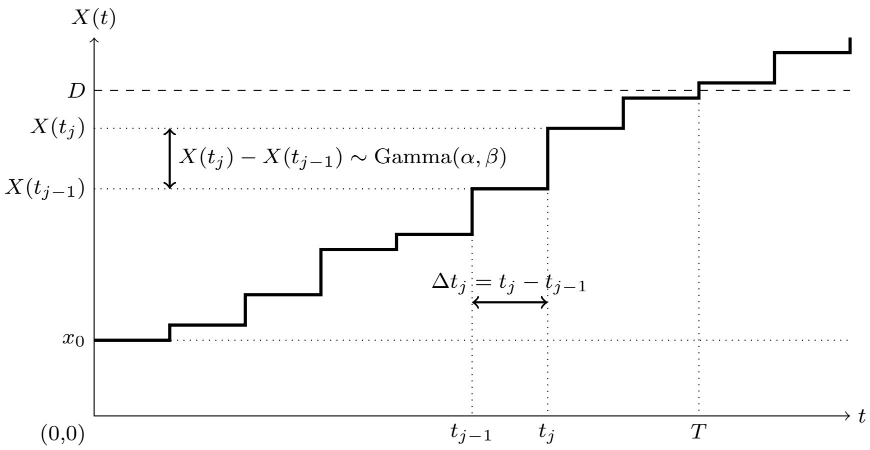


Fig. 3.4: *Stochastic Gamma process*

Let $X(t_j)$ be the degradation measure at instant t_j. The random increments $X(t_j)-X(t_{j-1})$ are independent and identically Gamma distributed with the shape parameter $\alpha(t_j - t_{j-1})$ and the inverse scale parameter β. It then follows that the distribution of $X(t)$, $t > 0$ follows a Gamma distribution with shape parameter αt, and inverse scale parameter β.

Let $D > x_0$ denote the critical failure limit. A component is supposed to have failed if the degradation level exceeds D. The associated lifetime is then $T = \inf\{t > 0 : X(t) \geq D\}$, and is distributed as [42, 29, 49]

$$F_T(t; D, x_0) = \frac{\Gamma(\alpha t, \beta(D - x_0)))}{\Gamma(\alpha t)} \tag{3.6}$$

where

$$\Gamma(\alpha t, \beta(D - x_0)) = \int_{\beta(D-x_0)}^{\infty} y^{\alpha t-1} \exp\{-y\}\, dy \tag{3.7}$$

is the incomplete gamma function. According to the work of Park and Padgett [44], the pdf of

T is expressed as

$$f_T(t;D,x_0) = \alpha\left[\Psi(\alpha t) - \log(\beta(D-x_0))\right]\left(1 - \frac{\Gamma(\alpha t, \beta(D-x_0))}{\Gamma(\alpha t)}\right) + \frac{\alpha}{\Gamma(\alpha t)}\frac{(\beta(D-x_0))^{\alpha t}}{(\alpha t)^2}\,{}_2F_2^{(D-x_0)} \tag{3.8}$$

where

$$\Psi(\alpha t) = \Gamma'(\alpha t)/\Gamma(\alpha t) \tag{3.9}$$

is the digamma function, and ${}_pF_q$ represents the generalized hypergeometric function defined by:

$${}_pF_q\left(\{a_1,\ldots,a_p\},\{b_1,\ldots,b_q\},z\right) = \sum_{k=0}^{\infty}\frac{(a_1)_k,\ldots,(a_p)_k z^k}{(b_1)_k,\ldots,(b_q)_k k!} \tag{3.10}$$

where $(a_i)_k = a_i(a_i+1)\ldots(a_i+k-1)$ and ${}_2F_2\left(\{\alpha t,\alpha t\},\{\alpha t+1,\alpha t+1\},-\beta(D-x_0)\right)$ is defined by:

$${}_2F_2 = 1 + \sum_{k=1}^{\infty}\left(\frac{\alpha t}{\alpha t+k}\right)^2\frac{(-\beta(D-x_0))^k}{k!} \tag{3.11}$$

The first two moments of the lifetime T can be evaluated numerically, using [64]

$$E[T;D,x_0] = \int_0^{\infty}(1-F_T(t;D,x_0))\,dt \tag{3.12}$$

$$E[T^2;D,x_0] = 2\int_0^{\infty}t(1-F_T(t;D,x_0))\,dt \tag{3.13}$$

Examples of the pdf and cdf of the lifetime T of the Gamma process are shown in Appendix D.

3.1.3. Parameter estimation and model selection

The purpose of this section is the estimation of the parameters (η,k) and (α,β) of the linear random coefficient model and the Gamma stochastic process model, respectively. In addition, an exploratory method is presented to select the model that best fits the degradation data.

The degradation data is collected in the following format

$$\begin{array}{ccccc}
t_1 & x_{1,1} & x_{1,2} & \cdots & x_{i,n} \\
t_2 & x_{2,1} & x_{2,2} & & \vdots \\
\vdots & \vdots & & \ddots & \vdots \\
t_m & x_{m,1} & \cdots & \cdots & x_{m,n}
\end{array}$$

where t_i are the time points and x_{ij}, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$ the degradation level at time point t_i for unit j. A unit is an instance of the degrading component. So every unit j produces a sample path that consists of the degradation levels $x_{i,j}$ at time points $i = 1,2,\ldots$. Based on this degradation data, the parameters of the degradation models are estimated.

Linear random coefficient model

The unit-to-unit variation of the degradation rate Θ is assumed to follow a Weibull distribution with scale parameter η and shape parameter k. First the degradation rate $\hat{\Theta}_i$ of each unit j is estimated using the least squares method [38, pp. 391–399]

$$\hat{\Theta}_j = \frac{\sum_{i=1}^{n} t_i \times x_{ij}}{\sum_{i=1}^{n} t_i^2}$$

Subsequently, we estimate the parameters of the Weibull distribution from the sample slopes $\hat{\Theta}_i$, using the maximum likelihood estimator as given in [1]. This estimation is attractive because the computations are relatively simple. This method is less appealing when degradation paths are nonlinear. This method may give adequate estimates if (i) there are enough data for precise estimation of the Θ_j values; and (ii) the amount of measurement error is small [37, p. 337].

Gamma process

Suppose that the time increments are uniform, i.e., $\Delta t = t_{i+1} - t_i$ for all $i \geq 1$. Then the observed increments Δx_k, $k = 1, 2, \ldots, m$ are i.i.d. random variables with a $\Gamma(\alpha \Delta t, \beta)$ distribution, and two classical estimators can be used: the empirical estimator, and the maximum likelihood estimator. However, it is known that the likelihood estimators for i.i.d. observations with finite moments are consistent [49]. Therefore, the maximum likelihood estimators are prefered over the empirical estimators. The maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ are the solution of the following equations [49]:

$$m \ln\left(\hat{\alpha} \frac{m \Delta t}{\sum_{k=1}^{m} \Delta x_k}\right) + \Delta t \sum_{k=1}^{m} (\ln(\Delta x_k) - \Psi(\hat{\alpha} \Delta t)) = 0$$

where $\Psi(\cdot)$ is the digamma function as given in eq. (3.9), and

$$\hat{\beta} = \hat{\alpha} \frac{n \Delta t}{\sum_{j=k}^{m} \Delta x_k}$$

This can be solved by the use of standard iterative procedures (e.g., the Newton-Raphson method).

Furthermore, the 95% confidence interval (CI) of the ML estimates is constructed using the *likelihood ratio* statistic as described in [50]. MATLAB computes the 95% CIs by default.

Model selection

Once the parameter estimates have been obtained (i.e., for the linear random coefficient model $\hat{\eta}$ and $\hat{k}$, and for the Gamma process $\hat{\alpha}$ and $\hat{\beta}$) it must be determined which degradation model fits the degradation data best. We propose a graphical method that is now described.

The distribution distribution of the fitted linear random coefficient model follows a Weibull distribution with scale parameter $\hat{\eta}t$ and shape parameter $\hat{k}$. The fitted Gamma process follows a Gamma distribution with scale parameter $\hat{\alpha}t$ and inverse shape parameter β. Using these results, the 5th, 50th, and 95th percentile can be computed at each time point t_j for the estimated degradation distribution ($\hat{F}_{X(t)}$) and compared to the empirical percentiles (i.e., of the degradation data). The nth percentile of the distribution of $X(t)$ is defined as:

$$\hat{p}_n(t) = \hat{F}^{-1}_{X(t)}\left(\frac{n}{100}\right) \tag{3.14}$$

where $F^{-1}_{X(t)}(\cdot)$ is the inverse cdf of $X(t)$.

If the percentiles of the fitted degradation model match the empirical percentiles, then this indicates that the model fits the degradation data well. For example, 90% of the degradation paths should lie between the 95th and 5th percentile of the fitted model. Based observation of the percentile plots of both models, the model that best fits the degradation data can be selected.

3.1.4. Summary

In this section we have introduced the concept of degradation modeling. We have given a brief overview of two types of degradation formulations, and analyzed one instance of both degradation formulations in greater detail. For both the linear random coefficient model and the Gamma stochastic process we have deduced closed form expressions of the lifetime distribution function. Furthermore, we have discussed maximum likelihood methods to estimate the model parameters, so that the chosen degradation model can be fitted to the available ASML condition data. Subsequently, the fitted model can be compared to the degradation data using percentile plots. In summary, with the results of this section, we can

(i) model the degradation process of a component as a linear random coefficient model or a Gamma process;
(ii) estimate the model parameters from the available degradation data and compute the 95% confidence intervals;
(iii) and select the best model on the basis of a percentile plot of the degradation paths and the fitted models.

The degradation models form the basis of the condition-based maintenance models that are developed in sections 3.2, 3.3, and 3.4.

3.2. Failure-based maintenance and condition-based maintenance

The purpose of this section is to evaluate the expected total maintenance cost and the expected downtime for one component in a single machine that is subject to parameter monitoring, and is thus eligible for condition-based maintenance.

As described in the previous section, the degradation of the condition of the monitored component can be described by a degradation model that is noted as $X(t)$, which is assumed to be be a monotonic function of time $t > 0$. The initial level of degradation x_0 is assumed to be zero.

In addition, recall from the analysis in section 1.3.5, that two critical limits are set by ASML engineers:

- A critical failure limit (noted as D). Once the condition crosses the critical limit at time T, the monitored component is supposed to have failed, and a system unscheduled down (USD) occurs, during which failure-based maintenance (FBM) is performed.
- In addition, a warning limit is set (noted as D_{pm}). Once the condition crosses this limit at time T_{pm}, a notification is triggered that a failure is imminent, and preventive maintenance

(PM) on this component must be performed as quickly as possible. This warning limit is set well below the critical failure limit, such that there is enough time available to schedule a PM action, before the component actually fails. The time that is required to schedule a PM action is noted as ℓ and is equal to 48 hours. It includes (i) the time that is required for an emergency shipment of a spare part (24 hours); plus (ii) the monitoring delay of 24 hours (cf. 1.3.5).

These two critical limits define the failure-based and condition-based maintenance policy, respectively. For the derivation of the expected maintenance cost (g) of the aforementioned policies, we use elementary renewal theory [9], which implies that

$$g = \lim_{t\to\infty} \frac{C(t)}{t} = \frac{ECC}{ECL}$$

where ECC and ECL denote the expected cycle cost and expected cycle length, respectively. In other words, g can be calculated by deriving the expected cycle length and cost. A cycle ends if the monitored component is renewed: it is assumed that any maintenance actions renews the monitored component to its initial state (i.e., $X(t) = x_0$).

In the following two subsections, we first evaluate the failure-based policy, which resembles the current ASML policy and serves as a reference policy for the other (pro-active or condition-based) policies. Subsequently, we evaluate a condition-based maintenance policy for this component. The latter policy forms the basis of a decision model for synchronizing condition-based maintenance in sections 3.3 and 3.4.

3.2.1. *Failure-based maintenance*

Under this policy, the monitored component is maintained upon failure, which is at time T when the degradation level has reached the critical failure limit D:

$$T = \min\{t : X(t) \geq D\}$$

Once the monitored component fails, an unscheduled down (USD) occurs, during which corrective maintenance (CM) is performed in order to restore the system as quickly as possible. Accordingly, the maintenance cost are divided into a USD setup cost (c_s^{usd}), and a CM cost (c_{cm}). In the case of ASML, these generic cost parameters can be specified, based on the description in 1.3.7, as follows:

$$c_s^{usd} = c_e + d_s^{usd}(c_l^{usd} + c_p) \tag{3.15}$$

and

$$c_{cm} = c_m^{cm} + d_{cm}(c_l^{usd} + c_p) \tag{3.16}$$

where

c_e : is the average emergency shipment cost of a spare part;
c_m^{cm} : is the average material cost that is consumed during a corrective repair;
d_s^{usd} : is the downtime (cf. section 1.3.2) that is related to the setup of a USD. That is, the customer interface delay, diagnostics time, parts & tools delay, stabilization time, and other delay;

d_{cm} : is the downtime that is related to corrective maintenance on the monitored component. That is, the repair time plus the metrology time;

c_p : is the downtime penalty cost per unit of downtime;

c_l^{usd} : is the labor cost per unit of *unscheduled* downtime.

Please note that downtime is accounted for in the maintenance costs through the downtime penalty cost (c_d). This is because it is assumed that system downtime is negligible with respect to the total up-time, which is reasonable since the minimum system availability is 95%.

In accordance with the description of the FBM policy, the expected total maintenance cost are given by

$$g_{\text{fbm}} = \frac{c_s^{usd} + c_{cm}}{E[T]} \tag{3.17}$$

and the corresponding expected downtime is

$$d_{\text{fbm}} = \frac{d_s^{usd} + d_{cm}}{E[T]} \tag{3.18}$$

This failure-based maintenance policy resembles the current ASML maintenance policy, and serves as a reference policy for the other pro-active or condition-based policies to be developed in this chapter.

3.2.2. Condition-based maintenance

Under this pro-active policy, a PM is scheduled with lead time ℓ once the degradation level hits the critical warning limit $D_{pm} < D$, at time T_{pm}:

$$T_{pm} = \min\{t : X(t) \geq D_{pm}\}$$

It is assumed that D_{pm} is set below the critical failure limit D, so that there is always enough time to perform PM at time $T_{pm} + \ell$ before the component actually fails at time T. That is,

$$\Pr\{T > T_{pm} + \ell\} = 1$$

The cost for performing a planned maintenance action consists of two parts: the setup cost of a scheduled down (SD) of the system (noted as c_s^{sd}), plus the cost for preventive maintenance (noted as c_{pm}). For ASML, these two generic cost parameters can be specified as follows:

$$c_s^{sd} = c_e + d_s^{sd}(c_l^{sd} + c_p) \tag{3.19}$$

and

$$c_{pm} = c_m^{pm} + d_{pm}(c_l^{sd} + c_p) \tag{3.20}$$

where

d_s^{sd} : is the downtime (cf. section 1.3.2) that is related to the setup of a SD. That is, the stabilization time, and other delay;

d_{cm} : is the downtime that is related to preventive maintenance on the monitored component;

$c_l^{sd:}$ is the labor cost per unit of *scheduled* downtime;

c_m^{pm} : is the average material cost that is consumed during a preventive repair.

It can be easily seen that the expected total maintenance cost is obtained by

$$g_{\mathrm{cbm}} = \frac{c_s^{sd} + c_{pm}}{E[T_{pm}] + \ell} \tag{3.21}$$

and the corresponding expected downtime is

$$d_{\mathrm{cbm}} = \frac{d_s^{sd} + d_{pm}}{E[T_{pm}] + \ell} \tag{3.22}$$

3.2.3. Improving the warning limit

The default critical warning limit D_{pm} is set by ASML engineers by trial and error. Because of the variation in the degradation of components, it is difficult to determine an accurate warning limit; one that is close to the critical failure limit, but not too close. More specifically, the time that expires between when a warning is triggered (T_{pm}) and the time of failure (T), also known as the *lead time to failure*, should always be higher than the time required to schedule a PM (ℓ). In mathematical terms, this is expressed as

$$\min\{T - T_{pm}\} \geq \ell$$

or equivalently

$$\Pr\{T - T_{pm} > \ell\} = 1$$

Interestingly, with the estimated degradation model $X(t)$ of the monitored component, we can derive the cdf for the lead time to failure ($T - T_{pm}$). Hence, this enables us to determine the maximum value of D_{pm} such that the above holds true. Let this improved warning limit be noted as D'_{pm}. The previous condition-based maintenance policy can be re-evaluated with the improved warning limit, in order to assess the expected cost as well as downtime savings that can be achieved.

3.3. Synchronizing condition-based maintenance with random system failures

In this section a mathematical model is developed for the evaluation of synchronizing condition-based maintenance (CBM) on a monitored system component with unscheduled downs due to random failures of other system components.

3.3.1. Model description

We consider one continuously monitored component of a single system, that is subject to degradation. The degradation of the condition of the monitored component is described by a degradation model that is noted as $X(t)$, and the initial level of degradation x_0 is assumed to be zero. Once the degradation level reaches the critical warning limit D_{pm} at T_{pm} a PM is scheduled with lead time ℓ. The cost of performing a scheduled PM action consists of two parts: the setup cost of a scheduled down (SD) of the system (noted as c_s^{sd}), plus the cost for preventive maintenance (noted as c_{pm}). These cost parameters are specified in great detail in the previous section.

The occurrence of a random system failure is modeled as a Poisson process with a constant failure rate λ. From the analysis of the random failures of the ASML AT/XT system (cf. Appendix H) it follows that it is possible to describe the random failure process with a constant

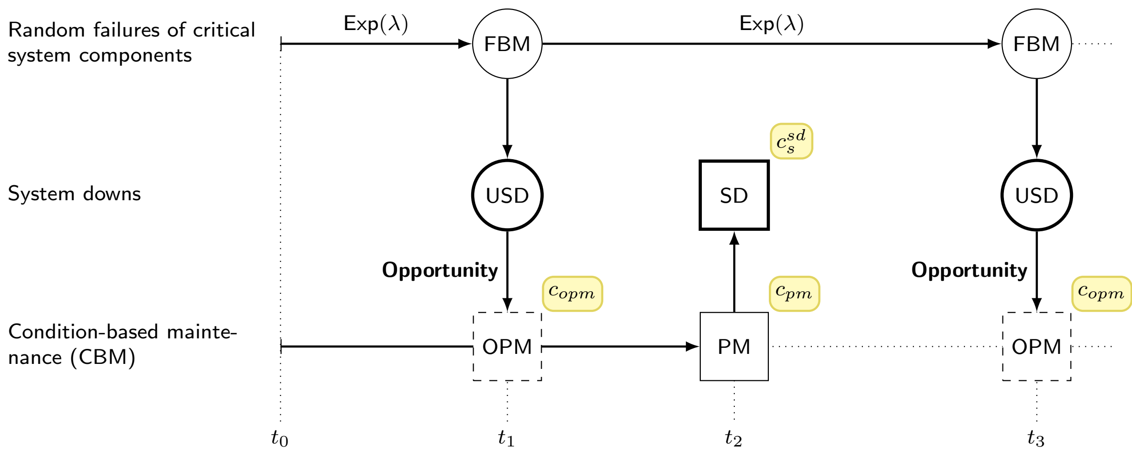


Fig. 3.5: *Synchronizing condition-based maintenance with failure-based maintenance*

failure rate. As described in section 1.3.2, as soon as any critical component in the system fails, the system breaks down. This is called an unscheduled down (USD) of the system.

Under the current policy, these system USDs are considered as free *opportunities* to perform preventive maintenance on the monitored component. This is known as opportunistic preventive maintenance (OPM). Performing maintenance opportunistically is interesting, since no additional setup cost are incurred as the system is down already due to a random failure. This is demonstrated in Fig. 3.5. If the opportunity to perform PM on monitored component is taken, no additional setup cost will be charged and there will only be an OPM cost for this monitored component. It must be noted that the cost for OPM and PM are usually not the same, because PM is performed during a system SD, whereas OPM is performed jointly with corrective maintenance of a faulty component, during a USD of the system. Therefore, we introduce a separate parameter for the OPM cost, that is denoted by c_{opm}. For ASML, the cost of OPM during a system USD, is specified as follows:

$$c_{opm} = c_m^{pm} + d_{pm} \cdot c_l^{usd} \tag{3.23}$$

where c_m^{pm}, d_{pm}, and c_l^{usd} are the cost of material for a PM, the required downtime for the repair, and the labor cost per unit of unscheduled downtime, respectively (as previously described in section 3.2).

Now, the consequences of (not) undertaking such an opportunity are discussed in greater detail:

- If the opportunity is undertaken, the monitored component can be repaired before its degradation level reaches the critical PM limit D_{pm}, and no additional down of the system is required. On the other hand, the monitored component is still operational at the time an opportunity occurs, and by preventively repairing it, its remaining useful life is wasted.
- If the opportunity is rejected, the risk is taken that the state of the component soon reaches the critical PM limit D_{pm} and a scheduled down must be arranged for PM on the monitored component, which incurs an additional SD setup cost (c_s^{sd}). On the other hand, less useful life of the monitored component is wasted. Please note that this depends on the setting of the critical PM limit (D_{pm}) and the critical failure limit D. For example, if D_{pm} is set well below the critical failure limit, then the time when the critical limit is

reached, might be long before the monitored component is supposed to really fail.

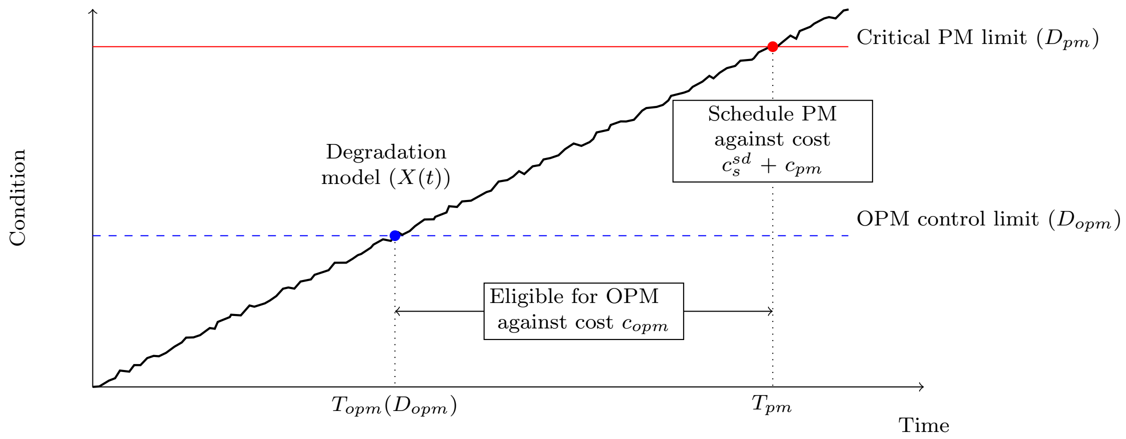


Fig. 3.6: *Condition control model for OPM and PM. On the basis of the degradation level and the OPM control limit D_{opm}, an opportunity is either undertaken or not.*

The decision to accept or reject such an opportunity is modeled as follows (see Fig. 3.6). Let $D_{opm} < D_{pm}$ be the control limit for the opportunistic preventive maintenance (OPM) of the monitored component. Once the degradation level crosses this OPM control limit, the monitoring component becomes eligible for OPM. That is, any opportunity to simultaneously perform PM on the monitored component is accepted. Otherwise, the opportunity is rejected. Setting the OPM control limit too low results in superfluous preventive repairs and a short life cycle of the monitored component. On the other hand, a high D_{opm} implies that hardly any opportunity is utilized and expensive setup costs are incurred. The objective is to minimize the expected total maintenance cost per unit time g. The problem can be formulated mathematically as follows:

$$\begin{array}{lrl} \text{(P1)} & \text{Min} & g(D_{opm}) \\ & \text{Subject to} & D_{opm} \in [0, D_{pm}] \end{array} \tag{3.24}$$

Input parameters

The model input parameters are now listed.

- λ : Random failure rate of the system. Arrival rate of opportunities for simultaneous maintenance of the monitored component due to FBM.
- c_s^{sd} : Set-up cost for a system SD for planned PM on the monitored component.
- c_{opm} : Cost for OPM on the monitored component (during a system USD).
- c_{pm} : Cost for PM on the monitored component.
- D_{pm} : Critical PM limit.
- ℓ : The minimum time that is required for arranging a PM action on the monitored component. In other words, the lead time for scheduling a PM.

Remember that in case of ASML the (generic) cost parameters can be specified as given in eq. (3.19), (3.23), (3.16), and (3.20).

Assumptions

All modeling assumptions are now listed.

- The condition of the monitored component is summarized as a continuous *monotonic* degradation model, and is noted as $\{X(t), t \geq 0\}$. Later, $X(t)$ takes on the form of a linear random coefficient model, or a stochastic Gamma process (cf. section 3.1).
- The failure-based maintenance process is modeled as a homogeneous Poisson process, noted as $\{N(t), t \geq 0\}$, with a constant failure rate λ. This implies that the failure process is *memoryless* with exponentially distributed inter-arrival times with mean $1/\lambda$. From analysis of the failure rate of AT/XT lithography systems (cf. Appendix H), it appears the there exists a considerable lifetime period during which the failure rate is constant and failures occur randomly over time.
- $N(t)$ and $X(t)$ are mutually independent, which means the failure process and the degradation process of the respective system components do not affect each other.
- A repair which can be anything from a simple cleaning or recalibration to a complete swap, returns the monitored component to its initial state of operation ("as-good-as-new").
- Maintenance is instantaneous, implying that system downtime is negligible with respect to the uptime of the system. This is reasonable since the uptime is at least 95% or more. Because downtime is expensive, system downtime is penalized via downtime penalty cost that is included in both the setup and maintenance cost parameters.
- Monitoring of the system is continuous and perfect, i.e., it reveals instantaneously the true state of the monitored component. This assumptions is reasonable if the ASML conditions models are precise, i.e., a low false alarm count, and a high failure coverage (cf. 1.3.5).
- Opportunistic maintenance of the monitored component is performed in parallel with the corrective maintenance of a randomly failed critical system component. It is assumed that there is sufficient time and labor available during the USD to simultaneously perform OPM.
- The critical failure limit D and the critical PM (warning) limit (D_{pm}) are set by ASML engineers. The performance of critical failure limit is validated with the entries of the confusion matrix (as explained in section 1.3.5). The critical PM limit is set such that the there is enough time (at least ℓ) left before the monitored component actually fails, to schedule a PM. In other words, the lead time to failure ($T - T_{pm}$) must always be higher than the time required to arrange a PM (ℓ). In mathematical terms: $T - T_{pm} \geq \ell$ w.p. 1, or equivalently $\Pr\{T - T_{pm} < \ell\} = 0$.
- At time T_{pm} a PM is scheduled for time $T_{pm} + \ell$. Any random failure that occurs between T_{pm} and $T_{pm} + \ell$ is not considered for OPM, because a PM is already arranged, which would otherwise be wasted.
- For ASML, the generic model cost parameters as listed in 3.3.1 can be further specified as given in in eq. (3.19), (3.23), (3.16), and (3.20).

3.3.2. Model evaluation

The opportunistic policy is evaluated in terms of the long-run expected cost per unit time, noted as g. Recall that from elementary renewal theory it follows that $g = ECC/ECL$, where ECC and ECL denote the expected cycle cost and expected cycle length, respectively. Under this policy, a cycle is ended upon either OPM or PM of the monitored component:

- Once the degradation level exceeds the OPM control limit D_{opm} at time $T_{opm}(D_{opm})$, the monitored component becomes eligible for PM together with FBM. The FBM process is described by a Poisson process with constant failure rate λ. This implies that the inter-arrival time of opportunities is exponentially distributed with mean $1/\lambda$. Let T_1 be the first opportunity that occurs after T_{opm}. Because of the memoryless property of the inter arrival times of random failures, T_1 follows an exponential distribution with mean $1/\lambda$. If any opportunity occurs, at time $T_{opm}(D_{opm}) + T_1$, before the critical PM limit is reached at time T_{pm}, the monitored component is preventively maintained together with FBM, against a cost of c_{opm}.

- If no opportunity arrives between T_{opm} and T_{pm}, which is the case if $T_{opm}(D_{opm}) + T_1 > T_{pm}$, a SD of the system is immediately scheduled, and PM is performed at time $T_{pm} + \ell$ against a cost of $c_s^{sd} + c_{pm}$.

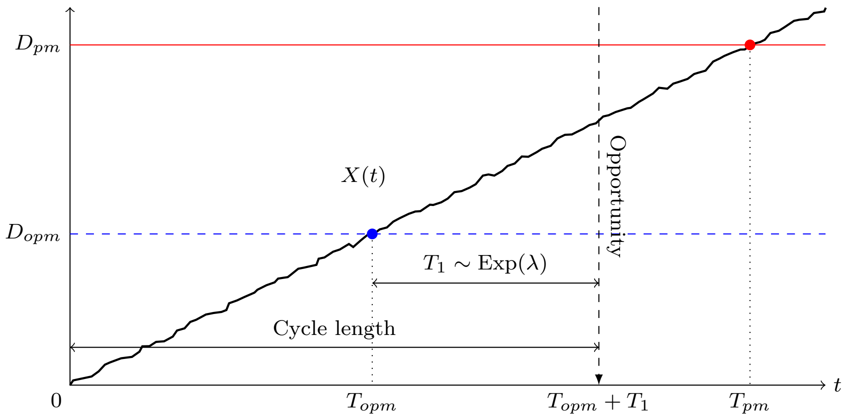


Fig. 3.7: *Graphical analysis of the renewal of the monitored component. If $T_1 < T_H - T_{opm}$ then OPM is performed at time T_1 against a cost of c_{opm}, else an PM is performed at time $T_{pm} + \ell$ against a cost of $c_s^{sd} + c_{pm}$*

This is summarized in Fig. 3.7. So, the cycle is ended either sometime during $(T_{opm}(D_{opm}), T_{pm})$, or at $T_{pm}+\ell$, depending on the occurrence of an opportunity for simultaneous maintenance of the monitored component (at time $T_{opm}(D_{opm})+T_1$). In mathematical terms, if $T_{opm}(D_{opm})+T_1 \leq T_{pm}$ then OPM is performed, otherwise PM is undertaken. Let

$$p_{opm}(D_{opm}) = \Pr\{T_{opm}(D_{opm}) + T_1 \leq T_{pm}\} \tag{3.25}$$

be the probability that OPM is performed on the monitored component. Then $1-p_{opm}(D_{opm})$ is the probability that no opportunity has been undertaken and a PM for the monitored component must be scheduled.

Now, in accordance with the renewal analysis, the long-term expected cycle cost rate can be obtained by

$$g(D_{opm}) = \frac{ECC(D_{opm})}{ECL(D_{opm})} \tag{3.26}$$

where the expected cycle cost is

$$ECC(D_{opm}) = p_{opm}(D_{opm}) \cdot c_{opm} + (1 - p_{opm}(D_{opm})) \cdot (c_s^{sd} + c_{pm})$$

and similarly, the expected cycle length is expressed as

$$\begin{aligned} ECL(D_{opm}) =& p_{opm}(D_{opm}) \cdot E[T_{opm}(D_{opm}) + T_1 | T_{opm}(D_{opm}) + T_1 < T_{pm}] \\ &+ (1 - p_{opm}(D_{opm})) \cdot E[T_{pm} + \ell | T_{opm}(D_{opm}) + T_1 > T_{pm}] \end{aligned}$$

This expression can be reduced to (cf. Appendix E.2)

$$ECL(D_{opm}) = E[T_{opm}(D_{opm})] + \frac{1}{\lambda} p_{opm}(D_{opm}) + \ell\,(1 - p_{opm}(D_{opm}))$$

where (cf. Appendix E.1)

$$p_{opm}(D_{opm}) = \int_0^\infty (1 - e^{-\lambda t}) f_T(t; D_{pm}, D_{opm})\, dt$$

Suppose that we have found the optimal OPM control limit that minimizes the expected total maintenance cost rate. Then, the corresponding expected downtime for maintenance on the monitored component is obtained by expressing the ECC in terms of downtime. Remember that an OPM is related to zero system downtime (i.e., $d_{opm} = 0$), since it is performed during a USD that is incurred by a random system failure. For PM however, an additional scheduled down must be arranged during which a preventive repair is performed, against a downtime of $d_s^{sd} + d_{pm}$ (see also eq. 3.22). Hence, the expected downtime that is incurred by the monitored component, given D^*_{opm}, is

$$ECC(D^*_{opm}) = (1 - p_{opm}(D^*_{opm}))(d_{pm} + d_s^{sd})$$

3.3.3. *Allocation of CBM*

The allocation of condition-based maintenance given the OPM control limit D_{opm}, is denoted by $p(D_{opm})$, and defines the allocation of maintenance (i.e., OPM and PM) that is performed on the monitored component in the long run. It is denoted as

$$p(D_{opm}) = [p_{pm}(D_{opm}),\, p_{opm}(D_{opm})] \tag{3.27}$$

where $p_{opm}(D_{opm}) + p_{pm}(D_{opm}) = 1$. Hence, $p_{pm}(D_{opm})$ follows from

$$p_{pm}(D_{opm}) = 1 - p_{opm}(D_{opm}) \tag{3.28}$$

Remember that $p_{opm}(D_{opm})$ is given in eq. (3.25). The CBM allocation can be used to *decompose* the expected total cost of maintenance over the respective types of maintenance. In addition, the CBM allocation defines which portion of maintenance is performed opportunistically. Hence, it shows how much maintenance can be performed without incurring additional system downs and thus downtime.

In the next section the current policy for synchronizing CBM with random failures, is extended with periodic maintenance.

3.4. Including periodic maintenance

The first model (cf. section 3.3) attempts to synchronize condition-based maintenance (CBM) of a monitored system component, and failure-based maintenance (FBM) of randomly failing system components. In this section the full maintenance operations are modeled. That is, the extended model also includes scheduled downs (SDs) of the system due to periodic maintenance (PerM), as described in section 1.3.6. These periodic scheduled downs of the lithography system provide additional opportunities for preventive maintenance on the monitored component.

First, the extended policy is described in section 3.4.1, followed by a listing of the additional assumptions and parameters. Then, the policy's evaluation model is discussed in section 3.4.2. Finally, the maintenance allocation of the monitored component is given in section 3.4.3.

3.4.1. Model description

As said, the extended model synchronizes condition-based maintenance, failure-based maintenance, *and* periodic maintenance. Periodic maintenance (PerM) is performed at fixed time intervals $n\tau$, $n = 1, 2, \ldots$ (e.g., weekly). So, the current model considers *two* sources of opportunities for simultaneous PM on the monitored component (see Fig. 3.8):

1. Scheduled downs (SDs) of the system every τ units of time, for periodic maintenance ; and
2. Unscheduled downs (USD) of the system due to random failures of critical system components (with failure rate λ).

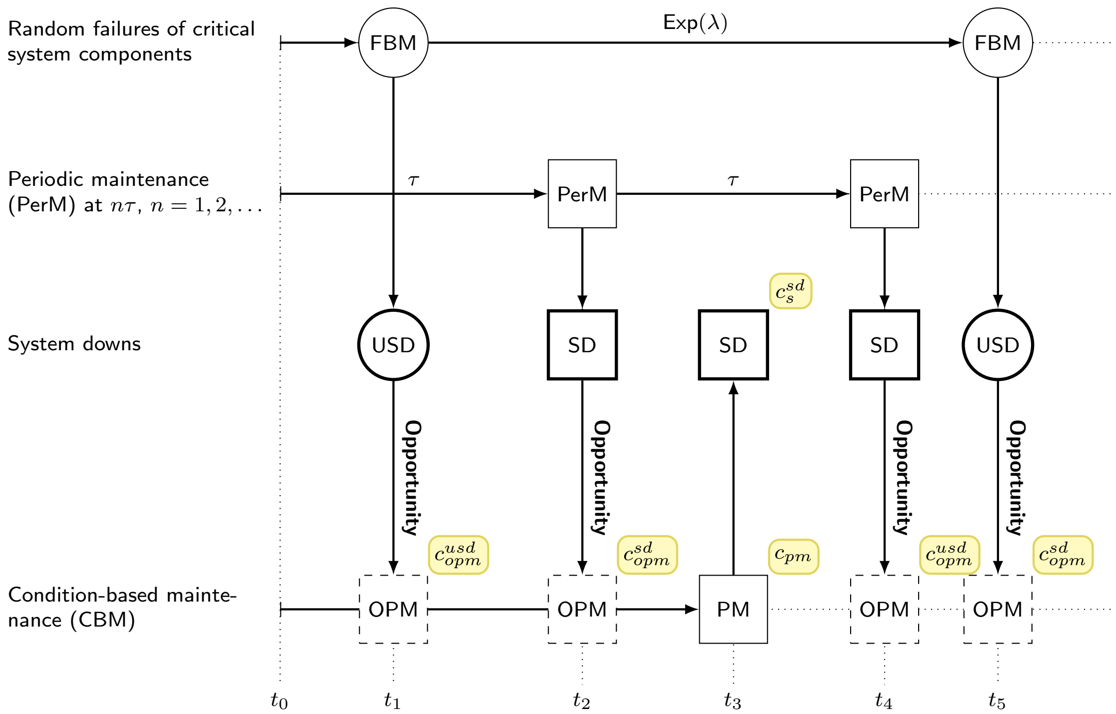


Fig. 3.8: *Structure of the extended maintenance policy. Besides opportunities that come from the FBM process, additional opportunities arise because the machine is shut down periodically in order to perform preventive maintenance actions*

In general, periodic maintenance is aimed at improving the reliability of a system. However, it is assumed that the inclusion of periodic maintenance schedule does not influence the failure rate of the system. In other words, the PerM interval τ is not a decision variable, but an input parameter. In order to model τ as a decision variable that can be optimized, the failure rate λ

must be modeled as a function of τ. So that a shorter PerM interval, results in a lower random failure rate of the system. We leave this for future research.

The decision to accept or reject an opportunity to preventively repair the monitored component (e.g. at t_1, t_2, t_3, and t_4 in Fig. 3.8) is again specified through the OPM control limit, D_{opm}. This implies that there is only one control limit for both types of opportunities: those that arise from FBM (e.g., at t_1 and t_5), and those from PerM (e.g., at t_2 and t_4). It is likely that by specifying separate control limits for both type of opportunities, better synchronization results can be achieved. However, this will likely increase the complexity of the problem formulation, evaluation, and optimization, as it becomes a two-variable problem.

The opportunistic PM cost is different for both type of opportunities. This is because for ASML the labor cost during a system SD is smaller than during a system USD. For this reason, we introduce two OPM cost parameters: c^{sd}_{opm} and c^{usd}_{opm}, where the latter is equal to c_{opm} as defined in the preliminary model, see eq. (3.23). The former can be specified as

$$c^{sd}_{opm} = c^{pm}_{m} + d_{pm} \cdot c^{sd}_{l} \tag{3.29}$$

where c^{sd}_{l} is the labor cost per unit of *scheduled* downtime, and c^{pm}_{m} and d_{pm} are the respective average material cost and average downtime for PM, as described in section 3.2.

If in the event of a USD or SD (due to resp. a random failure or a PerM action), the degradation level $X(t)$ exceeds the OPM control limit, then the opportunity to preventively repair the monitored component is undertaken; otherwise, the opportunity is rejected. If the degradation level exceeds the critical PM limit, then an SD of the system is immediately arranged for PM on the monitored component with lead time ℓ (this is 48 hours in the case of ASML). Just as in the preliminary model, it is assumed that once the critical PM limit is crossed, there is enough time left to schedule a PM, before the monitored component actually fails (i.e., more than 48 hours).

Again, the objective is to determine the OPM control limit D_{opm} that minimizes the expected maintenance cost per unit time, g. In other words, the model aims at finding the OPM control limit so that condition-based maintenance is optimally synchronized with both FBM and PerM. Mathematically, this problem is formulated as

$$\begin{aligned} \text{(P2)} \qquad \text{Min} \qquad & g(D_{opm}) \\ \text{Subject to} \qquad & D_{opm} \in [0, D_{pm}] \end{aligned} \tag{3.30}$$

Assumptions

Below you find a summary of the additional assumptions that have been made.

- The system is shut down for periodic maintenance every τ time units, at times $n\tau$, $n = 1, 2, \ldots$.
- The periodic maintenance process does not influence the FBM process. In other words, the periodic maintenance interval τ does not influence the failure rate of the system λ. Accordingly, τ is modeled as an input parameter instead of a decision variable. This assumption is justified, because PerM is performed on a relatively small portion of system components, whereas most system components are subject to FBM. However, it must be noted that this assumption is rather weak as it threatens the application of the model to more general settings.

- Only one control limit is specified for both types of of opportunities. This potentially limits the effectiveness of the model, but it reduces the complexity of the model and makes the analysis more convenient. It is likely that by specifying two control limits, one for OPM upon FBM and one for OPM upon PerM, better synchronization can be achieved.
- The cost for OPM during a system SD is smaller than the cost for OPM during a system USD. This is because the ASML labor cost are (approximately three times) cheaper for an SD than a USD.
- It is assumed that if OPM is performed on the monitored component at time t, the next SD of the system occurs again at time $t + \tau$, so that at time t the system is renewed. This is a reasonable assumption in case the lifetime of the monitored component is high in comparison with the periodic maintenance interval τ.

Input parameters

Below the input parameters of the extended model are listed.

- λ : Random failure of the system. Arrival rate of opportunities for simultaneous maintenance of the monitored component due to FBM.
- τ : Periodic maintenance interval. Each period an opportunity for simultaneous maintenance of the monitored component arrives due to PerM.
- c_s^{sd} : Set-up cost of a system SD for PM on the monitored component.
- c_{opm}^{sd} : Cost for OPM on the monitored component during a system SD (due to PerM).
- c_{opm}^{usd} : Cost for OPM on the monitored component during a system USD (due to FBM)).
- c_{pm} : Cost for PM on the monitored component.
- D_{pm} : Critical PM limit.
- D : Critical failure limit.
- ℓ : Lead time for scheduling a PM.

Remember that in case of ASML the (generic) cost parameters can be specified as given in eq. (3.19), (3.23), (3.29), (3.16), and (3.20).

3.4.2. Model evaluation

Analogues to the analysis in section 3.3.2, the expected total maintenance cost rate g is obtained by dividing the expected cycle cost (ECC) by the expected cycle length (ECL). The renewal analysis of the monitored component is as follows. Fig. 3.8 shows that the monitored component is renewed either

- opportunistically, during a USD simultaneously with failure-based maintenance;
- opportunistically, during an SD together with periodic maintenance; or
- if $X(t)$ hits the critical PM limit and a SD is arranged for PM.

Suppose at a time t between two scheduled downs $(n-1)\tau$ and $n\tau$, $n = 1, 2, \ldots,$ the condition of the monitored component crosses the OPM control limit. Recall that the inter-arrival time of USDs is exponentially distributed and thus memoryless. Let T_1 be the first arrival of a USD after T_{opm}, then if $t + T_1$ occurs before the critical limit is reached (at time T_{pm}) and before the next scheduled down at time $n\tau$, the monitored component is opportunistically repaired at time $t + T_1$ against a cost of c_{opm}^{usd}. If however the critical PM state is reached first (i.e., before $t + T_1$), then a SD is arranged (with lead time ℓ) and a PM is supposed to be performed at time $T_{pm} + \ell$, against a cost of $c_s^{sd} + c_{pm}$. Otherwise, if both $t + T_1$ and T_{pm} occur after $n\tau$, the monitored component is opportunistically repaired during an SD (due to PerM) at time $n\tau$, against a cost of c_{opm}^{sd}. This is mathematically formulated as follows (cf. Appendix F.1):

$$g(D_{opm}) = \frac{ECC(D_{opm})}{ECL(D_{opm})} \tag{3.31}$$

where

$$\begin{aligned} &ECC(D_{opm}) = \\ &\sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} \left\{ \int_{v=u}^{v=n\tau} \left(c_{opm}^{usd}(1 - e^{-\lambda(v-u)}) + (c_{sd} + c_{pm})e^{-\lambda(v-u)} \right) f_{T_{pm}}(v|T_{opm}(D_{opm}) = u)\, dv \right. \\ &\quad \left. + \left(c_{opm}^{usd} \left(1 - e^{-\lambda(n\tau-u)}\right) + c_{opm}^{sd} e^{-\lambda(n\tau-u)} \right) \int_{v=n\tau}^{v=\infty} f_{T_{pm}}(v|T_{opm}(D_{opm}) = u)\, dv \right\} f_{T_{opm}}(u; D_{opm})\, du \end{aligned} \tag{3.32}$$

and

$$\begin{aligned} &ECL(D_{opm}) = \\ &\quad \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} \left\{ u + \int_{v=u}^{v=n\tau} \left(\frac{1}{\lambda} \left(1 - e^{-\lambda(v-u)}\right) + \ell e^{-\lambda(v-u)} \right) f_{T_{pm}}(v|T_{opm}(D_{opm}) = u)\, dv \right. \\ &\quad \left. + \frac{1}{\lambda} \int_{v=n\tau}^{v=\infty} \left(1 - e^{-\lambda(n\tau-u)}\right) f_{T_{pm}}(v|T_{opm}(D_{opm}) = u)\, dv \right\} f_{T_{opm}}(u; D_{opm})\, du \end{aligned} \tag{3.33}$$

where $f_{T_{opm}}(\cdot)$ and $f_{T_{pm}}(\cdot)$ are the pdfs of T_{opm} and T_{pm}, respectively.

The corresponding expected downtime can be obtained by expressing the ECC in terms of downtime only. For opportunistic maintenance no additional downtime is incurred, whereas for preventive maintenance an additional scheduled down for PM must be arranged against a downtime of $d_s^{sd} + d_{pm}$ (see also eq. 3.22). Given that D_{opm}^* minimizes the expected total maintenance cost rate, the related expected downtime for the monitored component is

$$ECC(D_{opm}^*) = (d_s^{sd} + d_{pm}) \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} \left\{ \int_{v=u}^{v=n\tau} e^{-\lambda(v-u)} f_{T_{pm}}(v|T_{opm}(D_{opm}^*) = u)\, dv \right\} f_{T_{opm}}(u; D_{opm}^*)\, du$$

3.4.3. Condition-based maintenance allocation

Also for the extended model, we derive the CBM allocation. Let

$$p(D_{opm}) = [p_{sd}^{opm}(D_{opm}),\, p_{usd}^{opm}(D_{opm}),\, p^{pm}(D_{opm})]$$

denote the maintenance allocation of the monitored component, where

- $p_{sd}^{opm}(D_{opm})$ is the fraction of maintenance on the monitored component that has been performed opportunistically during a system SD, given D_{opm};
- $p_{usd}^{opm}(D_{opm})$ is the fraction of maintenance on the monitored component that has been performed opportunistically during a system USD, given D_{opm}; and
- $p^{pm}(D_{opm})$ is the fraction of maintenance on the monitored component that has been performed non-opportunistically during an extra system SD, given D_{opm}.

where

$$p_{sd}^{opm}(D_{opm}) + p_{usd}^{opm}(D_{opm}) + p^{pm}(D_{opm}) = 1$$

From the derivation of the ECC and ECL of the condition-based maintenance process in Appendix F.1, it is easily seen that

$$p^{pm}(D_{opm}) = \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} \left\{ \int_{v=u}^{v=n\tau} e^{-\lambda(v-u)} f_{T_{pm}}(v|T_{opm}(D_{opm}) = u)\, dv \right\} f_{T_{opm}}(u)\, du \quad (3.34)$$

and

$$p_{sd}^{opm}(D_{opm}) = \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} e^{-\lambda(n\tau-u)} \left(1 - F_{T_{pm}}(n\tau|T_{opm}(D_{opm}) = u)\right) f_{T_{opm}}(u)\, du \quad (3.35)$$

and then

$$p_{usd}^{opm}(D_{opm}) = 1 - (p^{pm}(D_{opm}) + p_{sd}^{opm}(D_{opm})) \quad (3.36)$$

3.5. Optimization

In this section we discuss the optimization of the evaluation models that have been developed in sections 3.3 and 3.4. Because of the complexity of the objective functions, the purpose of this section is to find appropriate heuristic methods for optimizing the evaluation models. That is, it is out of the scope of this thesis to find the mathematically exact optimal control limit D^*_{opm}. Instead, we analyze the objective functions by numerical exploration, based on which an appropriate optimization method is proposed.

The objective function of the preliminary model *appears* to be convex (see Fig. G.2). However, it is out of this thesis's scope to proof this. Regarding the extended decision model, the exploratory analysis in Fig. G.1 shows that the objective function of the extended model is not convex. However, it still appears to be an *unimodal* function. That is, it has one minimum m, for which it is monotonically decreasing for $x \leq m$ and monotonically increasing for $x \geq m$.

As both objective function appear to be (at least) unimodal, we propose *golden section search* for the optimization of both models. The golden section search is a method for finding the minimum or maximum of a strictly unimodal function, by successfully narrowing the range of values inside which the extremum is known to exist. A listing of this algorithm is available here[1]. This technique is readily available in MATLAB through `fminbnd()`.

3.6. Summary

In this chapter, we have executed the modeling process as depicted in Fig. 3.1. In section 3.1 the first step of the process was performed, namely describing the degradation process of a monitored component by a mathematical degradation model. Two types of degradation models were analyzed in detail: the linear random coefficient model and the stochastic Gamma process. With the help of a degradation model, evaluation models for (synchronizing) condition-based maintenance were developed (in sections 3.2, 3.3, and 3.4) to obtain the expected total maintenance cost rate and expected downtime. In short, the following policies have been evaluated:

- A failure-based policy that resembles the current ASML maintenance policy. Under this policy the (monitored) component simply runs to failure.
- A condition-based maintenance (CBM) policy under which a preventive maintenance action is scheduled before the monitored component really fails.

[1] http://en.wikipedia.org/wiki/Golden_section_search

- A preliminary policy for synchronizing CBM, that combines CBM of the monitored component with failure-based maintenance of randomly failing critical system components, such that no scheduled down of the system has to be planned for CBM.
- An extended policy for synchronizing CBM, that combines CBM with both FBM of randomly failing critical system components, and periodic maintenance.

In section 3.5, the optimization problem of both synchronization models is analyzed, and the golden search algorithm is proposed to minimize the objective function of both the preliminary and the extended evaluation model.

In the following chapter, the performance of the policies is demonstrated and applied to a real world setting, namely the ASML maintenance operations.

4. Case study

In the previous chapter we have derived the evaluation models of several maintenance policies. With these models we can evaluate the current ASML maintenance policy, as well as the (synchronized) condition-based maintenance (CBM) policies, in terms of both the expected total maintenance cost (ETMC) and expected downtime (EDT).

The purpose of this chapter is to demonstrate the benefits of condition-based maintenance for ASML through two case studies of the ASML AT/XT lithography system. Each case study (A and B), considers the (synchronized) CBM of a system component that is subject to parameter monitoring, and thus eligible for condition-based maintenance and synchronizing. By comparing the (synchronized) CBM policies with the current (reference) policy, the benefits of the pro-active policies can be determined.

First in section 4.1 the case studies for component A and B are prepared. The preparation is followed by a description of the results for both case A and B in section 4.2. In section 4.3 a sensitivity analysis performed in order to generate additional insights from the condition-based maintenance models. Finally, in section 4.4 the important results of both case studies and the insights of the sensitivity analysis are briefly summarized and discussed.

4.1. Preparation

The preparation consists of: (i) the estimation and selection of the degradation models; (ii) the determination of the maintenance cost parameters, and downtime parameters; and (iii) an analysis of the random failure rate, and the periodic maintenance interval.

4.1.1. Estimation and selection of degradation models

In this section two condition models of AT/XT system components (denoted as component A and B) are presented. Using the ASML condition model descriptions and the corresponding parameter data, a degradation model is selected for each component according to the methodology presented in section 3.1.

Condition models

The ASML model descriptions provide: (i) the system parameters that must be monitored to obtain the status of the monitored component; and (ii) the validated critical limits (cf. 1.3.5). The setting of the critical failure limit is validated by comparing the actual failure data with the failures implied by the failure limit. Subsequently, the warning limit is set such that there is enough time to schedule a preventive maintenance action before the actual failure occurs. By using the confusion matrix methodology (cf. Appendix B) the model precision can be calculated, that is defined as the proportion of the total number of correct failures predictions.

For component A and B the model precision is 100% and 59%. This means that model A signals only true positives, whereas model B is correct in 6 out of 10 times. A high model precision is required, since it is assumed that the condition degradation model ($X(t)$) describes the *true* state of monitored component. Obviously, this assumptions cannot be confidently justified for component B. Nevertheless, in the remainder of this case study we assume that the estimated degradation model of component B has a perfect precision, which implies that the results for component B represent the ideal case, and should thus be considered as an upper bound for the achievable results in practice. For this reason, we propose that in future research, the condition degradation model should account for false positives (i.e., imperfect models), so that also the consequences of pro-active maintenance with imprecise condition models can be analyzed.

Degradation model estimation and selection

For component A parameter monitoring data of 73 units from 1 December 2009 to 13 January 2010 (43 days) is used. For component B data of 18 units from 21 July 2008 to 29 January 2009 (192 days) is used. Both data sets describe the degradation of multiple units over time. In order to establish degradation models from parameter monitoring data, we use the estimation and selection method as described in section 3.1.3. That is, first both a linear random coefficient model and a Gamma process is fitted on the normalized degradation data. Then, the degradation model that best describes the degradation data is selected. For component A the degradation data of 73 units from 1 December 2009 to 13 January 2010 (43 days) is used. For component B the degradation data of 18 units from 21 July 2008 to 29 January 2009 (192 days) is used. Because the Gamma process is monotonic, the raw degradation paths must be "monotonized" (cf. Appendix I). For convenience, the parameter estimation procedure is implemented in MATLAB and available for reference in Appendix N.1. The selected degradation models for components A and B are shown in Table 4.1; the full results are shown in Appendix J. The degradation models for both component A and B are normalized so that the critical failure limit is $D = 100$ (%), and the initial degradation level $x_0 = 0$ (%). Then, a degradation level of e.g., $X(t) = 80$ means that the monitored component has degraded for 80%, or equivalently, has consumed 80% of its useful life.

Table 4.1: *Estimated and selected degradation models for components A and B*

	Model	**Critical failure limit** (%)	**Critical PM limit** (%)	**Parameters**	**ML Estimate**	**95% CI**	
A	Linear random coefficient	100	88	Scale η	.1591	.1398	.1819
				Shape k	3.732	2.668	5.221
B	Gamma process	100	71	Shape α	.08463	.07599	.09426
				Inverse scale β	.4114	.3512	.4820

4.1.2. Random system failure rate and periodic maintenance interval

The random failure process is modeled as a Poisson process, and must satisfy the assumption of a constant failure rate. Generally, the failure rate of a technical system consists of several periods. The failure rate of new systems is usually relatively high but decreasing due to early life failures. Once the early life failures have disappeared, the system reaches a mature state during which failures occur randomly over time. This is the useful life period of the system, during which the failure rate is constant. Finally, as the system reaches its end of life, the failure rate increases due to wear-out of system components. We are interested in this useful life period, since it is characterized by a constant failure rate, and thus satisfies the modeling assumption. In addition, we are interested in the random failures that result in significant downtime, so that the system downtime is not delayed due to the opportunistic maintenance on the monitored component. In Appendix H failure rate curve is analyzed. The analysis shows that a clear useful life period can be detected of approximately 7 years, that corresponds with a failure rate of $8.8550 \cdot 10^{-3}$ failures per day per machine (i.e., one random failure every 113 days).

For the AT/XT systems, several periodic maintenance (PerM) actions are proposed. Every month, quarter, year, and three years, a set of preventive maintenance actions should be performed by ASML customers. The precise actions are documented in the standard preventive maintenance schedules. From these schedules, the amount of required downtime that is required for this periodic maintenance is determined, and shown below. The average time to repair for

	Interval (days)	**Downtime** (hours)
Monthly	30	9.13
Quarterly	91	25.72
Yearly	365	143.72
Three-yearly	1095	209.38

component A and B is 11.3 and 24.1 hours, respectively (cf. Table K.1). The monthly preventive maintenance cannot be considered as an opportunity for OPM, since the monthly downs are too short, for both component A and B. The quarterly schedule requires roughly 26 hours of downtime on average, which is sufficient for OPM (for both component A and B). Hence, the (minimum) time between to opportunities for OPM due to PerM is a quarter or 91 days, i.e. $\tau = 91$ days.

4.1.3. Determining the maintenance cost parameters

The maintenance evaluation models in chapter 3, require various (generic) input cost parameters: the scheduled down setup cost, unscheduled down setup cost, corrective maintenance cost, preventive maintenance cost, opportunistic preventive maintenance (OPM) cost during an SD, and OPM cost during a USD. For ASML, and thus component A and B, these parameters can be obtained as given in eq. (3.19), (3.15), (3.16), (3.20), (3.29), and (3.23), respectively. Hence, with these cost specifications and the ASML maintenance and downtime parameters (see Appendix K), we can determine the generic cost parameters.

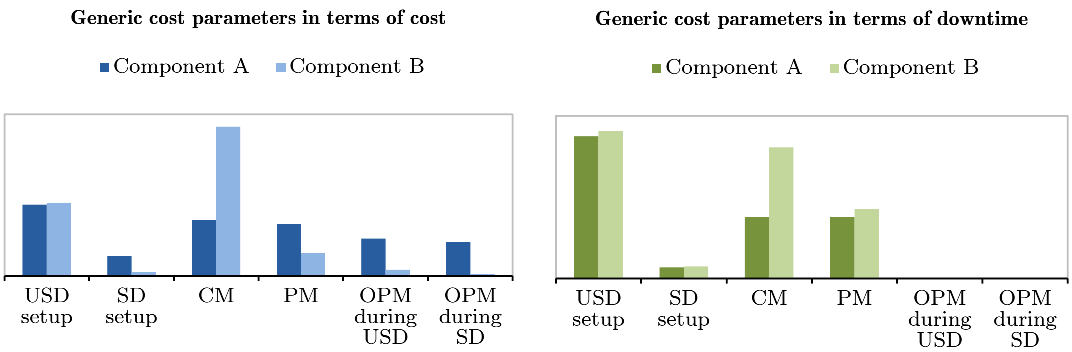


Fig. 4.1: *Plot of the generic cost parameters in terms of the total maintenance cost (left) and the corresponding system downtime (right)*

The generic cost parameters for both case A and B are computed in terms of maintenance cost as well as downtime, and are shown Fig. 4.1 and listed in Table 4.2. Please note the following that for component B the PM cost are almost negligible compared to the CM cost. This is because the CM action consists of replacing the faulty component by a new one, whereas a PM action consists of a "simple" recalibration of this component, that requires (i) less repair time (ii), no (expensive) part cost, and (iii) no emergency shipment cost. By contrast, in the case of component A the PM and CM cost about the same, since both CM and PM consists of the same repair action (only the labor cost are different). Also, remember that for OPM no extra system downtime is incurred, since OPM is simultaneously performed during a system (U)SD that inflicted by a random system failure or periodic maintenance.

Table 4.2: *Generic cost parameters*

Parameter	**Description**	**Specified in**	**A**		**B**	
			Cost^a	Downtime^b	Cost^a	Downtime^b
c_s^{usd}	Setup cost for a USD	eq. (3.15)	[redacted]	26.2	[redacted]	27.2
c_s^{sd}	Setup cost for an SD	eq. (3.19)	[redacted]	2.0	[redacted]	2.2
c_{cm}	Cost for CM	eq. (3.16)	[redacted]	11.3	[redacted]	24.1
c_{pm}	Cost for PM	eq. (3.20)	[redacted]	11.3	[redacted]	12.9
c_{opm}^{usd}	Cost for OPM during a USD	eq. (3.23)	[redacted]	0	[redacted]	0
c_{opm}^{sd}	Cost for OPM during an SD	eq. (3.29)	[redacted]	0	[redacted]	0

^{a}: In EURO, ^{b}: In hour

4.2. Results

In this section, the benefits of pro-active maintenance are evaluated for two independent cases: component A and B. With the help of the evaluation models in chapter 3, we assess the performance of each maintenance policy in terms of: the expected total cost of maintenance (ETMC), and the corresponding expected downtime (EDT). For convenience a brief summary of the current policy and the condition-based maintenance policies is now given.

- *Current*: the ASML reference policy, in which the monitored component is run to failure;
- *CBM*: the default condition-based maintenance (CBM) policy, in which preventive maintenance (PM) is performed once the monitored condition hits the PM warning limit;
- *Sync. CBM and FBM*: this improved/synchronized CBM policy attempts to perform the PM on the monitored component opportunistically, i.e., during the downtime that is incurred by a random system failure. We say that CBM is synchronized with random failure-based maintenance (FBM); and
- *Sync. CBM, FBM and PerM*: this policy attempts to synchronize CBM with both FBM and periodic maintenance (PerM). That is, this policy attempts to perform PM on the monitored component during: (i) the unscheduled downtime that is incurred by a random system failure, and (ii) the (scheduled) downtime that is incurred due to the periodic maintenance (PerM) of the system.

Besides the ETMC and EDT, we are also interested in the maintenance allocation for all policies (i.e., the fraction of CM, PM, and OPM that is performed under a given policy).

4.2.1. Expected total maintenance cost and downtime

The ETMC and EDT for each policy is depicted in Fig. 4.2[1] below. In both cases, the pro-active policies outperform the current policy both in terms of the ETMC and EDT. The absolute cost and downtime values are available for reference in Table L.1. In both cases, the CBM policy results in a reduction of the ETCM and EDT. Synchronizing CBM with FBM (and PerM) reduces the EDT and ETCM even further. The performance improvement is particularly high for component B. This can be explained by the fact that the cost of corrective maintenance (CM) during an unscheduled down (USD) is many times higher than the cost of preventive maintenance (PM) during an scheduled down (SD) (cf. Fig. 4.1 in section 4.1.3). To a lesser extent, this is also true for component A, hence a less dramatic improvement is seen there. Altogether, it appears that characteristic of case B are ideal synchronizing CBM.

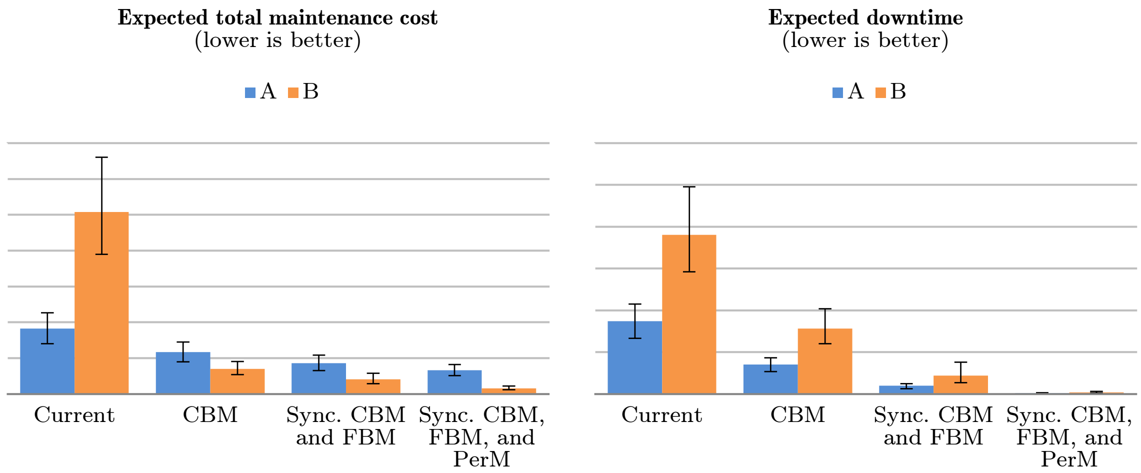


Fig. 4.2: *Performance of the reference and pro-active maintenance maintenance policies for cases A and B.*

Below in Fig. 4.2, the performance of the pro-active policies is expressed as a percentage of the current policy's performance. The plot clearly shows that the pro-active maintenance policies, especially the synchronized policies, yield a considerably lower cost and downtime, than the current ASML policy. For both cases, the pro-active policies yield a downtime reduction of roughly, 60, 90, and almost 100%, respectively. The ETMC is reduced with 36, 53, and 64%, respectively in the case of component A; whereas for component B, the pro-active policies reduce the ETMC by 86, 92, and 97% cost savings. Hence, the maximum benefits are achieved under the extended synchronized policy (CBM, FBM and PerM). The synchronized policies perform so well that they eliminate almost all downtime that is incurred by the monitored component. This is possible, because the synchronized policies attempt to *opportunistically* perform PM on the monitored component during existing system downtime.

[1] Perhaps, you have noticed the error bars; these represent the 95% confidence limits, because of the estimation error in the fitted degradation model parameters (cf. section 4.1.1).

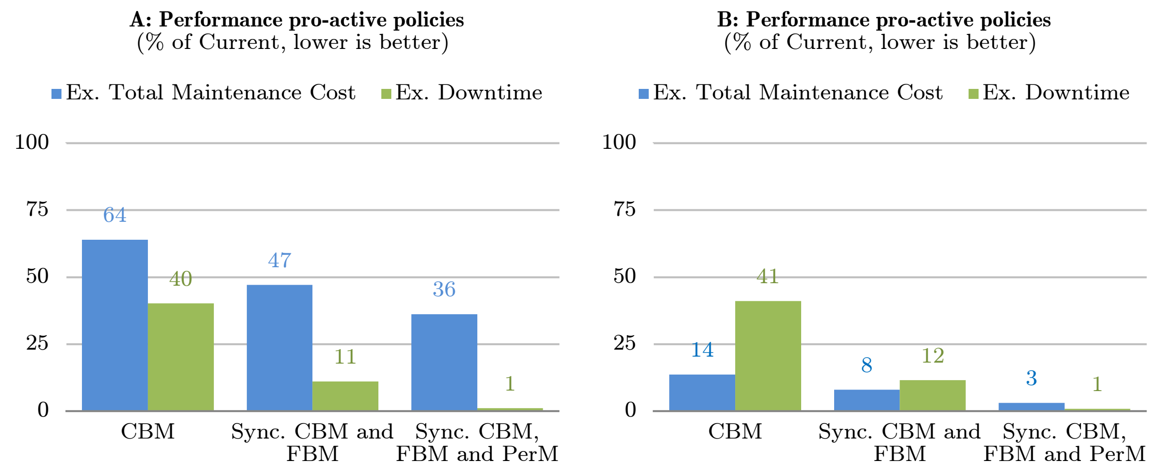


Fig. 4.3: *Relative performance of the pro-active maintenance maintenance policies for case A (left) and B (right).*

4.2.2. *Maintenance allocation*

To investigate what type of maintenance is performed in all of the policies, we have plotted the maintenance allocation per policy in Fig. 4.4. Obviously, the current policy is characterized by CM only, and the standard CBM policy as PM only. Under the synchronized CBM policies, PM can also be performed opportunistically (i.e., OPM). Recall that in the extended synchronized policy, OPM can be performed during an USD (due to a random system failure) or during a SD (for periodic maintenance). The figure shows that the preliminary synchronized CBM policy performs 3 out of 4 PM actions opportunistically in the case of component A, and 4 out of 5 in the case of B. In the extended synchronized policies, almost all maintenance (98%) is performed opportunistically for both cases. This explains the huge downtime savings under the synchronized policies, compared to the other policies.

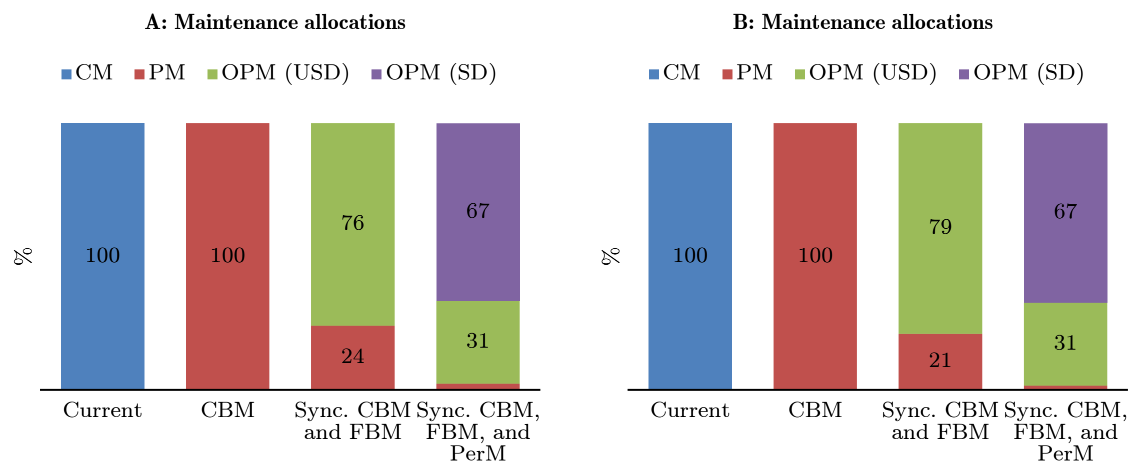


Fig. 4.4: *The maintenance allocation for all policies for cases A (left) and B (right).*

4.2.3. Impact on system availability

Using the EDT, we can compute the impact on the system availability per policy. The results are depicted in Fig. 4.5. The plot demonstrates the expected system availability can be considerably increased by subjecting both components to an extended synchronized CBM policy: roughly .2 % and .4% in case of A and B, respectively, or .6% in total.

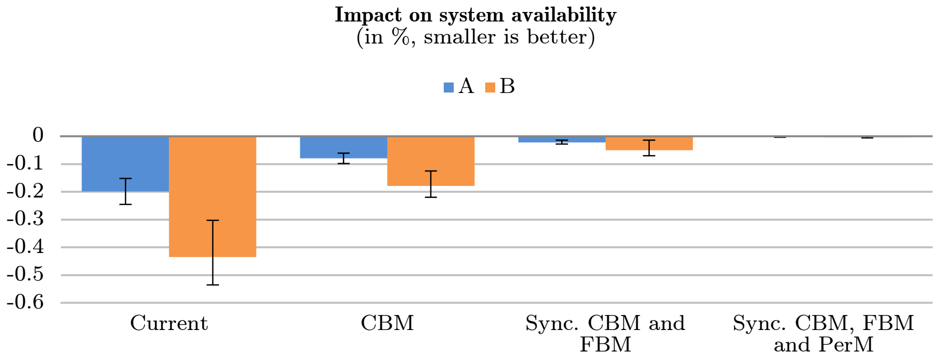


Fig. 4.5: *The impact on the system availability per policy for case A (left) and B (right)*

4.2.4. Improving the PM warning limit

The default PM warning limits are set at 88 and 71 for components A and B, respectively; and the critical failure limits are normalized at 100 (cf. Table 4.1). These are set by ASML engineers through trial and error, such that the lead time to failure once a warning has been triggered, is always greater than the PM lead. In accordance with the description in section 3.2.3, it follows that the warning limits can be improved as shown in Table 4.3 The table shows

Table 4.3: *Improved PM warning limits*

	PM warning limit (%)			**Ex. Increase of**	**Ex. lead time**	**Ex. failure**
	Default	Improved	Change	**useful life** (days)	**to failure** (days)	**time** (days)
A	88	97	+9	+71	24	786
B	71	75	+4	+25	127	492

that for both case A and B, the warning limit can be slightly increased, which results in an extension of the expected useful life, especially for component A. Please note that the average lead time to failure is now roughly 3 weeks for A, and almost 4 months for B. The latter is due to the high variance of the lifetime of component B: a high "safety time" is required to schedule PM, to ensure that the component does not fail during the lead-time of a PM. The performance of the improved pro-active CBM policies is compared to the default policies. The results are shown in Fig. 4.6 below. With the improved PM warning limit, the performance of the pro-active policies only slightly increases compared to the current ASML policy, especially for component B.

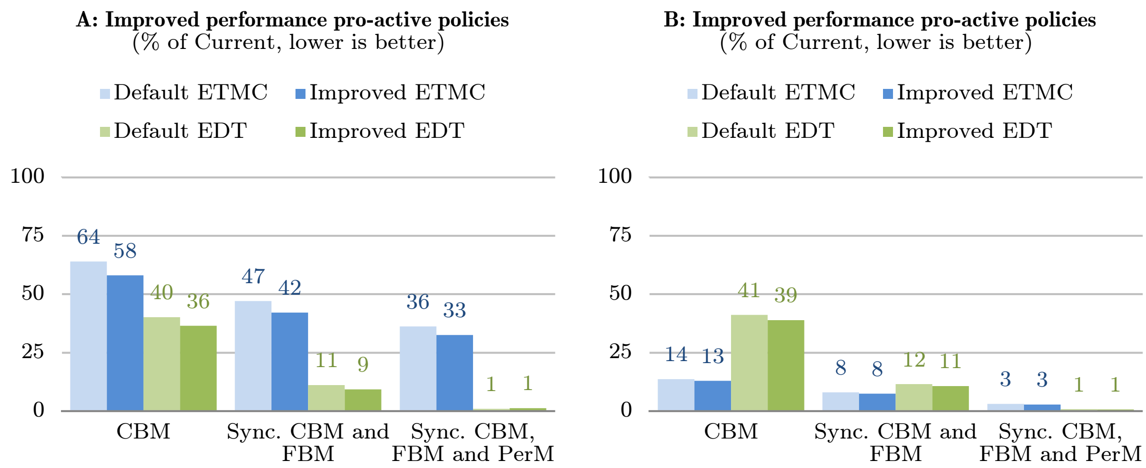


Fig. 4.6: *Performance of the improved pro-active policies for case A (left) and B (right).*

4.3. Sensitivity analysis

Now, we analyze the sensitivity of the preliminary and extended policies for synchronizing CBM. In particular, the influence of the OPM control limit, the system failure rate, and the downtime penalty cost are investigated.

4.3.1. Optimal OPM control limit

The preliminary and extended policy for synchronizing CBM with resp. FBM, and FBM and periodic maintenance (PerM), make use of an OPM control limit for undertaking PM opportunities (cf. sections 3.3 and 3.4, respectively). The OPM control limit is optimized such that it minimizes the ETCM. Now, we investigate how changes of the (optimal) OPM control limit influence the ETCM of both policies. The results are available in Fig. 4.7. For the case of com-

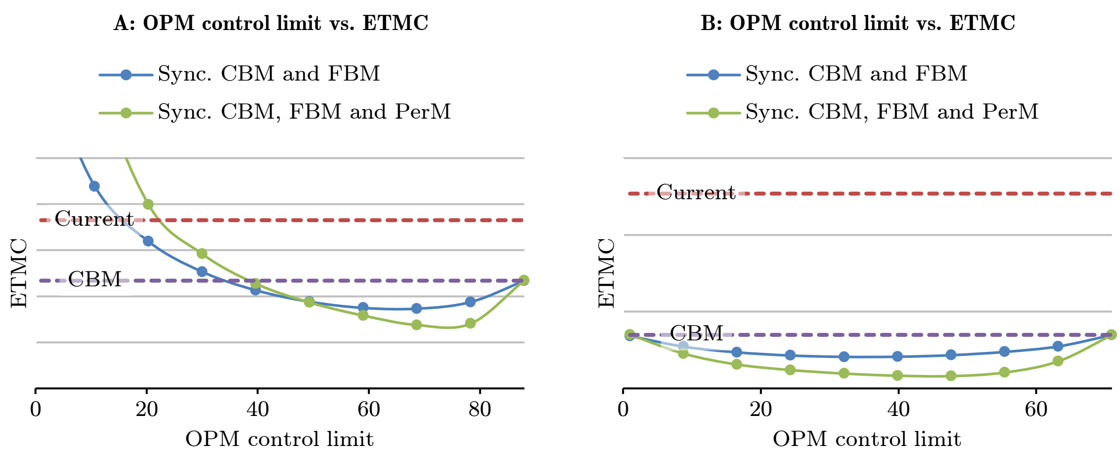


Fig. 4.7: *Sensitivity analysis of the OPM control limit for case A (left) and B (right).*

ponent A, the optimal OPM control limits are 65 and 73 for the preliminary and the extended policy, respectively. Both policies perform better than the CBM policy for a OPM control limit within $(35, D^A_{pm} = 88)$. However, for values outside this range (i.e., below 35), the performance of synchronizing CBM quickly deteriorates. A control limit of approximately 20 results in a higher ETMC than the current ASML performance. By contrast, component B is much less sensitive to changes in the OPM control limit. Every control limit within $(0, D^B_{pm} = 71)$ results in a performance that is equal to or less than the CBM policy. This can be explained by the fact that OPM cost of component B are very low in comparison to the PM plus SD setup cost. A relative low OPM control limit results in superfluous PM (cost) of the monitored component as it becomes available for PM opportunities too soon in its lifetime. However, since the OPM cost relatively low for component B, the superfluous PM actions have no significant effect on the ETMC.

4.3.2. System failure rate

Besides the OPM control limit, the performance of synchronizing CBM is also dependent on the arrival rate of opportunities for OPM. Recall that opportunities come from random failures and periodic maintenance. Fig. 4.8 shows that the preliminary policy performs better when the failure rate is high (i.e., 10 times the default failure rate, and 100 times the low failure rate), both in terms of the ETMC and EDTS. This is because random failures are the only source of OPM opportunities in the preliminary policy for synchronizing CBM. By contrast, in the extended policy an additional source of opportunities is PerM. We see that the extended policy, remains almost unaffected by changes in the system random failure rate. In fact, both the ETMC and EDT slightly increase (for both cases) for a high failure rate. This can be explained by the fact, that as the failure rate increases, the optimal OPM limit increases as well. This leads to a higher fraction of PM (+30% for A, +23% for B), OPM during a USD (24% for A, 26% for B), and a smaller fraction of OPM simultaneously with PerM (-60% for A, -50% for B); see Table L.2. Because PM and OPM during a USD is costlier than OPM during an SD, the ETMC and EDT slightly increase. This perhaps surprising effect can be solved by introducing separate OPM control limits for opportunities due to random failures, and periodic maintenance. Accordingly, we propose this is a subject for future research.

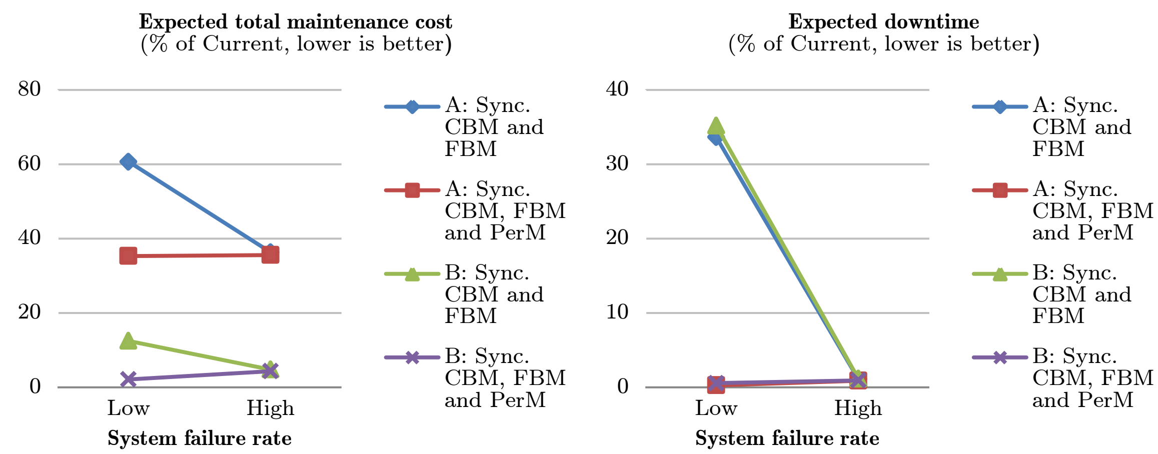


Fig. 4.8: *Sensitivity analysis of the ETMC (left) and the EDT (right) as a function failure rate.*

4.3.3. Downtime penalty cost

Recall that downtime is accounted for in the ETMC by means of a downtime penalty cost. The penalty cost for downtime is an artificial cost, since ASML is not directly affected by system downtime (except for labor cost), but the customer is because of lost production. Indirectly, through service level agreements, ASML is held responsible for (a portion) of the downtime of systems. However, this cost is not measurable. Therefore, it is important to understand how the downtime penalty cost influences the performance of the CBM policies. For this reason, we analyze two cases:

- A relatively high penalty cost ([redacted] EURO/hour), which implies that the cost of downtime for ASML is immense, and should be avoided at all cost; and
- No penalty cost, which implies that the only cost related to downtime is the labor cost for the ASML service engineers (cf. Appendix K).

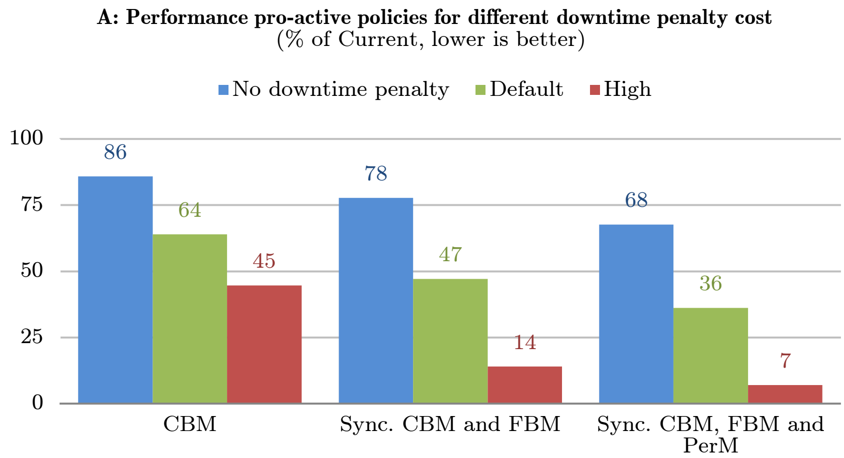


Fig. 4.9: *Influence of the downtime penalty cost on the pro-active maintenance policies for the case of component A*

From the analysis in Fig. 4.9 it follows that the performance of the pro-active policies increase as downtime becomes more costly for ASML. Earlier, it has been shown that the pro-active policies dramatically reduce the EDT that is required for the maintenance on the monitored component (see Fig. 4.2). For this reason, it is clear that the pro-active policies become more effective as the cost of downtime increase.

4.4. Discussion

In this section the performance of various pro-active policies have been compared to the current ASML policy for two cases: component A and B. In both cases the expected total cost of maintenance as well as the expected downtime can be dramatically reduced. Especially the policy for synchronizing CBM with both random failures and periodic maintenance proves to be extremely efficient, eliminating almost all system downtime that is required for the maintenance on the monitored component. Through a comparison of both cases, the characteristic of component B (very low (opportunistic) preventive maintenance and scheduled down setup cost, and very high corrective maintenance and unscheduled down setup cost) appear to be especially favorable for condition-based maintenance.

This brings us to the question of how these results translate to a situation where you have multiple monitored components subject to a pro-active policy. Can you simply add up the benefits? Unsurprisingly, it depends. First of all, the failure behavior (i.e., degradation) of the monitored components must be independent. And secondly, when the number of monitored components that is subjected to a pro-active policy increases, the number of random failures of the system decreases. This implies that there are less opportunities for parallel PM of the monitored components. This would influence the marginal benefits of the synchronized policies: For example, suppose that all system components are subject to perfect condition monitoring, and pro-active maintenance. Then, no random system failures occur anymore, which renders the preliminary policy for synchronizing CBM useless. Furthermore, in the case in which n components are subject to pro-active maintenance, then the condition-based PM actions can be synchronized with each other (i.e., synchronizing CBM with CBM). This is an important subject for future research.

5. Implementation

The results in chapter 4 shows that for components A and B the maximum benefits of (synchronizing) condition-based maintenance with both random system failures and periodic maintenance, are achieved under a given optimal OPM control limit, and improved PM warning limit. Therefore, the implementation of this policy consists of adjusted the PM warning limit, and adding the obtained control limit of undertaking opportunistic preventive maintenance on the monitored component. Subsequently, the condition-based maintenance decisions should be in accordance with these limits and the actual condition of the monitored component. The decision making based on the setup of these two limits, is depicted in Fig. 5.1 and described below.

- Once the monitored condition crosses the OPM control limit, it becomes eligible for OPM. This implies that preventive maintenance should be performed as soon as any opportunity arrives. An opportunity can either be a random system failure, or a periodic maintenance action; whichever comes first.
- If no opportunity has arrived once the monitored condition hits the PM warning limit, a scheduled down of the system should be immediately arranged for preventive maintenance on the monitored component, before it actually fails.

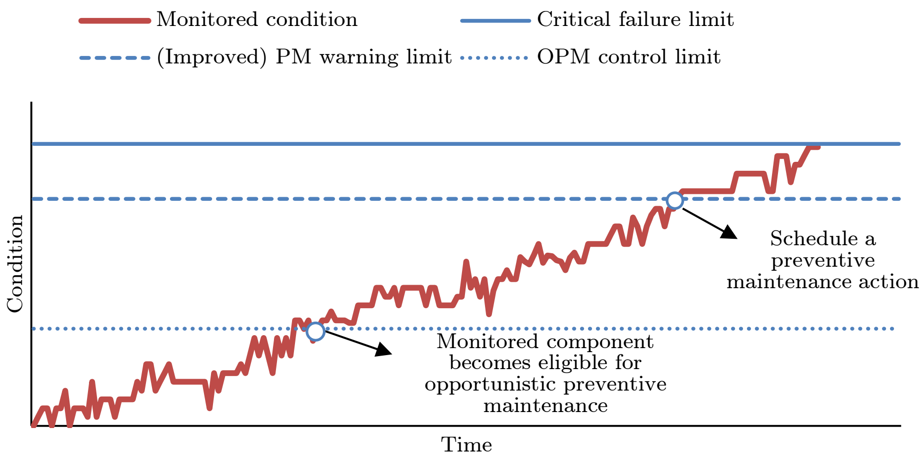


Fig. 5.1: *New condition-based control model for undertaking (opportunistic) preventive maintenance*

Generally, in order to implement a synchronized CBM policy for a monitored component, the following steps must be taken.

1. Fit a degradation model on the available degradation data of the monitored component. For the basic degradation models that are described in section 3.1, a MATLAB implementation is available in Appendix N.1.
2. Estimate the average failure rate of the system, and the periodic maintenance interval (if any) of the system (cf. section 4.1).
3. Determine the cost of corrective maintenance, (opportunistic) preventive maintenance, as well as the setup cost of a scheduled down (cf. Appendix K).

4. The critical failure limit can be determined either by design, or by comparing the actual failure data and the parameter monitoring data. The critical failure limit can be validated through the entries of the confusion matrix (cf. Appendix B).
5. Then, use the degradation model to derive the distribution of the lead time to failure, and determine an appropriate warning limit for preventive maintenance (cf. 3.2.3).
6. Then, use the evaluation models in section 3.3 (or 3.4) to determine the optimal OPM control limit for synchronizing CBM with FBM (and PerM).
7. Finally, condition-based maintenance decision making should occur in accordance with the obtained PM warning limit and the optimal OPM control limit, as described earlier in this chapter.

This procedure should be automated, and periodically executed, so that the degradation model, as well as the optimal OPM control limit and the PM warning limit are continuously adjusted, and it is ensured that maintenance decision making is up to date. This process is schematically summarized in Appendix M. In addition, this procedure can be extended to multiple monitored component on a single machine if the degradation processes of these components are mutually independent. One should remember that such a policy does not synchronize the CBM of one monitored component, with the CBM of other system components. In a true multi-component policy, this should be taken into account, and is subject for future research.

6. Conclusion and recommendations

Due to the increasing customer demands for higher precision in order to produce smaller and faster microchips, the ASML lithography systems become ever more complex. ASML customer service is challenged with satisfying the customer service levels for each new generation of lithography systems, through providing adequate service. Inevitably, in order to be able to satisfy service level agreements with respect to system downtime and availability in the future, a shift to pro-active (maintenance) service is necessary.

This project has focused on evaluating the benefits of a pro-active maintenance approach at ASML. In this pro-active approach, parameter monitoring data of the lithography systems is utilized to develop condition models that predict the status of system components. This way, the actual status of system components can be remotely monitored, and maintenance can be initiated based on the monitored condition of a system component. However, the availability of validated condition models for pro-active maintenance is still limited, and these models are only locally deployed. Therefore, most ASML systems are currently subjected to a run-to-failure policy, in which maintenance is performed reactively, only upon the random failure of a critical system component. Additionally, periodic maintenance schedules are advised to customers, that includes scheduled downs of the system for the execution of a predetermined set of preventive maintenance tasks.

In order to combine the effects of condition-based maintenance (CBM), failure-based maintenance (FBM), and periodic maintenance (PerM), we developed a model for synchronizing the CBM on one monitored system component with both FBM and PerM of other systems components. This synchronized policy considers the unscheduled downtime that is incurred by FBM, and the scheduled downtime due to PerM, as *opportunities* to perform preventive maintenance of the monitored component (also known as opportunistic preventive maintenance or OPM). This way both the cost of maintenance and the downtime that is incurred for the monitored component can be significantly reduced. Accordingly, the evaluation model concentrates on the performance in terms of the expected total cost of maintenance (ETMC) and the expected downtime (EDT).

The optimized evaluation model results in an optimal control limit for opportunistic preventive maintenance, as well as an improved warning limit for preventive maintenance (PM). With the setup of these two limits, condition-based maintenance decisions can be taken on an operational level. By comparing the performance of the according policy, to the standard (i.e., non-synchronized) policy, as well as the current reactive ASML policy, we can evaluate the impact of (synchronizing) condition-based maintenance in terms of the ETMC, EDT, and system availability. Overall, the results of two ASML case studies indicate that condition-based maintenance, especially synchronizing condition-based maintenance, greatly reduces the ETMC and the EDT, and significantly improves the system availability. Also, the results demonstrate that the cost and downtime reduction is especially worthwhile for components that are characterized by: a relatively high corrective maintenance cost, and a relatively low (opportunistic) preventive maintenance cost. Furthermore, the results show that the benefits of synchronizing condition-based maintenance are especially high for systems that are subject to a (i) relatively high failure rate, (ii) high downtime cost, and (iii) periodic maintenance.

Building on these results and insights, we recommend to

- use the decision model for synchronizing condition-based maintenance to improve the

performance of the current ASML maintenance operations;
- implement the OPM control limits in the current ASML condition models for scheduling opportunistic preventive maintenance of monitored component, in accordance with the described implementation plan;
- deploy this policy primarily on systems/components with high potential characteristics. The decision maker can use the evaluation models that have been developed in this project, to evaluate the impact of synchronizing CBM for a monitored component, and consequently select the “high potentials” for pro-active maintenance; and
- improve the default PM warning limit, for which currently no robust methodology is available. In fact, using the modeling results, one can determine the maximum PM warning limit, so that there is enough time available to arrange a scheduled down of the system for PM, before the system actually fails.

The limitations of this study have been mentioned throughout this thesis and are now listed.

- First of all, the decision model assumes that the condition of the monitored component can be perfectly predicted. However, in practice we see that most condition model are imperfect, in that they trigger false positives (e.g., false warnings for PM), as well as false negatives (no warning if PM is actually required). This obviously affects the potential of pro-active maintenance, and has not been accounted for in the current study. Therefore, the results of the case study should be considered as theoretically maximum achievable results.
- Secondly, the current decision model considers only one OPM control limit for two types of opportunities: that is, those that arrive due to random system failures; and the ones that arrive due to periodic maintenance. The results of the case study indicate that probably better results can be achieved with separate control limits.
- And finally, in this thesis, the periodic maintenance interval is considered an input parameter. But in practice customers often ignore the proposed PerM schedule by ASML.

For future research, we propose the following directions.

- Extend the decision model to facilitate the synchronizing of CBM for multiple monitored components on a single system. This opportunistic maintenance model for multiple components subject to condition monitoring, should take into account the opportunities for performing preventive maintenance on multiple monitored component simultaneously. In case the components are mutually independent, the current evaluation model can be applied to each component individually, and serve as a lower bound for the true multi-component model.
- Extent the current research with more general non-monotonic degradation formulations. The current decision model for synchronizing CBM only accounts for monotonic degradation. Although this has proven to be adequate in the ASML case study, it will enhance the model’s generic quality.
- incorporate so called multivariate degradation models for supporting multiple root causes of failure (cf. [65]). The current condition degradation models do not account for multiple dominant failure modes. Often in practice however, the deterioration of a component can be traced to various root causes.
- incorporate a time-dependent failure rate. Currently, the failure rate of the system is assumed to be constant. Although analysis shows that the failure rate of is indeed constant

during a considerable part of the system lifetime, it is noted that especially premature systems, are subject to a quickly decreasing failure rate. This is not accounted for in the current study, and is a direction for future research.

- Model the PerM interval as a decision variable that impacts both the ETMC ("adding PerM costs money"), as well as the system failure rate ("adding PerM improves the reliability of system component, and thus reduces the failure rate").
- Investigate the consequences of a (synchronized) CBM policy for the logistic spare parts planning.

References

[1] M A Al-Fawzan. *Methods for Estimating the Parameters of the Weibull Distribution*. http://interstat.statjournals.net/YEAR/2000/articles/0010001.pdf. King Abdulaziz City for Science and Technology (KACST), 2000.

[2] ASML. *About ASML*. http://www.asml.com/asml/show.do?ctx=271. ASML, Nov. 2011.

[3] ASML. *FAQ*. http://www.asml.com/asml/show.do?lang=EN&ctx=32894. ASML, Apr. 2012.

[4] ASML. *Fulfilling the potential of semiconductor lithography*. http://www.asml.com/doclib/press/misc/asml_20080404_CB_notprintable.pdf. ASML, Apr. 2012.

[5] ASML. *Sustainability - Performance and targets*. http://www.asml.com/asml/show.do?ctx=40715&rid=40723. ASML, Nov. 2011.

[6] M Bengston. "On condition based maintenance and its implementation in industrial settings". PhD thesis. Mlardalen University, 2007.

[7] A Christer and W Wang. "A simple condition based monitoring model for a direct monitoring process". In: *European Journal of Operational Research* 82.2 (1995), pp. 258–269.

[8] F.P.A Coolen and R Dekker. "Analysis of a 2-Phase Model for Optimization of Condition-Monitoring Intervals". In: *IEEE Transactions on Reliability* 44.3 (1995), pp. 505–511.

[9] D R Cox. *Renewal Theory*. Wiley, New York, 1962.

[10] M Cross. "Raising the value of maintenance in the corporate environment". In: *Management Research News* 11.3 (1988), pp. 229–240.

[11] R Dekker. "Applications of maintenance optimization models: a review and analysis". In: *Reliability Engineering and System Safety* 51.3 (1996), pp. 229–240.

[12] R Dekker, R E Wildeman, and F A van der Duyn Schouten. "A Review of Multi-Component maintenance Models with Ecnonomic Dependence". In: *Mathematical Methods of Operations Research* 45.3 (1997), pp. 411–435.

[13] G van Dijkhuizen and A van Harten. "Optimzal clustering of frequency-constrained maintenance jobs with shared set-ups". In: *European Journal of Operational Research* 99.3 (1997), pp. 552–564.

[14] F.A van der Duyn Schouten and S.G Vanneste. "Two Simple Control Policies for a Multicomponent Maintenance System". In: *Operations Research* 41.6 (1993), pp. 1125–1136.

[15] A.H Elwany and N.Z Gebraeel. "Sensor-driven prognostic models for equipment replacement and spare parts inventory". In: *IIE Transactions* 40.7 (2008), pp. 629–639.

[16] A Grall, C Brenguer, and L Dieulle. "A condition-based maintenance policy for stochastically deteriorating systems". In: *Reliability Engineering and System Safety* 76.2 (2002), pp. 167–180.

[17] F R S de Gusmo, E M M Ortega, and G M Cordeiro. "The generalized inverse Weibull distribution". In: *Springer Statistical Papers* 52.3 (2011), pp. 591–619.

[18] A Helle. “Development of prognostic concepts and tools”. In: *Prognostics for prognostic machinery availability.* Espoo, Finland: VTT Technical Research Centre of Finland, 2006, pp. 5–15.

[19] K Holmberg et al. *Safety and reliability - Technology review.* Research Report. Espoo, Findland: VTT Industrial Systems, 2004.

[20] A.K.S Jardine and J.A Buzacott. “Equipment reliability and maintenance”. In: *European Journal of Operations Management* 19.3 (1985), pp. 285–296.

[21] A.K.S Jardine, D Lin, and D Banjevic. “A review on machinery diagnostics and prognostics implementing condition-based maintenance”. In: *Mechanical Systems and Signal Processing* 20.7 (2006), pp. 1483–1510.

[22] R Jiang and A.K.S Jardine. “Health state evaluation of an item: A general framework and graphical representation”. In: *Reliability Engineering and System Safety* 93.1 (2008), pp. 89–99.

[23] D Juki and J Markovi. “On nonlinear weighted least squares fitting of the three-parameter inverse Weibull distribution”. In: *Mathematical Communications* 15.1 (2010), pp. 13–24.

[24] R Kohavi and F Provost. “Glossary of terms”. In: *Machine Learning* 30.2-3 (1998), pp. 271–274.

[25] R Kothamasu, S.H Huang, and W.H VerDuin. “System health monitoring an prognostics - a review of current paradigms and practices”. In: *International Journal of Advanced Manufacturing Technology* 28.9-10 (2006), pp. 1012–1024.

[26] A Kranenburg and G-J van Houtum. “A new partial pooling structure for spare parts networks”. In: *European Journal of Operational Research* 199.3 (2009), pp. 908–921.

[27] V G Kulkarni. *Modeling, Analysis, Design, and Control of Stochastic Systems.* New York: Springer-Verlag, 1999.

[28] C.T Lam and R.H. Yeh. “Optimal maintenance policies for deteriorating systems under various maintenance strategies”. In: *IEEE Transactions on Reliability* 43.3 (1994), pp. 423–430.

[29] J Lawless and M Crowder. “Covariates and Random Effects in a Gamma Process Model with Application to Degradation and Failure”. In: *Lifetime Data Analysis* 10 (2004), pp. 213–227.

[30] E.E Lewis. *Introduction to Reliability Engineering.* 2nd. John Wiley & Sons, Inc., 1996.

[31] T.Y Liang. “Optimum Piggyback Preventive Maintenance Policies”. In: *IEEE Transactions* 34.5 (1985), pp. 529–538.

[32] H Liao, E.A Elsayed, and Chan L.Y. “Maintenance of continuously monitored degrading systems”. In: *Proceedings of the European JOurnal of Operational Research* 175.2 (2006), pp. 821–853.

[33] J Lu, J Park, and Q Yang. “Statistical inference of a time-to-failure distribution derived from linear degradation data”. In: *Technometrics* 39.4 (1997), pp. 391–400.

[34] J C Lu and W Q Meeker. “Using Degradation Measures to Estimate a Time-to-Failure Distribution”. In: *Technometrics* 35.2 (1993), pp. 167–174.

[35] K.F Martin. "A review by discussion of condition monitoring and fault analysis in machine tools". In: *International Journal of Machine Tools and Manufacture* 34.4 (1994), pp. 527–551.

[36] S Mathew. "Optimal inspection frequency: A tool for maintenance planning/forecasting". In: *International Journal of Quality & Reliability Management* 21.7 (2004), pp. 763–771.

[37] W Q Meeker and L A Escobar. *Statistical methods for reliability data*. New York: Wiley & Sons, 1998.

[38] D C Montgomery and G C Runger. *Aplied Statistics and Probability for Engineers*. John Eiley & Sons, Inc., 2007.

[39] J Moubray. *Reliability centered maintenance*. New York: Industrial, 1997.

[40] P D N Murthy, M Xie, and R Jiang. *Weibull Models*. John Wiley & Sons, Inc., 2004.

[41] G Niu, B Yang, and M Pecht. "Development of an optmized condition-based maintenance system by data fusion and reliability-centered maintenance". In: *Reliability Engineering and System Safety* 95.7 (2010), pp. 786–796.

[42] J M van Noortwijk. "A survey of the application of gamma processes in maintenance". In: *Reliability Engineering and System Safety* 94 (2009), pp. 2–21.

[43] M Ohnishi, H Kawai, and H Mine. "An optimal inspection and replacement policy for a deteriorating system". In: *Journal of Applied Probability* 23.4 (1986), pp. 973–988.

[44] C Park and W Padgett. "Accelerated degradation models for failure based on geometric Brownian motion and gamma process". In: *Lifetime Data Analysis* 11 (2005), pp. 511–527.

[45] C Peng and S Tseng. "Mis-specification analysis of linear degradation models". In: *IEEE Transactions on Reliability* 58.3 (2009), pp. 444–455.

[46] H Peng, Q Feng, and D.W Coit. "Reliability and maintenance modeling for systems subject to multiple dependent competing failure processes". In: *IIE Transactions* 43.1 (2010), pp. 12–22.

[47] S-T. Tseng C-Y. Peng. "Mis-Specification Analysis of Linear Degradation Models". In: *IEEE Transactions on Reliability* 39.4 (2009), pp. 444–455.

[48] Y Peng, M Dong, and M.J. Zuo. "Current status of machine prognostics in condition-based maintenance: a review". In: *International Journal of Advanced Manufacturing Technology* 50.1-4 (2010), pp. 297–313.

[49] M Roussignol. *Gamma stochastic process and application to maintenance*. http://perso-math.univ-mlv.fr/users/roussignol.michel/CoursProcGamma.pdf. Universit Paris-Est, 2009.

[50] A B Slavkovi. *Loglikelihood and Confidence Intervals*. http://sites.stat.psu.edu/~sesa/stat504/Lecture/lec3_4up.pdf. Pen State University, 2005.

[51] T Tan. *Service Supply Chains for Capital Goods: Elementary Maintenance Models*. 1CM30 Course Handouts. Eindhoven University of Technology, Feb. 2011.

[52] B L R Timmermans. "Master thesis preparation — Literature review". MA thesis. Eindhoven University of Technology, 2012.

[53] M Todinov. "Reliability Analysis Based on the Losses from Failures". In: *Reliability Analysis* 26.2 (2006), pp. 311–335.

[54] M Treacy and F Wiersema. "Customer Intimacy and Other Value Disciplines". In: *Harvard Business Review* 71.1 (1993), pp. 88–93.

[55] A.H.C Tsang. "Condition-based maintenance: tools and decision making". In: *Journal of Quality Maintenance Engineering* 1.1 (1995), pp. 3–17.

[56] A.H.C Tsang. "Strategic dimensions of maintenance management". In: *Journal of Quality Maintenance Engineering* 8.1 (2002), pp. 7–39.

[57] J. Virtamo. *Queueing Theory: Poisson process*. English. http://www.netlab.tkk.fi/opetus/s38143/luennot/poisson.pdf. S-38.143 Course Handouts. Helsinki University of Technology, Jan. 2012.

[58] H Wang. "A survey of maintenance policies of deteriorating systems". In: *European Journal of Operational Research* 139.3 (2002), pp. 469–489.

[59] W Wang. "A model to determine the optimal citritical level and the monitoring intervals in condition-based maintenance". In: *International Journal of Production Research* 38.6 (2000), pp. 1425–1436.

[60] W Wang and A Christer. "Towards a general condition based maintenance model for a stochastic dynamic system". In: *Journal of Operational Research Society* 51.2 (2000), pp. 145–155.

[61] W Weibull. "A statistical distribution function of wide applicability". In: *Journal of Applied Mathematics and Mechanics* 18 (1951), pp. 293–297.

[62] R.E Wildeman, R Dekker, and A.C.J.M Smit. "A dynamic policy for gourping maintenance activities". In: *European Journal of Operations Research* 99.3 (1997), pp. 530–551.

[63] R.C.M Yam et al. "Intelligent Predictive Decision Support System for Condition-Based Maintenance". In: *International Journal of Advanced Manufacturing Technology* 17.5 (2001), pp. 383–391.

[64] X Yuan. "Stochastic Modeling of Deterioration in Nuclear Power Plant Components". PhD thesis. University of Waterloo, 2007.

[65] J Zhou, Z Pan, and Q Sun. "Bivariate Degradation Modeling Based on the Gamma Process". In: *Proceedings of the World Congress on Engineering (WCE)*. London, 2010.

Acronyms

CBM	Condition-Based Maintenance
cdf	cumulative distribution function
CFR	Constant Failure Rate
CI	Confidence Interval
CM	Corrective Maintenance
DFR	Decreasing Failure Rate
ECC	Expected Cycle Cost
ECL	Expected Cycle Length
EDT	Expected Downtime
ETCM	Expected Total Cost of Maintenance
EUV	Extreme Ultra Violet
FBM	Failure-Based Maintenance
i.i.d.	independently and identically distributed
IFR	Increasing Failure Rate
KPI	Key Performance Indicator
LD	Long Down
MSE	Mean Squared Error
MTTF	Mean Time To Failure
OPM	Opportunistic Preventive Maintenance
pdf	probability density function
PdM	Predictive Maintenance
PerM	Periodic Maintenance
PM	Preventive Maintenance
RCM	Reliability Centered Maintenance
SA	Simulated Annealing
SD	Scheduled Down
TBM	Time-Based Maintenance
USD	Unscheduled Down

Nomenclature

$X(t)$: Degradation model that describes the degradation of a monitored component at time t
x_0 : Initial amount of degradation. In other words, the degradation level at $t = 0$
D : Critical failure limit
T : Lifetime of the monitored component
$f_T(t; D, x_0)$: Pdf of the lifetime T given D and x_0
$F_T(t; D, x_0)$: Cdf of the lifetime T given D and x_0
Θ : Unit-to-unit variation in the linear random coefficient model
η, k : Scale and shape parameters of the Weibull distribution function
α, β : Shape and inverse scale parameters of the Gamma distribution
$\Gamma(\cdot, \cdot)$: The incomplete gamma function
$\Gamma(\cdot)$: The gamma function
$\Psi(\cdot)$: The digamma function
${}_pF_q(\cdot, \cdot, \cdot)$: Generalized hypergeometric function

t_i : Time point i
x_{ij} : Degradation level of unit j at time point t_i
$\hat{\Theta}_i$: Degradation rate of the degradation sample path of unit j
η, k : Maximum likelihood estimators for η and k
Δt : Uniform time increment
$\hat{\alpha}$, $\hat{\beta}$: Maximum likelihood estimators for α and β
$\hat{p}_n(t)$: nth percentile of the distribution of $X(t)$
$F^{-1}_{X(t)}$: Inverse cdf of $X(t)$

D_{pm} : Critical preventive maintenance warning limit
T_{pm} : The time at which the degradation of the monitored component crosses D_{pm}
D_{opm} : Control limit for opportunistic preventive maintenance of the monitored component
T_{opm} : The time at which the degradation of the monitored component crosses D_{opm}
ℓ : The minimum time that is required to arrange a PM of the monitored component
g : The total expected maintenance cost (per unit time)
ECC : Expected cycle cost
ECL : Expected cycle length

c_s^{usd} : Setup cost for a system unscheduled down for corrective maintenance on the monitored component
c_{cm} : Cost for corrective maintenance on the monitored component
c_{pm} : Cost for preventive maintenance on the monitored component
c_e : Average emergency shipment cost of a spare part
c_m^{cm} : Average material cost (parts & tools) that is consumed during a corrective repair
c_m^{pm} : Average material cost (parts & tools) that is consumed during a preventive repair

d_s^{usd} : Amount of downtime that is related to the setup of a unscheduled down

d_{cm} : Amount of downtime that is related to the corrective maintenance on the monitored component

d_{pm} : Amount of downtime that is related to the preventive maintenance on the monitored component

c_p : Downtime penalty cost per unit of downtime

c_l^{usd} : Labor cost per unit of unscheduled downtime

g_{fbm} : Expected total maintenance cost for failure-based maintenance of the monitored component

d_{fbm} : Expected downtime for failure-based maintenance of the monitored component

c_s^{sd} : Setup cost for a system scheduled down for preventive maintenance on the monitored component

d_s^{sd} : Amount of downtime that is related to the setup of a scheduled down. That is, the stabilization time plus other delay

c_l^{usd} : Labor cost per unit of unscheduled downtime

g_{cbm} : Expected total maintenance cost for condition-based maintenance of the monitored component

d_{cbm} : Expected downtime for condition-based maintenance of the monitored component

D'_{pm} : Improved critical preventive maintenance warning limit

λ : Random failure rate of the system

τ : Periodic maintenance interval

c_{opm}^{usd} : Cost for OPM on the monitored component during a system USD (due to FBM)

c_{opm}^{sd} : Cost for OPM on the monitored component during a system SD (due to PerM)

c_l^{sd} : Labor cost per unit of scheduled downtime

$g(D_{opm})$: Expected total maintenance cost as a function of the control limit for OPM

$ECC(D_{opm})$: Expected cycle cost as a function of the control limit for OPM

$ECL(D_{opm})$: Expected cycle length as a function of the control limit for OPM

$p(D_{opm})$: Maintenance allocation of the monitored component

$p_{sd}^{opm}(D_{opm})$: Fraction of maintenance on the monitored component that has been performed opportunistically during a system SD, as a function of the control limit for OPM

$p_{usd}^{opm}(D_{opm})$: Fraction of maintenance on the monitored component that has been performed opportunistically during a system USD, as a function of the control limit for OPM

$p_{pm}(D_{opm})$: Fraction of maintenance on the monitored component that has been performed non-opportunistically during an extra system SD, as a function of the control limit for OPM

Definitions

(Un)scheduled down process
: The steps that are taken during a(n) (un)scheduled down in order to bring back the system to operation.

(Un)scheduled downtime
: The time that is required to bring back the system to operation after the occurrence of a(n) (un)scheduled down.

Condition model
: A condition model indirectly assesses the state, condition, or health of a system component from (a combination of) system parameters.

Condition-based maintenance
: Preventive maintenance that is carried out on the basis of the condition of the system. This condition can be directly assessed by inspection, or remotely through parameter monitoring.

Condition-based maintenance allocation
: Defines the fraction of (opportunistic) preventive maintenance that is applied on the monitored component.

Critical limit
: The setup of this limit implies that once the condition of the monitored component exceeds this limit it is supposed to have reached a critical state and it thus requires maintenance as quickly as possible.

Degradation model
: A degradation model is a mathematical formulation that attempts to describe the deterioration of a component's condition over time. The two main classes of degradation models are: (i) random coefficient models, and (ii) the stochastic processes.

Downtime composition
: Defines the amount of downtime that is allocated to each of the steps in the (un)scheduled down process. Each component has a unique downtime allocation, because each step in the downtime process consumes a portion of time that is dependent on the component that caused system down.

Downtime penalty cost
: The (indirect) cost that is incurred per unit of system. This is mostly a cost that is not visible on the ASML balance, and takes into account the affects on system availability, customer lost sales, and customer satisfaction.

Failure-based maintenance
: Maintenance that is carried out only after the occurrence of a failure. Also known as corrective maintenance and run-to-failure maintenance.

Gamma process
: A stochastic process in which the continuous (degradation) increments follow a Gamma distribution.

Linear random coefficient model
: A model that formulates the degradation process as a linear function with a random slope.

OPM control limit
: The setup of this limit implies that once the condition of the monitored component exceeds

this limit is becomes eligible for opportunistic maintenance.

Opportunistic maintenance
Preventive maintenance that is performed opportunistically. That is, in combination with the maintenance of other (system) components. Performing PM together with failure-based maintenance is also known as piggybacking (cf. [31]).

Parameter monitoring
The monitoring of system parameters through sensors embedded in the system.

Periodic maintenance
Preventive maintenance that is carried out periodically.

Scheduled down
A down of the system that has been planned in advance for preventive maintenance.

Synchronizing
Aligning the execution of maintenance tasks of two or more maintenance processes through considering opportunistic maintenance, in order to save maintenance cost and/or downtime.

Time dependent variation
Implies that degradation is a gradual process with random increments over time.

Unit-to-unit variation
Implies that the rate of degradation varies from unit to unit.

Unscheduled down
A breakdown of the system due to a sudden failure of a critical system component.

List of Figures

List of Tables

A. Literature overview

Turbulence and dynamics are the hallmark of many industries today. For this reason, it is crucial to continuously enhance the capability to create value to customers, and improve the cost effectiveness of the operational processes. Maintenance, as an important support function in business with considerable investments in physical assets, plays a key role in fulfilling this order [56].

When a product or system ceases to perform its intended function, or according to its specifications, it is said to fail (e.g., an engine stops running, or a structure collapses). The occurrence of a cessation of function indicates a clear system failure [30]. Generally, the failure of high-tech long-lived systems has dramatic consequences, such as production losses, delays, and unplanned intervention to the system. Possible problems of failures in production systems that can incur huge costs, are [53, 6]: (i) lost sales due to stoppage of production time and volume; (ii) product quality issues, and possibly warranty payments; and (iii) cost of mobilization of emergency resources.

Today in industry, the unexpected stoppage costs are as high as up to 400,000 euros per day [18]. Research of Holmberg et al. [19] shows that the economic consequences of an unexpected 1-day stoppage may become as high as 100,000 to 400,000 euros. Besides unplanned maintenance interventions being costly, maintenance in general comprises a large part of the total operating costs; e.g. in the UK manufacturing industry, maintenance spending comprises between 12 and 23 per cent of the total operation costs [10]. In addition, the maintenance and operations departments are the largest and each comprises up to 30 per cent of total staffing according to the research of Dekker et al. [11]. Despite these figures, the importance of cost effective maintenance is not widely acknowledged in industry. In fact, maintenance activities have traditionally been regarded as a necessary evil, and maintenance frequently emerges in cost-reduction programs. Developments that have led to the increasing importance and required performance (in terms of availability and reliability) of maintenance are [56] emerging trends of operation strategies, toughening societal expectations, technological changes; and changes in the people and organizational systems.

A.1. Maintenance classification

In recent decades, maintenance and replacement problems have been widely investigated. A classification scheme of maintenance models that is amenable to current theoretical development is presented in Table A.1. In general, two broad classes of maintenance activities can be distinguished according to Niu et al. [41], namely corrective maintenance and preventive maintenance. Reliability centered maintenance is classified as predictive maintenance.

Corrective maintenance is the most basic form of maintenance, also known as failure-based maintenancne, breakdown maintenance, and run-to-failure maintenance. Corrective maintenance (CM) is "[...] carried out only after the occurrence of an obvious functional failure, malfunction, or breakdown of equipment" [63], and "[...] is intended to restore an item to a state in which it can perform its required function" [25]. It is a reactive approach because maintenance actions are triggered by the unscheduled event of a system or equipment failure, and is thus dominated by unplanned events. As a result of this policy, maintenance related costs are usually high due to [55]: (i) the high cost of restoring equipment to an operable condition under crisis situation; (ii) the secondary damage and safety/health hazards inflicted by the failure;

Table A.1: *Maintenance classification scheme*

Maintenance classes		
Preventive	*Predictive*	*Corrective*
· Age-based, usage-based	· Reliability-centered	· Failure-based
· Time-based, periodic		
· Condition-based		
· Imperfect		

and (iii) the penalty associated with lost production, i.e. opportunity costs.

Within the PM paradigm, items are replaced or returned to a good-as-new condition before a failure occurs to reduce the probability of failure or performance degradation [25]. Typically, PM involves system shutdown at fixed intervals for overhaul or predetermined repair activities on still functioning equipment. This is known as time-directed (TD) or time-based maintenance (TBM). This strategy helps preventing functional failures by replacing critical components "at regular intervals just shorter than their expected useful lifetime" [63]. In addition to a time-based schedule, PM can also be usage or age based. Hence, preventive maintenance can be classified into periodic, age-based, or imperfect maintenance.

- Periodic maintenance is performed at fixed intervals, in addition to maintenance prompted by failure that is performed correctively.
- Age-based maintenance is similar to periodic maintenance in that it carried out at fixed intervals. However, the time considered is the operational time of the equipment. So, maintenance is performed when the equipment has reached a specific age t. If failure occurs before t, corrective maintenance is performed.
- The above two strategies restore the system to its original condition, after a maintenance action. Imperfect maintenance, however, returns the system to a state between good (original) and bad (failure). It considers the uncertainty of the current system state while scheduling preventive maintenance activities.

Contrary to preventive maintenance which is performed on a fixed schedule, predictive maintenance (PdM) is adaptively determined [25]. Reliability-centered maintenance (RCM) is a PdM approach that "[...] utilizes reliability estimates of the system to formulate a cost-effective schedule for maintenance" [25]. It is formally defined by Moubray [39] as: "A process sued to determine what must be done to ensure that any physical asset continues to do what its users want it to do in its preset operating context".

A.2. Condition-based maintenance

Another type of PM is condition-based maintenance (CBM), also known as condition-directed maintenance. CBM is a preventive maintenance technique in which maintenance resources are employed based on the condition or "health" of the system that is assessed through condition monitoring. An actual preventive action is only initiated upon the detection of an incipient failure [55]. Therefore, unlike PM, a CBM program can significantly reduce maintenance cost by reducing the number of superfluous scheduled preventive maintenance operations [21]. In particular, CBM is better than age-based maintenance in terms of maintenance timeliness since the decision to intervene with a maintenance action is based on the state of health, rather than relying solely on age [22]. CBM is generally regarded as a proactive approach to reduce uncertainty of maintenance activities, whereas other classic PM approaches are often regarded as reactive approaches. The following definition nicely captures the above introduction to CBM: "[It] is a maintenance strategy where decisions are made depending on either continuously or regularly measured equipment states. It is often an efficient tool for cost-effective maintenance, since compares with time-based preventive maintenance, it reduces uncertainty with respect to actual states of equipment, and can thus avoid unnecessary repair or replacement." [11].

As the term implies, condition monitoring monitors the condition (or 'health') of the system under consideration [35]. This does not involve intrusion into the system, as condition monitoring is usually achieved via censoring technology embedded in the system. In general two types of data are collected: event data and condition monitoring data. Event data includes information of what happened to the system under consideration. Condition monitoring data includes the measurements related to the health of the physical asset, e.g. vibration data, acoustic data, oil analysis data, temperature, pressure, moisture, and humidity. Condition monitoring data is collected via various types of sensory and wireless technologies and consequently stored in (maintenance) information systems. Condition monitoring is an important part of the data acquisition step in the CBM process of maintenance decision making. Both Jardine et al. [21] and Martin [35] postulate that three main steps are needed to support maintenance decision making:

1. Data acquisition is the process of collecting and storing information from targeted physical assets for the purpose of CBM.
2. Data processing includes the cleaning, analysis, and processing of the condition data.
3. Maintenance decision-making which can be divided into diagnostic and prognostic techniques.

Two approaches to maintenance decision support can be distinguished: diagnostics and prognostics. Fault diagnostics focuses on the detection, isolation and identification of failures when they occur. Prognostics on the other hand, "addresses the use of automated methods to detect, diagnose, and analyze the degradation of physical system performance, calculating the remaining life in acceptable operating state before failure or unacceptable degradation of performance occurs" [48].

The key of a successful CBM program is precise and reliable prognostics. Well known prognostic methods are predicting the remaining useful life (RUL) and the probability that a failure occurs given the current machine condition and past operation profile (i.e. degradation path). For a complete and detailed overview of diagnostic and prognostic methods, see Jardine et al. [21] and Peng et al. [48].

A.3. Conventional maintenance models

Considering the modeling of conventional maintenance policies, different maintenance objectives can be distinguished, such as the reliability, availability, and the total cost of maintenance. In addition to reliability itself, other common (and equivalent) quantifications of reliability include the mean time to failure and the failure or hazard rate. One of the most used statistics to characterize random failure behavior is the mean time to failure (or MTTF). A random failure process is generally modeled according to a Poisson process [57], using failure characteristics, such as the mean time to failure [30, 20], or a hazard/failure rate function [36, 27]. The failure rate typically consists of three distinct areas, namely

1. the early life stage, characterized by high but decreasing failure rates (DFR), and often referred to as the period of infant mortality, or early failures;
2. the useful life contains nearly constant failure rates (CFR). This constant failure rate is caused by random events and thus referred to as random failures; and
3. the wear out stage, characterized by increasing failures rates (IFR). During this period aging failures become dominant.

Thus, a failure-based maintenance model for random failures is usually appropriate during the useful life period, where the failure rate is constant. The lifetime of a part or system is commonly represented by a Weibull distribution [61] because of its flexibility and thus wide applicability [51, 40, 20]. A useful result to evaluate the cost-effectiveness of a maintenance policy is Renewal Reward Theory [27], from which follows that the long-run expected cost can be obtained by calculating the expected cost during a renewal cycle of a system, divided by the expected cycle length. In addition to a policy's objective(s), one must consider the decisions variables of the policy, such as the inspection interval in time-based maintenance, and the critical failure limits in condition-based maintenance [55].

A.4. CBM modeling

Regarding the modeling of a CBM policy a considerable and growing variety of methods and techniques exists, (e.g. [46, 15, 32]). However, common use is the monitoring of the condition of a specific item regularly, and based on the obtained reading and the set-up of both the critical failure and preventive limit, maintenance decisions are taken. Obviously, the determination of the critical level will influence the consequences of CBM [59]. If the critical level is set at a relatively low level, the number of system failures may reduce, but at a loss of unexpired remaining life as most maintenance will be premature. On the other hand, setting the critical level too high may result in a prolonged use of an item, but at a loss of reliability.

The modeling and optimization of a condition-based maintenance policy can be roughly described by the identified five-step process: the modeling of the system deterioration process through degradation modeling [34, 37, 47], estimation of model parameters, estimation of the failure time distribution [32, 59, 34], decision modeling, and optimization of decisions. This modeling process is discussed in great detail in section 3.1. In literature the majority of models (cf. [8, 16, 28, 43, 7, 60]) assume a periodic inspection interval, whereas continuous monitoring is also applicable in practice.

A.5. Multi-component maintenance

Multi-component maintenance systems are of increasing importance, not only in traditional areas such as the aircraft industry, but also in the design and operation of computers and other service facilities [14]. The goal in maintenance optimization is to balance preventive maintenance and corrective maintenance. Usually, cost-saving opportunities arise from economic dependencies between components. However, single-unit maintenance does not consider these dependencies. On the other hand, multi-component maintenance policies do take into account the dependencies between components by performing either group, or opportunistic maintenance on a set of components. Accordingly, in literature two types of multi-component maintenance policies can be identified: group maintenance (e.g. [13] and [62]) and opportunistic maintenance. The former performs preventive maintenance simultaneously on a group of components when a failure occurs, which is preferable from the point of view of the system's reliability or operational cost [58]. The latter allows that corrective maintenance can be used to save set-up costs, "which is not possible in the former type of models, since there only preventive/plannable maintenance is grouped" [12, p. 424]. In other words, "due to economies of scale in maintenance cost functions, the unpleasant event of a failing component is at the same time considered as a opportunity for preventive maintenance of other components" [12, p. 424]. The idea of combining PM with CM is also known as piggybacking (cf. Liang [31]), where PM of a component can only be performed in combination with CM of another component, the piggybackee.

In a multi-component system subject to condition monitoring, economic dependencies between components can be exploited on the basis of component deterioration measures; that is, maintenance can scheduled, on the basis of the condition of components. Also, condition-based maintenance may be included in opportunistic maintenance policies, where PM of the monitored component can be combined with both corrective and preventive maintenance of other components. This, i.e., opportunistic CBM, is an area in multi-component literature that has not yet been researched. This is an area of maintenance literature that this master thesis attempts to explore and extend.

B. Confusion matrix

		Prediction	
		Failures	*Non-failures*
Actual	*Failures*	(a) True Positive	(b) False Negative
	Non-failures	(c) False Positive	(d) True Negative

Table B.1: *Confusion matrix*

A confusion matrix contains the actual and predicted classifications of a prediction model. The performance of the model is evaluated using the matrix entries (see Table B.1).

- a is the number of correct predictions that a failure occurs (True Positives);
- b is the number of incorrect predictions that a failure occurs (False Negatives);
- c is the number of incorrect predictions that no failure occurs (False Positives);
- d is the number of correct predictions that a no failure occurs (True Negatives).

The precision is the proportion of the total number of correct positive predictions

$$P = \frac{a}{a+c}$$

C. Derivation of lifetime distribution of linear random coefficient model

The linear random coefficient model is defined as:

$$X(t) = x_0 + \Theta t, \quad t \geq 0$$

where Θ denotes the unit-to-unit variation of the degradation rate, and x_0 is the initial degradation level. It is assumed that Θ follows a Weibull distribution with scale $\eta > 0$ and shape $k > 0$, and has the distribution function

$$F_\Theta(t) = 1 - e^{\left(\frac{t}{\eta}\right)^k}$$

Let $D > x_0$ denote the critical level. The part works until the process reaches D. Hence its lifetime T is then obtained by setting

$$T = \frac{D - x_0}{\Theta}, \quad \Theta > 0$$

The distribution function of T is then obtained as follows

$$\begin{aligned}
F_T(t; D, x_0) &= \Pr\left[T \leq t\right] = \Pr\left[\frac{D - x_0}{\Theta} \leq t\right] = \Pr\left[\frac{D - x_0}{t} \leq \Theta\right] \\
&= 1 - F_\Theta\left(\frac{D - x_0}{t}\right) = e^{\left(\frac{D - x_0}{\eta t}\right)^k}, \quad t > 0
\end{aligned}$$

The density function T is derived as follows

$$\begin{aligned}
\frac{\Delta F_T(t; D, x_0)}{\Delta t} &= -\frac{D - x_0}{\eta t^2} \cdot -k\left(\frac{D - x_0}{\eta t}\right)^{k-1} e^{-\left(\frac{D - x_0}{\eta t}\right)^k} \\
&= \frac{k}{t}\left(\frac{D - x_0}{\eta t}\right)^k e^{-\left(\frac{D - x_0}{\eta t}\right)^k} \\
&= \frac{\eta k}{D - x_0}\frac{D - x_0}{\eta t}\left(\frac{D - x_0}{\eta t}\right)^k e^{-\left(\frac{D - x_0}{\eta t}\right)^k} \\
&= \frac{\eta k}{D - x_0}\left(\frac{D - x_0}{\eta t}\right)^{k+1} e^{-\left(\frac{D - x_0}{\eta t}\right)^k}, \quad t > 0
\end{aligned}$$

This is better known as the Inverse Weibull distribution (see e.g. [17] and [23]).

D. Lifetime probability density and distribution plots

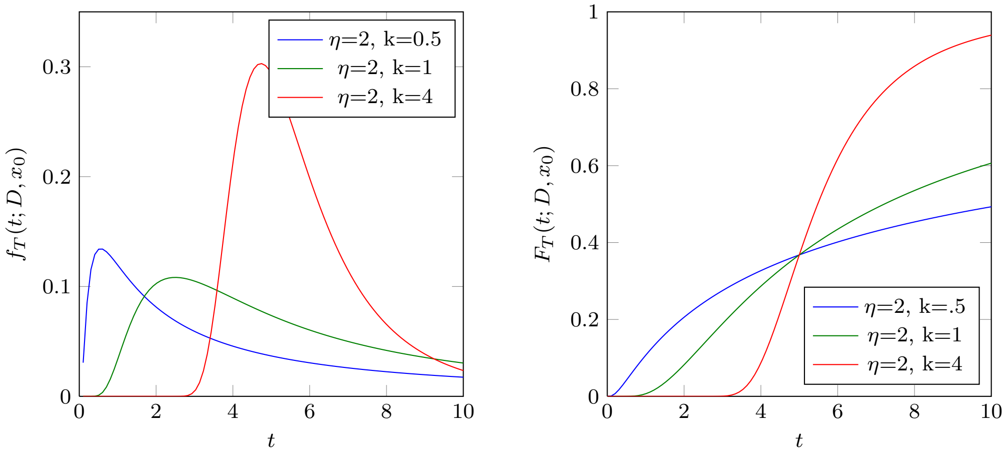


Fig. D.1: *Probability density (a) and cumulative distribution (b) functions of the lifetime T of the linear coefficient model with $D = 10$*

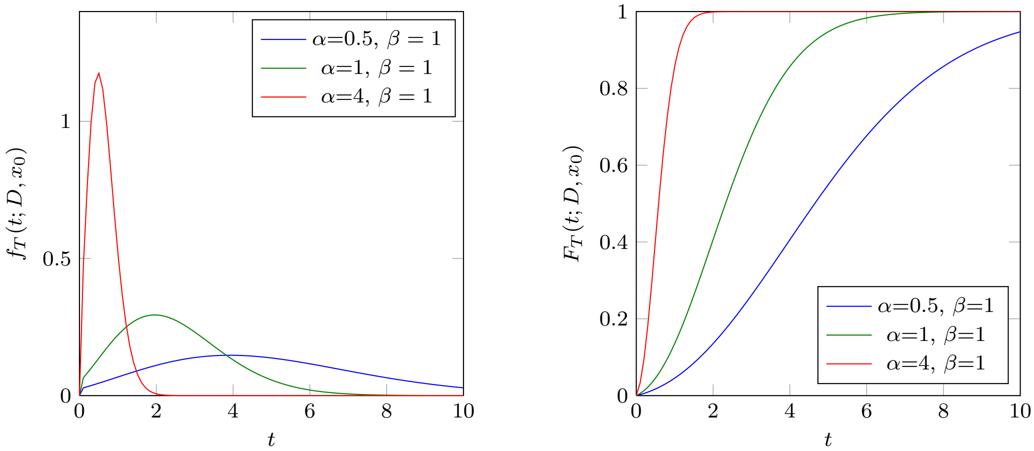


Fig. D.2: *Probability density (a) and cumulative distribution (b) functions of the lifetime T of the Gamma process with $D = 2$*

E. Mathematical derivations for the preliminary model

E.1. Probability of OPM

Let p_{opm} denote the probability that the monitored component is preventively repaired, upon a FBM opportunity. In mathematical terms, we get

$$p_{opm}(D_{opm}) = \Pr\{T_{opm}(D_{opm}) + T_1 \leq T_{pm}\}$$

which is equivalent to

$$\Pr\{T_1 \leq T_{pm} - T_{opm}(D_{opm})\}$$

where T_1 follows an exponential distribution with mean $1/\lambda$. The probability density functions of T_1 thus equals

$$f_{T_1}(t) = \lambda e^{-\lambda t}, \quad \lambda \geq 0,\, t \geq 0 \tag{E.1}$$

and pdf of $T_{pm} - T_{opm}(D_{opm})$ is written as

$$f_{T_{pm}-T_{opm}}(t; D_{opm}) = \Pr\{T_{pm} - T_{opm}(D_{opm}) = t\}, \quad t \geq 0 \tag{E.2}$$

Now, using eq. (E.1) and (E.2), we write

$$\begin{aligned} p_{opm}(D_{opm}) &= \int_{y=0}^{\infty} \int_{x=0}^{y} f_{T_1}(x) f_{T_{pm}-T_{opm}}(y; D_{opm})\, dx dy \\ &= \int_{y=0}^{\infty} \int_{x=0}^{y} \lambda e^{-\lambda x} f_{T_{pm}-T_{opm}}(y; D_{opm})\, dx dy \end{aligned}$$

After integration with respect to x then, we arrive at

$$p_{opm}(D_{opm}) = \int_{0}^{\infty} (1 - e^{-\lambda y}) f_{T_{pm}-T_{opm}}(y; D_{opm})\, dy \tag{E.3}$$

where

$$f_{T_{pm}-T_{opm}}(y; D_{opm}) = f_T(t; D_{pm}, D_{opm})$$

as given in either eq. (3.4) or (3.8), depending on the type of degradation model.

E.2. Expected cycle length

From the description of the renewal cycle, it follows that the expected cycle length is a function of D_{opm} and can be formulated as

$$\begin{aligned} ECL(D_{opm}) =& p_{opm}(D_{opm}) \cdot E[T_{opm}(D_{opm}) + T_1 | T_{opm}(D_{opm}) + T_1 < T_{pm}] \\ &+ (1 - p_{opm}(D_{opm})) \cdot E[T_{pm} + \ell | T_{opm}(D_{opm}) + T_1 > T_{pm}] \end{aligned}$$

We note that this is equivalent to

$$\begin{aligned} ECL(D_{opm}) =& E[T_{opm}(D_{opm})] + p_{opm}(D_{opm}) \cdot E[T_1 | T_1 \leq T_{pm} - T_{opm}(D_{opm})] \\ &+ (1 - p_{opm}(D_{opm})) \cdot E[T_{pm} - T_{opm}(D_{opm}) + \ell | T_1 > T_{pm} - T_{opm}(D_{opm})] \end{aligned}$$

The last two terms can be expressed as

$$\int_{y=0}^{y=\infty} \left(\int_{x=0}^{x=y} x f_{T_1}(x)\, dx + (y+\ell) \int_{x=y}^{x=\infty} f_{T_1}(x)\, dx \right) f_{T_{pm}-T_{opm}}(y; D_{opm})\, dx\, dy \tag{E.4}$$

where $f_{T_{pm}-T_{opm}}(y; D_{opm})$ is the pdf of $T_{pm} - T_{opm}(D_{opm})$. Then, using $T_1 \sim \text{Exp}(\lambda)$, we have that

$$\int_{x=0}^{x=y} x f_{T_1}(s)\, dx + (y+\ell) \int_{x=y}^{x=\infty} f_{T_1}(x)\, dx = \lambda \int_{x=0}^{x=y} x e^{-\lambda x}\, dx + (y+\ell) e^{-\lambda y}$$

Applying partial integration and rearranging of terms then yields

$$\begin{aligned} \lambda \int_{x=0}^{x=y} x e^{-\lambda x}\, dx + (y+\ell) e^{-\lambda y} &= \frac{1}{\lambda} \left(1 - (1 + y\lambda) e^{-\lambda y}\right) + (y+\ell) e^{-\lambda y} \\ &= \frac{1}{\lambda} \left(1 - e^{-\lambda y}\right) + \ell e^{-\lambda y} \end{aligned} \tag{E.5}$$

Substituting eq. (E.5) into eq. (E.4), yields

$$\int_0^{\infty} \left\{ \frac{1}{\lambda} \left(1 - e^{-\lambda y}\right) + \ell e^{-\lambda y} \right\} f_{T_H - T_{opm}}(y; D_{opm})\, dy$$

Finally, noting eq. (E.3), we can express $ECL(D_{opm})$ in terms of $p_{opm}(D_{opm})$ as

$$ECL(D_{opm}) = E[T_{opm}(D_{opm})] + \frac{1}{\lambda} p_{opm}(D_{opm}) + \ell \left(1 - p_{opm}(D_{opm})\right) \tag{E.6}$$

E.3. Degradation models

Note that $f_T(t; D, x_0)$ and $E[T; D, x_0]$ are respectively the probability density function and the mean of the degradation model lifetime and depend on the selected degradation model $X(t)$, as shown in Table E.1.

Table E.1: *Expressions for the lifetime of the monitored component for both degradation models*

Expression	Linear random coefficient model	Gamma process
$f_T(t; D, x_0)$	eq. (3.4)	eq. (3.8)
$E[T; D, x_0]$	eq. (3.5)	eq. (3.12)

F. Mathematical derivations for the extended model

F.1. Mathematical renewal analysis

The long-term expected cost rate of the monitored component is obtained by

$$g(D_{opm}) = \frac{ECC(D_{opm})}{ECL(D_{opm})} \tag{F.1}$$

First, $ECL(D_{opm})$ is formulated as

$$\begin{aligned}\sum_{n=1}^{\infty} \Bigg\{ &\int_{v=u}^{v=n\tau} \left(\int_{s=0}^{s=v-u} (u+s) f_{T_1}(s)\, ds + (v+\ell) \int_{s=v-u}^{s=\infty} f_{T_1}(s)\, ds \right) f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\, dv \\ &+ \int_{v=n\tau}^{v=\infty} \left(\int_{s=0}^{s=n\tau-u} (u+s) f_{T_1}(s)\, ds + n\tau \int_{s=n\tau-u}^{s=\infty} f_{T_1}(s)\, ds \right) f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\, dv \Bigg\} \\ &\times f_{T_{opm}}(u; D_{opm})\, du \end{aligned} \tag{F.2}$$

where $f_{T_{opm}}(\cdot)$, $f_{T_{pm}}(\cdot)$, and $f_{T_1}(\cdot)$ denote the probability density functions of $T_{opm}(D_{opm})$, T_{pm}, and T_1, respectively. Note that T_1 follows an exponential distribution with mean $1/\lambda$:

$$f_{T_1}(t) = \lambda e^{-\lambda t}, \quad t \geq 0$$

The general definitions of $f_{T_{opm}}(\cdot)$ and $f_{T_{pm}}(\cdot)$ are given as

$$f_{T_{opm}}(t; D_{opm}) = \Pr\{T_{opm}(D_{opm}) = t\}, \quad t > 0$$

and

$$f_{T_{pm}}(t) = \Pr\{T_{pm} = t\}, \quad t > 0$$

Now, we rewrite

$$\int_{s=0}^{s=v-u} (u+s) f_{T_1}(s)\, ds + (v+\ell) \int_{s=v-u}^{s=\infty} f_{T_1}(s)\, ds = \lambda \int_{s=0}^{s=v-u} (u+s) e^{-\lambda s}\, ds + (v+\ell)\lambda \int_{s=v-u}^{s=\infty} e^{-\lambda s}\, ds$$

After partial integration and rearranging terms, we arrive at

$$u\left(1 - e^{-\lambda(v-u)}\right) + \frac{1}{\lambda}\left(1 - (1 + (v-u)\lambda) e^{-\lambda(v-u)}\right) + (v+\ell) e^{-\lambda(v-u)}$$

which reduces to

$$u + \frac{1}{\lambda}\left(1 - e^{-\lambda(v-u)}\right) + \ell e^{-\lambda(v-u)} \tag{F.3}$$

Similarly, we have that

$$\int_{s=0}^{s=n\tau-u} (u+s) f_{T_1}(s)\, ds + n\tau \int_{s=n\tau-u}^{s=\infty} f_{T_1}(s)\, ds = u + \frac{1}{\lambda}\left(1 - e^{-\lambda(n\tau-u)}\right) \tag{F.4}$$

Substituting eq. (F.3) & (F.4) into eq. (F.1), yields

$$\sum_{n=1}^{\infty}\int_{u=(n-1)\tau}^{u=n\tau}\left\{\int_{v=u}^{v=n\tau}\left(u+\frac{1}{\lambda}\left(1-e^{-\lambda(v-u)}\right)+\ell e^{-\lambda(v-u)}\right)f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right.$$
$$\left.+\int_{v=n\tau}^{v=\infty}\left(u+\frac{1}{\lambda}\left(1-e^{-\lambda(n\tau-u)}\right)\right)f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right\}f_{T_{opm}}(u;D_{opm})\,du$$

which reduces to

$$\sum_{n=1}^{\infty}\int_{u=(n-1)\tau}^{u=n\tau}\left\{u+\int_{v=u}^{v=n\tau}\left(\frac{1}{\lambda}\left(1-e^{-\lambda(v-u)}\right)+\ell e^{-\lambda(v-u)}\right)f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right.$$
$$\left.+\frac{1}{\lambda}\int_{v=n\tau}^{v=\infty}\left(1-e^{-\lambda(n\tau-u)}\right)f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right\}f_{T_{opm}}(u;D_{opm})\,du$$

Analogues, for $ECC(D_{opm})$ we arrive at

$$\sum_{n=1}^{\infty}\int_{u=(n-1)\tau}^{u=n\tau}\left\{\int_{v=u}^{v=n\tau}\left(c_{opm}^{usd}(1-e^{-\lambda(v-u)})+(c_{s}^{sd}+c_{pm})e^{-\lambda(v-u)}\right)f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right.$$
$$\left.+\int_{v=n\tau}^{v=\infty}\left(c_{opm}^{usd}\left(1-e^{-\lambda(n\tau-u)}\right)+c_{opm}^{sd}e^{-\lambda(n\tau-u)}\right)f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right\}$$
$$\times f_{T_{opm}}(u;D_{opm})\,du$$

which reduces to

$$\sum_{n=1}^{\infty}\int_{u=(n-1)\tau}^{u=n\tau}\left\{\int_{v=u}^{v=n\tau}\left(c_{opm}^{usd}(1-e^{-\lambda(v-u)})+(c_{sd}+c_{pm})e^{-\lambda(v-u)}\right)f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right.$$
$$\left.+\left(c_{opm}^{usd}\left(1-e^{-\lambda(n\tau-u)}\right)+c_{opm}^{sd}e^{-\lambda(n\tau-u)}\right)\int_{v=n\tau}^{v=\infty}f_{T_{pm}}(v|T_{opm}(D_{opm})=u)\,dv\right\}f_{T_{opm}}(u;D_{opm})\,du$$

If the degradation model $X(t)$ is known, then both $f_{T_{opm}}(\cdot)$ and $f_{T_{pm}}(\cdot)$ and subsequently equations (3.32) and (3.33) can be conveniently simplified. For the linear random coefficient model and the stochastic Gamma process, this is discussed in sections F.2 and F.3, respectively.

F.2. Degradation: Linear random coefficient model

Suppose that $X(t)$ is a linear coefficient model with a Weibull distributed unit-to-unit variation of the degradation rate $\Theta\sim(\eta,k)$, and initial degradation level $x_0=0$:

$$X(t)=\Theta t,\quad t\geq 0$$

If at time u the degradation level hits the OPM control limit, i.e., $X(u)=D_{opm}$, the degradation rate is known to be

$$\Theta=\frac{D_{opm}}{u}$$

and then T_{pm} is known as

$$T_{pm}(u) = \frac{D_{pm}}{\Theta} = \frac{D_{pm}}{\frac{D_{opm}}{u}} = u\frac{D_{pm}}{D_{opm}}$$

This is graphically illustrated in Fig. F.1. Hence, the pdf of T_{pm} given that $T_{opm}(D_{opm}) = u$ may is rewritten as

$$f_{T_{pm}}(v|T_{opm}(D_{opm}) = u) = \begin{cases} 1, & \text{if } v = u\frac{D_{pm}}{D_{opm}} \\ 0, & \text{otherwise} \end{cases}$$

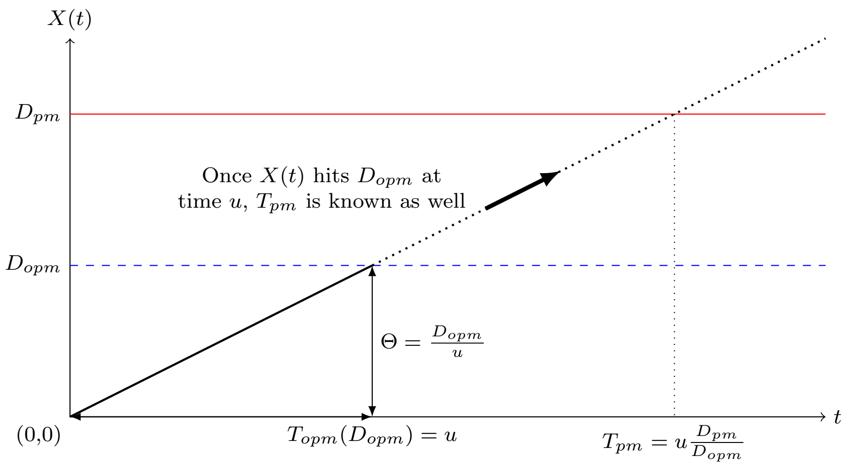


Fig. F.1: *Given that $X(t)$ is a linear random coefficient model with degradation rate Θ, T_{pm} follows from $T_{opm}(D_{opm})$*

Now, we use this result to simplify both eq. (3.32) and (3.33). First we determine the range of $u \in ((n-1)\tau, n\tau))$ for which $T_{pm}(u) \in (u, n\tau)$. We set

$$\begin{aligned} T_{pm}(u) &< n\tau \\ u\frac{D_{pm}}{D_{opm}} &< n\tau \\ u &< n\tau\frac{D_{opm}}{D_{pm}} \end{aligned}$$

In addition, because $u \in ((n-1)\tau, n\tau))$

$$\begin{aligned} u &< n\tau\frac{D_{opm}}{D_{pm}} \\ u &> (n-1)\tau \end{aligned}$$

from which follows that

$$n\tau\frac{D_{opm}}{D_{pm}} > (n-1)\tau$$

Hence, it follows that the range of u for which $T_{pm}(u) \in (u, n\tau)$, noted as $\mathbb{D}_1(n, D_{opm})$, is

$$\mathbb{D}_1(n, D_{opm}) = \left((n-1)\tau, \max\left\{ n\tau \frac{D_{opm}}{D_{pm}}, (n-1)\tau \right\} \right)$$

Similarly, it follows that the range of u for which $T_{pm}(u) \in (n\tau, \infty)$, noted as $\mathbb{D}_2(n, D_{opm})$, is

$$\mathbb{D}_2(n, D_{opm}) = \left(\max\left\{ n\tau \frac{D_{opm}}{D_{pm}}, (n-1)\tau \right\}, n\tau \right)$$

Finally, we can rewrite eq. (3.32) and (3.33) as shown below.

$$\begin{aligned} &ECC^{lin}(D_{opm}) = \\ &\quad \sum_{n=1}^{\infty} \int_{u \in \mathbb{D}_1(n, D_{opm})} \left(c_{opm}^{usd} \left(1 - e^{-\lambda u \left(\frac{D_{pm}}{D_{opm}} - 1 \right)} \right) + (c_s^{sd} + c_{pm}) e^{-\lambda u \left(\frac{D_{pm}}{D_{opm}} - 1 \right)} \right) f_T(u; D_{opm}, 0)\, du \\ &\qquad + \int_{u \in \mathbb{D}_2(n, D_{opm})} \left(c_{opm}^{usd} \left(1 - e^{-\lambda(n\tau - u)} \right) + c_{opm}^{sd} e^{-\lambda(n\tau - u)} \right) f_T(u; D_{opm}, 0)\, du \end{aligned}$$

$$\begin{aligned} &ECL^{lin}(D_{opm}) = \\ &\quad \sum_{n=1}^{\infty} \int_{u \in \mathbb{D}_1(n, D_{opm})} \left(u + \frac{1}{\lambda} \left(1 - e^{-\lambda u \left(\frac{D_{pm}}{D_{opm}} - 1 \right)} \right) + \ell e^{-\lambda u \left(\frac{D_{pm}}{D_{opm}} - 1 \right)} \right) f_T(u; D_{opm}, 0)\, du \\ &\qquad + \int_{u \in \mathbb{D}_2(n, D_{opm})} u + \frac{1}{\lambda} \left(1 - e^{-\lambda(n\tau - u)} \right) f_T(u; D_{opm}, 0)\, du \end{aligned}$$

Furthermore, given that $X(t)$ is a linear random coefficient model, we can obtain the condition-based maintenance allocation by rewriting eq. (3.34) and (3.35) to

$$p^{pm}(D_{opm}) = \sum_{n=1}^{\infty} \int_{u \in \mathbb{D}_1(n, D_{opm})} e^{-\lambda u \left(\frac{D_{pm}}{D_{opm}} - 1 \right)} f_T(u; D_{opm}, 0)\, du$$

and

$$p_{sd}^{opm}(D_{opm}) = \sum_{n=1}^{\infty} \int_{u \in \mathbb{D}_2(n, D_{opm})} e^{-\lambda(n\tau - u)} f_T(u; D_{opm}, 0)\, du$$

where $f_T(\cdot)$ and $F_T(\cdot)$ are the pdfs and cdfs of the first passage times of the linear random coefficient model. The general formulas are available in eq. (3.4) and (3.3), respectively.

F.3. Degradation: Gamma process

Suppose $X(t)$ is described by a Gamma process. Then, using the results of section 3.1 and the properties of the gamma process:

- Monotonic; and
- Independently and identically distributed $\Gamma(\alpha, \beta)$ increments;

we may rewrite

$$f_{T_{opm}}(t) = f_T(t; D_{opm}, 0),$$

and

$$\int_{v=u}^{v=n\tau} (\dots) f_{T_{pm}}(v|T_{opm}=u)\,dv = \int_{x=0}^{x=n\tau-u} (\dots) f_{T'}(x; D_{pm}, D_{opm})\,dx, \quad x = v-u$$

and

$$\int_{v=n\tau}^{v=\infty} f_{T_{pm}}(v|T_{opm}=u)\,dv = 1 - F_{T'}(n\tau-u; D_{pm}, D_{opm})$$

where $f_{T'}(\cdot)$ and $F_{T'}(\cdot)$ are the pdfs and cdfs of the first passage times of the Gamma degradation process with $x_0 = D_{opm}$. The general formulas are available in eq. (3.8) and (3.6), respectively. Substitution of these terms in eq. (3.33) and (3.32) yields

$$\begin{aligned} &ECC^{gp}(D_{opm}) = \\ &\quad \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} \left\{ \int_{x=0}^{x=n\tau-u} \left(c_{opm}^{usd}(1-e^{-\lambda x}) + (c_s^{sd}+c_{pm})e^{-\lambda x}\right) f_{T'}(x; D_{pm}, D_{opm})\,dx \right. \\ &\qquad \left. + \left(c_{opm}^{usd}\left(1-e^{-\lambda(n\tau-u)}\right) + c_{opm}^{sd} e^{-\lambda(n\tau-u)}\right)(1-F_{T'}(n\tau-u; D_{pm}, D_{opm}))\right\} f_T(u; D_{opm}, 0)\,du \end{aligned}$$

and

$$\begin{aligned} &ECL^{gp}(D_{opm}) = \\ &\quad \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} \left\{ u + \int_{x=0}^{x=n\tau-u} \left(\frac{1}{\lambda}\left(1-e^{-\lambda x}\right) + \ell e^{-\lambda x}\right) f_{T'}(x; D_{pm}, D_{opm})\,dx \right. \\ &\qquad \left. + \frac{1}{\lambda}\left(1-e^{-\lambda(n\tau-u)}\right)(1-F_{T'}(n\tau-u; D_{pm}, D_{opm}))\right\} f_T(u; D_{opm}, 0)\,du \end{aligned}$$

Furthermore, given that $X(t)$ is a Gamma process, we can obtain the condition-based maintenance allocation by rewriting eq. (3.34) and (3.35) to

$$p^{pm}(D_{opm}) = \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} \left\{ \int_{x=0}^{x=n\tau-u} e^{-\lambda x} f_{T'}(x; D_{pm}, D_{opm})\,dx \right\} f_T(u; D_{opm}, 0)\,du$$

and

$$p_{sd}^{opm}(D_{opm}) = \sum_{n=1}^{\infty} \int_{u=(n-1)\tau}^{u=n\tau} e^{-\lambda(n\tau-u)} \left(1 - F_{T'}(n\tau; D_{pm}, D_{opm})\right) f_T(u; D_{opm}, 0)\,du$$

G. Exploratory optimality analysis

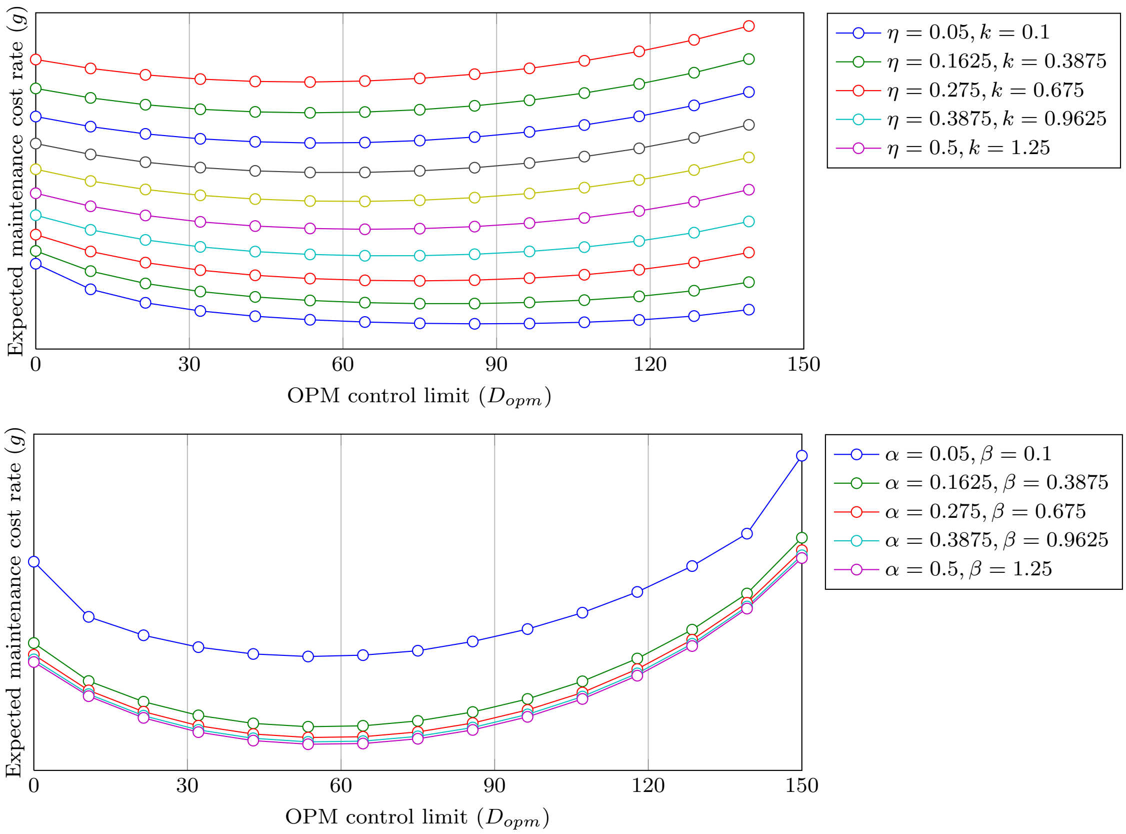
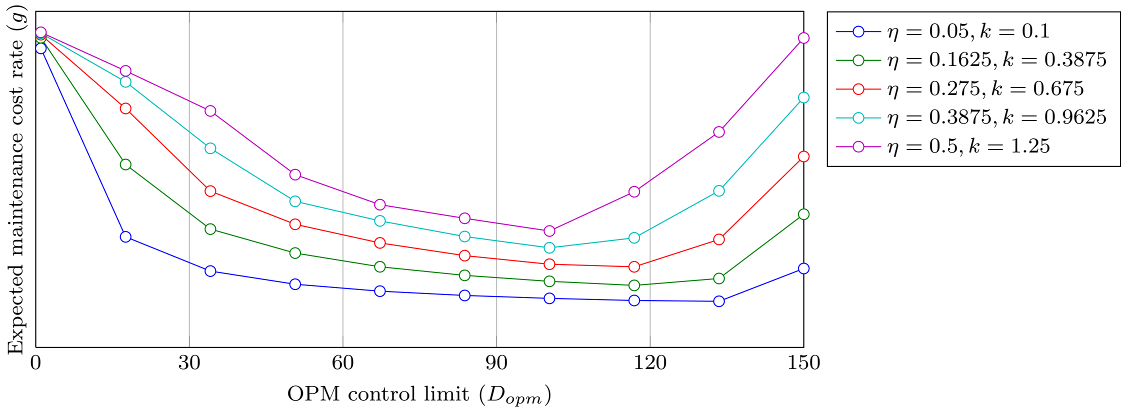


Fig. G.1: *Plot of the objective function of the preliminary model for the linear random coefficient model and the stochastic Gamma process, respectively*



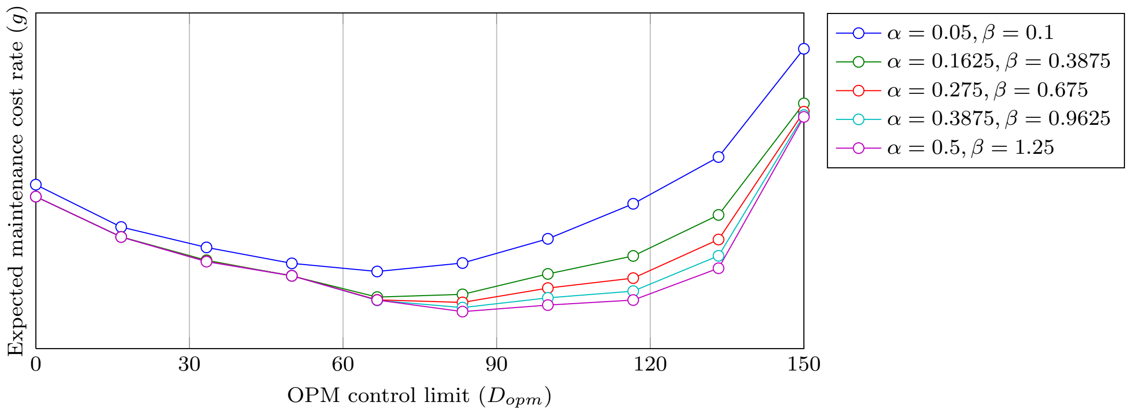


Fig. G.2: *Plot of the objective function of the extended model for the linear random coefficient model and the stochastic Gamma process, respectively*

H. Analysis of the random failure rate

For this analysis the failure data of the AT/XT lithography machines from January 2010 to September 2012 is used. The dataset contains over 91,000 failure records of 1,183 machines. Each failure corresponds with unscheduled downtime, ranging from only a couple of minutes to hundreds of hours. Because short interrupts are too short to provide joint maintenance opportunities, we restrict our analysis to long unscheduled downs, or simply *long downs* (LDs). For example, a downtime of only 10 minutes is too short to perform PM on any other system component, and would unnecessarily delay the unscheduled down process.

Additionally, a dataset is available with the "birth" date of the 1,183 AT/XT machines. The birth date of a system, is its first production date. By merging the two datasets, we obtain the time of failure in terms of the machine lifetime. That is, the date of failure on machine i minus the birth date of machine i. This way, the lifetime of failure can be obtained for all 8,627 failures. Before we proceed, please note that the dates of machine birth as well as machine failures/USD time are given in real time (instead of production time).

[redacted]

Fig. H.1: *Number of failures per day that result in LDs*

In Fig. H.1 the number of failures per day is plotted. By dividing this by 1,183 (i.e., the number of machines that account for these failures), we get an estimate for the failure rate per AT/XT machine. However, there is one major problem with this approach. Namely, the introduction of new AT/XT machines is very inconsistent between 2002 to 2012. In other words, the lifetime of the machines in our dataset is not uniformly distributed. For example, from our preliminary analysis in Fig. H.1 it appears that the failure rate is sharply increasing between $t = 1,000$ days and $t = 2,000$ days, but this likely due to the fact that the machines that account for these failures are overrepresented.

Table H.1: *Selection of machine birth dates for the analysis of each lifetime interval*

Lifetime interval		**Machine birth dates**	
From	*To*	*From*	*To*
0	365	January-10	September-11
365	730	January-09	September-10
⋮	⋮	⋮	⋮
2920	3285	January-02	September-03
3285	3650	January-01	September-02

For this reason, the failure rate for a specific lifetime must be corrected for the number of machines that contributed to the failures in that lifetime period. To do this, the lifetime is divided into several intervals during which number of machines that are responsible for these failures, is *constant*. Therefore, each lifetime interval is analyzed using a different set of failures that belong to machines within a specified range of birth dates, see Table H.1.

The corrected failure rate (*failures per machine per unit time*) is plotted in Fig. H.2. Note that the peak at $t = 1500$ days has completely disappeared. Now, in order to estimate the failure

[redacted]

Fig. H.2: *Number of failures per day per machine, that result in LDs*

rate function of a AT/XT lithography machine, we fit an appropriate function (Gaussian) to the failure rate data. The function that best describes the failure rate data, without becoming too complicated and without the loss of too much detail ($R^2_{adj} = .2613$) is depicted in Fig. H.3.

[redacted]

Fig. H.3: *Fit of the failure rate per machine, including the 95% confidence interval bounds*

I. Monotomize degradation data

The Gamma process is a monotonic function of time. This implies that the degradation increments must be positive as the Gamma distribution only supports $x \in (0, \infty)$. Therefore, we must manipulate the raw degradation pahts, so that the assumption of monoticity is justified.

Suppose we have a set $S = s_i, s_{i+1}, \ldots, s_n$ of degradation data. In order to make this degradation path monotonic (i.e., strictly increasing), the following algorithm is employed.

Algorithm 1 Algorithm monotomize S

for i=2:n **do**
  if $s_i \leq s_{i-1}$ **then**
    $s_i \leftarrow s_{i-1} + \Delta$
  end if
end for

where Δ is an infinitesimal number. It makes sure that the degradation path is strictly increasing, and that the Gamma process parameters can be estimated from the degradation path increments. An example is available in Fig. I.1.

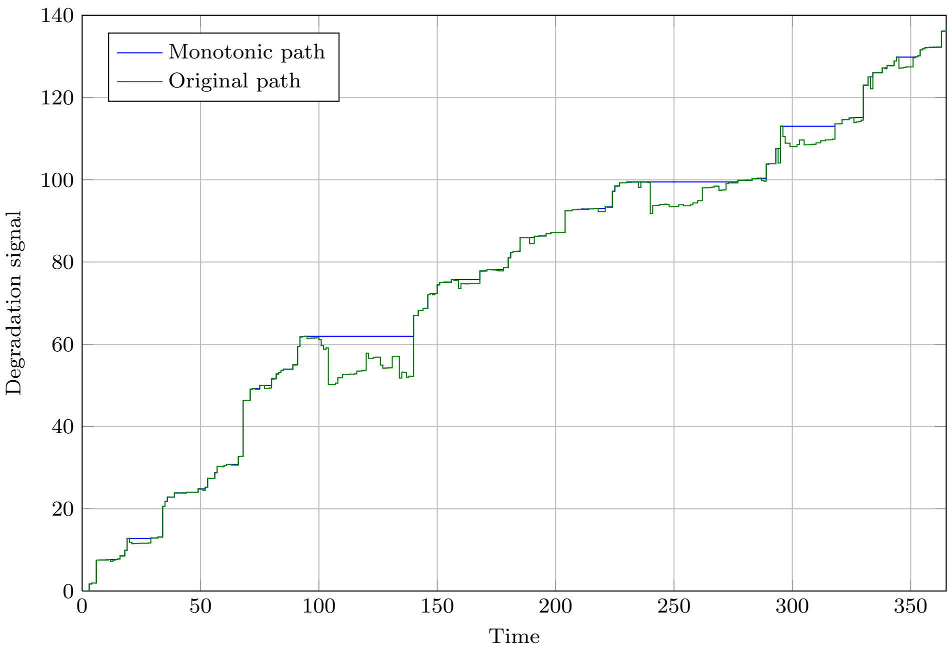
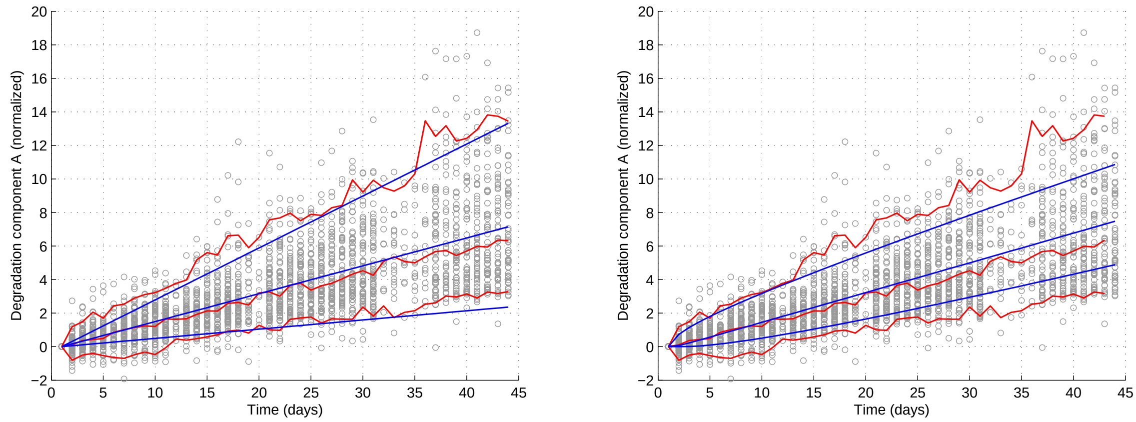


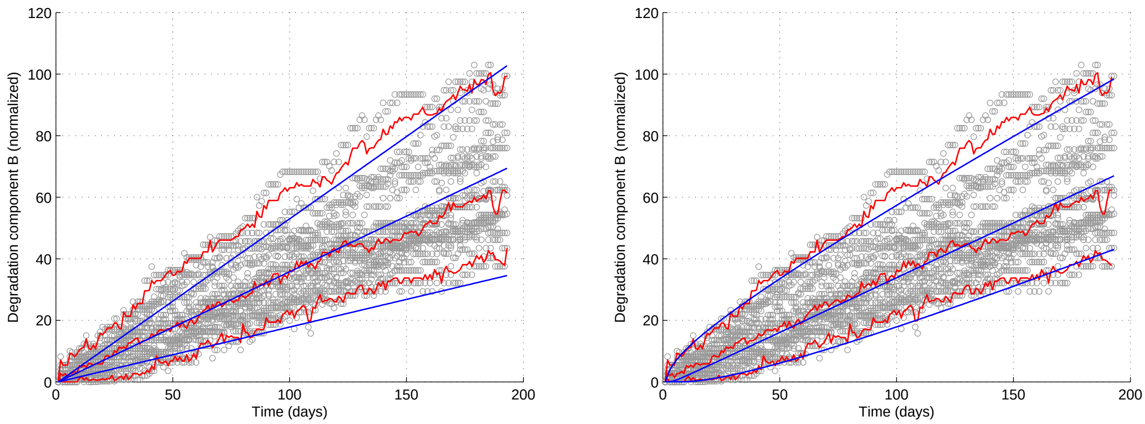
Fig. I.1: *Original degradation path and the manipulated (monotonic) path*

J. Degradation model estimation and selection



(a) *Linear random coefficient model* **(b)** *Gamma process*

Fig. J.1: *95th, 50th and 5th percentile plots of both the fitted degradation model and the degradation paths of component A*



(a) *Linear random coefficient model* **(b)** *Gamma process*

Fig. J.2: *95th, 50th and 5th percentile plots of both the fitted degradation model and the degradation paths of component B*

Table J.1: *Estimated degradation rates of the degradation paths of component A*

Path i	Degradation rate $\hat{\Theta}_i$	R_i^2	Path i	Degradation rate $\hat{\Theta}_i$	R_i^2
1	0.1929	0.8741	37	0.2514	0.9511
2	0.1954	0.9589	38	0.1044	0.8283
3	0.1330	0.8730	39	0.1878	0.9304
4	0.2027	0.9837	40	0.2979	0.9885
5	0.2440	0.9639	41	0.1109	0.5916
6	0.2341	0.7504	42	0.0823	0.7512
7	0.1271	0.9236	43	0.3132	0.9824
8	0.1003	0.8452	44	0.2264	0.9843
9	0.2112	0.6806	45	0.0838	0.9418
10	0.1086	0.9343	46	0.1256	0.5189
11	0.0613	0.4871	47	0.1388	0.7561
12	0.1588	0.9824	48	0.2388	0.9569
13	0.1172	0.7286	49	0.0699	0.7708
14	0.0974	0.8662	50	0.1145	0.9214
15	0.1509	0.9411	51	0.2243	0.9666
16	0.1741	0.8479	52	0.0977	0.9324
17	0.2295	0.7963	53	0.0975	0.7067
18	0.2132	0.9626	54	0.2197	0.8415
19	0.1133	0.8742	55	0.1452	0.9333
20	0.1144	0.8610	56	0.1784	0.9208
21	0.1134	0.5778	57	0.2119	0.8885
22	0.1492	0.8691	58	0.2427	0.9867
23	0.1940	0.9727	59	0.2502	0.9748
24	0.1376	0.9697	60	0.0775	0.4710
25	0.1385	0.9503	61	0.3845	0.8288
26	0.1187	0.9530	62	0.3423	0.8586
27	0.1088	0.9869	63	0.1494	0.9653
28	0.0820	0.6466	64	0.1431	0.9601
29	0.1445	0.9646	65	0.0817	0.9414
30	0.0919	0.7603	66	0.1378	0.5280
31	0.1131	0.9578	67	0.2834	0.9541
32	0.0840	0.9800	68	0.1793	0.9300
33	0.2413	0.8723	69	0.1410	0.9102
34	0.3991	0.9094	70	0.3246	0.8483
35	0.1371	0.8080	71	0.3508	0.9923
36	0.1909	0.8208	72	0.1727	0.9601
			73	0.1133	0.8534

Table J.2: *Estimated degradation rates of the degradation paths of component B*

Path i	Degradation rate $\hat{\Theta}_i$	R_i^2
1	0.23568	0.9790
2	0.29567	0.9828
3	0.60928	0.9716
4	0.33445	0.8742
5	0.52106	0.9593
6	0.48351	0.9827
7	0.35387	0.7668
8	0.33788	0.8485
9	0.45505	0.8923
10	0.30039	0.9467
11	0.35462	0.9592
12	0.27550	0.9485
13	0.35705	0.9561
14	0.42095	0.9859
15	0.21359	0.9227
16	0.27509	0.9557
17	0.36070	0.9644
18	0.30973	0.9378

K. ASML maintenance cost and downtime parameters

In this chapter the ASML specific maintenance cost parameters are determined and described, both in terms of cost and downtime. These parameters can be used to determine the generic maintenance cost that are required for the evaluation models in chapter 3.

The total spare part consumption of component A during February 2011 to February 2012 was [redacted] EURO and 93 parts. Hence, the expected material cost is [redacted] EURO for component A. The material cost of component A are the same for corrective and preventive maintenance ($c_m^{cm} = c_m^{pm}$). For component B the total part cost were [redacted] EURO during May 2011 to May 2012. In total 36 parts were consumed, so the average material cost are [redacted] EURO for component B. Interestingly, a corrective repair of component B means a complete swap, whereas a preventive repair consists of a simple recalibration and thus no material cost. Likewise, the downtime related to a corrective repair of component B is much higher, than for a preventive "repair". Furthermore, the hourly labor cost during an unscheduled down (c_l^{usd}) is [redacted] EURO per hour (for a scheduled down the hourly labor cost, c_l^{sd}, is only [redacted] EURO per hour). The cost for an emergency shipment via plane (c_e) is [redacted] EURO. The penalty cost of downtime (c_p) are assumed to be [redacted] EURO per hour. Recall that this is an *artificial* cost, since it is not visible on the ASML balance sheet. However, downtime definitely hurts ASML indirectly. For example, downtime reduces the system availability which must be higher than 95%, otherwise ASML violates the customer service contract. Furthermore, system downtime reduces customer satisfaction, which could ultimately lead to a loss of market share, and future sales. In the sensitivity analysis, the effect of this parameter will be analyzed. An overview of the ASML maintenance cost parameters and the related downtime parameters is shown in the table below.

Table K.1: *ASML maintenance cost and downtime parameters*

Parameter	Description	Component A	Component B	
c_m^{pm}	Average material cost (parts & tools) that is consumed during a (opportunistic) preventive repair	[redacted]	[redacted]	EURO
c_m^{cm}	Average material cost (parts & tools) that is consumed during a corrective repair	[redacted]	[redacted]	EURO
c_e	Average emergency shipment cost of a spare part	[redacted]	[redacted]	EURO
c_p	Downtime penalty cost per unit of downtime	[redacted]	[redacted]	EURO/hour
c_l^{usd}	Labor cost per unit of unscheduled downtime	[redacted]	[redacted]	EURO/hour
c_l^{sd}	Labor cost per unit of scheduled downtime	[redacted]	[redacted]	EURO/hour
d_s^{usd}	Amount of downtime that is related to the setup of a unscheduled down	26.2	27.2	hour
d_s^{sd}	Amount of downtime composition that is related to the setup of a scheduled down	2.0	2.2	hour
d_{cm}	Amount of downtime that is related to the corrective maintenance on the monitored component	11.3	24.1	hour
d_{pm}	Amount of downtime that is related to the preventive maintenance on the monitored component	11.3	12.9	hour

L. Case study results

Table L.1: *Performance of all (pro-active) maintenance policies for both cases*

Case	Policy	ETMC[a]	Lower	Upper	EDT[b]	Lower	Upper	IoA[c] (%)
A	Current	[redacted]	[redacted]	[redacted]	17.4	13.3	21.6	0.2
	CBM	[redacted]	[redacted]	[redacted]	7.0	5.4	8.7	0.1
	Sync. CBM and FBM	[redacted]	[redacted]	[redacted]	1.9	1.4	2.6	0.0
	Sync. CBM, FBM, and PerM	[redacted]	[redacted]	[redacted]	0.2	0.1	0.4	0.0
B	Current (FBM)	[redacted]	[redacted]	[redacted]	38.1	29.2	49.6	0.4
	CBM	[redacted]	[redacted]	[redacted]	15.7	12.0	20.3	0.2
	Sync. CBM and FBM	[redacted]	[redacted]	[redacted]	4.4	2.7	7.6	0.1
	Sync. CBM, FBM, and PerM	[redacted]	[redacted]	[redacted]	0.3	0.2	0.6	0.0

[a]: Expected Total Maintenance Cost in EURO/year/machine

[b]: Expected Downtime in hour/year/machine

[c]: Impact on Availability

Table L.2: *Sensitivity analysis failure rate*

Failure rate[a]	**Case: Policy**	**Optimal control limit** (%)	**ETMC**[b]	**EDT**[b]	CBM allocation: % of PM	CBM allocation: % of OPM (USD)	CBM allocation: % of OPM (SD)
Low	A: Preliminary	54.7	60.7	33.6	97.8	2.2	—
High	A: Preliminary	81.6	36.3	1.1	65.2	34.8	—
Low	A: Extended	70.4	35.4	0.3	0.6	0.4	99
High	A: Extended	82.2	35.6	0.9	36.3	24.1	39.6
Low	B: Preliminary	22.7	12.5	35.2	98	2	—
High	B: Preliminary	59.3	4.8	1.2	58.8	41.2	—
Low	B: Extended	41.2	2.2	0.6	1	0.4	98.6
High	B: Extended	59.7	4.4	0.9	24	26.8	49.1

[a]: Low is $\frac{1}{10}$th of the default failure rate; high is 10 times the default failure rate

[b]: % of Current

M. Implementation

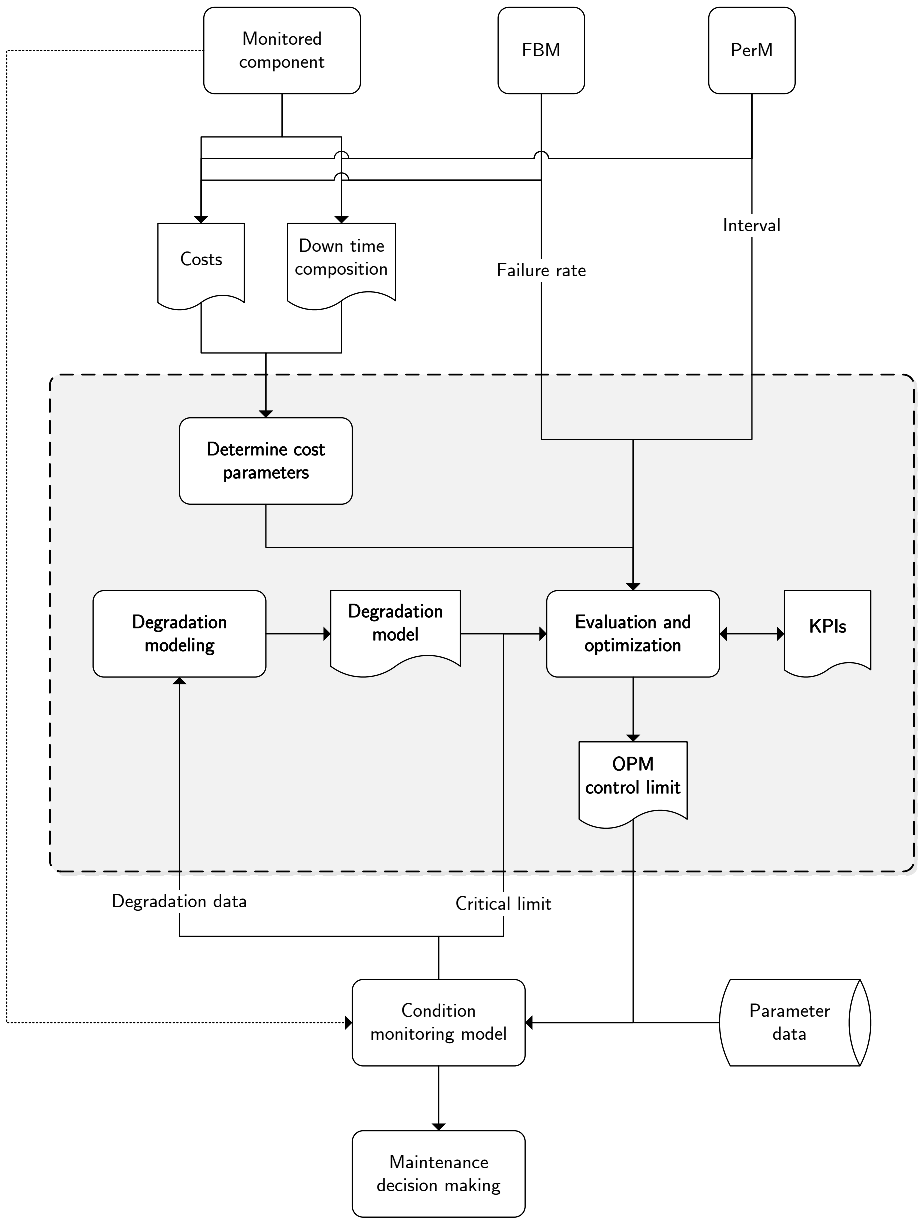


Fig. M.1: *Implementation of the new policy in the current maintenance operations*

N. MATLAB implementation

N.1. Script for degradation modeling

Listing N.1: *Estimate linear random coefficient model parameters and generate percentile plot*

```
% Matlab code to estimate parameters of the
% Weibull unit-to-unit variation of the degradation rate
% in the linear random coefficient model
%
% Input data format
%
% Time       Unit1    Unit2    ...     Unit n
% 1          .1       .2       ...     .11
% 2          .19      .25      ...     .19
% ...        ...      ...      ...     ...
% n          10.12    10.22    ...     15.05

% Input settings
file = 'data.xlsx'; % Excel file path
sheet = 'Abs_path';% Sheet name
range = 'A3:S195'; % Range (rows = time points, columns = units)
deltat = 1; % Time between degradation points

data = xlsread(file,sheet,range); % Get data
[n m] = size(data);
times = data(:,1);
data = data(:,2:m);

slopes = zeros(m-1,2);
for i=1:m-1
    [b,bint,r,rint,stats] = regress(data(:,i),times); % Obtain slopes through Least ↩
        Squares
    slopes(i,1) = b;
    slopes(i,2) = stats(1);
end

[wbl_p,wbl_ci] = wblfit(slopes(:,1)); % ML estimate of eta and k

% Print fit results
fprintf('\n\n** OLS estimates for degradation rate\n\n');
fprintf('Unit\t Degradation rate\t R2\n');
for i=1:m-1
        fprintf('%i\t\t %g\t\t\t %g\n',[i slopes(i,1) slopes(i,2)]);
end
fprintf('\n\n** ML estimates of Weibull distribution representing\n** the unit-to-unit↩
     variation of degradation rate\n\n');
fprintf('Parameter\t ML est. \t CI\n');
fprintf('eta\t\t\t %g \t %g\t %g\n',[wbl_p(1) wbl_ci(1,1) wbl_ci(2,1)]);
fprintf('k\t\t\t %g \t %g\t %g\n',[wbl_p(2) wbl_ci(1,2) wbl_ci(2,2)]);

% Percentiles of data
pt = zeros(n,3);
for i=2:n
  pt(i,:) = prctile(data(i,:),[5 50 95]);
end

% Plot
figure
hold on;
plot(data,'o','MarkerSize',4,'Color',[.6,.6,.6]);
```

```
p1 = plot(pt,'r','LineWidth',1);
p2 = plot(times*wblinv(.05,wbl_p(1),wbl_p(2)),'-','LineWidth',1); % Percentiles model
p3 = plot(times*wblinv(.50,wbl_p(1),wbl_p(2)),'-','LineWidth',1); % Percentiles model
p4 = plot(times*wblinv(.95,wbl_p(1),wbl_p(2)),'-','LineWidth',1); % Percentiles model
hold off;
grid on;
xlabel('Time (days)');
ylabel('Degradation component B (normalized)');
```

Listing N.2: *Estimate Gamma process parameters and generate percentile plot*

```
% Matlab code to estimate parameters of the Gamma process
%
% Input data format
%
% Increment  Unit1   Unit2   ...    Unit n
% 1          .1      .2             .11
% 2          .09     .25     ...    .19
% ...        ...     ...     ...    ...
% n          .12     .22     ...    .05
%
% Time       Unit1   Unit2   ...    Unit n
% 1          .1      .2      ...    .11
% 2          .19     .25     ...    .19
% ...        ...     ...     ...    ...
% n          10.12   10.22   ...    15.05

% Change settings
file = 'data.xlsx'; % Excel file path
sheet_inc = 'Abs_inc'; % Sheet name increments
sheet_path = 'Abs_path'; % Sheet name deg paths
range = 'A3:S195'; % Range (rows = time points, columns = units)
deltat = 1; % Time between degradation points
critical_limit = 41; % Critical degradation limit

data = xlsread(file,sheet_inc,range); % Get data
[n m] = size(data);
data = data(:,2:m);

vdata = data( : ); % Make column vector
vdata = vdata(~isnan(vdata)); % Remove NaN
cdata = cumsum(vdata); % Cumulative

% Monotomize data
mondata = [];
for i = 2:numel(cdata)
    if cdata(i-1) > cdata(i)
        cdata(i) = cdata(i-1);
    elseif cdata(i) > cdata(i-1)
        mondata = [mondata cdata(i)-cdata(i-1)];
    end
end
deltat_cor = deltat*(length(vdata)./length(mondata)); % Correct for non-positive ↩
    increments

[gp_p,gp_ci] = gamfit(mondata,.05); % ML estimates of alpha and beta
gp_p(1) = gp_p(1)/deltat_cor; % Correct for deltat
gp_ci(:,1) = gp_ci(:,1)./deltat_cor; %Correct for deltat

% Print results
fprintf('\n\n** ML estimates of the Gamma process\n\n');
fprintf('Parameter\t ML est. \t CI\n');
fprintf('alpha\t\t %g \t %g\t %g\n',[gp_p(1) gp_ci(1,1) gp_ci(2,1)]);
fprintf('beta\t\t %g \t %g\t %g\n',[1/gp_p(2) 1/gp_ci(2,2) 1/gp_ci(1,2)]);
```

```
path_data = xlsread(file,sheet_path,range); % Get paths
times = path_data(:,1);
path_data = path_data(:,2:m);

% Percentiles of data
pt = zeros(n,3);
for i=2:n
  pt(i,:) = prctile(path_data(i,:),[5 50 95]);
end

% Plot
figure
hold on;
plot(path_data,'o','MarkerSize',4,'Color',[.6,.6,.6]);
plot(pt,'r','LineWidth',1);
plot(gaminv(0.05,gp_p(1)*times,gp_p(2)),'-','LineWidth',1); %Percentiles model
plot(gaminv(0.50,gp_p(1)*times,gp_p(2)),'-','LineWidth',1); %Percentiles model
plot(gaminv(0.95,gp_p(1)*times,gp_p(2)),'-','LineWidth',1); %Percentiles model
hold off;
grid on;
xlabel('Time (days)');
ylabel('Degradation component B (normalized)');
```

N.2. Script for running case study

Listing N.3: *Code for preparing and running the evaluation models for the ASML case study*

```
% Clear
clear

% Parameters
csusdA   = ###;
cssdA    = ###;
cpmA     = ###;
ccmA     = ###;
copmA    = ###;
copmsdA  = ###;

dsusdA   = 26.2;
dssdA    = 2;
dcmA     = 11.3;
dpmA     = 11.3;
dopmA    = 0;
dopmsdA  = 0;

csusdB   = ###;
cssdB    = ###;
ccmB     = ###;
cpmB     = ###;
copmB    = ###;
copmsdB   = ###;

dssdB    = 2.2;
dsusdB   = 27.2;
dpmB     = 12.9;
dcmB     = 24.1;
dopmB    = 0;
dopmsdB   = 0;

D        = 100;
DpmA     = 88; %97;
DpmB     = 71; %75;
```

```
eta      = [.1398 .1591 .1819];
k        = [2.668 3.732 5.221];
alpha    = [.07599 .08463 .09426];
beta     = [.3512 .4114 .4820];
eta      = eta(2);
k        = k(2);
alpha    = alpha(2);
beta     = beta(2);

lambda   = 8.8550e-3;
tau      = 91;
LA        = 2;
LB        = 1;

% Init. model classes
Ap       = m1lin2(lambda,DpmA,cssdA,cpmA,copmA,eta,k,LA);
Apd      = m1lin2(lambda,DpmA,dssdA,dpmA,dopmA,eta,k,LA);
Bp       = m1gp2(lambda,DpmB,cssdB,cpmB,copmB,alpha,beta,LB);
Bpd      = m1gp2(lambda,DpmB,dssdB,dpmB,dopmB,alpha,beta,LB);
Ae       = m2lin2(lambda,DpmA,cssdA,cpmA,copmA,copmsdA,eta,k,tau,LA);
Aed      = m2lin2(lambda,DpmA,dssdA,dpmA,dopmA,dopmsdA,eta,k,tau,LA);
Be_hat   = m2gpe2(lambda,DpmB,cssdB,cpmB,copmB,copmsdB,alpha,beta,tau,LB);
Be       = m2gp2(lambda,DpmB,cssdB,cpmB,copmB,copmsdB,alpha,beta,tau,LB);
Bed      = m2gp2(lambda,DpmB,dssdB,dpmB,dopmB,dopmsdB,alpha,beta,tau,LB);

% Calculate ETMC, EDT, and CBMA
% FBM
gfbmA = 365*(csusdA+ccmA)/Ap.exFPT(D);
dfbmA = 365*(dsusdA+dcmA)/Ap.exFPT(D);
gfbmB = 365*(csusdB+ccmB)/Bp.exFPT(D);
dfbmB = 365*(dsusdB+dcmB)/Bp.exFPT(D);
% CBM
gcbmA = 365*(cssdA+cpmA)/(Ap.exFPT(DpmA)+LA);
dcbmA = 365*(dssdA+dpmA)/(Ap.exFPT(DpmA)+LA);
gcbmB = 365*(cssdB+cpmB)/(Bp.exFPT(DpmB)+LB);
dcbmB = 365*(dssdB+dpmB)/(Bp.exFPT(DpmB)+LB);
% Sync. CBM
[DopmAp,gsyncAp]=fminbnd(@(x) 365.*Ap.g(x),0,DpmA,optimset('TolX',1e-2,'MaxFunEvals'↩
    ,100));
[DopmAe,gsyncAe]=fminbnd(@(x) 365.*Ae.g(x),0,DpmA,optimset('TolX',1e-2,'MaxFunEvals'↩
    ,100));
[DopmBp,gsyncBp]=fminbnd(@(x) 365.*Bp.g(x),0,DpmB,optimset('TolX',1e-2,'MaxFunEvals'↩
    ,100));
[DopmBe,gsyncBe]=fminbnd(@(x) 365.*Be_hat.g(x),0,DpmB,optimset('TolX',1e-2,'↩
    MaxFunEvals',100));
dsyncAp = 365*Apd.g(DopmAp);
dsyncAe = 365*Aed.g(DopmAe);
dsyncBp = 365*Bpd.g(DopmBp);
dsyncBe = 365*Bed.g(DopmBe);

% Output results
 xlswrite('results.xlsx',[gfbmA dfbmA ; gcbmA dcbmA ; gsyncAp dsyncAp ; gsyncAe ↩
     dsyncAe ; ...
     gfbmB dfbmB ; gcbmB dcbmB ; gsyncBp dsyncBp ; gsyncBe dsyncBe],...
     'A')
```

N.3. Classes for evaluation and optimization

Listing N.4: *Preliminary model given linear degradation*

```
classdef m1lin2
    % Class for evaluation of Model 1
```

```
% Degradation is described as a linear random coefficient model
% with a Weibull(eta,k) distributed degradation rate
%
%   /--MODEL PARAMETERS--/
%
%   cssd    : Set-up cost of SD
%   cpm     : Preventive maintenance cost
%   copm    : Opportunistic maintenance cost
%   Dpm     : Critical PM limit
%   D       : Critical failure limit
%   lambda  : Failure rate of FBM process
%   alpha   : Shape parameter of Gamma degradation process
%   beta    : Inverse scale parameter of Gamma degradation process
%   L       : PM lead time
%
properties
    % Default parameters
    lambda  = 0;
    cssd    = 0;
    cpm     = 0;
    copm    = 0;
    Dpm     = 0;
    eta     = 0;
    k       = 0;
    L       = 0;
end
methods
    function obj = m1lin2(lambda,Dpm,cssd,cpm,copm,eta,k,L)
        if nargin > 0
            obj.cssd    = cssd;
            obj.cpm     = cpm;
            obj.copm    = copm;
            obj.Dpm     = Dpm;
            obj.lambda  = lambda;
            obj.eta     = eta;
            obj.k       = k;
            obj.L       = L;
        end
    end
    function y = pdfFPT(m1,x,C)
        % Lifetime density of linear RCM
        y = ((m1.eta.*m1.k)./C).*(((C)./(m1.eta.*x)).^(m1.k+1)).*exp(-((C)./(m1.↩
            eta.*x)).^m1.k);
    end
    function y = cdfFPT(m1,x,C)
        % Lifetime distribution of linear RCM
        y = exp(-1.*(C./(m1.eta.*x)).^m1.k);
    end
    function y = exFPT(m1,C)
        % First moment of lifetime of linear RCM
        y = (C./m1.eta).*gamma(1-1./m1.k);
    end
    function y = varFPT(m1,C)
        % Variance of lifetime of linear RCM
        y = (C./m1.eta).^2.*gamma(1-2./m1.k)-(C./m1.eta).*gamma(1-1./m1.k);
    end
    function y = popm(m1,Dopm)
        % Probability of OPM
        y = zeros(1,length(Dopm));
        for i=1:length(Dopm)
            if Dopm(i) < m1.Dpm
                f = @(x) (1-exp(-1.*m1.lambda.*x)).*m1.pdfFPT(x,m1.Dpm-Dopm(i));
                y(i) = quadgk(f,0,Inf);
            end
        end
    end
```

```
        function y = g(m1,Dopm)
            % Long-term expected system cost rate
            y = zeros(1,length(Dopm));
            for i = 1:length(Dopm)
                p = m1.popm(Dopm(i));
                y(i) = (p.*m1.copm+(1-p).*(m1.cssd+m1.cpm))/(m1.exFPT(Dopm(i))+(1./m1.↩
                    lambda).*p+(1-p).*m1.L);
            end
        end
    end
end
```

Listing N.5: *Preliminary model given stochastic degradation*

```
classdef m1gp2
    % Class for evaluation of Model 1
    % Degradation is described as a stochastic Gamma degradation process
    %
    %   /--MODEL PARAMETERS--/
    %
    %   cssd    : Set-up cost of SD
    %   cpm     : Preventive maintenance cost
    %   copm    : Opportunistic maintenance cost
    %   Dpm     : Critical PM limit
    %   D       : Critical failure limit
    %   lambda  : Failure rate of FBM process
    %   alpha   : Shape parameter of Gamma degradation process
    %   beta    : Inverse scale parameter of Gamma degradation process
    %   L       : PM lead time
    %
    properties
        % Default parameters
        lambda  = 0;
        cssd    = 0;
        cpm     = 0;
        copm    = 0;
        Dpm     = 0;
        alpha   = 0;
        beta    = 0;
        L       = 0;
    end
    methods
        function obj = m1gp2(lambda,Dpm,cssd,cpm,copm,alpha,beta,L)
            if nargin > 0
                obj.cssd    = cssd;
                obj.cpm     = cpm;
                obj.copm    = copm;
                obj.Dpm     = Dpm;
                obj.lambda  = lambda;
                obj.alpha   = alpha;
                obj.beta    = beta;
                obj.L       = L;
            end
        end
        function y = pdfFPT(m1,x,C)
            % First Passage Time pdf of degradation process
            delta = 1e-8;
            y = zeros(size(x));
            for i=1:length(y)
                if C>0 && x(i)>delta
                    y_2 = m1.cdfFPT(x(i)+delta,C);
                    y_1 = m1.cdfFPT(x(i)-delta,C);
                    y(i) = (y_2-y_1)/(2*delta);
                elseif C==0 && x(i)<=delta
```

```
                    y(i) = 1;
                end
            end
        end
        function y = cdfFPT(m1,x,C)
            % First Passage Tim cdf of degradation process
            y = 1-gammainc(C*m1.beta,m1.alpha*x);
        end
        function y = exFPT(m1,C)
            % Mean First Passage Time of degradation process
            int = @(x) (1-m1.cdfFPT(x,C));
            y = quadgk(int,0,inf,'MaxIntervalCount',100,'AbsTol',1.e-2);
        end
        function y = varFPT(m1,C)
            % Variance FPT of degradation process
            int = @(x) x.*(1-m1.cdfFPT(x,C));
            y = 2*quadgk(int,0,inf,'MaxIntervalCount',100,'AbsTol',1.e-2) - m1.exFPT(C↩
                )^2;

        end
        function y = popm(m1,Dopm)
            % Probability of OPM
            y = zeros(1,length(Dopm));
            for i=1:length(Dopm)
                if Dopm(i) < m1.Dpm
                    f = @(x) (1-exp(-1.*m1.lambda.*x)).*m1.pdfFPT(x,m1.Dpm-Dopm(i));
                    y(i) = quadgk(f,0,Inf,'MaxIntervalCount',100,'AbsTol',1.e-2);
                end
            end
        end
        function y = g(m1,Dopm)
            % Long-term expected system cost rate
            y = zeros(1,length(Dopm));
            for i=1:length(Dopm)
                p = m1.popm(Dopm(i));
                y(i) = (p.*m1.copm+(1-p).*(m1.cssd+m1.cpm))/(m1.exFPT(Dopm(i))+(1./m1.↩
                    lambda).*p+(1-p).*m1.L);
            end
        end
    end
end
```

Listing N.6: *Extended model given linear degradation*

```
classdef m2lin2
    % Class for evaluation of Model 2
    % Degradation is described as a stochastic Gamma degradation process
    %
    %   /--MODEL PARAMETERS--/
    %
    %   cssd    : Set-up cost SD
    %   cpm     : Preventive maintenance cost
    %   copmsd  : Opportunistic maintenance cost during USD
    %   copmusd : Opportunistic maintenance cost during SD
    %   cperm   : Expected periodic maintenance cost
    %   Dpm     : Critical PM threshold
    %   lambda  : Failure rate of FBM process
    %   tau     : Periodic maintenance interval
    %   eta     : Parameter of Weibull degradaiton rate
    %   k       : Parameter of Weibull degradation rate
    %   L       : PM lead time
    %
    properties
        % Default parameters
```

```
        lambda  = 0;
        cssd    = 0;
        cpm     = 0;
        copmusd = 0;
        copmsd  = 0;
        Dpm     = 0;
        eta     = 0;
        k       = 0;
        tau     = 0;
        L       = 0;

        tol_pr  = 1e-10;
    end
    methods
        function obj = m2lin2(lambda,Dpm,cssd,cpm,copmusd,copmsd,eta,k,tau,L)
            if nargin > 0
                obj.cssd    = cssd;
                obj.cpm     = cpm;
                obj.copmusd = copmusd;
                obj.copmsd  = copmsd;
                obj.Dpm     = Dpm;
                obj.lambda  = lambda;
                obj.eta     = eta;
                obj.k       = k;
                obj.tau     = tau;
                obj.L       = L;
            end
        end
        function y = pdfFPT(m2,x,C)
            % Lifetime density of linear RCM
            y = ((m2.eta.*m2.k)./C).*(((C)./(m2.eta.*x)).^(m2.k+1)).*exp(-((C)./(m2.↩
                eta.*x)).^m2.k);
        end
        function y = cdfFPT(m2,x,C)
            % Lifetime distribution of linear RCM
            y = exp(-1.*(C./(m2.eta.*x)).^m2.k);
        end
        function y = exFPT(m2,C)
            % First moment of lifetime of linear RCM
            y = (C./m2.eta).*gamma(1-1./m2.k);
        end
        function y = varFPT(m2,C)
            % Variance of lifetime of linear RCM
            y = (C./m2.eta).^2.*gamma(1-2./m2.k)-(C./m2.eta).*gamma(1-1./m2.k);
        end
        function y = ECC_n(m2,n,Dopm)
            % ECC during ((n-1)\tau,n\tau)
            f_1 = @(u)(m2.copmusd.*(1-exp(-1.*m2.lambda.*u.*(m2.Dpm./Dopm-1)))...
                        +(m2.cssd+m2.cpm).*exp(-1.*m2.lambda.*u.*(m2.Dpm./Dopm-1)))...
                        .*m2.pdfFPT(u,Dopm);
            f_2 = @(u) ((m2.copmusd.*(1-exp(-1.*m2.lambda.*(n.*m2.tau-u)))...
                        +m2.copmsd.*exp(-1.*m2.lambda.*(n.*m2.tau-u)))...
                        .*m2.pdfFPT(u,Dopm));
            a = (n-1)*m2.tau;
            b = max(n*m2.tau*(Dopm/m2.Dpm),(n-1)*m2.tau);
            c = n*m2.tau;
            y = quad(f_1,a,b)+quad(f_2,b,c);
        end
        function y = ECL_n(m2,n,Dopm)
            % ECL during ((n-1)\tau,n\tau)
            f1 = @(u)(u+(1./m2.lambda).*(1-exp(-1.*m2.lambda.*u.*(m2.Dpm./Dopm-1)))...
                +m2.L.*exp(-1.*m2.lambda.*u.*(m2.Dpm./Dopm-1))).*m2.pdfFPT(u,Dopm);
            f2 = @(u)(u+(1./m2.lambda).*(1-exp(-1.*m2.lambda.*(n.*m2.tau-u))))...
                        .*m2.pdfFPT(u,Dopm);
            a = (n-1)*m2.tau;
            b = max(n*m2.tau*(Dopm/m2.Dpm),(n-1)*m2.tau);
```

```
            c = n*m2.tau;
            y = quad(f1,a,b) + quad(f2,b,c);
        end
        function [p_pm,p_sd_opm] = p_n(m2,n,Dopm)
            % Probabilities during ((n-1)\tau,n\tau)
            f_pm = @(u)exp(-1.*m2.lambda.*u.*(m2.Dpm./Dopm-1))...
                        .*m2.pdfFPT(u,Dopm);
            f_sd_opm = @(u)exp(-1.*m2.lambda.*(n.*m2.tau-u))...
                        .*m2.pdfFPT(u,Dopm);
            a = (n-1)*m2.tau;
            b = max(n*m2.tau*(Dopm/m2.Dpm),(n-1)*m2.tau);
            c = n*m2.tau;
            p_pm = quad(f_pm,a,b);
            p_sd_opm = quad(f_sd_opm,b,c);
        end
        function y = ECC(m2,Dopm)
            if Dopm < m2.Dpm
                z = zeros(1,999);
                pr = zeros(1,999);
                for n=1:length(z)
                    pr(n) = m2.cdfFPT(n*m2.tau,Dopm) - m2.cdfFPT((n-1)*m2.tau,Dopm);
                    if pr(n) > m2.tol_pr
                        z(n) = m2.ECC_n(n,Dopm);
                    end
                    if n>1 && pr(n) < pr(n-1) && pr(n) < m2.tol_pr
                        break;
                    end
                end
                y = sum(z);
            elseif Dopm == m2.Dpm
                % Dopm == Dpm --> No OPM, only PM --> ECC = csusd + cpm
                y = m2.cssd + m2.cpm;
            end
        end
        function y = ECL(m2,Dopm)
            if Dopm < m2.Dpm
                pr = zeros(1,999);
                z = zeros(1,999);
                for n=1:length(z)
                    pr(n) = m2.cdfFPT(n*m2.tau,Dopm) - m2.cdfFPT((n-1)*m2.tau,Dopm);
                    if pr(n) > m2.tol_pr
                        z(n) = m2.ECL_n(n,Dopm);
                    end
                    if n>1 && pr(n) < pr(n-1) && pr(n) < m2.tol_pr
                        break;
                    end
                end
                y = sum(z);
            elseif Dopm == m2.Dpm
                % Dopm == Dpm --> Topm == Tpm
                %  No OPM, only PM --> ECL = E[Tpm]
                y = m2.exFPT(m2.Dpm)+m2.L;
            end
        end
        function [p_usd_opm,p_sd_opm,p_pm] = p(m2,Dopm)
            % Maintenance distribution of monitored component
            pr = zeros(1,999);
            p_sd_opm = zeros(1,999);
            p_pm = zeros(1,999);
            for n=1:999
                pr(n) = m2.cdfFPT(n*m2.tau,Dopm) - m2.cdfFPT((n-1)*m2.tau,Dopm);
                if pr(n) > m2.tol_pr
                    [p_pm(n),p_sd_opm(n)] = m2.p_n(n,Dopm);
                end
                if n>1 && pr(n) < pr(n-1) && pr(n) < m2.tol_pr
                    break;
```

```
                end
            end
            p_sd_opm = sum(p_sd_opm);
            p_pm = sum(p_pm);
            p_usd_opm = 1-p_sd_opm-p_pm;
        end
        function y = g(m2,Dopm)
            % Long-term expected system cost rate
            y = zeros(1,length(Dopm));
            for i=1:length(Dopm)
                y(i) = m2.ECC(Dopm(i))/m2.ECL(Dopm(i));
            end
        end
    end

end
```

Listing N.7: *Extended model given stochastic degradation*

```
classdef m2gp2
    % Class for evaluation of Model 2
    % Degradation is described as a stochastic Gamma degradation process
    %
    %   /--MODEL PARAMETERS--/
    %
    %   cssd    : Set-up cost SD
    %   cpm     : Preventive maintenance cost
    %   copmsd  : Opportunistic maintenance cost during USD
    %   copmusd : Opportunistic maintenance cost during SD
    %   cperm   : Expected periodic maintenance cost
    %   Dpm     : Critical PM threshold
    %   lambda  : Failure rate of FBM process
    %   tau     : Periodic maintenance interval
    %   alpha   : Parameter of Gamma degradation process
    %   beta    : Parameter of Gamma degradation process
    %   L       : PM lead time
    %
    properties
        % Default parameters
        lambda  = .01;
        cssd    = 200;
        cpm     = 200;
        copmusd = 100;
        copmsd  = 100;
        Dpm     = 100;
        alpha   = 1;
        beta    = 2;
        tau     = 90;
        L       = 10;

        tol_pr  = 1e-6;
    end
    methods
        function obj = m2gp2(lambda,Dpm,cssd,cpm,copmusd,copmsd,alpha,beta,tau,L)
            if nargin > 0
                obj.cssd    = cssd;
                obj.cpm     = cpm;
                obj.copmusd = copmusd;
                obj.copmsd  = copmsd;
                obj.Dpm     = Dpm;
                obj.lambda  = lambda;
                obj.alpha   = alpha;
                obj.beta    = beta;
                obj.tau     = tau;
```

```
            obj.L       = L;
        end
    end
    function y = pdfFPT(m2,x,C)
        % First Passage Time pdf of degradation process
        delta = 1e-8;
        y = zeros(size(x));
        for i=1:length(y)
            if C>0 && x(i)>delta
                y_2 = 1-gammainc(C*m2.beta,m2.alpha*(x(i)+delta));
                y_1 = 1-gammainc(C*m2.beta,m2.alpha*(x(i)-delta));
                y(i) = (y_2-y_1)/(2*delta);
            elseif C==0 && x(i)<=delta
                y(i) = 1;
            end
        end
    end
    function y = cdfFPT(m2,x,C)
        % First Passage Time cdf of degradation process
        y = zeros(size(x));
        if C>0
            y = 1-gammainc(C.*m2.beta,m2.alpha.*x);
        end
    end
    function y = exFPT(m2,C)
        % Mean First Passage Time of degradation process
        int = @(x) (1-m2.cdfFPT(x,C));
        y = quadgk(int,0,inf,'MaxIntervalCount',100,'AbsTol',1.e-2);
    end
    function y = ECC_n(m2,n,Dopm)
        % ECC during ((n-1)\tau,n\tau)
        function z = f_outer(u)
            z = zeros(size(u));
            for j=1:length(u)
                a=0;
                b=n*m2.tau-u(j);
                f_inner = @(x)(m2.copmusd.*(1-exp(-1.*m2.lambda.*x))...
                    +(m2.cssd+m2.cpm).*exp(-1.*m2.lambda.*x))...
                    .*m2.pdfFPT(x,m2.Dpm-Dopm);
                z(j) = (quadgk(f_inner,a,b,'MaxIntervalCount',100,'AbsTol',1.e-2)↩
                    ...
                    + (m2.copmusd*(1-exp(-1*m2.lambda*(n*m2.tau-u(j))))...
                    +m2.copmsd*exp(-1*m2.lambda*(n*m2.tau-u(j))))...
                    *(1-m2.cdfFPT(n*m2.tau-u(j),m2.Dpm-Dopm)))...
                    *m2.pdfFPT(u(j),Dopm);
            end
        end
        y = quadgk(@f_outer,(n-1)*m2.tau,n*m2.tau,...
            'MaxIntervalCount',100,'AbsTol',1.e-2);
    end
    function y = ECL_n(m2,n,Dopm)
        % ECL of cbm process during ((n-1)\tau,n\tau)
        function z = f_outer(u)
            z = zeros(size(u));
            for j=1:length(u)
                a=0;
                b=n*m2.tau-u(j);
                f_inner = @(x)((1./m2.lambda).*(1-exp(-1.*m2.lambda.*x))...
                    +m2.L.*exp(-1.*m2.lambda.*x)).*m2.pdfFPT(x,m2.Dpm-Dopm);
                z(j) = (u(j)+quadgk(f_inner,a,b,...
                    'MaxIntervalCount',100,'AbsTol',1.e-2)...
                    + (1/m2.lambda)*(1-exp(-m2.lambda*(n*m2.tau-u(j))))...
                    *(1-m2.cdfFPT(n*m2.tau-u(j),m2.Dpm-Dopm)))...
                    *m2.pdfFPT(u(j),Dopm);
            end
        end
```

```
            y = quadgk(@f_outer,(n-1)*m2.tau,n*m2.tau,...
                'MaxIntervalCount',100,'AbsTol',1.e-2);
        end
        function [p_pm,p_sd_opm] = p_n(m2,n,Dopm)
            % Probabilities during ((n-1)\tau,n\tau)
            function z = f_outer(u)
                z = zeros(size(u));
                for j=1:length(u)
                    a=0;
                    b=n*m2.tau-u(j);
                    f_inner = @(x)exp(-1.*m2.lambda.*x)...
                        .*m2.pdfFPT(x,m2.Dpm-Dopm);
                    z(j) = quadgk(f_inner,a,b,...
                        'MaxIntervalCount',100,'AbsTol',1.e-2)...
                        *m2.pdfFPT(u(j),Dopm);
                end
            end
            f_int = @(u) exp(-1.*m2.lambda.*(n*m2.tau-u))...
                        .*(1-m2.cdfFPT(n.*m2.tau-u,m2.Dpm-Dopm))...
                        .*m2.pdfFPT(u,Dopm);
            p_pm = quadgk(@f_outer,(n-1)*m2.tau,n*m2.tau,...
                'MaxIntervalCount',100,'AbsTol',1.e-2);
            p_sd_opm = quadgk(f_int,(n-1)*m2.tau,n*m2.tau,...
                'MaxIntervalCount',100,'AbsTol',1.e-2);
        end
        function y = ECC(m2,Dopm)
            if Dopm < m2.Dpm
                z = zeros(1,999);
                pr = zeros(1,999);
                for n=1:length(z)
                    pr(n) = m2.cdfFPT(n*m2.tau,Dopm) - m2.cdfFPT((n-1)*m2.tau,Dopm);
                    if pr(n) > m2.tol_pr
                        z(n) = m2.ECC_n(n,Dopm);
                    end
                    if n>1 && pr(n) < pr(n-1) && pr(n) < m2.tol_pr
                        break;
                    end
                end
                y = sum(z);
            elseif Dopm == m2.Dpm
                % Dopm == Dpm --> No OPM, only PM --> ECCcbm = csusd + cpm
                y = m2.cssd + m2.cpm;
            end
        end
        function y = ECL(m2,Dopm)
            if Dopm < m2.Dpm
                pr = zeros(1,999);
                z = zeros(1,999);
                for n=1:length(z)
                    pr(n) = m2.cdfFPT(n*m2.tau,Dopm) - m2.cdfFPT((n-1)*m2.tau,Dopm);
                    if pr(n) > m2.tol_pr
                        z(n) = m2.ECL_n(n,Dopm);
                    end
                    if n>1 && pr(n) < pr(n-1) && pr(n) < m2.tol_pr
                        break;
                    end
                end
                y = sum(z);
            elseif Dopm == m2.Dpm
                % Dopm == Dpm --> Topm == Tpm
                %  No OPM, only PM --> ECLcbm = E[Tpm]
                y = m2.exFPT(m2.Dpm)+m2.L;
            end
        end
        function [p_usd_opm,p_sd_opm,p_pm] = p(m2,Dopm)
            % Maintenance distribution of monitored component
```

```
            pr = zeros(1,999);
            p_sd_opm = zeros(1,999);
            p_pm = zeros(1,999);
            for n=1:999
                pr(n) = m2.cdfFPT(n*m2.tau,Dopm) - m2.cdfFPT((n-1)*m2.tau,Dopm);
                if pr(n) > m2.tol_pr
                    [p_pm(n),p_sd_opm(n)] = m2.p_n(n,Dopm);
                end
                if n>1 && pr(n) < pr(n-1) && pr(n) < m2.tol_pr
                    break;
                end
            end
            p_sd_opm = sum(p_sd_opm);
            p_pm = sum(p_pm);
            p_usd_opm = 1-p_sd_opm-p_pm;
        end
        function y = g(m2,Dopm)
            % Long-term expected system cost rate
            y = zeros(1,length(Dopm));
            % Dopm > 1.e-4
            for i=1:length(Dopm)
                if Dopm(i)==0
                    Dopm(i) = 1.e-4;
                end
                y(i) = m2.ECC(Dopm(i))/m2.ECL(Dopm(i));
            end
        end
    end
end
```