Doctoral theses at NTNU, 2020:188

Adriaen Verheyleweghen

Control Degrees of Freedom for Optimal Operation and Extending Remaining Useful Life

Application to subsea production and processing

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Natural Sciences
Department of Chemical Engineering

NTNU
Norwegian University of
Science and Technology

Adriaen Verheyleweghen

Control Degrees of Freedom for Optimal Operation and Extending Remaining Useful Life

Application to subsea production and processing

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2020

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Chemical Engineering

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Natural Sciences
Department of Chemical Engineering

ISBN 978-82-326-4724-8 (printed ver.)
ISBN 978-82-326-4725-5 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:188

Printed by NTNU Grafisk senter

To my parents

Abstract

The global demand for energy is increasing. While the transition to renewable energy sources is happening faster than ever, traditional energy sources such as oil and gas still make up a large part of the energy portfolio, and will continue to do so for many years to come. However, most new discoveries of oil and gas are more challenging to produce, and old fields are getting depleted. Being able to efficiently operate mature brown fields and challenging green fields will be crucial for oil companies in the coming years. Subsea technology looks very promising, as it enables production from fields which were previously deemed infeasible, either because they were too deep, too remote, or in too hostile environments such as the arctic. Some companies even envision all-subsea production and processing of oil and gas, as this would eliminate the need for expensive topside facilities. However, placing production and processing equipment on the seabed comes with a number of challenges tied to reliability and automation.

Operating a subsea production and processing facility remotely and potentially autonomously is more challenging than a corresponding topside facility for a number of reasons. Firstly, due to the inaccessibility of the plant, reliability of the plant is key. Maintenance costs can be astronomically high, since maintenance intervention of a subsea installation usually involves the retrieval and replacement of entire modules with specialized lifting ships, as well as lost production. It is therefore extremely important to minimize the downtime of the system and to maximize the reliability. However, this objective is in opposition with the production objective. From a production perspective, it is often desirable to run the plant at the capacity limit, which degrades the system faster. Furthermore, it is desired to use the manipulated variables of the plant actively to reject disturbances. Subsea chokes, however, are often not designed to be operated in this fashion, and degrade faster if the choke opening is changed repeatedly and abruptly.

Another issue that arises when operating subsea production and processing equipment is that of uncertainty. Measurements are inaccurate and scarce, due to cost limitations. Equipment degradation models are often poor or non-existing, due to the highly stochastic nature of degradation processes. However, if not dealt with properly, this uncertainty renders useless any attempt at planning production and maintenance of the plant. Such a plan is not acceptable due to the possibility of an unexpected breakdown and maintenance intervention.

This thesis is a collection of papers produced during my PhD.

The papers address the problems raised above by developing strategies for integrating equipment health and reliability information into the control and production planning. By doing so, we obtain a so-called health-aware control structure, which achieves the control objectives without jeopardizing the equipment reliability. We apply our ideas to industrially relevant case studies such as subsea compression, subsea choke control, subsea separation and LNG liquefaction. In order to ensure that the solution is optimal despite model uncertainty, we apply robust problem formulations. Various methods are tested and compared, with the conclusion that for most cases, a scenario-based approach gave a desired trade-off between optimality of the solution and robustness against uncertainty.

We also developed a new approach to combined optimization of maintenance scheduling and production planning. We show that for certain processes, we can formulate the combined problem as a non-linear program (NLP) instead of a formulating it as a mixed-integer program (MIP), and solve it using standard, off-the-shelf solvers. The solution obtained this way is near globally optimal, and performs much better than a clock- or age-based maintenance schedule, which is common in industry today.

Contents

List of figures

List of tables

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of philosophiae doctor (PhD) at the Norwegian University of Science and Technology (NTNU).

The work performed during this PhD was carried out at the Department of Chemical Engineering from August 2015 to October 2019, under the supervision of associate professor Johannes Jäschke and co-supervision of professor Sigurd Skogestad. The original project duration was 3 years, but the project was extended due to teaching duties.

During my PhD studies I had the privilege to visit professor Paul Barton's lab at the Massachusetts Institute of Technology (MIT), USA, for two weeks during January 2017. I am very grateful for the hospitality I was shown there during my short stay, and I would like to thank professor Barton and Rohit Kannan for the fruitful collaboration.

I also had the privilege to visit professor Galo Le Roux's lab at the University of São Paulo (USP), Brazil, from April 2019 to June 2019. I would like to thank professor Galo and his group for their exceptional hospitality and for the fruitful collaboration.

This project is part of the SUBPRO center for research-based innovation (SFI), which is a research program on subsea production and processing funded by NTNU, the Norwegian Research Council (project number 237893) and industrial partners.

Acknowledgments

It feels strange writing these acknowledgments almost ten years after I first came to Trondheim to pursue a degree in Chemical Engineering. The last four and a half years were spent working on this PhD. It has been a long and interesting journey. There are certainly a lot people who I need to thank at this point, and without whom this PhD would not have been possible, or at least not very enjoyable.

First and foremost, I want to thank my supervisor Johannes Jäschke, for his constant support and guidance, and for believing in me and giving me the opportunity to work on this project. I have learned a lot from our countless discussions, and I really appreciate that he has been so encouraging and patient. Johannes is not only a great supervisor, teacher and mentor, but also an overall great human being and friend. This project certainly would not have been possible without him.

Secondly, I want to thank my cosupervisor Sigurd Skogestad for his immense knowledge, which he eagerly shares with everyone. Despite being busy with directing SUBPRO, being the head of the PSE group, and a professor with lots of teaching duties, he manages to ensure that everyone in the group feels welcome, and that there is always a social event to look forward to. Speaking of SUBPRO, I am grateful for the financial support, the travelling opportunities and the great social climate SUBPRO has provided.

During one of my travels I had the privilege to visit Prof. Paul Barton at MIT. I am very thankful for being given challenging tasks and opportunities while I stayed there. A special thanks goes to Rohit Kannan, for patiently teaching me, helping me out, and for letting me crash his code.

A big thanks goes to Prof. Galo Le Roux and his group at USP. Thanks for the fruitful collaboration and the many interesting talks at the coffee machine. Shoutout to José, Rafael, Maria, Roberta, Henrique, Priscilla, Alexandre, Matheus and Kaccnny, for giving me such warm welcome and making me feel right at home in the group! Thanks also to Suelen and everyone else who took care of me while I was in São Paulo. I had a great time in Brazil, and I will definitely be back to have a cachaça or two with you guys in the future! The financial support by INTPART that enabled this trip is gratefully acknowledged.

Working on a PhD is not always easy, so I am very grateful for having had such wonderful officemates. Thanks to Eka, Tamal, Christoph, Dinesh, Allyne, José, Timur and Ana for creating a relaxed and fun work environment. Thinking back

at all the (sometimes bizarre) conversations puts a smile on my face. Furthermore, I want to thank the rest of the PSE group, including Julian, Adriana, Tobias, Cansu, Sigve, Mandar, Cristina, Pablo, Vlad, Haakon, David, Lucas, Fabienne, Andrea, Bahareh, Niloufar, Robert, Zawadi, Fernando and Pedro. I hope I did not forget anyone, but I hope you can forgive me if I did. Thanks for the many coffee breaks, lunches and extracurricular activities. I hope I will see you all at Dinesh's next "innflytningsfest". To all the other people who I have met on and off campus, whether it be at band practice, on the slope in Åre, at a karaoke nachspiel in PFI, dancing in Downtown (Are and Kristin, this is your shout-out!) or drinking beer in Circus; thanks for brightening my mood at those times when PhD-life was more difficult. That being said, being a PhD at IKP has been nothing but a good experience, so thanks to everyone in the department for creating such a great environment to be a researcher in.

Since October 2019 I have been working in Cybernetica, while completing my thesis in my spare time. I am very lucky to be able to work with so many bright and nice people as my new collogues, and that I get to continue working on exciting and challenging MPC problems.

Dinesh and Cristina helped proofreading this manuscript on short notice. Thank you very much!

I also want to thank the committee members Prof. John Bagterp Jørgensen, Dr. Olav Slupphaug and Prof. Magne Hillestad for giving their time and effort to evaluating my thesis.

Finally, I want to thank my family. I am eternally grateful to my parents Daniela and Ferdinand, for everything they have done for me. I also want to thank my wonderful sister Thérèse, for being someone I can always look up to and learn from.

Chapter 1

Introduction

1.1 Motivation

With most easy to produce oil and gas fields being exhausted, the petroleum industry has to resort to increasingly challenging fields to satisfy the global energy demand. Water depths increase, tie-backs become longer and fields with less ideal compositions have to be operated. Subsea processing enables fields that were previously considered infeasible. Equinor[1] envisions a full subsea factory where all topside equipment is moved to the seabed [166]. The advantages of a subsea factory are evident: higher efficiency, increased safety, enhanced production and enablement of fields in remote locations with harsh climate conditions. In [140], Equinor defines the following ambitions for the subsea factory of tomorrow:

- Operate with **longer** distances between shore and production facilities, and longer step-outs from wells.
- Produce from **deeper** wells, leading to higher pressures and increased inaccessibility for maintenance.
- Operate in **colder** environments such as in the arctic, with potentially more viscous fluids.
- Increased recovery from **older** fields.

However, a plethora of challenges have to be overcome first before the subsea factory becomes reality [140].

1.2 Main research questions

Traditional subsea installations, have very few degrees of freedom (DOFs) for control, apart from the flow rate through the Christmas tree. Optimal operation is consequently achieved when the production rate is at its given set-point. Utilizing novel technologies to enable the ambitions of a subsea factory, means that processing equipment such as separators, pumps and compressors are put on the seabed.

[1] Then Statoil

With these new and untested kind of equipment, more degrees of freedom (DOFs), e.g. compressor speed, level set-points in the separator, etc., become available. Restrictions for subsea operation are also somewhat different than for similar topside installations. For example, noise limits are less strict. Optimal operation becomes non-trivial, and a systematic operating strategy is needed. This leads to the first and main research question that is answered in this thesis:

Research question 1. *How do we best utilize the novel degrees of freedom introduced by advanced processing equipment such as compressors, pumps and control valves?*

- This question is the underlying theme of the thesis, and is addressed in all chapters.

Another major concern regarding subsea processing is the reliability of the system. Unexpected production stops result in the loss of valuable production time. Additionally, the cost of intervention in the case of a module breakdown is very high due to the need for specialized recovery vessels and remotely operated vehicles (ROVs) [64]. It is for these reasons that the equipment is commonly designed and operated in such a fashion that the chance of failure becomes marginally small. However, this can lead to a very conservative design and operation strategy. Prognostics and health monitoring (PHM) is used to assess the health of the system. The second research question is:

Research question 2. *What is the current status of subsea PHM technology?*

- This question is addressed in Chapter 2.

The idea behind the current project is that PHM can be combined with advanced control methods to ensure reliable operation without having to operate excessively conservatively. The goal is to ensure that the remaining useful life (RUL) of the equipment is not exhausted before the next planned maintenance stop, whilst maximizing the profit. This leads to the third research question that is answered in this thesis:

Research question 3. *How can we systematically integrate PHM data into our existing control structures to ensure that we have a control structure that adjusts production such that equipment RUL is not exhausted prematurely?*

- This question is addressed in Chapter 3 and Chapter 4.

Going one step further, one might ask if it is possible to use optimization to optimize the maintenance schedule simultaneously with the production profile between maintenance interventions. Especially if a clock-based or a periodic maintenance schedule is used, as it is common in the petroleum industry, a lot of expenses can be saved by potentially avoiding unnecessary maintenance stops. In theory, by adding decision variables to our optimization problem, we can simultaneously find the optimal schedule of maintenance interventions. However, the resulting problem

becomes larger and more complex, and requires novel methods and approaches to solve efficiently. The forth research question is:

Research question 4. *Knowing that equipment will have to be maintained in the future, what is the optimal way of operating between maintenances, and what is the optimal schedule for maintenance of the equipment?*

- This question is addressed in Chapter 7 and Appendix B.

In order to use automation and models of complex novel processing equipment, it is important to develop methods and tools which are robust to uncertainty and stochasticity. Since many of the unit operations required to realize a subsea factory have yet to be developed and installed in subsea conditions, operational data is scarce or non-existent. Models will therefore have a large degree of uncertainty associated with them. Furthermore, as sensors are expensive to install and maintain in subsea settings, instrumentation of the subsea plant may be lacking when compared to a comparable topside or onshore facility. Last, but not least, the final research question is:

Research question 5. *How can we ensure that our strategy is still optimal in the presence of uncertainty and stochasticity?*

- This question is addressed in all chapters, except for Chapter 5, in which uncertainty is not explicitly considered.

1.3 Thesis outline

The main aim of this PhD was to investigate the research questions posed above. The chapters in this thesis are all based on conference and journal articles published during my PhD, with exception of Chapter 2, which is based on a gap analysis performed at the start of my PhD, and Chapter 8, which concludes the thesis. For a full list of publications, see Section 1.5.

- **Chapter 2** is based on a gap analysis performed in the early stages of my PhD. The aim was to get an overview over the academic and industrial status in the fields of PHM and advanced control, and thereby answering research question 2. The gaps identified were used to formulate the other research questions from the introduction. I revisited the gap analysis towards the end of the PhD to supplement it with articles that have been published in the 4 years since the start of the PhD.
- **Chapter 3** investigates research question 3. We developed a systematic approach for integrating reliability information into the control domain. We applied the approach to a subsea compression station.
- **Chapter 4** builds upon Chapter 3. We further developed the systematic approach for integrating reliability information into the control domain. We applied the approach to a LNG liquefaction plant. In this chapter, we also

discussed the importance of proper time-scale separation for these kind of problems. This is an important issue, since degradation processes usually occur over months and years, whereas operational decisions are made on a minute-to-minute basis.

- **Chapter 5** focuses more on the time-scale separation issue. The system studied there is an active surge controller, which stabilized operation of a subsea compressor close to an unstable operating point. The surge phenomena is characterized by oscillations with frequency in the range of 1Hz, whereas the accumulated damage on the compressor does not become significant until after some years. We therefore devised a three-layer structure to control this system.

- **Chapter 6** deals with a different kind of degradation, namely particle erosion in production chokes. This problem was pitched by industry partners in SUBPRO. In this chapter, we develop a controller to find the optimum production rates to keep the erosion within acceptable limits, while simultaneously maximizing hydrocarbon production.

- In **Chapter 7**, we primarily address research question 4. We investigate how a relatively complex system with multiple components should be operated. We also develop a new method for optimizing maintenance intervals. Contrary to most approaches from literature, we do not have to solve a mixed integer program (MIP). We formulate the problem as a non-linear program (NLP), so that we can solve it in short time without specialized solvers. We show that the maintenance schedule from our optimization-based approach is superior to the clock-based approach to maintenance scheduling that is predominantly used in industry.

- In **Appendix A** we study the operation of a subsea compression station under uncertainty. Two different formulations for handling uncertainty were compared.

- In **Appendix B**, we primarily address research question 4. The paper presented in this chapter is a precursor to the paper presented in Chapter 7. Many of the concepts later built upon in that chapter were introduced here.

Research question 1, being the most basic of all posed research questions, is addressed in all of the chapters. Since the focus of question 1 is on novel process concepts using advanced process equipment such as compressors, most of the case studies throughout the thesis are focusing on these kind of processes.

The gap analysis showed that model uncertainty is a fundamental limitation of PHM systems. Discussion with SUBPRO industry partners also confirmed that various kinds of uncertainty, whether it be measurement uncertainty, model inaccuracy, or lacking instrumentation, is one of the biggest issues when it comes to optimal operation of oil and gas processes. Research question 5 was therefore deemed so important that every Chapter deals with this issue in one form or another, with exception of Chapter 5.

1.4 Thesis contributions

In the author's view, the two main contributions of the thesis are

1. The application of reliability-aware operations, optimization under uncertainty, and advanced process control to novel subsea case studies.

2. A new method for simultaneously optimizing production and maintenance schedules by reformulating the problem as a non-linear problem with complementarity constraints.

1.5 List of publications

Over the course of the last four and a half years, I have authored or co-authored seven peer-reviewed papers published in the proceedings of international conferences, and four journal papers. Two of these papers were based on my master's thesis, and are unrelated to the topic of my PhD. I also presented my work in the form of oral or poster presentations at a number of occasions.

1.5.1 Conference papers

1. A. Verheyleweghen and J. Jäschke. Framework for combined diagnostics, prognostics and optimal operation of a subsea gas compression system (IFAC World Congress, Toulouse, France). *IFAC-PapersOnLine*. volume 50, pages 15916—15921, 2017 - **Chapter 3**

2. A. Verheyleweghen, J.M. Gjøby and J. Jäschke. Health-Aware Operation of a Subsea Compression System Subject to Degradation (28th European Symposium on Computer Aided Process Engineering (ESCAPE), Graz, Austria). *Computer Aided Chemical Engineering*. volume 43, pages 1021—1026, 2018 - **Chapter 5**

3. A. Verheyleweghen and J. Jäschke. Oil Production Optimization of Several Wells Subject to Choke Degradation (IFAC Workshop on Automatic Control in Offshore Oil and Gas Production (OOGP), Esbjerg, Denmark). *IFAC-PapersOnLine*. volume 51(8), pages 1—6, 2018 - **Chapter 6**

4. A. Verheyleweghen and J. Jäschke. Robust Health-aware operation of a subsea gas compression system. *Proceedings of the 2017 Conference of Foundations of Computer Aided Process Operations / Chemical Process Control (FOCAPO/CPC)*. 2017 - **Appendix A**

5. A. Verheyleweghen, H. Srivastav, A. Barros and J. Jäschke. Combined Maintenance and Production Optimization (European Safety and Reliability Conference (ESREL), Hannover, Germany). *Proceedings of the 29th European Safety and Reliability Conference*. pages 499—506, 2018 - **Appendix B**

1.5.2 Journal papers

1. A. Verheyleweghen and J. Jäschke. Combined Reliability and Optimal Operation: Application to an LNG Liquefaction Plant. *Journal of Process Control*. (Under review), 2020 - **Chapter 4**
2. A. Verheyleweghen, H. Srivastav, A. Barros and J. Jäschke. A Unified Approach for Simultaneous Optimization of Production and Maintenance Schedules. *IEEE Transactions on Reliability*. (Under review), 2020 - **Chapter 7**

1.5.3 Papers published during my PhD studies, but not included in the thesis

1. A. Verheyleweghen and J. Jäschke. Self-optimizing control of an LNG liquefaction plant. *Journal of Process Control*. volume 74, pages 63—75, 2019
2. A. Verheyleweghen and J. Jäschke. Self-optimizing control of a two-stage refrigeration cycle (11th IFAC Symposium on Dynamics and Control of Process Systems, Including Biosystems (DYCOPS), Trondheim, Norway).*IFAC-PapersOnLine*. volume 49(7), pages 845—850, 2016
3. J. Rúa, A. Verheyleweghen, J. Jäschke and L.O. Nord. Optimal scheduling of flexible thermal power plants with lifetime enhancement under uncertainty. *Applied Energy*. (Under review), 2020
4. T. Das, S.J. Heggheim, M. Dudek, A. Verheyleweghen and J. Jäschke. Optimal operation of a subsea separation system including a coalescence based gravity separator model and a produced water treatment section. *Industrial & Engineering Chemistry Research*. volume 58(10), pages 4168—4185, 2019
5. C.J. Backi, D. Krishnamoorthy, A. Verheyleweghen and S. Skogestad. Combined nonlinear moving horizon estimation and model predictive control applied to a compressor for anti surge control (IEEE Conference on Control Technology and Applications (CCTA), Copenhagen, Denmark). *Proceedings of the IEEE Conference on Control Technology and Applications (CCTA)*. pages 1552—1557, 2018

1.5.4 Presentations and posters (invited, or with abstract only)

1. A. Verheyleweghen. Conference presentation at Nordic process control workshop (NPCW), Stockholm, Sweden, 2016: "Health-aware operation and control of a subsea gas compression station"
2. A. Verheyleweghen. Presentation at the Subsea Valley Conference, Oslo, Norway, 2017: "Process control for extending component life"
3. A. Verheyleweghen. Presentation at the OG21 Forum, Oslo, Norway, 2017: "Smart production optimization"
4. A. Verheyleweghen. Presentation at PhD Grand Prix, Trondheim, Norway, 2018: "Selvkjørende prosessanlegg: Hva kan prosessindustrien lære av Tesla og co.?" (in Norwegian)

5. A. Verheyleweghen. Presentation at 7th Oil and Gas Production Optimization Workshop, Rio de Janeiro, Brazil 2018: "Health-aware production optimization to improve remaining useful life"
6. A. Verheyleweghen. Presentation at INTPART Brazil-Norway Subsea Operation Consortium Workshop, Trondheim, Norway, 2018: "Health-aware production optimization to improve remaining useful life"
7. A. Verheyleweghen. Presentation at CAPE workshop, Bucharest, Romania, 2018: "Risk-averse health-aware control of subsea plants"
8. A. Verheyleweghen. Conference presentation at American Institute of Chemical Engineers Annual meeting (AIChE), Pittsburgh, PA, USA, 2018: "Risk-averse health-aware optimization"
9. A. Verheyleweghen. Presentation at Nordic process control workshop (NPCW), Turku, Finland, 2018: "Risk-Based Health-Aware Control of Subsea System"
10. A. Verheyleweghen. Presentation at IIX Oil and Gas Production Optimization Workshop, Rio de Janeiro, Brazil, 2019: "Combined maintenance and production optimization"
11. A. Verheyleweghen. Presentation and poster at PRONTO Workshop, Trondheim, Norway, 2019: "Combined maintenance scheduling and production optimization"
12. A. Verheyleweghen. Conference poster at Nordic process control workshop (NPCW), Lyngby, Denmark, 2019: "Health-Aware Control - Current status and future outlook"

Chapter 2

Academic and industrial status

2.1 Academic status

This section is meant to give an overview of the academic status on the topic of prognostics and control.

First, the general concept of condition-based maintenance is introduced in Subsection 2.1.1, where also the concept of prognostics and health monitoring (PHM) is explained, and some examples of applications of PHM in literature are given. PHM in subsea is treated explicitly in Subsection 2.1.2. Several different failure mechanisms of various subsea equipment are covered.

Subsection 2.1.3 deals with how PHM and control can be united in one common framework to achieve reliability-aware control and operation.

2.1.1 Condition-based maintenance

Historically, maintenance was performed when equipment broke down or performance became unacceptably low. This run-to-failure approach is also known as *reactive/corrective maintenance*. In more recent times, focus has shifted from reactive to proactive maintenance. The simplest form of proactive maintenance is based on inspecting the system at constant intervals. This policy is sometimes referred to as *clock-based* or *constant-interval maintenance* [87]. A slightly more advanced policy is to inspect the system more frequently depending on its age and knowledge of some underlying stochastic lifetime expectancy. This approach is typically referred to as *age-based maintenance* [87]. The disadvantage of these approaches is that the interval between maintenance is based on statistical information such as the mean time between failures. Due to the insufficiency of the statistical information, these approaches cannot ensure that the system will function satisfactory until the next planned maintenance stop, even with a built-in safety margin.

So-called condition-based maintenance (CBM) has largely taken over as the standard approach in industry and aviation [80]. The advantage of CBM is that it enables planned maintenance stops, which usually lead to significantly higher availability compared to a reactive maintenance policy.

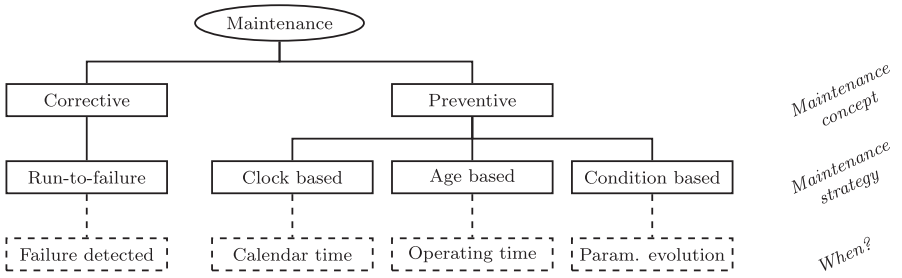

Figure 2.1: Overview over different approaches to maintenance. Adapted from [142, Ch. 9]

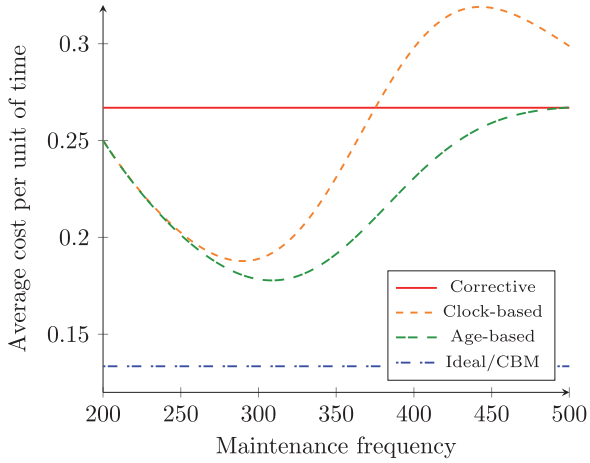

Figure 2.2: Comparison of average cost per time for different maintenance policies. Adapted from [94, 142].

Figure 2.1 shows an overview over different approaches to maintenance. As argued above, the maintenance approaches result in different degrees of availability, which results in different average operation costs. Figure 2.2 shows a comparison of the average cost per time unit for different maintenance policies for a simple example with fixed maintenance costs and inspection costs. The average costs are calculated based on an example from [94, 142]. It can be seen that corrective maintenance is generally associated with the highest cost. Clock-based maintenance provides a lower cost, unless the parameters are poorly chosen. Age-based maintenance is always as good as, or better than, a corrective maintenance policy. Assuming we have perfect information about the degradation state of the system, a CBM policy will result in ideal cost.

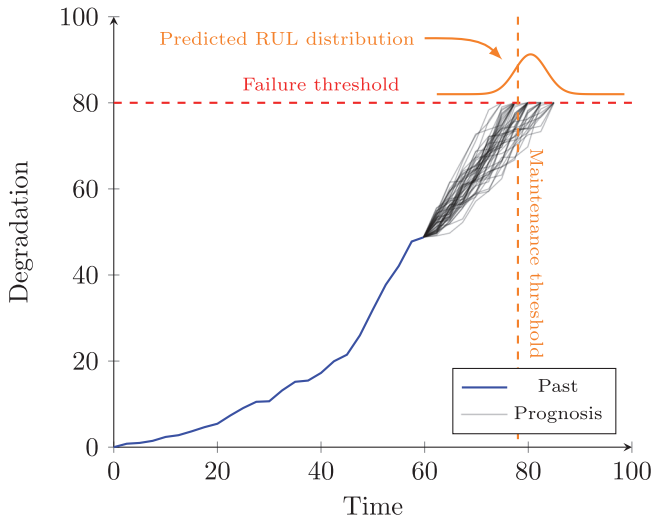

Figure 2.3: Prognosis of RUL. Adapted from [174, Ch. 7.5]

Prognostics and health monitoring

In order to optimize the maintenance and repair schedule of the equipment, the condition or health of the system has to be known. Usually the condition cannot be measured directly, but must be estimated based on measurements of other key variables in the process. Typically, environmental parameters such as temperature, vibration, pressure and shock, as well as operation parameters such as current, power consumption and heat dissipation can be used to estimate the health state of the equipment [142, Ch. 9.6]. Not only is the failure state estimated at the current time (diagnosis), but predictions for the future (prognosis) are generated as well. Diagnosis revolves around the detection and isolation of failures, whereas prognosis is used to determine if the equipment can perform sufficiently until the next scheduled maintenance stop.

Diagnosis and prognosis are commonly collectively known as prognosis and health management (PHM) [87]. The estimation of the remaining useful life (RUL) is central in PHM [169]. Being able to do accurate prognosis of the performance measures is an integral part of PHM systems, and consequently a substantial amount of literature on the subject exists. See for instance review papers by Katimapula, Jardine, Peng, Zhang and corresponding coauthors [80, 83, 133, 192]. This prognosis is also the main drawback of CBM compared to time-based maintenance, due to the difficulty of finding accurate models for prognosis.

A wide variety of approaches to PHM can be found in literature. Generally, they can be divided into one of three categories:

1. *Model-based approaches*
   Physical model-based approaches are based on a mathematical (first princi-

ples) model representation of the actual system. Deriving these models can be a challenging task, as it requires a deep understanding of the underlying process and failure modes. Furthermore, the model-based approach results in inaccurate predictions if the occurring failure mode was not considered in the model of the system. It is therefore important to consider all relevant failure modes when deriving the model. For complex systems, the model can thus become very large, and consequently numerically expensive to solve and difficult to calibrate. The advantages of physics-based approaches are that they can be very accurate if the model is good, and that the same model can be used for different systems simply by re-parameterizing it. Model-based approaches are sometimes also known as deep-knowledge expert systems in industry, since they require a deep understanding of the modelled system. In contrast, shallow-knowledge expert systems (covered under point 2) require less insight into the underlying physics of the system [37, Ch. 12].

Faults are detected by evaluating the residuals of the difference between the model output and the output of the actual system. Typically, a mathematical observer is used to estimate the health state of the system, as the variable related to the health is not directly measurable. Approaches using Luenberger observers or Kalman filters fall in this category. For recent applications, see for example [66, 125, 145, 193].

One important aspect which is worth addressing, is prognostics in the presence of uncertainty. Since the RUL-prediction is usually based on incomplete information about the state and the model of the system, it is usually distributed rather than a single value. This is illustrated in Figure 2.3. In this case the maintenance threshold is defined as a percentile of the RUL distribution. Calculating the percentile can be challenging in practice unless the model is linear. For the general case, it is therefore necessary to calculate an approximation of the RUL percentile based on Monte Carlo (MC) simulations, linear approximations such as the first-order reliability method (FORM) [154] or Bayesian networks [98].

2. *Knowledge-based approaches*
Because physics-based models can be hard to derive, model-free approaches are sometimes preferred. The class of knowledge-based approaches are characterized by being very "human-like" [133]. Expert systems (ESs) are combining human reasoning and the speed of computers. ESs consist of large sets of rules which are implemented as logic and IF-THEN commands. The rules are based on experience and heuristics. In that way, they can be compared to medical diagnosis, which can detect diseases without a complete model of the human, based on common symptoms alone. ESs are sometimes also referred to as shallow-knowledge expert systems [37, Ch. 12]. The disadvantage of ESs is that the ability to detect faults strongly depends on the quality of the incorporated knowledge. If a certain fault is new to the expert and consequently not incorporated into the PHM system, it cannot be detected. Another issue is that the ES is unique to each process, and cannot easily be adopted for other processes. Developing the ES can be a costly and time consuming endeavor, which is why pure ES applications are not commonly found

in recent literature. They have found some use in industry, see for instance [26, 32].

3. *Data-driven approaches*
   Data driven approaches are either statistical methods or machine learning methods based on pattern recognition. These methods are widely used and employ a range of different techniques. For a comprehensive survey, see the two papers by Schwabacher and Goebel [157, 158]. All techniques use historical operating data to fit models of the system. Daigle argues that it is often difficult or even impossible to obtain large sets of run-to-failure training data, which makes data-driven approaches to PHM rather limited in use [40]. Physics-based approaches do not rely on data, and are consequently a better-suited alternative in these cases. Nonetheless, data-driven approaches are widely used. Three commonly used methods include Bayesian networks (BNs), hidden Markov models (HMMs), and artificial neural networks (ANNs).

   Bayesian networks (BNs) are acyclic networks of connected nodes [132, Ch. 3.3]. All edges are directed and have associated probabilities of transition. BNs therefore form an extension of graph theory. BNs are widespread in biomedicine and health-care [101], but are also common in reliability engineering [91, 108]. Dynamic Bayesian networks (DBNs) applies the same principle to time-progressing systems.

   Markov models (MMs) are very similar to BNs, but differ in that the edges are undirected, and that the graph does not have to be acyclic. It is assumed that the nodes in MMs are Markovian, i.e. that the the state $x(t)$ at time t does only depend on the state $x(t-1)$ at time $t-1$. This assumption limits the application somewhat, since real world systems are seldom Markovian [133]. BNs do not have this assumption, and are able to represent the dependencies between nodes. Hidden Markov models (HMMs) are extensions of ordinary MMs where the states are not directly visible, but must be observed through the output [138]. HMMs and MM must be trained with run-to-failure data from experiments.

   Artificial neural networks (ANNs) are networks of connected nodes inspired by biological neural networks. The nodes are connected in the hidden layer such that the ANN is able to fit non-linear functions and recognize patterns. ANNs are very flexible and can be used for a large number of different applications [155]. The core feature of ANNs is the "training" of the network, i.e. the assignment of edges between the nodes. Learning can be unsupervised, meaning that the network constructs features with which the observed variables can be predicted, or unsupervised, which is essentially some form of regression with target output values. The necessity of sufficiently large sets of training data is often problematic. ANNs can be used for RUL prediction, taking the age and the condition monitoring measurements as the inputs to the network [172].

A comparison of the applicability and complexity of the three approaches is nicely summarized by Vachtsevanos et al. [174, Ch. 6], see Figure 2.4.

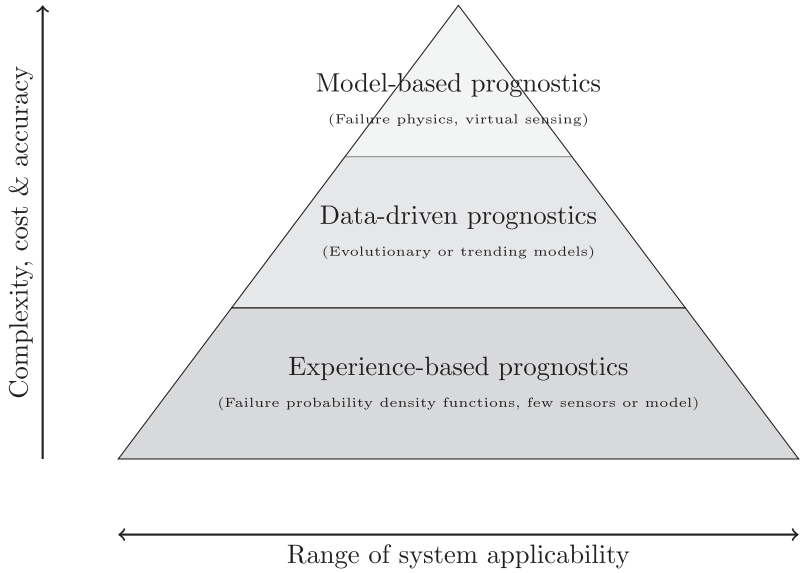

Figure 2.4: Different approaches to prognostics. Adapted from [174, Ch. 6]

At this point it should be noted that although the reader may have been led to believe that the three methods are inherently different, it is uncommon in practice that a PHM system relies entirely on one approach alone. Rather, a combination of all three methods is typically used.

2.1.2 Prognostics and health monitoring in subsea

Due to the high intervention cost for subsea processes, unplanned equipment breakdown and consequent stop of production is very undesirable. As a consequence, CBM is very applicable to subsea industry. The stringent reliability requirements are shared by other, better established, industries such as aviation, energy or rail. This means that some of the mature technologies from those areas can be applied subsea. Indeed, in one of the early works on the topic of subsea condition monitoring, Friedemann et al. draw many parallels between rail and energy, and subsea [61]. They identify several areas of subsea engineering where PHM can be applied, such as in in-flow diagnostics, equipment monitoring and flow assurance. They stress the need for PHM systems which not only estimate the health of single components, but which take the entire system into consideration.

Vaidya points out the insufficiency of the classical approach of RUL estimation based on parametric failure models [175]. She argues that the lack of real-life time-to-failure field data is limiting the applicability of the classical approach, and that laboratory experiments result in unsatisfactory fits due to unrealistic operating conditions. Furthermore, the multitude of simultaneous failure mechanisms which are acting on a component are difficult to replicate in a lab. Vaidya proposes to use

BNs to perform prognosis of subsea equipment. In a follow-up paper by Vaidya and Rausand, the applicability of statistical models, ANN-based models, physics-based models and Bayesian methods for prognosis is discussed [176]. They illustrate their proposed method by applying it to a physics-based model of a subsea pump for raw seawater injection subject to seal leakage failure. No simulation results are shown, but the proposed method is explained in detail. They also highlight the need of expert judgment.

Recently, Chze stresses the need for subsea CBM [39]. She describes how a large majority of installations do not utilize all available plant data efficiently, especially in the context of CBM. Operators are focusing on production monitoring, rather than equipment integrity, which leads to suboptimal operation. A field case study of a failing subsea control module is shown to illustrate the need for prognostics. The paper by Chze shows that despite CBM being used in some installations, it still represents a major industry gap.

For the remainder of this subsection, we will discuss several different failure modes for some pieces of equipment commonly found in subsea production and processing, and how their condition can be monitored.

Sand erosion in chokes and bends

The mechanisms of erosion in tight bends and tees [18, 123, 124, 131, 195], and choke valves [43, 135] are well known. In its simplest and most commonly used form, the model for sand erosion in pipe bends can be written as [150]

$$V_{\max} = \frac{C}{\sqrt{\rho_m}} \tag{2.1}$$

where V is the allowable velocity, C is a constant and ρ_m is the fluid density. This is also known as the API RP 14E standard. Typical values of C are in the range of 100 to 200 $(\mathrm{m}^3/\mathrm{kg})^{1/2}$, but a lower value should be used when solids are anticipated. However, several authors are critical to API RP 14E [150]. Salama suggests a slight alteration of the above equation for sand-ladden fluids [150]. The maximum allowable velocity is then

$$V_{\max} = \frac{D\sqrt{\rho_m}}{20\sqrt{W}} \tag{2.2}$$

Salama defines acceptable erosion rates to be in the order of 0.1 mm/yr for a reservoir containing sand with an average size of 250 µ [150].

For chokes, a general recommendation is to not operate at choke capacities less than 20% of the maximum choke capacity, c_v, for prolonged periods of time if the well is producing sand [43]. Failing to do so can cause erosion wear to choke trim components, which negatively impacts the controllability of the choke. Especially in gas fields, erosion may be a real threat, due to the low density and viscosity of the fluid. DNV-GL suggest that choke vendors specify the allowable operating ranges at different sand loadings and fluid velocities, similar to Figures 4-8 in [43]. In general, however, it is difficult to establish a model which is applicable to a wide

variety of chokes. Since the geometries are different from vendor to vendor, it is advisable that CFD models are used to predict the erosion.

Saether studied the design of wear-resistant chokes for subsea use [148]. The erosion was predicted using computational fluid dynamics. A PHM system was also developed. Estimates of the flow coefficient c_v were used to predict the erosion rate using a physics-based model. Nystad et al. developed a model to assess the RUL of choke valves subject to erosion damage using an empirical c_v-model [121]. Gola and Nystad further developed the model to define a PHM framework for optimal mainenance scheduling of choke valves [65]. Due to the lack of data, they mention that the proposed model is not generally applicable, and that more data needs to be collected. In a later paper [122], they developed a framework based on a non-stationary gamma process with gamma-distributed reliability constraints. Zhang developed a hybrid prognostic approach combining ANNs for estimation of the degradation indicator, and a gamma process to calculate the RUL [191]. Like most other authors, Zhang also used the flow coefficient to as a degradation indicator.

Corrosion in pipes

Internal corrosion might be a problem if the field produces water. In oil and gas streams, the main corrosive agents are carbon dioxide (CO_2), hydrogen sulfide (H_2S), and (free) water. A good overview of various corrosion mechanism can be found in a recent paper by Popoola et al. [136]. For an in-depth summary of sweet corrosion, see Kermani & Morshed [85]. Equivalently, sour corrosion is treated in a recent work by Goodwin et al. [68]. The most common remedies against corrosion damage are chemical inhibitors, anodic/cathodic protection, and protective coatings.

Caleyo et al. list several different physics-based failure pressure models to assess the reliability of a corroded pipe [34]. Due to uncertainties in the load and resistance parameters, stochastic methods were utilized to estimate the probability of the pressure to be lower than a threshold safety pressure. Three different stochastic methods were studied, namely Monte Carlo (MC) simulations, the first-order reliability method (FORM), and the first-order second-moment (FOSM) method. The authors write that all methods gave similar results when a linearized version of the limit state function was used. However, for non-linear limit state functions (as is the case in a pipe with a large number of corrosion defects), one should be wary of the results from the FORM. The authors also observe that the models are sensitive to the load and resistance parameters, which means that careful attention must be paid when the pipeline safety is evaluated over a long time horizon. Caleyo et al. found that the Shell-92 corrosion model [86] and the ASME B31G corrosion model [76] (two of the most commonly used corrosion models) result in the highest and lowest failure probabilities, respectively. In a study conducted by Mokhtar et al., the results from the FORM were compared to a simple degradation analysis based on pipe wall thickness [111]. It was found that the FORM results in a less conservative estimate than the degradation analysis. They also claim that the results from the FORM are comparable to those found in literature. Vinod et al. present a data-driven model for assessment of erosion-corrosion damage utilizing

Markov models [183]. Again, FORM is utilized to perform the reliability analysis. Recently, Ossai used data from 11 different fields to estimate pipeline corrosion using a data-driven linear and power models [128]. Uncertainties are taken into account by formulating the model as a Brownian random walk model. The reliability is evaluated using Monte Carlo simulations.

Due to the wide range of different corrosion mechanisms, a multitude of different corrosion monitoring techniques exist [136, 171]. Typically, one does not rely on just a single of the monitoring techniques, but rather a combination of several different techniques. In addition to the above mentioned physics-based estimation methods, a variety of different sensing technologies exist. They can be divided into intrusive and non-intrusive methods.

One option is to periodically inspect the wall thickness on exposed locations along the pipe. Eddy current testing is a technique where the impedance of an electrical coil is measured [144]. The impedance of the coil changes if the coil is in the vicinity of a metallic surface, such as a pipeline. Furthermore, any surface imperfections, as caused by erosion or corrosion, will influence the impedance. Another option is to use an ultrasonic thickness measuring device [36]. This method is based on the fact that cracks bounce back acoustic waves at a different angle than the perfect wall surface. Both of these techniques are non-intrusive. The sensor can be permanently installed or deployed by an ROV. However, these techniques fail to predict unexpected corrosion and subsequent failure at locations along the pipe which are not being monitored.

There are also intrusive techniques for estimating corrosion damage. These typically have higher sensitivities than the non-intrusive techniques [137]. Weight loss corrosion coupons are simple chunks of metal which are being placed inside the pipe for a certain period of time. Subsequently the coupons are recovered and measured. Based on the weight loss of the coupon, the corrosion rate can be estimated. The disadvantage of coupons is that they do not provide real-time estimates of the corrosion, since results are typically only available after a period of 90 or 180 days. Electrical resitance (ER) probes are better in this respect because they do not have to be physically recovered and evaluated periodically. Instead, they provide near real-time data of the corrosion based on measurements of the resistance. To retrieve the data, a physical connection with the probe has to be made, but the probe itself does not have to be physically removed. ER probes are also much faster at detecting corrosion. In as little as four days from the onset of corrosion, some ER probes will be able to detect the change. For comparison, coupons need to be exposed for at least 30 days to be able to detect the onset of corrosion. Coupons are however still widely used, primarily because of their low cost [137]. Another intrusive method is the use of so-called intelligent pigs. These pigs are autonomous devices equipped with a variety of different sensors for detecting pipeline flaws [173].

Rotating equipment

The first booster pumps were installed subsea in the mid-90s [116]. Subsea compressors are even more recent developments. The first subsea gas compression plant was installed by Equinor [1] on the Åsgard field in 2015 [167]. The same year, Equinor also installed the first wet gas compressor on the Gullfaks field [126]. The available literature of prognostics of subsea pumps, compressors, and other rotating equipment is scarce, mainly because of the novelty of the research area and because companies do not share their proprietary data. Some examples of condition monitoring for on-shore pumps and compressors are listed below.

A method for predicting the performance degradation due to fouling of a compressor onboard a Naval vessel was presented by Kacprzynski & Caguiat [82]. They developed a simplified data-driven model to predict the compressor efficiency. Due to the lack of pressure measurements, they combined flow measurements and empirical rules relating the observed pressure bleeds to the load level. The prediction model was subsequently used to predict the optimal washing intervals.

A hybrid monitoring approach was developed by Miyoshi et al. [110]. They combined phenomenological models based in part on thermodynamic head estimates and in part on empirical compressor curves, with quality control charts.

The topic of PHM for subsea multiphase pumps was studied in detail by Liu [99]. He performs a failure mode, effects, and criticality analysis (FMECA) to identify the critical components of the pump. It was found that the main concern regarding the integrity of the pump was corrosion on the pump screws, followed by leakage in the mechanical seals. Several other critical failure modes were also identified, but due to time considerations, the thesis focused on the two most critical failure modes. Monitoring of seals is usually performed by measuring the temperature (i.e. the contact friction) at the seal faces, measuring the alignment of the rotor, measuring the film thickness between the seals using ultrasonic techniques, or measuring disturbances in the vibration / sound emission frequencies. Liu compares the different methods, and finds the two latter methods to have the greatest potential in a PHM setting. Physics-based models and data-driven models for the prognostics of seals are also described. Liu used a HMM to estimate the health of the pump screws by relating the wear to the pressure difference over the pump.

Monitoring of active magnetic bearings was discussed by Gouws [69]. He differentiates between external faults, including rotor impacts, rotor mass loss, base motion, rotor deformation, overhung rotor, rotor rub, bent rotor and misalignment, and internal faults, including power and electronics failures, transducer malfunctions, loss of I/O board channel, bearing magnet coil failures, computer software errors, computer hardware failures and rotor faults. He also lists how the various faults can be detected using different monitoring techniques.

[1] Then Statoil

Separators

The condition monitoring of a topside gravity separator is discussed by Houmstuen [75]. An FMECA is performed, and he concludes that the critical failure modes are abnormal instrument readings, external leakage and plugging. The state of the separator can be assessed by monitoring various process variables, including the wall thickness, the liquid level and the amount of foaming. Houmstuen lists neutron backscatter-, passive acoustic-, ultrasonic-, gamma-, microwave- and IR-monitoring as possible technologies for monitoring the relevant process variables.

Haugan et al. study non-intrusive methods for diagnosis of a topside three-phase separators [72]. They conclude that gamma transmission and passive acoustic measurements can be used to detect defects in the cyclone, presence of foreign particles and clogging of the separator. However, they also note that the simplified experimental setup results in unrealistic process conditions. In practice, the monitoring might not be as straightforward.

Various equipment

- Ang et al. list most failure modes for electrical subsea equipment, including transformers, compressors, pumps, electrical motors, control valves, valve actuators, wet mateable connectors, power cables and switchgear [13]. They also briefly list existing methods for condition monitoring. No models are presented, but the paper gives a good overview of failure modes that should be considered.
- Nguyen et al. discuss the RUL estimation of a stochastically deteriorating actuators [119]. They assume that the actuators are deteriorating due to random shocks with independent and identically distributed shock damage at discrete times, giving a Poisson process. They assess the RUL using a two-step technique based on particle filtering of the output measurements and Monte Carlo simulations for reliability assessment. It is assumed that the actuator degradation is proportional to the control action, meaning that the degradation can be observed without any additional measurements. [90]
- Jaoude, El-Tawil and coworkers studied the prognostics of offshore pipelines subject to fatigue damage [50, 79]. The damage was assumed to be caused by crack formation due to cyclic stress, and was estimated using a physics-based model [51].

Summary

Literature on subsea PHM focuses mainly on diagnosis and health monitoring of the equipment. Literature about prognostics and prognostic models can also be found, but it is less common.

2.1.3 Combining health prognostics and control

The task of this project is to combine health prognostics and control. Currently practice is to ignore the effects of aging, fatigue and damage in the control hierarchy

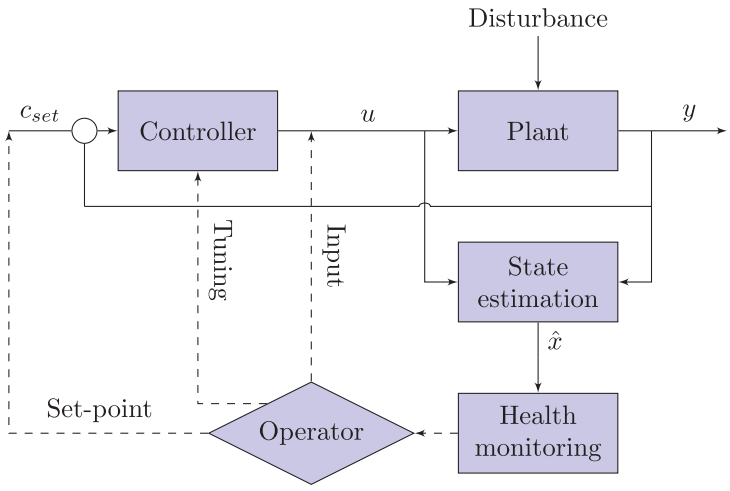

Figure 2.5: Block diagram for a system with health monitoring, but where the operator has to adjust set-points manually, based on information from the PHM system.

[56]. Data from PHM is only used for decision making by the operators. When the alarm goes off, the operators use their knowledge and experience to determine whether to adjust the set-points, schedule maintenance or whether it was a false alarm. A simplified block diagram is shown in Figure 2.5. Here, the operator can interact by adjusting the set-points or the controller tunings, or by overriding the control and set the inputs directly.

The disadvantage of leaving all the decision making to the operators is that there is delay due to the limited operator reaction time. More importantly, the decisions made the operator may not be optimal. By closing the loop and leaving the decision making to the controller, the response time could be significantly improved. The closed-loop system may also perform more optimally than the open-loop system. This requires the inclusion of PHM in the calculation of the optimal control trajectories. The reliability of the system is taken into account by introducing additional constraints in the optimal control problem. A simplified block diagram of the resulting system can be seen in Figure 2.6.

Fault-tolerant control

The possibility of combining control and diagnostics was first discussed in the late 70s / early 80s [38, 113, 114], mainly with application to control of airplanes. The idea was to include a supervisory layer which adjusts the control structure based on fault detection and identification (FDI) techniques. This way, the control structure would still perform satisfactory despite biased sensors or faulty actuators. Recovery of performance is only possible if the system has inherent actuator redundancies. For example, a pilot is able to maintain pitch control by using the ailerons if the

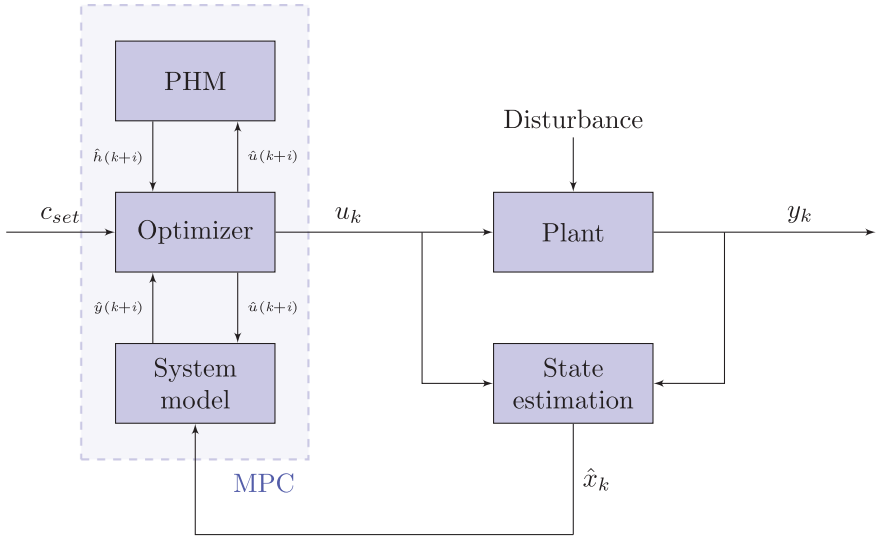

Figure 2.6: Block diagram for a system with health monitoring, where the PHM system has been integrated into the control system, so that it can operate autonomously. The operator is no longer actively in the loop, but is only required to supervise and monitor the total system performance.

elevator of his plane fails [97]. Chizeck and Willsky coined the term *fault-tolerant control* (FTC) to describe the resulting control structure. FTC has been a focus of research ever since. Not surprisingly, the aviation industry has been spearheading the research on FTC. The term automated contingency management (ACM) is also frequently used to describe FTC in the context of aviation [62, 170]. According to NASA [117], ACM is defined as:

> "Automated Contingency Management or ACM can be considered the ultimate technological goal of a health management system. The ability to confidently and autonomously adapt to fault conditions with the goal of still achieving mission objectives is a significant technical challenge that is dependent on the proper performance of several supporting technologies as well as the ACM system itself..."
> "...The ACM technology in general performs a multi-objective constrained optimization to accommodate impending failure conditions..."

An overview of existing efforts in the field of FTC/ACM can be found in the extensive review paper by Zhang & Jiang [194]. They list 376 references and divide them into classes depending on design approaches and field of application. Most of the listed references are dealing with fault-tolerant flight control systems.

Eterno et al. divide FTC systems into two classes: active and passive [57]. Passive FTC systems are robust towards disturbances and can tolerate actuator or sensor failures to some extent. Active FTC systems are able to detect failures and re-

configure the control structure on-line, such that control can be recovered and the system is able to continue operation with satisfactory performance.

- *Active fault-tolerant control systems*
  The aim of active FTC is to detect, identify and accommodate for faults in a system. When a fault occurs, the system either switches to a predetermined control structure [147], or calculates a new one on-line [109]. This requires real-time monitoring of the system. In most cases where FTC is employed, the available time for fault recovery is limited. It is therefore important that the fault detection, diagnosis and controller reconfiguration is quick and that the effective delay is small.

  Active FTC is sometimes described as adaptive, self-repairing, reconfigurable, restructurable, or self-designing [194]. According to Blanke et al., adaptive control has been shown to only work efficiently for linear systems [30, Ch. 1]. In practice, this requirement is seldom met, since faulty system often behave nonlinearly due to sudden parameter changes.
- *Passive fault-tolerant control systems*
  Passive FTC does not require fault detection, but relies instead on built-in robustness towards expected faults. This can be achieved with multi-objective optimization [97], quantitative feedback theory method [120], H_∞ control [96], and passivity-based cascade control [24], amongst other methods. Since passive FTC does not rely on fault-detection, it does not suffer from time delay, like active FTC. This means that the control system can react much faster to faults, which can potentially be the difference between a catastrophic failure and recovery. The disadvantage of passive FTC is that it is not necessarily robust towards unexpected disturbances or faults. Furthermore, passive FTC systems are sub-optimal under nominal conditions due to the trade-off between robustness and performance [30, Ch. 1].

In practice, a combination of passive and active FTC is used. The passive FTC ensures stability of the faulty system while the active FTC isolates the fault and synthesizes the appropriate post-fault control law [23].

Fault-tolerant control of subsea systems usually entails a so-called 2-out-of-3 sensor redundancy. This means that 3 independent sensors are used to measure the state of the system, and sensor bias or failure is accounted for by having a logic solver chose the two remaining sensors. Hardware redundancy is also reported. McLin proposes a system with two hot swappable subsea eletronic modules (SEMs) [106]. However, reconfigurable control structures for subsea systems have not been reported in literature.

Health-aware control

Escobet points out that traditional approaches to FTC only adjust the controller when a fault has occurred, but not when the system is in a non-faulty state [56]. This is not in line with industry needs, who are more concerned with avoiding faults altogether, argues Escobet. They propose a framework for so-called *health-aware control*, in which the strengths of PHM and control are combined. By considering

the system health in the control objective, fault-free operation can be ensured. The proposed framework was implemented to find the optimal trade-off between system reliability and power production of a wind turbine in [152].

A similar idea was previously discussed by Pereira et al., who propose an MPC which includes PHM information[134]. They derive a control structure which distributes the control efforts among redundant actuators to keep the actuator degradation above a safe threshold. Several simplifying assumptions are made, such as assuming that the system is linear and that the actuator degradation is proportional to the exerted effort and its first derivative.

Langeron et al. propose a framework for health-aware control of actuators subject to stochastic actuator degradation [90]. By combining the deterministic dynamics and the stochastic degradation, they derive a control structure which can extend its RUL. The issue with the proposed method is that it is assumed that the deterioration is measurable, which might not be the case.

In a very recent paper by Salazar et al., an MPC framework for system reliability is presented [151]. They model the reliability of the system using a data-driven approach with DBNs. As an example, they show how the framework can be used to control a drinking water network subject to pump failure.

A common issue in health-aware control is that the deterioration and the control take place on very different time scales. While control is performed on a timescale varying from milliseconds to minutes, the deterioration of the system typically happens on a timescale of days to years. This time horizon is too long for optimization, meaning that the problem has to be split into two different layers. The upper layer optimizes the set-points for the lower layer, while the lower layer implements the control. Consequently, it makes more sense to speak of health-aware real-time optimization (HARTO) than HAC.

Summary

No applications of health-aware control have been reported in literature for subsea systems. The literature on fault-tolerant control is abundant, but mostly with application to aviation.

2.2 Industrial status

This section gives an overview of the industrial status on the topic of prognostics and control for subsea systems.

2.2.1 Challenges facing the subsea oil & gas industry today

Control valves

Control valves are typically designed to last approximately 1 million cycles, whereas gate valves typically last approximately 1000 cycles [55]. The maturity of subsea control valves is low, and only a few valves have been qualified [112, 127]. Subsea

control valves by Mokveld are currently in use on the Ormen Lange and Åsgard fields.

Sand erosion can speed up this process significantly, choke erosion can be a large problem on certain fields. On other fields, such as the Troll field, chokes have been in use for over 20 years without degrading to failure [3]. The huge differences in erosion conditions from field to field mean that different strategies must be applied on a case-to-case basis.

Other issues that influence performance of valves and chokes are signal faults, stiction and clogging [78]. The valve performance is determined by comparing the actual pressure drop, valve opening and flow rate to the expected values.

Compressors

Compressors are large consumers of energy, so efficient and economical operation is very important [78]. Performance monitoring of compressors is required to detect decreases in efficiency, but also to detect mechanical wear, corrosion, leakage or electrical failures. Vibration measurements can be used to detect issues with the compressor, but performance monitoring could detect issues before they occur [78]. Compressor charts are used to relate the flow-rate and pressure ratio to the compressor efficiency. Currently there exist methods to estimate the compressor head, which can be used to calculate the efficiency. However, although the mechanisms which cause the decrease in head are known (e.g. liquid content in the gas), they are not easily measurable [55].

The control and safety systems of the qualification prototype of the Ormen Lange compression station are presented in [52]. Though no CBM routine is in place yet, some performance indicators are measured and could be used for CBM in the future. Specifically, actuators, magnetic bearings and compressor efficiencies were monitored. The travel time and power consumption of all actuators was measured and used to predict the condition of the valve. An increase in travel time or power consumption could indicate valve stiction. It is also possible to estimate the wear of the valve using mechanistic models based on the overall travel distance of the valve stem. Electromagnetic bearings are used to hold the shaft of the compressor in place. If the compressor blades are subject to erosion, or the axle somehow develops an unbalance, this can be detected from the increased power consumption of the bearings [53]. It is also possible to monitor the compressor, pump and variable speed drive efficiencies based on the knowledge of the power input and estimation of the compressor work. Using pressure, temperature and flow measurements, the compressor work can be estimated fairly accurately.

Heat exchangers

Heat exchangers are subject to fouling and particle deposition. This reduces the heat transfer and increases the pressure drop. Models of the heat exchangers are used to compare the actual performance to the expected performance. If the deviation becomes too large, maintenance might be necessary. One issue facing industry

is how to accurately model heat exchangers and how to quantify model uncertainty [49].

Pumps

Similarly as for compressors, vibration measurements can be used to monitor the health of pumps. The actual performance of the pump is measured or estimated and compared to the predicted performance of the pump. Model uncertainty and measurement uncertainty are again big issues in industry [49]. The consumption of barrier fluid is also a good indicator of pump performance, as increased consumption indicates a leak in the pump module. If the leak is too large, the pump must be replaced. By monitoring the barrier fluid consumption, the RUL can be predicted [54]. Lube oil consumption is also indicative of the health of the pump module. Large consumption of lube oil is caused by leakage due to worn out pump seals [53].

Typically, the wear of pumps and compressors scales cubically with the effort it expends. This means that a pump will only last $\frac{1}{8}^{\text{th}}$ the time if run at twice the power, but up to 8 times longer if run at half power [53].

Summary

The main challenges facing industry today are the lack of detailed predictive models, and the physical limitations of the equipment. The limited number of strokes on control valves can especially pose a problem for automatic control.

2.2.2 Prognostics and health monitoring

The area of prognostics and health monitoring of subsea systems is relatively new, in comparison with more established industries such as rail and energy. However, industrial interest in the topic is large, due to the importance of reliable operation. Below follows a selection of companies which offer condition monitoring solutions of various kinds. The list is not exhaustive.

- GE/NAXYS offers solutions for condition monitoring, including vibration analysis, speed tracking and leak detection [118].
- FMC implemented the worlds first subsea condition and performance monitoring system at the Gjøa field in 2012 [59]. Their system includes surveilance of sensors, multiphase flow meters, chokes and actuators, in addition to leak detection [165]. Their FlowManager tool for flow assurance also contains tools for monitoring erosion, corrosion and wax deposition along the pipeline [60].
- Aker provides Insight, a tool for online erosion monitoring [1]. Insight is currently in use in the Gullfaks and Statfjord fields [12].

For many applications, tailored CBM systems have been developed. Often these systems are proprietary and not published. Aker has shared their experience from developing the Ormen Lange and Åsgard subsea compression station in a number of publications, see for instance [52, 53].

2.2.3 Health-aware control

Currently, there is no systematic framework to combine condition monitoring and optimal operation for subsea systems [55].

2.3 Gap analysis

Based on the academic status from Section 2.1 and the industrial status from Section 2.2, the following gaps are identified.

2.3.1 Academic

There is need for further development of health-aware control. Most literature on the field is published in the the last 6 years, and only a handful of papers have been published.

There is also a knowledge gap when it comes to deriving first-principles models for failure mechanism for all kinds of subsea equipment. Currently, the most common approach to prognosis is using data-driven models. This can be a critical issue, as there is little run-to-failure data available for subsea equipment.

2.3.2 Industrial

There is currently no systematic way of combining health monitoring and control. This is the largest industrial gap identified.

A challenge associated with implementing health-aware control, or any automated control for that matter, is the limited number of control moves in the lifetime of a valve. As mentioned previously, control valves are typically designed to last 1 million moves. This can be imposed as a constraint in the optimization problem (so-called move blocking [33]). The time scale separation between the layers two layers in the HARTO also poses a problem. The optimal allocation of control moves over the entire lifetime must be determined in the upper layer and added as a constraint in the lower layer.

Chapter 3

Framework for combined diagnostics, prognostics and optimal operation of a subsea gas compression system

This chapter is from the article

- A. Verheyleweghen and J. Jäschke. Framework for combined diagnostics, prognostics and optimal operation of a subsea gas compression system (IFAC World Congress, Toulouse, France). *IFAC-PapersOnLine*. volume 50, pages 15916—15921, 2017

3.1 Abstract

The efficient and safe operation of subsea gas and oil production systems sets strict requirements to equipment reliability to avoid unplanned breakdowns and costly maintenance interventions. Because of this, condition monitoring is employed to assess the status of the system in real-time. However, the condition of the system is usually not considered explicitly when finding the optimal operation strategy. Instead, operational constraints on flow rates, pressures etc., based on worst-case scenarios, are imposed. This can lead to unnecessarily restrained operation and significant economic losses. To avoid sub-optimal operation, we propose to integrate diagnostics and prognostics with the optimal decision making process for operation to obtain an operational strategy which is optimal subject to the expected system degradation. This allows us to proactively steer the system degradation, rather than simply reacting to it. We use the operation of a subsea gas compressor subject to bearing degradation as a case example.

3.2 Introduction

Subsea processing is an enabling technology for fields that were previously deemed too remote, too deep or far away from existing infrastructure. However, several industrial challenges arise when moving topside equipment to the seabed. One of the potentially most prohibitive challenges is the inaccessibility of the plant for large parts of the year, and the need for specialized intervention ships. Consequently, unplanned shut-downs can be very costly and must be avoided as far as possible. In order to achieve this, strict reliability constraints are imposed on design and operation of the plant. While these safety margins provide a method to ensure reliable operation, they might be overly restrictive. One reason for this is because the information from the health monitoring system is often not utilized directly in the decision making process. Instead, a "worst-case" approach is often used to determine production set-points.

In this paper we propose a method for integrating health monitoring, prognostics and control to obtain an operational strategy that ensures maximum economic profit without jeopardizing the plant reliability. In particular, we include a health degradation model in our optimization routine, resulting in a model-predictive control (MPC)-like framework where we impose constraints on the remaining useful life (RUL) of the equipment.

MPC has gained increasing popularity in industry in recent years due to its ability to deal with constrained, multivariate, and nonlinear control problems, is based on the repeated optimization of the objective function, subject to constraints [115]. The first input of the optimized input trajectory is implemented in the plant, before new measurements are taken and the model is re-optimized.

The concept of health-aware control has been investigated by a few authors in recent years. The term "health-aware control" itself was first used by [56] to describe a control structure which through the combination of prognostics and health monitoring (PHM) and feedback control simultaneously can fulfill the control objectives and extend the component RUL. The method was applied to a conveyor belt system, and later to wind turbines [152]. Similar ideas of combining PHM and MPC were previously discussed by [134] and [151], with application to control effort distribution, and pumps in drinking water networks, respectively.

3.2.1 General description of framework

In this paper we propose an integrated framework for combining diagnostics, prognostics, and optimal operation using MPC. Our framework contains the following steps:

In the following sections, we will show how these steps can be applied to the case of a subsea compression station. Following a description of the process in Section 3.3, we cover step 1 and 2 in Section 3.4.1, step 3 in Section 3.4.2, and step 4 and 5 in Section 3.5. The results from the case study are presented in Section 3.6.

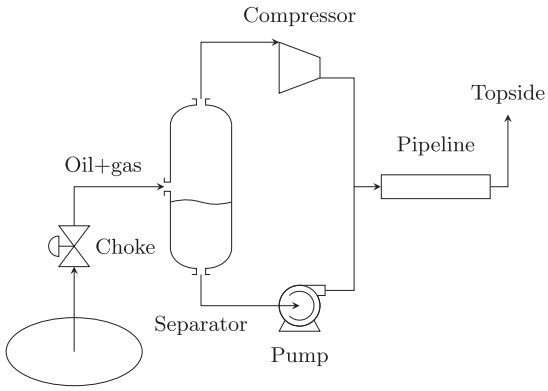

Figure 3.1: Subsea gas compression station

3.3 Process description

Our case study is a subsea gas compression station, similar to installations on the Åsgard field and the Ormen Lange pilot. The purpose of the gas compression station is to boost the pressure of the stream so that it is sufficiently high to overcome the pressure drop in the transportation pipeline and arrive at the receiving facility topside with the desired outlet pressure. A multiphase boosting pump could be used for this purpose, but since the maturity level of the technology is limited, it is chosen to split the well stream into its gas and liquid parts before increasing the pressure of each individual stream. An illustration of the process is shown in Figure 3.1.

The system consists of a well choke with which the flow of the hydrocarbons from the reservoir can be controlled. From the reservoir, the stream enters a gas-liquid separator, whose purpose it is to separate the gas from the oil and water. Due to imperfect separation, liquid droplets can be carried over to the gas outlet of the separator. The separator efficiency is modelled as a function of the gas velocity and the average fluid density [16]. The pressure of the liquid outlet is boosted by a pump before being recombined with the gas. Meanwhile, the pressure of the gas outlet is increased in a compressor. The compressor is modelled as a wet-gas compressor which can handle moderate amounts of liquid carry-over [4]. Suction gas-volume-fractions of 0.95 to 1.0 can be tolerated at the compressor inlet.

3.4 Diagnostic and prognostic modelling

Diagnostics and prognostics form the backbone of any PHM system [74]. In order to make meaningful decisions about future production, it is not only necessary to know what the health state of the equipment is at the current time, but one must also be able to predict how the condition of the equipment will develop in the future. Diagnostics is about the detection and monitoring of faults, whereas prognostics is about the prediction of health evolution and estimation of equipment

RUL.

Prognostics and diagnostics of a large system such as a gas compression station is a challenging task, due to the high complexity and large number of components. Condition monitoring systems should be able to detect a wide range of faults, including everything from signal failure to external impact of foreign objects. A variety of methods are used to monitor subsea production systems in industry. For example, sand erosion and corrosion rates are monitored in vulnerable parts of the pipeline, such as in bends. Erosion and corrosion rates are estimated through measurements of electrical resistance or by periodic inspection of coupons. Detection of leaks is also an important topic. Leaks are usually monitored through a combination of visual surveillance, electrical resistance measurements of the seawater, and temperature/pressure measurements of seals.

In order to limit the scope of the remainder of the paper, we make the simplifying assumption that only the most crucial faults of the system need to be considered. It is known that rotating machinery such as compressors and pumps are prone to faults due to their many moving parts and mechanical complexity [74]. This means that for the studied process, the compressor, the pump and the well choke need to be monitored closely due to the relatively high likelihood of critical faults occurring here.

Furthermore, in this paper, we exclude failures which cannot be influenced directly by manipulation of the inputs. This excludes a large number of important fault modes. Since the purpose of this paper is to combine condition monitoring and control, we chose to neglect faults which are independent of operational decisions for now. These kinds of faults will have to be addressed in future work.

3.4.1 Diagnostics

Vibration monitoring of rotating machinery is commonly used to assess their health. Imbalance caused by the onset of a fault will result in a periodic force with a characteristic periodicity and magnitude, which can be detected as vibrations. This technique can be used to detect defects on the shaft, bearings and impeller blades. Current subsea gas compression stations use magnetic bearings to stabilize the impeller, but since this technology is relatively new, not many degradation models are available in the open literature. Ball bearings, on the other hand, are widely used for a multitude of applications, including on-shore gas compressors. We will therefore use the case of a subsea compressor with ball bearings in this paper to demonstrate our framework.

Ball bearings, which are commonly found in pumps and compressors, are subject to large stresses due to their constant load and high rotational speed. At the same time, their survival is crucial for the operation of the machine. Therefore, condition monitoring of bearings is important to ensure high availability of the pump or compressor. The inner workings of a ball bearing are shown in Figure 3.2.

For a full overview of bearing fault diagnostics, see e.g. [186]. To make this paper self-contained, we give a short summary below. A surface defect on a bearing results

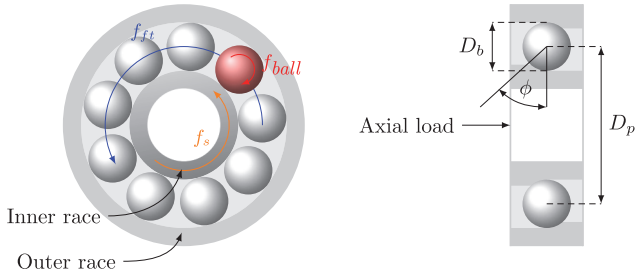

Figure 3.2: Illustration of a ball bearing

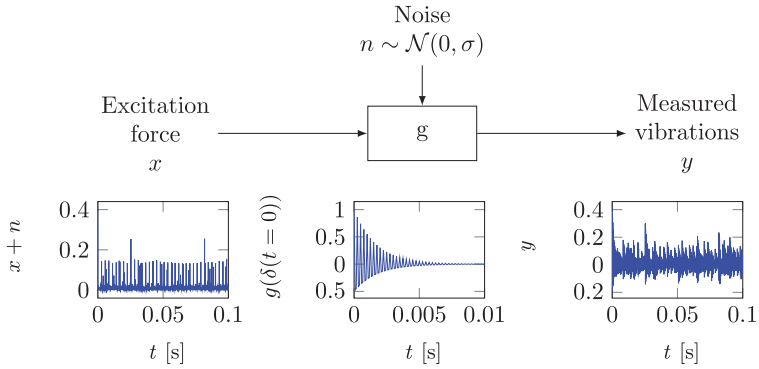

Figure 3.3: Signal path from excitation force x to measured vibrations y, via impulse response model g.

in a periodic excitation force with characteristic frequency f_{fault}. The excitation force can be described as an impulse train, and the severity of the fault can be estimated by looking at the magnitudes of the impulses. In addition to the periodic impulses, random vibrations (noise) act on the bearing. The sum of these periodic impulses and the random noise is shown on the left in Figure 3.3. The resulting force is modulated by the impulse response function of the equipment to create the measured vibrations. This impulse response can be seen as the modulation of the original signal from the fault location to the vibration sensor, due to e.g. the resonance vibrations of the bearing housing. The impulse response model g is a damped harmonic oscillator, as illustrated in the middle plot in Figure 3.3. Finally, the measured vibrations, i.e. the modulated signal, are shown on the right in Figure 3.3.

The fault frequency f_{fault} is dependent on the location of the fault and the specific geometry of the bearing, but is ultimately a function of the shaft frequency f_s. Let us first define the fundamental train frequency f_{ft} as

$$f_{ft} = \frac{f_s}{2}\left[1 - \left(\frac{D_p}{D_b}\right)^{-1}\cos(\phi)\right]. \tag{3.1}$$

In the above expression, D_p is the pitch diameter, D_b is the ball diameter, and ϕ is the contact angle. ϕ is the angle between the raceway and the ball, which is larger than 0 for bearings with axial loads. See [186] for details.

For an inner race (IR) fault, the fault frequency is then

$$f_{IR\ fault} = n_b \cdot (f_s - f_{ft})\,. \tag{3.2}$$

Similarly, for outer race (OR) and rolling element (RE) failures, the fault frequencies are

$$f_{OR\ fault} = n_b \cdot f_{ft} \tag{3.3}$$

and

$$f_{RE\ fault} = \frac{f_s}{2}\left(\frac{D_p}{D_b}\right)\left[1 - \left(\frac{D_p}{D_b}\right)^{-2}\cos(\phi)\right], \tag{3.4}$$

respectively, where n_b is the number of balls.

Furthermore, the amplitude of the excitation force is modulated by a sine wave with characteristic periodicity depending on the transmission path and the loading conditions of the bearing. For instance, under stationary loading an OR fault will be without periodicity, while the amplitudes of the impulse train in an IR fault will have periodicity f_s due to the varying distance to the vibration sensor. See e.g. [186] for a full overview of the periodic characteristics of the faults.

Knowing how the vibration signal is created, we can now take the reverse path to recover the original fault-induced impulse train x from the vibration measurements y by demodulating the signal. Assuming an estimate of g can be found experimentally, the demodulation is performed by solving

$$\bar{x} = G^{-1}y, \tag{3.5}$$

where G is the Toeplitz convolution matrix of g.

From the estimated excitation force $\bar{x}$, the original fault-induced impulse train can be recovered by removing the additive noise. A Wiener filter can be used for this purpose if the signal-to-noise ratio is known from experiments. Alternatively, the properties of the Wiener filter can be identified blindly by maximizing the spectral kurtosis (fourth moment) of the output of the filter [14]. In this work, we use a standard Wiener filter the signal-to-noise ratio assumed to be known.

3.4.2 Prognostics

A widely applied prognostic model for surface defects is Paris' crack propagation model [130], which states that the crack length a will develop according to

$$\frac{da}{dn_{cycles}} = D \cdot (\Delta K)^n\,, \tag{3.6}$$

where n_{cycles} is the number of cycles, D is a material constant, ΔK is the range of strain and n is an exponent. In the case of bearing faults, Paris' law can be

Figure 3.4: Illustration of estimated degradation model based on "measurements" of the crack length (which themselves are estimated from vibrational data)

reformulated as

$$\frac{da}{dt} = c_{Paris} \cdot \left(T^2 \cdot f_s\right) = c_{Paris} \cdot \left(\frac{P^2}{f_s}\right), \tag{3.7}$$

by assuming that the motor torque can be used as a health indicator for gross strain [20]. In the above equation, c_{Paris} is a lumped parameter, T is the motor torque and P is the motor power.

The true value of c_{Paris} is not known exactly, so c_{Paris} must be estimated from past measurements. A moving horizon estimator is used for this purpose. The "measurements" utilized in this case are the estimated crack lengths based on the past vibrational data.

Confidence limits for the parameters are obtained from the covariance matrix of the parameter estimation [100], which in turn can be used to predict the RUL distribution with Monte Carlo sampling. An illustration can be seen in Figure 3.4.

3.5 Optimizing economic performance subject to health constraints

The estimated system health and the health degradation model can now be integrated in the decision making process by imposing constraints relating to the maximum allowable degradation. The optimal control problem (OCP) can then be solved with state-of-the art nonlinear programming (NLP) solvers such as IPOPT [185]. Since information about the probability distribution of the parameter estimates is available, this should be embedded in the optimization problem to obtain a better solution. This gives rise to a stochastic NLP due to uncertainty in the parameter values, which can be written as

$$\min_{\mathbf{x}_k, \mathbf{u}_k} \sum_{k=1}^{N} \phi\left(\mathbf{x}_k, \mathbf{u}_k, \pi\right) \tag{3.8a}$$

$$\text{s.t.} \quad f(\mathbf{x}_k, \mathbf{u}_k, \pi) \leq 0 \quad \forall k = 1...N \tag{3.8b}$$

$$g(\mathbf{x}_k, \mathbf{u}_k, \pi) = 0 \quad \forall k = 1...N \tag{3.8c}$$

where $\mathbf{x}$ are the states, $\mathbf{u}$ are the inputs, π are the stochastic parameters, N is the horizon length, ϕ is the objective function, f and g are the inequality- and equality constraints, respectively. Since π is continuously distributed, finding the analytic solution of the resulting infinite-dimensional optimization problem maybe impossible. Listed below are three methods to deal with this.

1. The most naive approach for solving the stochastic problem is to substitute all uncertain parameters by their expected values. The solution obtained

Figure 3.5: Scenario tree with robust horizon $N = n$ and prediction horizon $S = 4$.

through this approach is likely to be sub-optimal, or even infeasible in the case where some constraints are active.

2. Another approach is to substitute the uncertain parameters by their worst-case realizations. The rationale is that if the solution holds for the worst-case scenario, it should hold for any scenario. This approach is also known as the min-max approach in literature. While it works for many practical applications, the min-max approach usually results in a very conservative, possibly even infeasible, solution.
3. A third approach is to employ a scenario-based method to explicitly deal with the parametric uncertainty. The idea stems from multistage stochastic programming, in which the uncertainty is discretized into a finite number of possible realizations, subject to which the optimization must be performed. Since the possibility of future recourse is taken explicitly into consideration, this approach is usually less conservative than the min-max approach. Due to the repeated measurement updates and input adjustments, stochastic optimal control problems are well suited to be solved with a scenario-based approach.

In this paper, we will use the scenario-based approach to solve the stochastic problem. The scenario-based deterministic equivalent of the stochastic OCP reads as

$$\min_{\mathbf{x}_{i,k},\mathbf{u}_{i,k}} \sum_{i=1}^{S} p_i \sum_{k=1}^{N} \phi\left(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}\right) \tag{3.9a}$$

$$\text{s.t.} \quad f(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}) \leq 0 \quad \forall i = 1...S\,,\ k = 1...N \tag{3.9b}$$

$$g(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}) = 0 \quad \forall i = 1...S\,,\ k = 1...N \tag{3.9c}$$

$$\sum_{i=1}^{S} \mathbf{A}_{i,k}\mathbf{u}_{i,k} = 0 \quad \forall k = 1...N \tag{3.9d}$$

where S is the number of scenarios p_i is the probability associated with scenario i and $\mathbf{A}$ are the non-anticipativity constraints. Figure 3.5 shows a scenario tree corresponding to $N = n$ and $S = 4$.

In order to reduce the number of scenarios, and thus the size of the OCP, it is common to define a robust-horizon $N_{robust} < N$ [102]. The robust horizon is defined as the stage up until which branching occurs in the scenario tree. Since branching represents the availability of new information in the future, shortening the robust horizon means disregarding future state information. The justification for doing so is that additional branching at later stages results in a much larger dimensionality of the NLP, with little improvement in the objective function.

Creating a scenario tree which captures the true nature of the uncertainty is a difficult task in and of itself, but is out of the scope of this paper. On one hand, the scenario tree should as detailed as possible to be a good approximation of the

probability distribution. On the other hand, the scenario tree should be as small as possible due to the curse of dimensionality. We refer the interested reader to [48]. In the current work, we do as proposed by [103], which is to generate the scenario tree by using combinations of the maximum, minimum, and the nominal uncertain parameters, as identified in the parameter estimation step.

3.6 Results

The proposed framework for integrating diagnostics, prognostics and control was applied to the subsea compression system. The goal is to optimize production while making sure that the wet-gas compressor remains operational until the time scheduled for maintenance. In our case study, a maintenance stop is planned after 5 years after initial startup. An outer race bearing fault was simulated under stationary compressor loading. The fault was initiated at time $t = 0$, with an initial crack length of 0.01 mm. The degradation threshold is defined as the time when the crack length exceeds 1mm. The ball bearing consists of 10 rolling elements, and has a $\frac{D_p}{D_b}$-ratio of 5.45. The compressor runs at a nominal speed of 60 Hz, with an operating window from 45-63 Hz. The operational objective is to maximize the net present value (NPV) of the gas production. Additionally, excessive control movement is penalized to avoid oscillatory or jumping solutions. The health-aware control problem reads

$$\min_{\mathbf{x}_{i,k},\mathbf{u}_{i,k}} \sum_{i=1}^{S=3} p_i \sum_{k=1}^{N=20} \left(-\frac{\dot{m}_{gas_{i,k}}}{(1+r)^{t_k}} + w\Delta\mathbf{u}_{i,k}^2 \right) \tag{3.10a}$$

where the discount factor $r = 0.015$, and the control movement penalty $w \geq 0$. We chose $w = 100$, resulting in approximately twice as much weight on the gas production term as on the control penalty term. The constraints are

$$f(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}) \leq 0 \quad \forall i = 1...3\,,\ k = 1...20 \tag{3.10b}$$

$$g(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}) = 0 \quad \forall i = 1...3\,,\ k = 1...20 \tag{3.10c}$$

$$\mathbf{u}_{i=1,k=1} = \mathbf{u}_{i=2,k=1} = \mathbf{u}_{i=3,k=1} \tag{3.10d}$$

The constraints f contain upper and lower bounds on the inputs $\mathbf{u}$ (choke opening $0 < \mathbf{u}_{choke} < 1$ and compressor speed $45 < \mathbf{u}_{compressor} < 63$ Hz), as well as constraints relating to the allowable operating region, i.e. to prevent compressor surge and compressor choke. A minimum discharge pressure of $P_{discharge} = 150$ bar is imposed after the compressor to ensure flow through the long pipeline to the topside, as well.

The uncertainty in the parameter c_{Paris} in the crack propagation model, Equation (3.7), is included in the problem formulation through the three scenarios. Each scenario represents a discrete realization of c_{Paris}, namely the 5% percentile, the 95% percentile and the expected value. A robust horizon of length $N_{robust} = 1$ is used, making the problem effectively a two-stage problem from a stochastic programming perspective.

The OCP described in Equation (3.10) is solved repeatedly in a shrinking horizon fashion. That is, the OCP is solved with the initial values for the states being the latest state estimates from the plant. After a solution is obtained, only the inputs corresponding to the current time step are implemented. This is also illustrated in the flow diagram in Section 3.2.1.

Between each optimization step, the vibration measurements are added by generating an impulse train for the excitation force and adding white noise. The force is modulated through the impulse response model of the compressor. We assume that g from Section 3.4.1 can be written as a damped sinusoid

$$g = \exp(-\lambda t) \cdot \cos(\omega t) \tag{3.11}$$

with decay factor $\lambda = 600\ \mathrm{s}^{-1}$ and frequency $\omega = 4$ kHz, resulting in the impulse response model shown in the middle graph in Figure 3.3. The magnitude of the original impulse is found by the method described in Section 3.4.1. Unfortunately, no real vibration data was available, so we used the same model for both generating the measurements and estimating the states. However, due to the added noise, the estimates were not perfect and the method can still be used to showcase the approach.

The closed-loop solution of the health-aware controller with a fixed maintenance horizon of 5 years is shown in Figure 3.6. The figure shows the evolution of the past inputs and the states and the optimal trajectories for the three scenarios at three different points in time, at $t = 0$, $t = 2$ and $t = 2.8$ years.

3.7 Discussion

The success of our proposed approach hinges on the quality of the degradation model and condition monitoring capabilities. Our approach relies on the equipment vendors to provide models and data for performing the diagnostics and prognostics. Since the objective of this paper is to demonstrate our framework, we have chosen to use a relative simple degradation and diagnostics model, that we adapted to our purposes.

From Figure 3.6, it can be seen that the predicted trajectories differ from the real trajectories. This is due to the NPV-term in the cost function and the presence of uncertainty. The closed-loop solution has two different operating regions, the first from $t = 0$ to $t = 2.5$ years where the compressor runs with maximum speed, and the second from $t = 2.5$ to $t = 5$ years where production has to be choked back in order to meet the required bearing health constraint. The abrupt change between operating regions occurs because of the NPV term. Since future production is valued less than present production, the optimizer will attempt to keep production at maximum as long as possible. In the second operating region, gas production is lower in order to meet the outlet pressure constraint and the minimum health requirement.

Overall, the results are as expected. It seems reasonable to require that the subsea installation is as profitable as possible while operating, i.e. that production

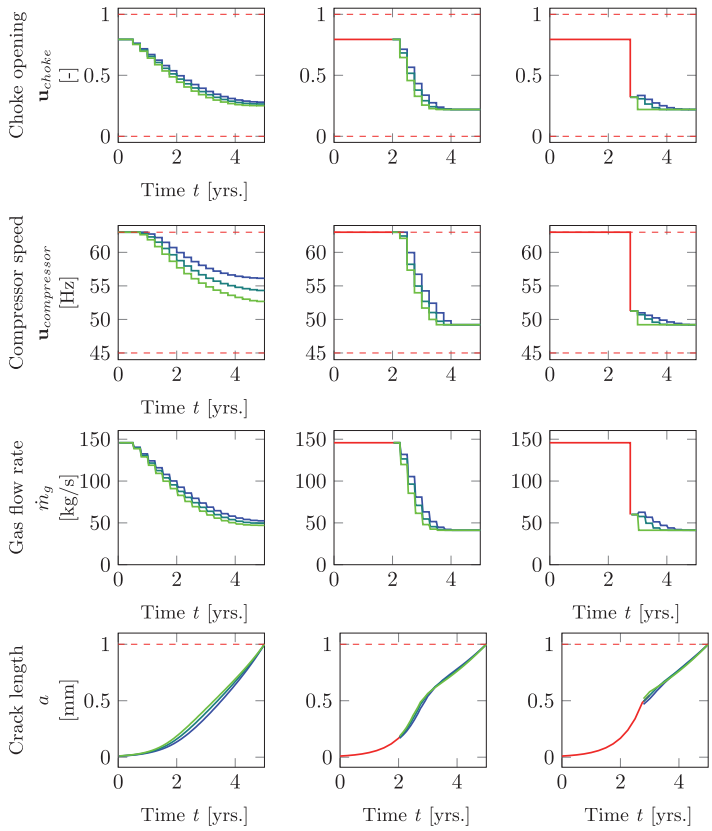

Figure 3.6: Three snapshots of the closed-loop solution at $t = 0$, $t = 2$ and $t = 2.8$ yrs. The blue, turquoise and green scenarios are for the low, expected and high realizations of the stochastic parameter c_{Paris}, respectively. The maintenance horizon is fixed at 5 years.

is at maximum. However, the fact that the production has to be throttled down after a while indicates that the specified maintenance horizon of 5 years may have been too long. Preferably, maintenance intervention should have been scheduled earlier, so that maximum throughput could have been achieved the entire time. In future work, we will consider the possibility of adapting the maintenance time during operation to make sure that maximum throughput can be achieved. In this case maintenance is scheduled a year in advance, and will be decided by the optimizer. Nevertheless, the control structure successfully meets the constraints while maximizing the production.

3.8 Conclusion and future work

We presented a framework to combine diagnostics, prognostics and control of a subsea gas compression plant subject to compressor bearing failure. By including measurements of fault indicators and fault prognostic models in the MPC framework, we can ensure that the operation is both economically optimal and safe. In the case of bearings, vibration measurements can be used to detect- and estimate the severity of faults. Paris' law for crack propagation can be used to predict fault development.

In future work, we will consider multiple failure mechanisms, not only those which can only be influenced through input manipulation. For example, the production strategy will look different when a seal fault has been detected and failure is eminent. Careful operation in order to "save" the bearings will be suboptimal, since intervention is required in the near future to replace the faulty seal. In similar vein, we will look at the entire subsea plant as a whole, to ensure that operation is optimal not only for a single unit (e.g. the compressor), but for all units in the plant.

Chapter 4

Combined Reliability and Optimal Operation: Application to an LNG Liquefaction Plant

This chapter is from the article

- A. Verheyleweghen and J. Jäschke. Combined Reliability and Optimal Operation: Application to an LNG Liquefaction Plant. *Journal of Process Control*. (Under review), 2020

4.1 Abstract

Unplanned process shutdowns due to equipment failure are costly and time consuming. Therefore, conservative operational strategies are often adapted to prevent these unplanned shutdowns from happening, which leads to suboptimal performance and lost revenues. To reduce the operational conservatism, we propose a framework for health-aware operation in this paper. The main idea is to reduce the conservatism by integrating diagnosis and prognosis of critical equipment, into process operation decisions without compromising process safety. We propose a step by step approach for designing a control structure for health-aware operation.

As a case example, we study the operation of a cascaded LNG refrigeration plant, where the refrigerant compressors are subject to load-induced bearing degradation. The designed control structure consists of two new decision making layers in addition to the stabilizing control layer: a top level real-time optimization layer where the aim is to keep bearing degradation below a threshold, and a lower level model-predictive control layer that rejects uncertainties and disturbances. In our approach, the constraints on the operation are updated in real-time to reflect the current reliability state of the plant. This allows us to proactively steer the system degradation, rather than simply reacting to it. As a result, we are able to operate less conservatively and use our equipment to its full potential.

4.2 Introduction

In parallel with the increase in complexity of industrial processes, the optimal operation of these industrial processes has become more and more difficult to define. In today's chemical plants, the decision making process regarding production, operation and maintenance decisions is typically decentralized and left up to the individual divisions. In particular

> *Culture has slowly been replaced with a new mindset about the appropriate division of labor — that "operations personnel run the equipment" and "maintenance personnel fix the equipment"* [146]

This can lead to a slow and suboptimal decision-making process. Long-term targets and strategies are defined first, imposing conservative constraints on the plant reliability and maintenance scheduling, which in turn impose constraints on day-to-day production and control. Such a sequential top-to-bottom decision making process will typically be very conservative, due to the inclusion of safety margins and back-offs. These margins and delays are especially necessary when communication and exchange of information between the different decision-makers is poor or non-existing. Due to the back-offs from the optimal operating point, the plant is operated sub-optimally. An illustration of this decentralized decision-making process is shown on the left in Figure 4.1.

Efforts have been made to reduce the sub-optimality stemming from decentralized decision-making. So-called operator-driven reliability (ODR) has existed for a while [184], and aims to blur the lines between maintenance decisions and operational decisions by tasking the operators with basic maintenance tasks, thereby challenging the operator to adopt a more "reliability-centered operations" mindset. However, industrial adoption of the practice has been slow. Lack of operator training and experience, unclear goals and responsibilities, and lack of clear key performance indicators (KPIs) is among the main reasons why ODR has not found footing in industry [8]. We seek to address some of these issues by introducing a systematic framework within which reliability and control decisions can be effectively made, and we propose to use optimization and automation to lessen the operator load.

4.2.1 Summary of our proposed approach

The main aim of this paper is to provide a systematic framework for developing decision making strategies which integrate reliability- and maintenance-based decisions, control- and process optimization-based decisions. Let us first establish how our proposed decision making strategy differs from a traditional approach.

In the traditional operation approach (Figure 4.1 left), there are two hierarchical decision structures, which are usually decoupled from one another; one for maintenance decisions, and one for operational and control decisions. In this paper, we propose to connect the maintenance decisions and the layers of the hierarchical control structure, to obtain an integrated health-aware control system (Figure 4.1 right). The idea is that the decision layers are horizontally interconnected, such

that the control structure has access to the equipment monitoring data and take this information into account to make overall optimal operation decisions.

Note that to a large degree, the previously discussed conservatism can be attributed to time scale separation. In the traditional approach (Figure 4.1 left) only the top layers of the control and maintenance hierarchies communicate. In this setting, conservatism is necessary because of the dynamics and the time delay in the lower layers. Decisions and information from the control domain must first go through the operator at the topmost level of the hierarchy before entering the reliability and maintenance domain, or vice versa. Our proposed approach reduces this time delay by integrating the domains horizontally, making relevant information available automatically across the domains without the need for manual intervention, and also allowing decisions to be taken on the relevant time scale, without needing to be propagated all the way to the top. Note that in both cases, the safety systems are left unchanged, as safety systems are redundant and independent.

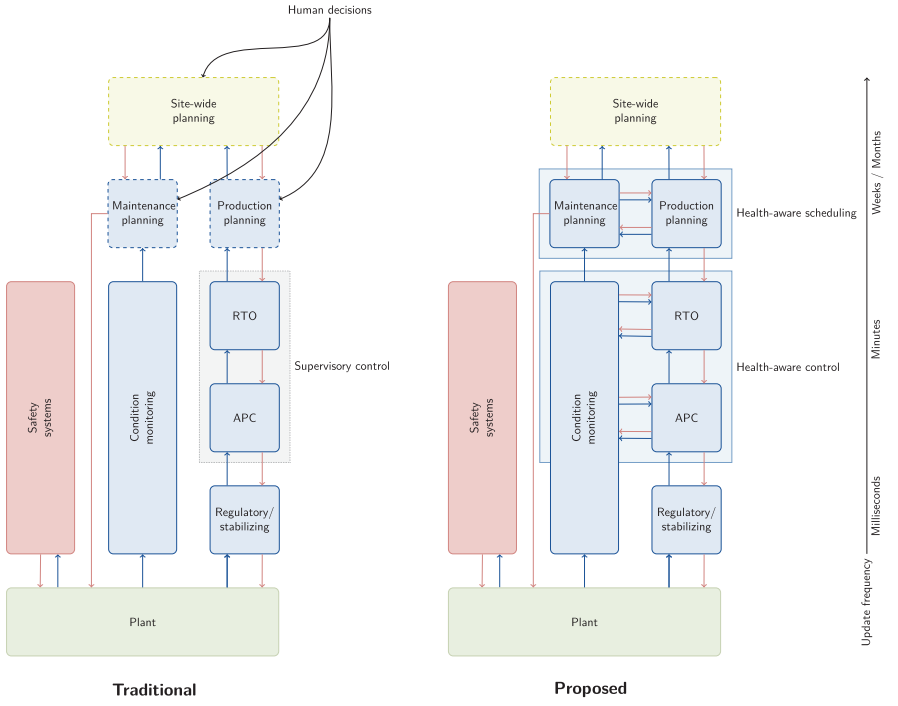

Figure 4.1: Traditional unconnected decision making hierarchies (left) and our proposed integrated approach (right). Blue arrows indicate the flow of information, whereas red arrows indicate the flow of decisions between the different blocks. Manual decision making levels are indicated by blocks with dashed outlines, as opposed to automatic decision making levels, which are indicated by blocks with a solid outline. Note that in both cases, the safety systems are left unchanged.

To set up the interconnected health-aware structure on Figure 1 right, and to reduce the conservatism of current industrial practice, we propose the following systematic approach:

Step 1: **Economic and process analysis:** Define desired plant behaviour as objectives and constraints. Objectives typically relate to the plant economics or the plant reliability, while constraints relate to the product quality, equipment safety or physics of the plant.

Outcome:

- Overall operational objectives and constraints

Step 2: **Defining sensible decision-making layers layers:** Identify which time-scales are involved, and divide the decision making problem into suitable, hierarchical layers (see e.g. the top-down approach in [163]).

Outcome:

- Time-scale separated decision layers
- Degrees of freedom (inputs and outputs) for the layers

Step 3: **Developing problem-specific / fit-for-purpose models for each of the layers:** For each layer, develop fit-for-purpose models of the plant with all important KPIs, for use in process monitoring, plant diagnostics and prognostics.

Outcome:

- Fit-for-purpose models for each layer

Step 4: **Optimization problem formulation and dealing with uncertainty**: Handle inevitable model uncertainty in a systematic manner, if necessary, by formulating a robust version of the decision making problem for the layers.

Outcome:

- Optimal operation as an optimization problem for each layer

Step 5: **Implementation** of the solution in the plant, and on line updating of the models when necessary. The on-line implementation can be broken down into the following sub-steps that are repeated performed

a) Data acquisition: collect measurements from the plant

b) Diagnostics: estimate current states

c) Prognostics: estimate prediction model

d) Optimal operation: use newest estimates to predict evolution of health and process states

Outcome from this layer:

- Implementation strategy for each layer (online optimization (real-time optimization, model predictive control), offline optimization (self-optimizing control)

4.2.2 Previous work

Other authors have addressed the issue of combined prognostic and diagnostic information into the decision making process. The descriptor "health-aware" itself was first introduced in [56] to describe a control structure which satisfies control objectives while extending the remaining useful life of the system, by combining prognostics and health monitoring (PHM) and feedback control. As a motivating case study, the authors looked at a conveyor belt system and wind turbine motors [152].

In [5], the authors developed a medium-term mixed-integer linear program (MILP) formulation for optimizing production planning, scheduling and maintenance simultaneously, and [187] developed a framework for integrating data-driven stochastic degradation models into a mixed-integer optimization formulation. Their resulting framework provides an efficient way to calculate an operational strategy that compromises between maintenance cost and equipment availability.

In [25], a health-aware controller was developed to extend the lifetime of a subsea CO2 separation system, which they showed improved the reliability of the system, compared to a traditional MPC structure.

The optimal operation of a polymerization batch reaction subject to fouling was investigated in [189]. The problem was formulated as a continuous time problem, with the aim to find a sequence of batches that lead to acceptable fouling rates.

The idea presented in this paper builds on previous work [179], in which we presented a similar approach for a small subsea case study. In this paper, we extend the idea by proposing a more general framework, and applying it to a more extended case study. A multi-layer control structure, similar to [181], was used to deal with the different time scales of the problem. The contributions of this paper are the following:

1. We further develop the idea from [179] to present a systematic step-by-step framework for developing a health-aware control structure.
2. We demonstrate the method on a cascaded LNG refrigeration case study

4.3 Development of a health-aware operations strategy

4.3.1 Step 1: Economic and process analysis

The first step that should be taken when developing a unified decision making strategy for control and maintenance (or indeed any decision making strategy), is to clearly define the overall goals and limitations, as this will determine the required complexity of the model and optimization in the steps to follow.

Typically, the operational goal is to maximize profit, minimize energy consumption or waste, or to minimize batch time. This goal is to be achieved subject to constraints such as product quality constraints, maximum allowable temperatures, maximum input rates, etc.

Equally importantly, it is necessary to define targets and limitations for the reliability of the plant. These can be described in terms of expected availability, allowable degradation, number of components failed, etc. At this stage it is also important to define the boundary conditions for the problem. Is the decision making strategy to include an entire plant, or just a small section of it? What kind of decisions are included in the decision making strategy, e.g. do we want to optimize the times of maintenance interventions, or do we assume that these are given? All these questions must be answered before proceeding, as their answers will determine not only the complexity of the models and optimization problems formulated later, but also the way we should go about formulating them. A useful step is performing a criticality analysis, so that less critical components of the plant may be omitted from the analysis, in order to reduce complexity.

At the end of this step, one should have a (potentially multi-objective) optimization problem with appropriate constraints.

4.3.2 Step 2: Defining sensible decision-making layers layers

Ideally, we would like consider the entire decision making problem at once, but due to the complexity and the range of time scales involved, the resulting optimization problem is not practically feasible for most applications due to limited computational power. Instead, approximations are introduced so that the problem can be split into multiple time scale separated problems. A hierarchical decision making structure results, where each layer represents a different time scale of the process. But in contrast to the traditional approach to decision making, our framework assumes that there is an exchange of information between the process control domain and the reliability domain. Operation can therefore be less conservative, and sub-optimality is reduced. Note that there are no strict rules for which layer covers which part of the time scale, and that consequently some layers are overlapping.

In contrast to the standard control hierarchy, where decisions are made based on economic key performance indicators, a health-aware control layer has access to and uses information from the condition monitoring system for decision-making. Analogously to the traditional control hierarchy, we define the layers of the health-aware control system:

- Health-aware control: Obtained by integrating condition monitoring data into the advanced process control (APC) / real-time optimization (RTO) layers. The time scale on which this layer operates ranges from milliseconds to hours. The advantage of combining these layers into one is that high-frequency disturbances that have an impact on the RUL of the plant can be counteracted quickly, without having to rely on manual intervention. By monitoring the KPIs of the plant, correct actions can be taken immediately. Typically the actions taken in this layer are purely reactive, though there might be some benefit to also having proactive controllers in this layer. Depending on the complexity of the process, these actions could either be achieved by a health-aware MPC, a health-aware feed-forward controller or a simple PI controller. Examples of actions taken in this layer: opening and closing of valves, com-

pressor speeds, etc.

- Health-aware scheduling: Obtained by unifying the decision making processes in the higher layers of the hierarchy. Typically, the time scale of these layers range from hours to years. Decisions in this layer pertain to long-term production strategies. Implementation may be in the form of an MPC or a dynamic RTO, or a more sophisticated scheduling problem. Actions taken in this layer are typically proactive, in addition to being reactive. Examples of actions taken in this layer: load distribution among parallel units, sequence of batches for a day of production, etc.
- Health-aware long term planning: Obtained by integrating the very top layers of the decision making hierarchy. The time scale of these decisions range from weeks to years. Often these decisions are made manually, though there could be some applications where it would be beneficial to make these decisions automatically. Examples of actions taken in this layer: how frequently to exchange catalyst in a reactor, when to perform washing of a gas turbine, when to retrieve and maintain a subsea installation.

4.3.3 Step 3: Developing problem-specific / fit-for-purpose models for each of the layers

Once we have defined the desired plant behaviour in terms of a decision-making hierarchy, the next crucial step is to develop problem specific or fit-for-purpose models for each of the layers. By fit-for-purpose we mean that the models are sufficiently detailed to capture the phenomena occurring in the relevant time scale with sufficient accuracy, but without being too numerically challenging to be run in real-time. Furthermore, contrary to what is used in the conventional (decentralized) approach to operations, where separate models are used for production optimization and maintenance optimization, we need a unified model covering both domains [31].

Integration of the existing models from the different domains may prove difficult, due to the use of different modelling approaches in the two domains. Usually, it is necessary to build the models from the bottom up, to ensure compatibility.

4.3.4 Step 4: Optimization problem formulation and dealing with uncertainty

How the decision making process in each of the layers is implemented will be situation dependent. But since we defined objectives and constraints in the previous steps, it is natural to use a model-predictive control (MPC)-based control scheme, or something similar. MPC is a optimization-based control strategy that has gained traction in industry lately due to its ability to deal with constrained, multivariate problems [115]. At each time step, an optimal control problem (OCP) is solved and the first inputs of the optimal solution are applied to the plant. New measurements are collected periodically and the OCP is resolved repeatedly in a moving- or shrinking-horizon fashion. An illustration of the MPC principle is shown in Figure 4.3.4.

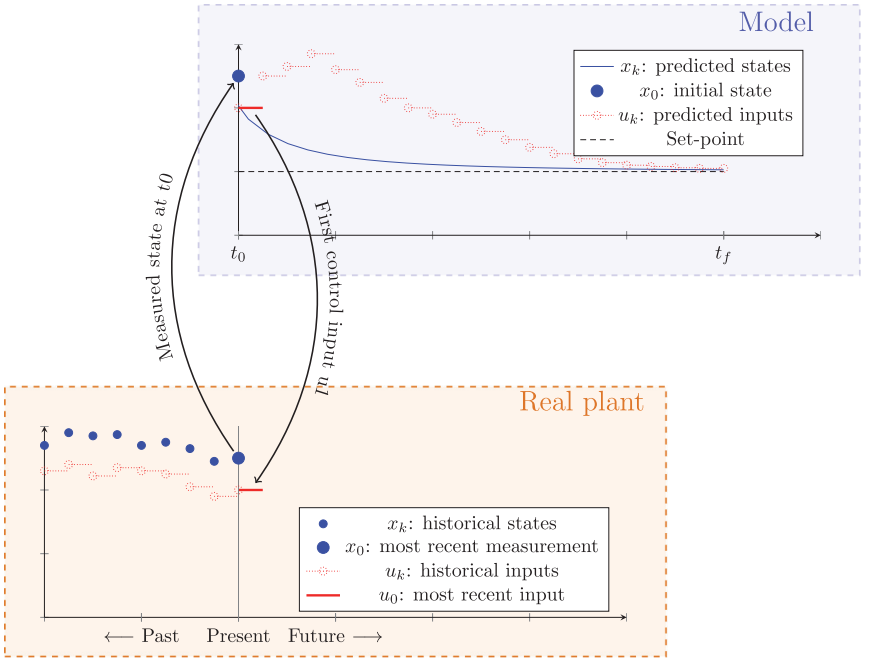

Figure 4.2: Illustration of the MPC principle.

The term "model-predictive control" is usually reserved to described controllers acting in the advanced process control (APC) layer, i.e. controllers either manipulating the inputs to the plant directly, or setting the set-points to the controllers in the stabilizing control layer. However, the term real-time optimization (RTO) is also used to describe the same methodology. The main difference between MPC and RTO is that RTO usually acts on slower time-scales, higher up in the decision hierarchy.

We mention MPC and RTO here in particular because they are both model-based optimization approaches to decision making, which makes them particularly well suited for the horizontal integration of the process control domain and the reliability/maintenance domain. One may "simply" include models of the plant degradation and maintenance into the existing MPC or RTO formulation, and thus optimize both at the same time. This requires little modification of the existing MPC methodologies, and existing theory can be used. Other approaches, such as feedback or feed-forward control structures may also be used if found to be sufficient.

Dealing with uncertain or stochastic parts of the model requires a slight reformulation of the MPC problem. Several different formulations exist. So called worst-case or min-max formulations have been popular approaches, as they ensure that the solution is feasible even if the worst-case realization of some bounded uncertain parameter should occur [21, 92, 159]. Since this approach requires the solving of

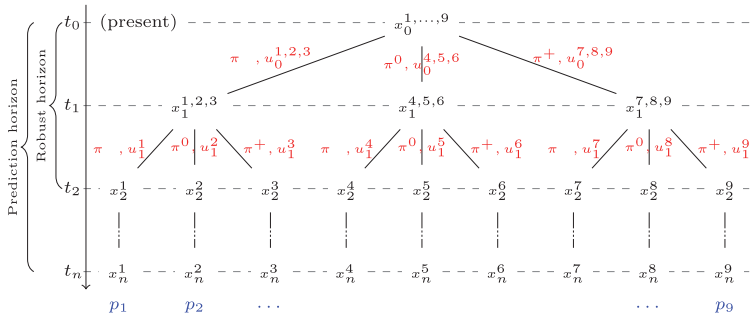

Figure 4.3: Scenario tree with prediction horizon n, robust horizon $n_{robust} = 2$ and 3 possible parameter realizations (π^-, π^0 and π^+) at each branching node. Every scenario has an associated probability p_i.

nested optimization problems, researchers have focused on making the MPC problem tractable, e.g. by finding simplifications or solving the problem offline and applying an online correction [41, 88, 104].

An alternative to the min-max formulation is tube-based MPC. In min-max MPC, the effect of feedback is not explicitly taken into the problem formulation. Consequently, the solution is a sequence of control moves which are open-loop robust. Naturally, this approach may perform very poorly due to the large amount of conservatism. Tube MPC seeks to reduce this conservatism by optimizing over control policies or control laws rather than just sequences of control moves. By doing so, the effect of feedback is explicitly considered [105, 139, 182].

A third approach which is popular is scenario-based or multi-stage MPC. Based on multi-stage stochastic programming, the main idea is to discretize the probability distribution of the uncertain parameter into scenarios [29]. Each scenario represents a path of possible future parameter realizations. Scenarios are bundled together in scenario trees, as shown in Figure 4.3. Each new branch in the tree represents the possibility of recourse, e.g. the possibility to adjust the inputs in a control setting. While constraint satisfaction can usually not be guaranteed except in the convex case and when the uncertainty set is bounded, the scenario-tree approach performs well for nonlinear systems as well, as long as the nonlinearity is reasonably small [102]. Due to the exponential growth of the scenario tree with the number of possible parameter realizations at each node, we usually only consider branching up until a point called the branching horizon [102]. The parameters are kept constant after the branching horizon. Since branching represents the availability of new information in the future, shortening the robust horizon means disregarding future state information.

4.3.5 Step 5: Implementation

Solving the optimization problem online requires appropriate software and hardware solutions. Typically, the fast-acting, low-level controllers require embedded

solutions, whereas the higher-levels can be solved remotely on a server or in the cloud.

Communication protocols must also be established to allow the control system to talk to the existing condition monitoring system.

Estimators such as the Kalman filter or a moving horizon filter must be run in parallel with the controllers and continuously update the models to reflect the ever-changing plant conditions.

4.4 Case study: process description

In order to study the framework discussed in this paper, we follow the step by step procedure to develop a health-aware decision making strategy for the refrigeration section of a cascaded liquid natural gas (LNG) liquefaction plant. An illustration of the full process is shown in Figure 4.4.

The advantage of a cascaded design is that the mean temperature difference in the heat exchangers is relatively small, thus resulting in a lower overall energy consumption. The drawback of the cascaded design over a mixed-refrigerant one is that the design requires more unit operations, thus potentially requiring a larger footprint. More importantly in the context of this paper: since each individual refrigerant cycle must be functional for the overall system to satisfactory perform its intended purpose, it is vital that the reliability of key equipment is ensured. The application of a health-aware control philosophy is therefore necessary.

The particular plant studied in this work has three intertwined refrigerant cycles; a high temperature propane cycle, a medium temperature ethane cycle and a cold temperature methane cycle. As is standard in refrigeration, each of the three cycles has a low pressure and high pressure zone. Pressure is increased in a single stage compressor driven by a turbine, allowing the compressor speed and consequently the compressor head, to be adjusted.

Upon exiting the compressor, the superheated refrigerant vapor is cooled and condensed by exchanging heat with cold sea water and, if applicable, the other liquid refrigerant streams, in the so-called condensing section. The condensing section of the propane cycle consists of a water-cooled condenser only. On the low-pressure side of the cycle, the refrigerant is evaporated in a series of cross-cycle heat exchangers and an LNG heat exchanger. For the propane cycle, the evaporation section consists of a propane/ethane heat exchanger followed by a propane/methane heat exchanger and a propane/LNG heat exchanger. Analogously, for the ethane cycle, the condensing section consists of a water-cooled condenser followed by the propane/ethane heat exchanger. The evaporation section consists of the ethane/methane heat exchanger and the ethane/LNG heat exchanger. Finally, methane is condensed by exchanging heat with water, propane and ethane, in that order. In the methane/LNG heat exchanger, the natural gas is brought down to its desired temperature of $-150\ ^\circ\mathrm{C}$.

Note that the representation of the condensing and evaporation section of the individual refrigerant cycles is somewhat unrealistic, as in reality, these small exchanger

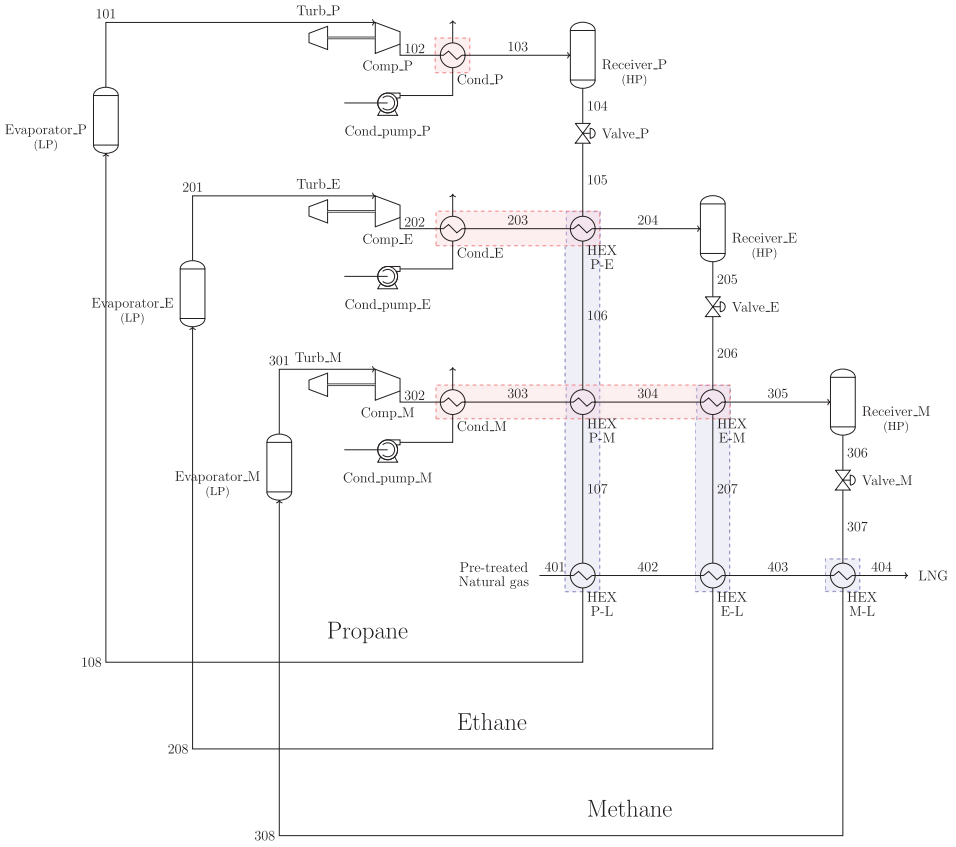

Figure 4.4: Process flow diagram of the studied cascaded refrigeration cycle LNG liquefaction plant. The plant consists of three interconnected refrigeration cycles, containing the refrigerants propane, ethane and methane, which are referred to as P, E and M, respectively. The abbreviations HP and LP refer to the high and low pressure sides of the cycles, respectively. The evaporation section of each cycle is highlighted in blue and the condensing section is highlighted in red.

would be replaced by a large, multi-stream heat exchanger in order to reduce equipment investment cost and footprint, and to reduce heat loss. However, we chose to model the multi-stream exchangers as series of two-stream heat exchangers to simplify mathematical model. As our intention is to use the model to develop a novel health-aware control strategy, and not to estimate investment cost or footprint, we assume that this simplification does not influence the results obtained in this paper.

4.4.1 Step 1: Economic and process analysis

The purpose of the cascaded LNG refrigeration plant is to subject a pretreated natural gas stream to multiple, increasingly cold, closed heat exchangers. After passing through the final heat exchanger, the natural gas has reached a sufficiently low temperature that it remains in liquid phase when subjected to ambient pressure, since transport and storage of natural gas in liquid phase is preferable to gas phase. An outlet temperature of no higher than -150 °C is required to ensure that the natural gas remains in liquid phase.

The LNG liquefaction plant is expected to operate for long periods of time without breaking down or needing maintenance. A criticality analysis was performed, and it was found that the rotating equipment, i.e. the compressors, are the most critical components of the system. Due to their complexity, they are more likely to degrade than the static components of the system. In particular, the compressor bearings degrade due to load-induced wear and tear. It is important to keep this wear and tear below a certain threshold to ensure that the compressor can operate reliably.

To summarize, the desired plant behaviour is the following:

- Operate at minimum energy consumption while still satisfying that the outlet temperature of -150 °C is not exceeded for any given load of natural gas.
- Ensure that the compressor bearing degradation does not exceed a critical level between two scheduled maintenance interventions.

4.4.2 Step 2: Defining sensible decision-making layers layers

Because the dynamics of the the LNG plant range from the range of milliseconds (measurement of bearing vibrations) to months (bearing degradation), it is natural to devise a hierarchical control structure. We propose the following layers

1. **Stabilizing control layer:** Acts on a very fast time-scale (seconds). Since degradation is assumed to take place on a much slower time scale, this layer does not take the degradation information into account. Degrees of freedom in this layer: valve openings.
2. **Optimizing control layer:** An economic NMPC whose objective it is to reject disturbances while simultaneously minimizing the energy consumption of the plant. In this layer, the dynamics relating to the long-term degradation can be assumed to be constant, so their differential states are replaced by constants.

The MPC is constrained by the maximum allowable outlet temperature, and the maximum allowable degradation rate set by the top-layer scheduling controller. Since this controller is operating on a somewhat fast time scale, multiple realizations of the uncertainty (in this case, we assume the inlet temperature and flow-rate, as well as the condenser water temperature, to be uncertain). Degrees of freedom in this layer: set-point to stabilizing control layer, compressor speed, condenser duty.

3. **Scheduling control layer:** A shrinking-horizon dynamic optimal scheduling controller whose objective it is to minimize the overall energy consumption of the plant, subject to a maximum allowed degradation limit between two maintenance interventions. In this layer, the fast dynamics, i.e. the mass and energy balances for the refrigerants, can be assumed to be instantaneous. Degrees of freedom in this layer: set-point to optimizing control layer, compressor speed, condenser duty.

An illustration of the proposed control structure is shown in Figure 4.5.

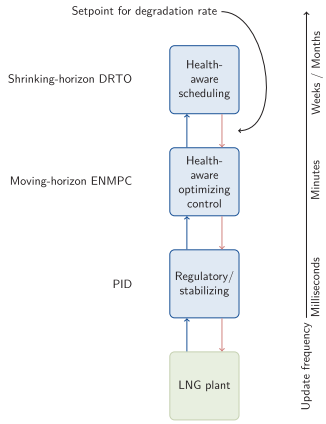

Figure 4.5: Proposed control structure for the LNG case study. Blue arrows indicate the flow of information, whereas red arrows indicate the flow of decisions between the different blocks.

4.4.3 Step 3: Developing problem-specific / fit-for-purpose models for each of the layers

The full model of the cascaded LNG refrigeration plant can be found in the appendix. The process model was previously described in [177].

Process model

The process model consists of mass and energy equations for the unit operations, including the evaporators, receivers, valves and compressors. The thermodynamic

properties of the fluids are calculated by surrogate models, which were fitted to thermodynamic data.

Degradation model

In addition to the regular conservation equations for mass and energy, the model describes the degradation of the system over time. For the sake of simplicity, only a single degradation mechanism is included in the model, namely the degradation of the compressor bearings due to load-induced wear and tear. Paris-Erdogans law of crack propagation [130] is used to describe the evolution of the bearing imperfections.

4.4.4 Step 4: Optimization problem formulation and dealing with uncertainty

A three-layer decision making hierarchy was proposed for this plant. The three layers are discussed below

Stabilizing control layer

This layer acts on a fast time scale, and its main purpose is to keep important variables (e.g. levels) at their set-points and stabilize the process. Therefore one could formulate optimal operation as minimizing the deviation from their set-points. However, since minor variations do not significantly effect the economics, it is common to use a decentralized control structure based on PI controllers.

Level controllers for the receivers were implemented as shown in Figure 4.6. The controllers are simple PI controllers tuned reasonably tightly, so that the disturbances are rejected reasonably quickly. Three degrees of freedom are used in this control layer, namely the valve opening of the receiver outlet valve. The remaining degrees of freedom for the above control layer are the compressor speeds and the condenser duties. Additionally, one could adjust the set-points of the level controllers, but these were left constant for the sake of robustness. Since the levels only have a transient effect, the level set-points have little effect on the overall economic

Optimizing control layer

A scenario-based moving-horizon economic NMPC [103] was chosen for this layer, as it was found that a non-linear, online method was necessary to deal with the disturbances. Since many realizations of the stochastic parameters would be observed, a scenario-based approach was chosen in order to ensure that the solution is acceptable no matter the realization.

The optimization problem solved at each time step can be formulated as

$$\min_{\mathbf{x}_{i,k},\mathbf{u}_{i,k}} \sum_{i=1}^{S} p_i \sum_{k=1}^{N} \phi\left(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}, \boldsymbol{\pi}_{i,k}\right) \tag{4.1a}$$

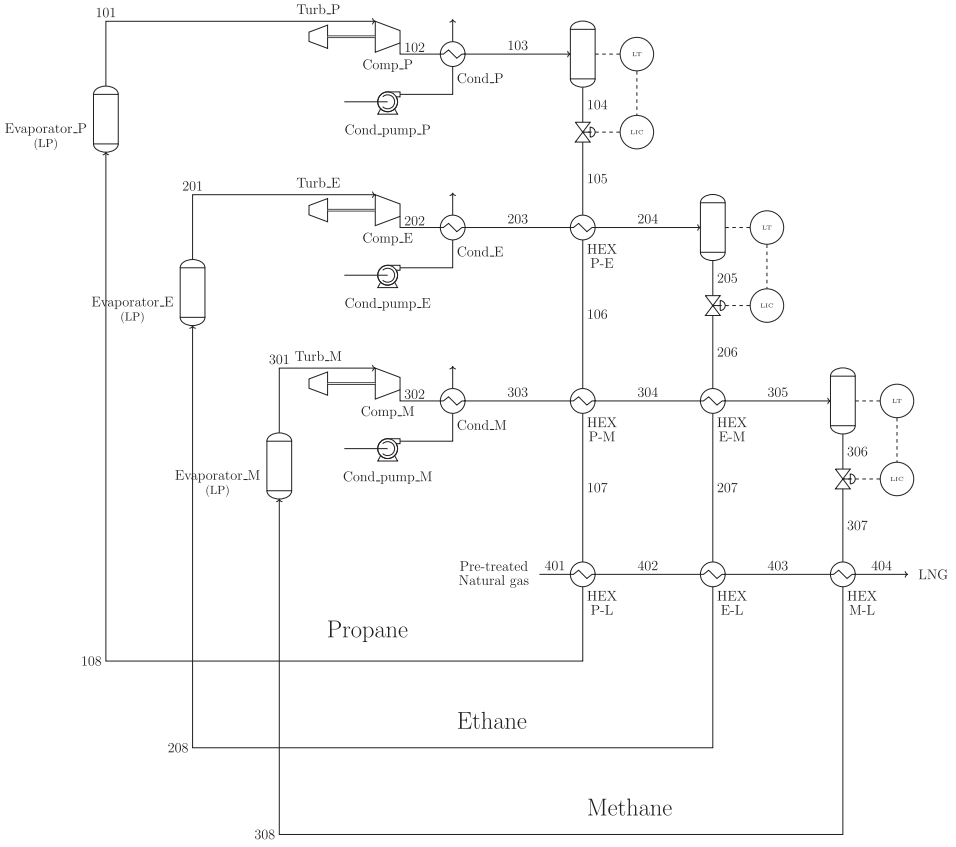

Figure 4.6: Process flow diagram of the studied cascaded refrigeration cycle LNG liquefaction plant, with added level controllers for the receivers.

$$\text{s.t.} \quad f(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}, \boldsymbol{\pi}_{i,k}) \leq 0 \quad \forall i = 1...S\,,\, k = 1...N \tag{4.1b}$$

$$g(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}, \boldsymbol{\pi}_{i,k}) = 0 \quad \forall i = 1...S\,,\, k = 1...N \tag{4.1c}$$

$$\sum_{i=1}^{S} \mathbf{A}_{i,k}\mathbf{u}_{i,k} = 0 \quad \forall k = 1...N \tag{4.1d}$$

where S is the number of scenarios p_i is the probability associated with scenario i and $\mathbf{A}$ are the non-anticipativity constraints. The variables $\mathbf{u}$ are the inputs, i.e. the compressor speeds and the condenser duty.

Here,

$$\phi = W_{comp,P} + W_{comp,E} + W_{comp,M} + c_{penalty}\boldsymbol{\Delta}\mathbf{u}^{\intercal}\boldsymbol{\Delta}\mathbf{u} \tag{4.2}$$

$$f_1 = T_{404} + 150 \tag{4.3}$$

$$f_2 = \Delta a - \Delta a_{setpoint}, \tag{4.4}$$

where the objective ϕ consists of the sum of the energy consumption of the three

compressors, and a term penalizing excessive change in input from one time step to the next. Two inequality constraints f_1 and f_2 are constraining the optimization. The former constraint sets the maximum allowable outlet temperature, the latter ensures that the compressor degradation rate Δa does not exceed a threshold $\Delta a_{setpoint}$, which is determined by the scheduling control layer, discussed below.

Each of the scenarios represents one possible realization of the uncertain parameters, in this case the natural gas flow-rate and temperature, as well as the condensing water temperatures. Each of these three disturbances was assumed to be normally distributed, to represent day-to-day variations in natural gas supply and weather conditions. In order to avoid a large scenario tree which impacts performance of the MPC, we chose to represent the uncertainty by seven possible combinations of realizations: one nominal case and a high realization case and a low realization case for each of the three uncertain parameters. To reduce the size of the scenario tree even further, it was chosen to use a robust horizon of length one, meaning that all uncertain parameters were assumed to be constant following the initial branching. This meant that the scenario three had a total of seven branches.

Dynamic scheduling control layer

A worst-case shrinking-horizon economic DSC was chosen for this layer, since the . Since few observations of the uncertain parameters would be made, a worst-case realization approach was chosen in order to ensure that the degradation would not exceed the acceptable threshold at the end of the horizon, no matter the true value of the degradation rate parameter.

The uncertainty here relates to the degradation rate parameter c_{Paris} from 4.7.5, which determines how quickly the cracks on the compressor bearings grow in size, given a certain load. As more observations are obtained, the true value of the degradation rate parameter can be estimated more accurately. Consequently, as the uncertainty is reduced, the worst-case prediction will become less and less conservative. Periodic model re-fitting is performed to get better estimates of the degradation rate parameter c_{Paris}

The optimization problem solved at each time step can be formulated as

$$\min_{\mathbf{x}_k, \mathbf{u}_k} \sum_{k=1}^{N} \phi\left(\mathbf{x}_k, \mathbf{u}_k, \pi\right) \tag{4.5a}$$

$$\text{s.t.} \qquad \pi = \arg\max_{p} f(\mathbf{x}_k, \mathbf{u}_k, p) \tag{4.5b}$$

$$f(\mathbf{x}_k, \mathbf{u}_k, \pi) \leq 0 \quad \forall k = 1...N \tag{4.5c}$$

$$g(\mathbf{x}_k, \mathbf{u}_k, \pi) = 0 \quad \forall k = 1...N \tag{4.5d}$$

$\mathbf{u}$ are the inputs, i.e. the compressor speeds and the condenser duty.

Note that the constraints f and g, as well as the objective ϕ is not the same in problem (4.5) and problem (4.1). Here,

$$\phi = W_{comp,P} + W_{comp,E} + W_{comp,M} + c_{penalty}\mathbf{\Delta u}^{\intercal}\mathbf{\Delta u} \tag{4.6}$$

$$f_1 = T_{404} + 150 \tag{4.7}$$

$$f_2 = a - a_{max} = a - 1. \tag{4.8}$$

The last constraint indicates that at the end of the period, when the next maintenance is scheduled, the maximum crack length on the compressor bearings must not exceed 1 mm.

The above optimization problem is solved repeatedly in a shrinking-horizon fashion.

4.4.5 Step 5: Implementation

Because the case is illustrative only and no real operational data exists, all results were obtained from simulation studies.

The model and MPC controllers were implemented in MATLAB and CasADi 3.4.1 [11]. Both the NMPC in the optimizing control layer and the DRTO in the scheduling control layer were transformed into non-linear programming problems (NLPs) using direct orthogonal collocation. IPOPT 3.12.3 [185] was used to solve the problems.

To illustrate how the condition monitoring system may be integrated into the operational decisions, assume that a vibration sensor exists on the housing of the compressors. In order to predict the bearing crack magnitude, the vibration signals are analysed. Commercial software and hardware solutions exist that do this already, but assuming that such a commercial solution does not exist for our LNG case study, one must develop an estimator to successfully "link" the two domains of reliability and control, so that the MPC controller knows what the degradation is at any time. (It can be argued that the following part belongs in section 2: modelling, but we chose to include it here). The bearing diagnostic model was previously presented in [179]. For a full overview of bearing fault diagnostics, see e.g. [186]

Bearing diagnostics - estimating the degradation

Figure 4.7 shows a frontal and axial cross sections of a ball bearing. When a ball bearing rotates, its inner race and balls will rotate at different frequencies, denoted f_s and f_{ft} respectively. Additionally, each ball will rotate around itself with frequency f_{ball}. These three frequencies are all dependent on the dimensions of the ball bearing, shown on the right of Figure 4.7. The frequency with which the balls rotate around the race, the *fundamental frequency*, can be calculated as

$$f_{ft} = \frac{f_s}{2}\left[1 - \left(\frac{D_p}{D_b}\right)^{-1}\cos(\phi)\right]., \tag{4.9}$$

where f_s is the shaft frequency, D_p is the pitch diameter, D_b is the ball diameter, and ϕ is the contact angle, as shown in Figure 4.7.

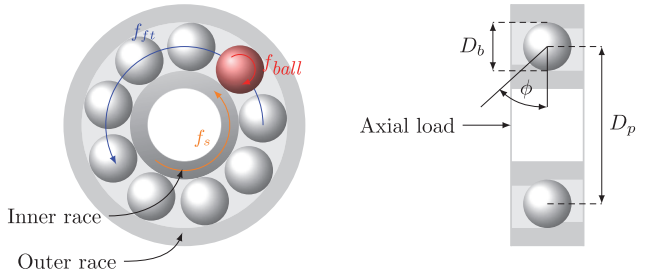

Figure 4.7: Frontal and axial cross sections of a ball bearing on the left and right, respectively.

When a ball bearing has a defect, this defect will create a periodic excitation force. Depending on if the defect is situated on the inner race, outer race or on one of the balls / rolling elements, this excitation force will have a particular characteristic frequency f_{fault}. [186] showed that for an inner race fault, f_{fault} can be expressed as

$$f_{IR\ fault} = n_b \cdot (f_s - f_{ft}), \tag{4.10}$$

where n_b is the number of balls in the bearing. For outer race (OR) and rolling element (RE) failures, the fault frequencies can be calculated as

$$f_{OR\ fault} = n_b \cdot f_{ft} \tag{4.11}$$

and

$$f_{RE\ fault} = \frac{f_s}{2}\left(\frac{D_p}{D_b}\right)\left[1 - \left(\frac{D_p}{D_b}\right)^{-2}\cos(\phi)\right]. \tag{4.12}$$

The magnitude of the excitation is proportional to the surface defect size. Furthermore, we know that seen from the perspective of a stationary vibration sensor, the excitation pulse will be weaker if the defect is on the far side of the bearing, and stronger if it is on the close side of the bearing. Consequently, the excitation pulses are magnitude modulated with a characteristic frequency depending on which kind of fault is occurring. The magnitude of an OR fault will be constant since the outer ring does not move in relation to the stationary vibration sensor. The magnitude of an IR fault will be modulated with frequency f_s because the distance between the fault and the sensor is not constant. Lastly, we know that vibrations are not transferred without energy losses from the vibration sensor and the fault location. Some energy goes into the system, due to elasticity etc. When the excitation signal arrives at the vibration sensor, it has been convoluted by the *impulse response function* g, which is a damped harmonic oscillator.

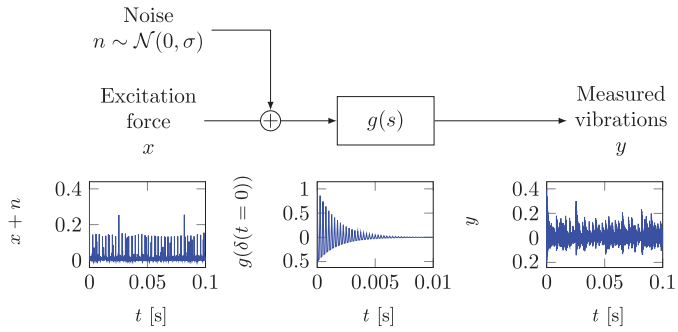

Figure 4.8: Periodic excitation force x (plus noise) is generated by the bearing defect, modulated by the impulse response function g, and measured as the vibration signal y, via impulse response model g.

Knowing all of this, we can predict how the excitation force will manifest itself as measured vibrations. This is illustrated in Figure 4.9. By going the reverse path and using spectral analysis, we can generate an estimate of the excitation force on the bearing, $\bar{x}$ from the vibration signal, y. Assuming that g can be found experimentally (by exciting the system and measuring the resulting vibration signal), we have that

$$\bar{x} = G^{-1} y, \tag{4.13}$$

where G is the Toeplitz convolution matrix of g. The additive noise is removed by using a Wiener filter.

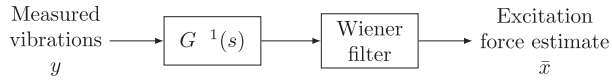

Figure 4.9: Estimating the excitation force from the measured vibrations

4.5 Case study: results

The simulation results are summarized in Figure 4.10. The upper row of plots shows the degradation of the bearings in the three compressors on the z axis, as a function of time on the x axis. On the y axis are the DRTO iterations. The black, dashed line shows the actual evolution (history) of the degradation. Around the black, dashed line, the predicted degradation evolution is shown in blue. Note that the prediction includes the 90% confidence interval around it. The confidence interval becomes larger the further in the future the prediction lies, as is to be expected. On the yz-plane behind the plots, the confidence interval at the final mission time $t_f = 360$ days is shown. Note that this confidence interval shrinks exponentially in size. The reason for this is twofold. Firstly, later DRTO iterations have starting time t_0 closer to the final time t_f, since t_f is fixed. Consequently, since the initial degradation state is known at the beginning of each iteration, the

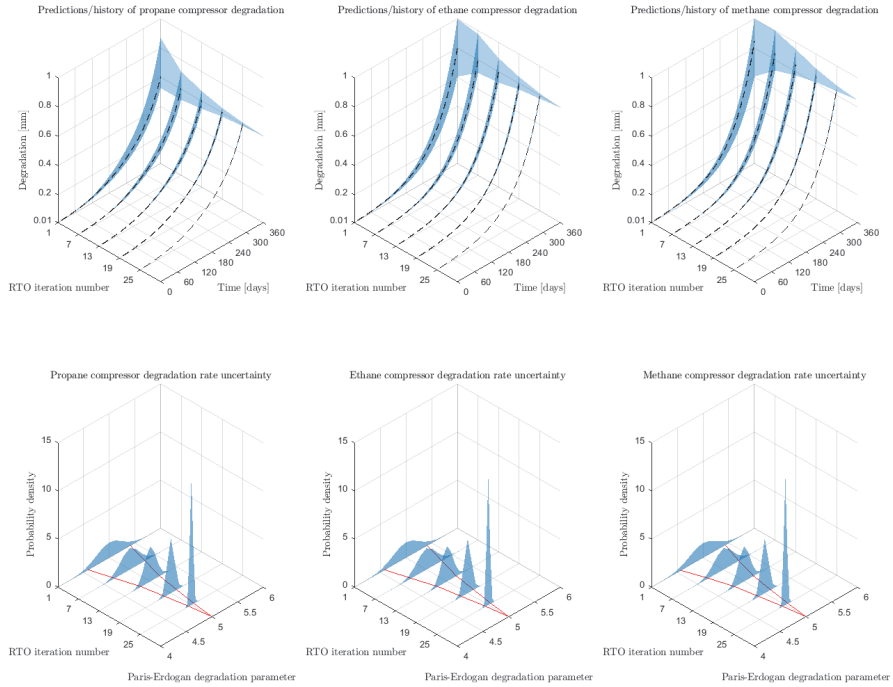

Figure 4.10: Top: predicted (blue) and actual (dashed black) degradation of the compressor bearings for the propane, ethane and methane compressors (left to right) at five different points in time. The 90% confidence interval of the prediction is also shown around the predicted degradation path. The confidence interval shrinks in size because the degradation rate parameter (bottom row of plots) becomes more accurately known as more observations of the uncertainty are obtained. Note that as the RTO iteration number increases, the starting time of the prediction gets later and later.

final confidence interval shrinks in size since the uncertainty does not have to be propagated so long in time. Secondly, more observations of the degradation have been observed for later DRTO iterations, which are used to update the probability distribution of the uncertain degradation parameter. This is shown in the bottom row of Figure 4.10. Starting from a relatively wide apriori distribution for the degradation parameter, the distribution narrows down as more observations are obtained. The red lines on the xy-plane indicate the 95% and 5% percentiles for the plotted probability distributions.

We note that for the first 3 iterations, the upper bearing degradation constraint is active for the ethane compressor. Similarly, for the first 9 iterations, the upper bearing degradation constraint is active for the methane compressor. In order to make sure that the constraint is not violated if the worst case realization happens, the methane compressor and the ethane compressor is throttled down a little. In

order to still satisfy the −150 °C LNG outlet temperature constraint, the load on the propane compressor is increased. The increased load on the propane compressor, and the fact that the ethane and the methane compressors are running at non-optimal operating points, increases the overall energy consumption of the system.

4.6 Conclusion and future work

In this paper we presented a framework for the optimal operation of process plants in the presence of degradation. This concept was dubbed health-aware operation. We argued that explicitly including the plant degradation in the decision making process leads to improved flexibility and more optimal operation than the conventional approach. Common practice today is that the allowable operating region of the plant is reduced by introducing additional constraints originating from a worst-case reliability consideration. Without continuous adjusting of these constraints, operation is sub-optimal. Due to the lack of feedback, conservatism must be introduced to ensure that all the constraints are satisfied at all times. By including the plant degradation directly in the control problem, feedback is introduced.

Due to the vastly different timescales on which degradation usually occurs, it is useful to divide the decision making process into hierarchical layers, where each layer follows a different time scale. Depending on the process in consideration, we discussed how one might go about designing a health-aware control strategy.

To illustrate the ideas discussed here, we applied the method to the case study of the refrigeration section of an LNG liquefaction plant. As LNG liquefaction is hugely energy intensive, the main aim of our control structure was to minimize the energy consumption of the refrigerant compression section. A secondary objective was to ensure that compressors were able to perform their intended task until the next maintenance intervention.

In the planning layer of our control scheme, a DRTO was tasked with balancing the load-induced degradation of the compressor bearings such that it would not exceed a predefined threshold at the next scheduled maintenance intervention. In a shrinking horizon fashion, the operation strategy was continuously adjusted by the DRTO to reflect the current plant conditions. To account for the inevitable plant-model mismatch in the prognostic degradation model, a worst-case dynamic optimization method was used. As more observations of the actual degradation were obtained, these were used to update the confidence interval of the degradation prediction.

In the supervisory control layer of our control structure, an NMPC was used to reject process disturbances on a minute-to-minute basis and while simultaneously ensuring that the maximum degradation rate prescribed by the DRTO was not exceeded. To account for the mismatch between the NMPC model and the plant, a scenario-based optimization was used. This way, the NMPC was near-optimal independent of the actual disturbance realization, while simultaneously ensuring that the LNG outlet temperature constrain was not violated.

In future work, we would like to apply the proposed method on a real-world case.

For example, in a real-world process, multiple failure mechanisms would have to be considered, rather than just a single one, like was done in this study. More sophisticated ways of handling uncertainty, such as risk-averse formulations for the upper layers of the decision hierarchy, might be useful, especially when constraint violation has large economic impact (e.g. equipment failure).

4.7 Appendix A - Model for the cascaded LNG refrigeration cycle

In this section, we describe the system model for the LNG refrigeration plant. In Subsection 4.7.1 we describe the thermodynamic model / equations of state of the refrigerants. In Subsections 4.7.2, 4.7.3 and 4.7.4, we describe the models of the compressors, valves and evaporators, respectively. The system is mostly static, with the exception of the energy balances of the receivers, which are given in section 4.7.5.In the following sections, we will use the subscript $i \in \{\text{propane, ethane, methane}\}$ to refer to the three refrigerants to generalize the notation and avoid cluttered equations.

Model parameters for the various units are given in the Appendix B.

4.7.1 Thermodynamics

The thermodynamic properties of propane, ethane and methane are calculated by polynomial surrogate models. AllProps [95], a software package for calculating thermodynamic properties of selected refrigerants using the Helmholtz equations of state, was used to fit the surrogate model.

Saturation temperatures, liquid enthalpies, vapour enthalpies and heat capacities of the refrigerants are calculated as

$$T_{\text{sat},i} = \sum_{j=0}^{2} c_{T_{\text{sat},i},j} \cdot \log(P_i)^j \tag{4.14}$$

$$\hat{H}_{\text{liq},i} = \sum_{j=0}^{6} c_{\hat{H}_{\text{liq},i},j} \cdot P_i^j \tag{4.15}$$

$$\hat{H}_{\text{vap},i} = \sum_{j=0}^{6} c_{\hat{H}_{\text{vap},i},j} \cdot P_i^j \tag{4.16}$$

$$C_{p,i} = \sum_{j=0}^{2} c_{C_{p,i},j} \cdot T_i^j, \tag{4.17}$$

where P_i is the pressure of refrigerant i, and $c_{*,i}$ are the model parameters, which can be found in Tables 4.1, 4.2, 4.3 and 4.4, respectively.

Dranchuk and Abou-Kassems equation of state [45] is used to calculate the compressibility factor Z_i of the three refrigerants in the vapour phase, as a function of

the reduced temperatures and pressures $T_{r,i}$ and $P_{r,i}$

$$\begin{aligned} Z_i = 1 &+ \left(c_{Z_i,1} + \frac{c_{Z_i,2}}{(T_{r,i})} + \frac{c_{Z_i,3}}{(T_{r,i})^3} + \frac{c_{Z_i,4}}{(T_{r,i})^4} + \frac{c_{Z_i,5}}{(T_{r,i})^5} \right) B \\ &+ \left(c_{Z_i,6} + \frac{c_{Z_i,7}}{(T_{r,i})} + \frac{c_{Z_i,8}}{(T_{r,i})^2} \right) B^2 \\ &- \left(\frac{c_{Z_i,7}}{(T_{r,i})} + \frac{c_{Z_i,8}}{(T_{r,i})^2} \right) c_{Z_i,9} B^5 \\ &+ c_{Z_i,10} \left(1 + c_{Z_i,11} B^2\right) \left(\frac{B^2}{T_{r,i}^3} \right) \exp\left(-c_{Z_i,11} B^2\right), \end{aligned} \tag{4.18}$$

where

$$B = \frac{0.27 P_{r,i}}{Z_i T_{r,i}} \tag{4.19}$$

$$T_{r,i} = \frac{T_i}{T_{c,i}} \tag{4.20}$$

$$P_{r,i} = \frac{P_i}{P_{c,i}}. \tag{4.21}$$

The coefficients for Equation (4.18) can be found in Table 4.5

4.7.2 Compressors

The compressors can be assumed to be polytropic compressors, i.e. we can assume that

$$\left(\frac{T_{discharge,i}}{T_{suction,i}} \right) = \left(\frac{P_{discharge,i}}{P_{suction,i}} \right)^{1/k_i}, \tag{4.22}$$

where the polytropic coefficient k_i is

$$k_i = \eta_i \frac{\gamma_i}{\gamma_i - 1} \tag{4.23}$$

$$\gamma_i = \frac{1}{2} \left(\frac{C_{p,suction,i}}{C_{p,suction,i} - R} + \frac{C_{p,discharge,i}}{C_{p,discharge,i} - R} \right). \tag{4.24}$$

η_i is the polytropic efficiency, and γ_i is the ratio between the average heat capacity at constant pressure C_p and the average heat capacity at constant volume $C_v = C_p - R$, where R is the universal gas constant.

The compressor head is

$$h_{poly,i} = \frac{k_i Z_{suction,i} R}{g M_{m,i}} \left(T_{discharge,i} - T_{suction,i} \right), \tag{4.25}$$

where g is the gravitational constant and $M_{m,i}$ is the molar weight.

Equipment-specific compressor maps are used to calculate the efficiency η_i and compressor head $h_{poly,i}$ for each of the compressor.

$$\eta_i = c_{\eta_i,1} h_{poly,scaled,i} + c_{\eta_i,2} - 10^{(c_{\eta_i,3} h_{poly,scaled,i} - c_{\eta_i,4})} \tag{4.26}$$

and

$$\frac{q_{suction}}{(u_i^{comp.})^{c_{comp,1}}} = \frac{c_{comp,2} h_{poly,scaled,i} - c_{comp,3}}{c_{comp,4}} \tag{4.27}$$

$$q_{suction,i} = \frac{\dot{n}_{suction,i} R T_{suction,i}}{P_{suction,i}} \tag{4.28}$$

$$h_{poly,scaled,i} = \frac{h_{poly,i}}{(u_i^{comp.})^{c_{h_{poly,scaled,i},1}}} \tag{4.29}$$

where $q_{suction}$ and $\dot{n}_{suction,i}$ are the suction volumetric and molar flow rates. $u_i^{comp.}$ is the compressor speed normalized, i.e. the speed in RPM divided by the operating speed in RPM.

Coefficients used in Equations (4.26)-(4.29) are shown in Tables 4.6-4.8 in Appendix B.

4.7.3 Valves

The flow through the system is given by the valve opening. The molar flow is calculated as

$$\dot{n}_i = u_i^{valve} c_{choke,i} \sqrt{P_{out} - P_{in}} \tag{4.30}$$

where u_i^{valve} is the fractional valve opening and $c_{choke,i}$ is the valve constant. The valve constants for the system can be found in Table 4.9.

4.7.4 Evaporators and condensers

The temperature change of the refrigerant in the evaporators and condensers is given by the energy balance

$$\dot{Q}_i = U A_i \frac{\Delta T_1 - \Delta T_2}{\log \Delta T_1 - \log \Delta T_2} \tag{4.31}$$

$$= \dot{H}_{cold,out} - \dot{H}_{cold,in} \tag{4.32}$$

$$= \dot{H}_{hot,in} - \dot{H}_{hot,out}, \tag{4.33}$$

where

$$\Delta T_1 = T_{hot,in} - T_{cold,out}, \tag{4.34}$$

$$\Delta T_2 = T_{hot,out} - T_{cold,in}. \tag{4.35}$$

4.7.5 Dynamics

For the sake of the two optimization-based control layers, we can assume that the levels in the receivers are perfectly controlled on the time scale we are interested in.

On the time scale of the MPC, the dynamics of the system are caused by thermal inertia of the three receivers. Since the receivers are relatively large, the exchange of energy is relatively slow, and cannot be assumed to be instantaneous.

The dynamic energy balances of the three receivers can be written as

$$\frac{dT_{receiver,P}}{dt} = \frac{\dot{n}_{103}}{n_{receiver}}\left(T_{103} - T_{104}\right) \tag{4.36}$$

$$\frac{dH_{receiver,E}}{dt} = \dot{H}_{204} - \dot{H}_{205} = \dot{n}_{204}\hat{H}_{204} - \dot{n}_{205}\hat{H}_{205} \tag{4.37}$$

$$\frac{dH_{receiver,M}}{dt} = \dot{H}_{305} - \dot{H}_{306} = \dot{n}_{305}\hat{H}_{305} - \dot{n}_{306}\hat{H}_{306} \tag{4.38}$$

4.7.6 Compressor bearing degradation

Paris' crack propagation model [130] is used to predict how the bearing defect will develop over time. Given a surface crack of size a, it will increase in size exponentially when force is applied. Paris' law states that

$$\frac{da}{dn_{cycles}} = D \cdot (\Delta K)^n , \tag{4.39}$$

where n_{cycles} is the number of cycles, ΔK is the strain applied to the system, and D and n are constants. Assuming that the motor torque can be used as an indicator for the strain [20], the above equation can be reformulated as

$$\frac{da}{dt} = c_{Paris} \cdot \left(T^2 \cdot f_s\right) = c_{Paris} \cdot \left(\frac{P^2}{f_s}\right), \tag{4.40}$$

where c_{Paris} is a lumped parameter, T is the motor torque and P is the motor power. In this case study, we assume that the true value of c_{Paris} is not known, so that it must be estimated online.

4.8 Appendix B - Design parameters

Table 4.1: Design parameters for calculating the saturation temperature in Equation (4.14)

Parameter	Value		
	Propane	Ethane	Methane
$c_{T_{\text{sat},i},0}$	3.0053	2.4141	1.4966
$c_{T_{\text{sat},i},1}$	0.93582	0.77135	0.51530
$c_{T_{\text{sat},i},2}$	0.24817	0.21080	0.13110

Table 4.2: Design parameters for calculating the specific liquid saturation enthalpy in Equation (4.15)

Parameter	Value		
	Propane	Ethane	Methane
$c_{\hat{H}_{\text{liq},i},0}$	-9992.7	-9447.4	-15187
$c_{\hat{H}_{\text{liq},i},1}$	15991	12605	5974.2
$c_{\hat{H}_{\text{liq},i},2}$	-11110	-12132	-5488.7
$c_{\hat{H}_{\text{liq},i},3}$	4838.4	7061.9	3138.6
$c_{\hat{H}_{\text{liq},i},4}$	-970.54	-2146.5	-943.33
$c_{\hat{H}_{\text{liq},i},5}$	50.621	319.1	138.42
$c_{\hat{H}_{\text{liq},i},6}$	4.8879	-18.124	-7.6457

Table 4.3: Design parameters for calculating the vapor saturation enthalpy in Equation (4.16)

Parameter	Value		
	Propane	Ethane	Methane
$c_{\hat{H}_{\text{vap},i},0}$	9036.9	5906	-6622.6
$c_{\hat{H}_{\text{vap},i},1}$	11005	6200.8	2117.7
$c_{\hat{H}_{\text{vap},i},2}$	-12980	-7884.9	-1650.9
$c_{\hat{H}_{\text{vap},i},3}$	9271.9	5412.9	518.47
$c_{\hat{H}_{\text{vap},i},4}$	-3644.6	-1973	-59.04
$c_{\hat{H}_{\text{vap},i},5}$	721.52	356.06	0
$c_{\hat{H}_{\text{vap},i},6}$	-56.405	-25.103	0

Table 4.4: Design parameters for calculating the heat capacity in Equation (4.17)

Parameter	Value		
	Propane	Ethane	Methane
$c_{C_{p,i},1}$	14.1504	9.4031	10.0858
$c_{C_{p,i},2}$	0.07550	0.1598	0.23931
$c_{C_{p,i},3}$	$-1.799 \cdot 10^{-5}$	$-4.623 \cdot 10^{-5}$	$-7.336 \cdot 10^{-5}$

Table 4.5: Design parameters for calculating the compressibility factor in Equation (4.18)

Parameter	Value
$c_{Z_i,1}$	0.3265
$c_{Z_i,2}$	-1.0700
$c_{Z_i,3}$	-0.5339
$c_{Z_i,4}$	0.01569
$c_{Z_i,5}$	-0.05165
$c_{Z_i,6}$	0.5475
$c_{Z_i,7}$	-0.7361
$c_{Z_i,8}$	0.1844
$c_{Z_i,9}$	0.1056
$c_{Z_i,10}$	0.6134
$c_{Z_i,11}$	0.7210

Table 4.6: Design parameters for calculating the compressor efficiency in Equation (4.26)

Parameter	Value		
	Propane	Ethane	Methane
$c_{\eta_i,1}$	0.061733	0.03448682	0.0251571
$c_{\eta_i,2}$	-0.074	-0.074	-0.074
$c_{\eta_i,3}$	0.338767	0.18925109	0.138053
$c_{\eta_i,4}$	6.5963	6.5963	6.5963

Table 4.7: Design parameters for calculating the polytropic head in Equation (4.27)

Parameter	Value		
	Propane	Ethane	Methane
$c_{comp,1}$	1.79	1.79	1.79
$c_{comp,2}$	1.9928	1.9928	0.27146
$c_{comp,3}$	35.1962	63.00	11.765
$c_{comp,4}$	18.3546	32.855	45.040

Table 4.8: Design parameters for calculating the polytropic head in Equation (4.29)

Parameter	Value		
	Propane	Ethane	Methane
$c_{h_{poly,scaled,i},1}$	2.11	2.11	2.11

Table 4.9: Design parameters for calculating the molar flow through the valve in Equation (4.30)

Parameter	Value		
	Propane	Ethane	Methane
$c_{choke,i}$	13.0	8.4	3.4

Chapter 5

Health-Aware Operation of a Subsea Compression System Subject to Degradation

This chapter is from the article

- A. Verheyleweghen, J.M. Gjøby and J. Jäschke. Health-Aware Operation of a Subsea Compression System Subject to Degradation (28th European Symposium on Computer Aided Process Engineering (ESCAPE), Graz, Austria). *Computer Aided Chemical Engineering*. volume 43, pages 1021—1026, 2018

5.1 Abstract

We propose an health-aware operation approach for combining short-term control objectives with long-term profit and reliability targets. In particular, we present a hierarchical approach for operating a compressor subject to degradation. We consider a case study of a subsea compressor, where the goal is to maximize the gas throughput, while ensuring that the compressor can be operated continuously until a planned maintenance stop. In the top layer, we repeatedly solve a dynamic optimization problem to find the optimal long-term operation strategy, subject to load-induced compressor degradation. The supervisory control layer below receives the computed setpoints and operational parameters, and applies them in a self-optimizing control structure to ensure near-optimal operation in the presence of disturbances. The regulatory control layer in the bottom stabilizes operation in an otherwise unstable operating region (surge). We show the efficacy of our health-aware operation approach by comparing it to traditional control structures where the equipment health is not explicitly considered as part of the production optimization. Our approach results in higher average production, without jeopardizing the health of the system.

5.2 Introduction

Unplanned maintenance intervention of subsea systems are costly, so it is necessary to ensure that operation does not reduce the system reliability to unacceptable levels. Traditionally, this has been achieved by introducing large safety margins and enforcing conservative operational strategies. Better economical performance can be achieved by employing prognostics and health monitoring (PHM), which means that the system state is monitored and projected into the future. A natural extension of PHM is health-aware control, in which we combine control and reliability objectives, yielding a control structure that maximizes plant profitability while keeping the plant health within acceptable limits [152, 179].

Health-aware control is achieved by repeatedly solving a shrinking horizon dynamic optimization problem to find an operating strategy based on the current compressor health and its predicted development. The time horizon is from the present until the next planned maintenance intervention, and the objective is to maximize the profit subject to health constraints. The dynamics of this layer are on the timescale of weeks to months. On a more frequent basis, disturbances are rejected by a supervisory control layer in order to keep operation close to the desired (optimal) operating point. We use self-optimizing control ideas [161] to achieve this. The lowest and fastest control layer is in charge of surge control. Surge is an unwanted mode of operation characterized by limit-cycle oscillations in flow and pressure, which can harm the internals of the compressor [107]. Traditionally, operation is restricted by a generous safety margin from the surge line. However, it is often desirable to operate closer to the surge line, as this leads to increased efficiency and lower operating costs. An alternative to surge avoidance is active surge control. For this purpose, a close-coupled valve (CCV) is introduced to the system. Using the CCV, we can control the compressor characteristic, thereby stabilizing operation in an otherwise unstable region [70]. A feedback linearizing controller proposed by [17], is used for this purpose. An illustration of the proposed control structure is shown in Figure 5.1.

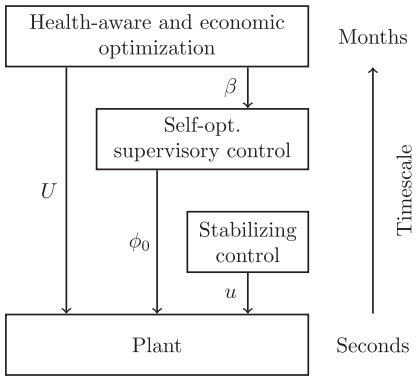

Figure 5.1: Multi-layer control structure for stable, health-aware operation.

The main contributions of this paper are the following: 1) We propose a three level

control structure for health-aware control of a compression system. 2) We show that the method outperforms traditional control methods

5.3 Model description

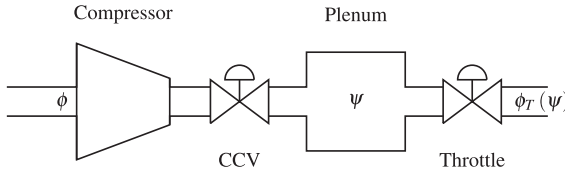

Figure 5.2: Flowsheet of the Greitzer compressor model

5.3.1 Short timescale dynamics: Surge

The surge model used here is that of a centrifugal compressor with an added CCV for surge control, which as described by [160]. We use the transformed version of the model presented by [70], by which the system can be described in terms of the non-dimensional compressor mass flow ϕ and the non-dimensional pressure rise across the plenum, ψ. A detailed description and derivation of the model is given in [70], but a summary is given below for completeness. An illustration of the system is given in Figure 5.2. Three degrees of freedom are available for control in the system: the compressor speed, the CCV opening and the throttle opening.

The two-state Greitzer model is given as:

$$\dot{\hat{\phi}} = B\left[\hat{\Psi}_C\left(\hat{\phi}\right) - \hat{\psi} - u\right] \tag{5.1}$$

$$\dot{\hat{\psi}} = \frac{1}{B}\left[\hat{\phi} - \hat{\Phi}_T\left(\hat{\phi}\right)\right], \tag{5.2}$$

where the $(\hat{\cdot})$-symbol is used to denote deviation from the specified operating points, $\hat{\phi} = \phi - \phi_0$ and $\hat{\psi} = \psi - \psi_0$. (ϕ_0, ψ_0) is the specified operating point. In the above expression, B is the Greitzer parameter, which is proportional to the compressor blade tip speed U, $B = kU$, where k is a geometry-dependent constant. $\hat{\Psi}_C$ is the cubic approximation of the axisymmetric compressor characteristic, $\hat{\Phi}_T$ is the throttle characteristic, and the input u is the pressure drop across the CCV (as determined by its opening). The compressor and throttle characteristics is shown in Figure 5.3.

The compressor characteristic $\hat{\Psi}_C$ indicates the pressure rise for a given flow, and is unique for every compressor. The characteristic is approximated by the cubic

$$\hat{\Psi}_C\left(\hat{\phi}\right) = -k_3\hat{\phi}^3 - k_2\hat{\phi}^2 - k_1\hat{\phi}, \tag{5.3}$$

where $k_1 = \frac{3H\phi_0}{2W^2}\left(\frac{\phi_0}{W} - 2\right)$, $k_2 = \frac{3H}{2W^2}\left(\frac{\phi_0}{W} - 1\right)$ and $k_3 = \frac{H}{2W^3}$. H and W are equipment-specific parameters relating to the peak and valley points of the com-

pressor characteristic. The peak point, $(\phi^*, \psi^*) = (2W, \psi^*)$, is assumed to be the surge point, with all points left of the peak being unstable.

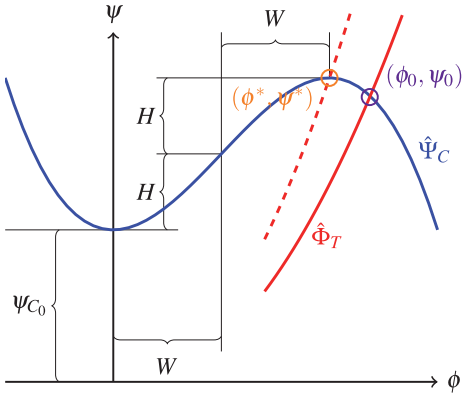

Figure 5.3: Cubic approximation of the axisymmetric compressor characteristic (blue) and the throttle characteristic (red). The operating point (ϕ_0, ψ_0) is shown in purple and the surge point $(\phi^*, \psi^*) = (2W, \psi^*)$ is shown in orange. The surge line (surge point at various compressor speeds) is shown as the dashed line.

The throttle characteristic is given as

$$\hat{\Phi}_T\left(\hat{\psi}\right) = \frac{\phi_0}{\sqrt{\psi_0}}\left(\operatorname{sign}(\psi)\sqrt{|\psi|} - \sqrt{\psi_0}\right). \tag{5.4}$$

The intersection between $\hat{\Psi}_C$ and $\hat{\Phi}_T$ gives the operating point (ϕ_0, ψ_0), shown in purple in Figure 5.3.

5.3.2 Long timescale dynamics: Compressor degradation

Variations in pressure and flow rate (as caused by surge) lead to radial vibrations, axial thrust displacement and a large temperature rise. This will in turn damage bearings, blades and other internal components [107]. We lump the accumulated damage on all internal components into a health indicator state x, whose propagation is modelled as

$$\dot{x} = \mathbb{E}\left(p_1\phi_0 + \int_0^\infty \left(p_2\left|\hat{\phi}(d)\right| + p_3\left|\dot{\hat{\phi}}(d)\right|\right)dt\right), \tag{5.5}$$

where $\mathbb{E}$ is the expected value operator, $d = \begin{bmatrix} W & H & \psi_{C_0} \end{bmatrix}^\intercal$ are independently normal distributed disturbances and p_i are weights. The p_1-term is the damage caused by regular operation, which is proportional with the throughput. The harder the compressor is run (in terms of throughput), the more rapidly it degrades. The p_2-term is damage caused by oscillations in pressure and flow, caused by surge. The p_3-term accounts for high-frequency oscillations, as these are thought to be more harmful to the compressor than low-frequency oscillations.

5.4 Hierarchical control structure for the subsea compressor

Due to the large difference in time scales for the problem, it is natural to divide it into several timescale-separated layers. The lowest layer stabilizes operation, the middle layer rejects disturbances, and the top layer is used to ensure reliable operation. Three degrees of freedom are available to achieve this: the pressure drop over the CCV, u, the blade tip speed, U, and the flow through the compressor, ϕ_0, as determined by the throttle. The three layers are described in more detail in the following subsections.

5.4.1 Stabilizing control layer

The purpose of the lowest control layer is to stabilize the compressor beyond the surge line. For this purpose, we use a feedback linearizing controller which adjusts the CCV. Feedback linearizing control enables controlling non-linear systems with a linear control law, allowing for higher sampling frequencies due to the reduced computational complexity. Since the surge phenomenon happens on a short time scale, while simultaneously being non-linear, the use of feedback linearization is appropriate. We use the feedback linearizing controller presented by [17]. A full description and derivation of the control law is given there.

The proposed feedback linearizing controller for the CCV is:

$$u = \mu_1 \hat{\phi} + \mu_2 \hat{\psi}, \tag{5.6}$$

where u is the pressure drop across the CCV and μ_1 and μ_2 are controller tuning parameters.

5.4.2 Supervisory control layer (Local disturbance rejection: Self-optimizing control)

After stabilizing the system with the CCV, we can optimize operation by adjusting the flow through the system. The operational objective is to maximize the compressor efficiency, but operation too close to the surge point is penalized.

$$\min_{\phi_0, U} \qquad J^{SOC} = -\eta(\phi_0) + \beta\,(2W - \phi_0)\,. \tag{5.7}$$

In the above expression η is the efficiency and β is the penalty weight. Using self-optimizing control [81, 161], we can keep the operation such that it is near optimal in the sense of Problem (5.7) by controlling a combination of carefully chosen plant measurements y, to a predetermined set-point:

$$c = H^{SOC} y \tag{5.8}$$

In this case, the plant measurements are augmented by disturbance measurements, $d = \begin{bmatrix} W & H & \psi_{C_0} \end{bmatrix}^\intercal$, such that $y = \begin{bmatrix} \phi & \psi & d \end{bmatrix}^\intercal$. A measurement combination

matrix H^{SOC} that can be shown to minimize the average loss $L = J(\phi_0, d) - J^{opt}\left(\phi_0^{opt}, d\right)$ is [190]

$$\left(H^{SOC}\right)^{\intercal} = \left(YY^{\intercal}\right)^{-1} G^{y}, \tag{5.9}$$

where

$$Y = \left[FW_d \quad W_{n^y}\right], \tag{5.10}$$

and $G^y = \frac{\partial y}{\partial \phi_0}\Big|_{\phi_0^{nom}}$ is the linearized system model evaluated at the nominal operating point, $F = \frac{dy^{opt}}{dd}$ is the optimal sensitivity matrix, and W_{n^y} and W_d are diagonal matrices of appropriate sizes with the variances of the measurement errors / noise n^y and the variances of d.

5.4.3 Optimal economic and reliable operation

In the top control layer, we devise a dynamic real-time optimization (DRTO) scheme to calculate the optimal compressor speed U and penalty weight β for the SOC layer. The purpose of this layer is to adjust operation for the other layers to ensure both economic optimality and satisfaction of operational and reliability constraints. At each time step we solve the following dynamic optimization problem

$$\min_{\beta, U} \quad J^{DRTO} = -\int_0^{t_f} NPV(\phi_0)dt = -\int_0^{t_f} \frac{\phi_0}{(1+i)^t}dt \tag{5.11}$$

$$s.t. \quad x < x_{max} \tag{5.12}$$

$$\psi_{out} > \psi_{out,min}, \tag{5.13}$$

where NPV signifies the net present value with discount rate i and x is the degradation from Equation (5.5). $\psi_{out} = \hat{\psi} + \psi_{res}$ is the outlet pressure from the compressor.

5.5 Simulations

The system described in Section 5.3 and the control structure described in Section 5.4, are implemented in MATLAB/Simulink and Casadi 3.0.0 [10]. IPOPT 3.12.3 [185] is used to solve the optimization problems Problem (5.7) and Problem (5.11).

5.5.1 Stabilizing control

Figure 5.4 shows the response of the system with the surge controller turned off (solid blue line) and with the surge controller turned on (dashed blue line) to a step change in ϕ_0. After the step, the new set-point lies within the unstable operating region, causing the limit cycle behavior in the uncontrolled case.

5.5.2 Local disturbance rejection (Self-optimizing control)

Figure 5.5 shows the response of the SOC structure (open loop (OL) and closed loop (CL)). As can be seen, the CL structure drives operation back to the optimal

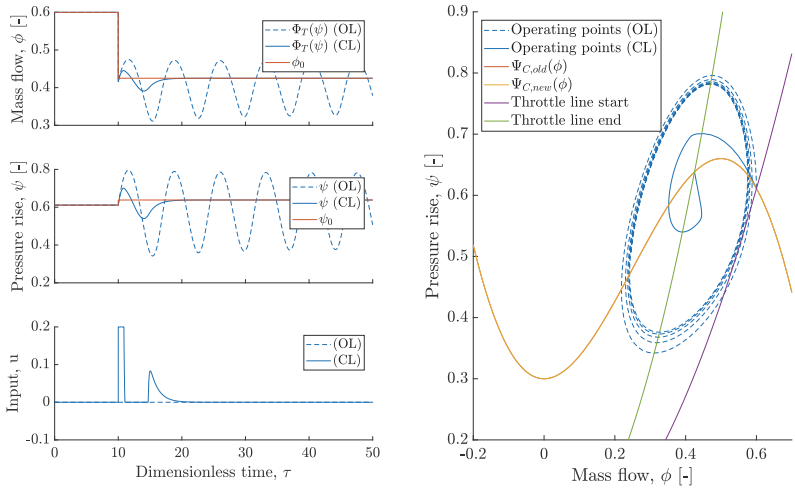

Figure 5.4: Closed-loop (CL) and open-loop (OL) responses to a set-point change in ϕ_0 int

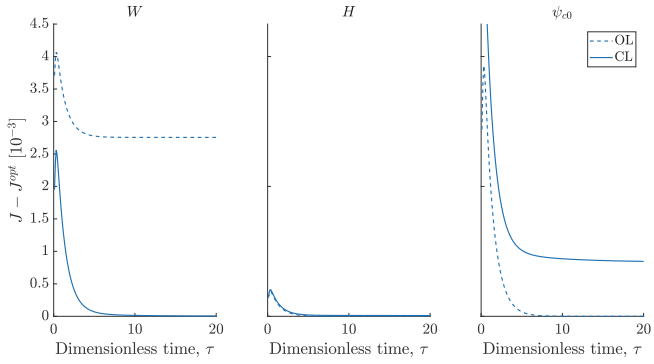

Figure 5.5: Open loop (OL) and closed loop (CL) losses for the SOC structure for the disturbances W, H and ϕ_{c_0}

point. The OL structure, while stable thanks to the surge controller, does not. Operation continues at a sub-optimal operating point, resulting in higher cost. Note that for the simulated disturbance in H and ψ_{c_0}, the steady state loss for the CL structure is higher than that of the OL structure. On the other hand, a step in W results in a lower loss, illustrating that it is the average loss that is minimized by Problem (5.9), not the loss for each individual disturbance.

5.5.3 Optimal economic and reliable operation

We consider three cases of DRTO. The DRTO1 and DRTO2 do not take the degratation constraint into account, and differ in terms of the maximum allowable shaft speed. DRTO1 allows higher shaft speed. DRTO2 is more conservative with a lower maximum allowable shaft speed. The DRTO3 is health-aware and does not have constraints on the shaft speed, but instead ensures that the degradation is not exceeded. The closed-loop responses of the DRTOs are shown in Figure 5.6.

It can be seen that operation is adjusted to maximize the NPV of the production in all three cases by gradually reducing the production over time. However, only the "conservative" DRTO2 with the lower maximum allowable speed and the health aware DRTO structures satisfy the reliability constraints. The non-health-aware DRTOs do not "see" the compressor degradation. The system is disturbed at around $t = 1.5$ and again at $t = 2.5$, by stepping first up, then down in the degradation speed, to show that the health-aware control structure takes into account the updated health information.

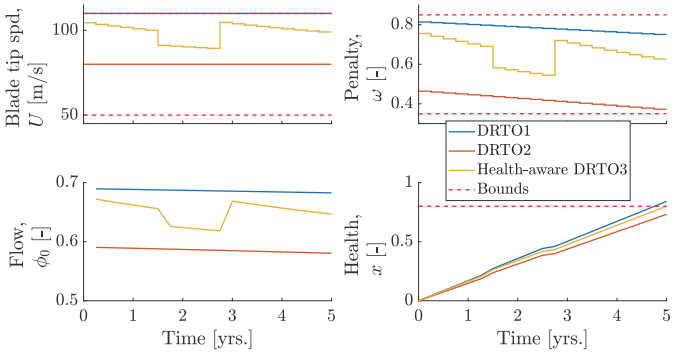

Figure 5.6: Closed loop responses of the regular DRTOs and the health-aware DRTO to disturbances in degradation speed.

5.6 Concluding remarks and future work

We have proposed a control structure for a compression system subject to long-term load-induced degradation. By using time scale separation it is possible to counteract surge and reject disturbances, while also achieving long term optimality and satisfaction of reliability constraints. We have shown that the proposed method is better than a "regular" DRTO scheme, in which reliability considerations are not taken into account when planning future production.

Several assumptions have been made in this work: we assume perfect state feedback for the DRTO, meaning that we can measure the health indicator state directly and without errors. This is somewhat unrealistic. In practice, we need to estimate the health indicator from other measurements. In the DRTO, we did not use parameter estimation to adapt the model when the operating conditions changed. Finally, the DRTO has to be made robust towards model uncertainty by formulating a robust/stochastic optimization problem. This will be addressed in future work.

Chapter 6

Oil Production Optimization of Several Wells Subject to Choke Degradation

This chapter is from the article

- A. Verheyleweghen and J. Jäschke. Oil Production Optimization of Several Wells Subject to Choke Degradation (IFAC Workshop on Automatic Control in Offshore Oil and Gas Production (OOGP), Esbjerg, Denmark). *IFAC-PapersOnLine*. volume 51(8), pages 1—6, 2018

6.1 Abstract

Unplanned maintenance interventions of subsea oil and gas production systems are very expensive, which leads to strict requirements to equipment reliability. Without a systematic way to ensure reliable operation however, a very conservative operational strategy is often chosen, which can lead to sub-optimal operation and the loss of large potential profits. We propose to integrate condition monitoring and prognostics into the production planning problem to reduce conservativeness by actively steering plant degradation and preventing violation of health-critical constraints. We achieve this by combining equipment degradation models with regular process models and solving a shrinking horizon real-time optimization problem until the next planned maintenance horizon. A network of oil and gas producing wells with artificial gas lift, subject to particle induced choke erosion is used as a case example.

6.2 Introduction

In this paper we consider an oil and gas production network consisting of multiple wells. The wells are connected to a common manifold, from which the combined flow goes through a riser to a topside receiving facility. As the field matures, the reservoir pressure decreases. Eventually, the pressure might drop to such low levels

that fluids can no longer overcome the resistance in the riser, and production comes to a stop. Artifical gas lift can be used to reduce the pressure drop and increase the flow, prolonging the lifetime of the field.

However, increased volume flows (and consequently velocities), in addition to the decreased density, may lead to accelerated degradation of vulnerable parts of the system. In particular, erosion of chokes and bends may be a problem, especially if the sand production from the reservoir is high. Particle erosion can severely limit the remaining useful life of exposed equipment. In rare cases, sand erosion has been known to erode away critical components such as chokes in as little as a few hours [73]. Choke replacement frequencies of 3-4 months, though having significant costs associated with them, are not unheard of in the subsea industry.

Sand production generally tends to increase as the field matures and reservoir pressure decreases, though it can also pose problems in some green fields. It is consequently vital to consider potential sand erosion when deciding on a production strategy, in order to prevent breakdowns which require costly unplanned maintenance intervention. Common industrial practice is to define an acceptable sand rate (ASR) above which operation is not permitted. The ASR is often conservatively defined in order to account for worst-case erosion scenarios. The operational degrees of freedom for production optimization are consequently severely constrained, leading to sub-optimal operation.

The conservativeness can be reduced by monitoring the rate of erosion on critical components real time and adjusting operation to reflect equipment integrity. Monitoring usually involves periodic inspection of weight loss coupons. Real-time erosion monitoring systems, such as ABBs INSIGHT [2], exist, but are not yet widespread in industry. These systems are usually not integrated with the control system. Set-points of the control system must still be manually adjusted by the operator. This dependency on the operator can lead to delays, manual overrides and overall reduced efficiency of the production system.

In this paper, we use a *health-aware real-time optimization* (RTO) approach, in which health monitoring and prognostics is included in the decision making process to find the optimal operational strategy without jeopardizing equipment health [179]. Specifically, we formulate the problem of optimal operation as a dynamic optimization problem where the objective is to maximize the overall profit of the plant, without violating constraints on the maximum allowable choke erosion. We also show how uncertainties in the model parameters can be taken into account by formulating the problem of optimal operation as a worst-case / min-max optimization problem or a multi-stage stochastic optimization problem. We implement both methods and solve the problem repeatedly in a shrinking-horizon, RTO-like fashion.

The remainder of the paper is structured as follows: In Section 6.3 we give a process description for the gas lifted well network. In Section 6.4 we formulate the optimization problem and explain how uncertainty is treated. Simulation results are presented and discussed in Section 6.5. Finally, concluding remarks are given and future work is described in Section 6.6.

6.3 Process description

The model for the oil and gas production system used in this work is based on the model by [89]. An illustration of the process is given in Figure 6.1. A full description of the model is given there, but for the sake of completeness, we provide a summary below. The model was modified slightly in the following ways:

1. The model uses a larger time horizon since our aim is to do health-aware RTO, which requires the time horizon to capture the degradation dynamics. We therefore assume that changes in mass flow rates are instantaneous, resulting in constant mass hold ups. The dynamics in our work are instead dictated by gradual choke degradation and slow decline of reservoir pressure.

2. The model considers a three-phase system consisting of oil, gas and water.

3. The model is extended to include three wells and a riser

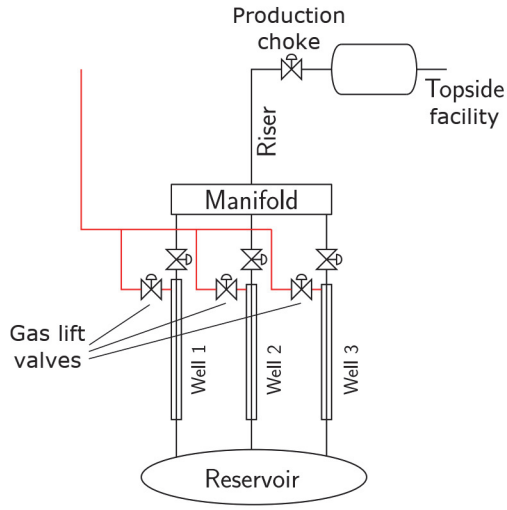

Figure 6.1: Illustration of the oil and gas network with artificial gas lift.

Gas injection at the bottom of the well lowers the average fluid density, thereby reducing the hydrostatic pressure drop in the well. As a result, the bottom hole pressure and consequently the flow from the reservoir increases, until a certain point. Too large gas injection rates result in increased frictional pressure drop due to increased velocities. We define the short term optimal gas injection rate (with respect to oil and gas production), as the point at which the marginal frictional pressure drop is balanced by the marginal hydrostatic pressure drop. As we shall see later, the increased velocities lead to more rapid degradation, which might force us to operate at lower-than-short-term-optimal gas injection rates.

6.3.1 Process model

The steady-state mass balances in each well are

$$\dot{m}_{pg} = \dot{m}_{rg} + \dot{m}_{lg} \tag{6.1}$$
$$\dot{m}_{pl} = \dot{m}_{rw} + \dot{m}_{ro} \tag{6.2}$$
$$\dot{m}_{p} = \dot{m}_{pg} + \dot{m}_{pl}, \tag{6.3}$$

where $\dot{m}_{lg}$ is the flow rate of lift gas through the annulus, $\dot{m}_{rg}$ is the flow rate of gas from the reservoir, and $\dot{m}_{pg}$ and $\dot{m}_{pl}$ are the flow rates of produced gas and liquid respectively. The liquid flow $\dot{m}_{rl}$ is the sum of the flow of water $\dot{m}_{rw}$ and the flow of oil $\dot{m}_{ro}$ from the reservoir. Finally, the total flow rate through the production choke is $\dot{m}_{p}$. Adjusting the gas lift rate and the total flow through the production choke is achieved by opening and closing the valves. The flow rates can then be expressed in terms of the valve equation:

$$\dot{m}_{p} = C_{pc}\sqrt{\rho_w \left(p_{wh} - p_m\right)} \tag{6.4}$$
$$\dot{m}_{lg} = C_{lg}\sqrt{\rho_a \left(p_a - p_{wi}\right)}. \tag{6.5}$$

Here, C_{pc} and C_{lg} are the valve coefficients of the production choke and the lift gas valve respectively, and ρ_w and ρ_a are the fluid densities in the well tubing and in the annulus. Pressures p driving the flow are denoted by wh for wellhead, m for manifold, a for annulus and wi for well injection point.

Assuming that the ideal gas law can be applied here, we express the density of the gas in the annulus as

$$\rho_a = \frac{M p_a}{T_a R} \tag{6.6}$$
$$= \frac{m_a}{\left(A_a^2 - A_w^2\right) L_a}, \tag{6.7}$$

where M is the molar mass of the lift gas, T_a is the temperature in the annulus and R is the universal gas constant. The average density in the well tubing is

$$\rho_w = \frac{m_{gt} + m_{lt} - \rho_l L_r A_r}{L_w A_w} \tag{6.8}$$
$$\rho_l = WC\rho_w + (1 - WC)\rho_o. \tag{6.9}$$

In the above expressions m_a, m_{gt} and m_{lt} are the hold-ups of gas in the annulus, and hold-ups of gas and liquid in the tubing, L_r and A_r are the length and cross-sectional area of the tubing above the gas injection point, and L_w and A_w are the length and cross-sectional area of the tubing below the gas injection point.

The flow from the reservoir is given by

$$\dot{m}_{rl} = PI \cdot \left(p_r - p_{bh}\right) \tag{6.10}$$
$$WC = \frac{\dot{m}_{rw}}{\dot{m}_{rl}} \tag{6.11}$$
$$\dot{m}_{rg} = GOR \cdot \dot{m}_{ro}, \tag{6.12}$$

where

$$p_r = \frac{m_{rg} RT}{V_r}. \tag{6.13}$$

Above, PI is the productivity index, WC is the water cut, GOR is the gas-oil-ratio, and p_r is the reservoir pressure. These are well-specific parameters.

Finally, the well pressures are decreasing as the reservoir is slowly depleting. We model the reservoir as a storage tank, yielding

$$\frac{dm_{rg}}{dt} = -\dot{m}_{rg} \tag{6.14}$$

6.3.2 Choke degradation model

Choke erosion rates depend on a number of different factors, such as physical properties of the fluid and the impacting particle. In addition, erosion rates are heavily dependent on the choke geometry, as this will influence the flow patterns. It is therefore a challenging task to predict the erosion rates for a given choke, without expensive computational fluid dynamics (CFD) simulations. [44] give an overview over some erosion prediction models for simple choke geometries, based on which they recommend ASRs. We use the erosion model presented in [44], which is a variation of the model presented in [73]. The erosion rate is given as

$$\frac{dE}{dt} = \frac{K \cdot F(\alpha) \cdot U_p^n}{\rho_t \cdot A_t} \cdot G \cdot C_1 \cdot GF \cdot \dot{m}_{sand} \cdot C_{unit} \tag{6.15}$$

where $\frac{dE}{dt}$ is the erosion rate in mm/yr., K, n, C_1, GF and C_{unit} are various constants. $\dot{m}_{sand}$ is the sand production rate and G is defined as

$$G = \frac{d_p \cdot \beta \cdot (1.88 \cdot \log(A) - 6.04)}{D_{pipe}} \tag{6.16}$$

where d_p is the particle diameter, and D_p is the pipe diameter. β and A are dimensionless parameters

$$A = Re \cdot \frac{\tan(\alpha)}{\beta} \tag{6.17}$$

$$\beta = \frac{\rho_p}{\rho_f} \tag{6.18}$$

where Re is the Reynolds number of the flow, ρ_p is the particle density and ρ_f is the fluid density.

The sand production rate $\dot{m}_{sand}$ is assumed to be proportional to the overall mass flow rate from the reservoir:

$$\dot{m}_{sand} = SR \cdot \dot{m}_r, \tag{6.19}$$

where SR is the sand rate parameter. Furthermore, in Equation (6.15), F is the ductility of the choke gallery material, which is

$$\begin{aligned} F = 0.6 \cdot \left[\sin(\alpha) + 7.2\left(\sin(\alpha) - \sin^2(\alpha)\right)\right]^{0.6} \\ \cdot \left[1 - \exp(-20\alpha)\right] \end{aligned} \tag{6.20}$$

for ductile materials. Here, α is the particle impact angle, which is given as

$$\alpha = \arctan\left(\frac{1}{\sqrt{2R}}\right), \tag{6.21}$$

with R being the radius of the choke gallery. U_p is the particle impact velocity, which is determined by

$$U_p = \frac{3 \cdot Q}{4 \cdot A_g} = \frac{3 \cdot Q}{8 \cdot H \cdot D}, \tag{6.22}$$

where Q is the actual volumetric flow rate, A_g is the effective gallery area, H is the effective height of the gallery and D is the gap between the choke cage and choke body.

6.4 Optimizing economic performance subject to health constraints

By combining the process model and the health degradation model described in Section 6.3, the combined DAE model can be used to formulate an optimization problem in which the economic performance is maximized subject to constraints on the maximum allowable health degradation. In previous work [178], we have shown that failing to include the constraints on health degradation will lead to unreliable operation, since this constraint always will be active in the optimal solution for the operation strategy.

The health state of the plant is assumed to be known at any given time, meaning that real-time erosion monitoring systems are installed and working.

The optimization problem which is solved at each RTO iteration can be written as:

$$\min_{\mathbf{u}} \int_{t_0}^{t_f} \phi\left(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{p}\right) dt \tag{6.23a}$$

$$\text{s.t.} \quad f(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{p}) \leq 0 \tag{6.23b}$$

$$g(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{p}) = 0 \tag{6.23c}$$

where ϕ is the objective function which is to be minimized, and f and g are the inequality constraints and equality constraints. The variables $\mathbf{x}$, $\mathbf{z}$ and $\mathbf{u}$ denote the differential states, algebraic states and inputs, respectively. $\mathbf{p}$ is used to denote the uncertain parameters.

The dynamic problem from Problem (6.23) is discretized and solved with orthogonal collocation with three collocation points for each finite element [27]. The discretized problem can be written as

$$\min_{\mathbf{u}} \sum_{k=1}^{N} \phi\left(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{p}_k\right) \tag{6.24a}$$

$$\text{s.t.} \quad f(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{p}_k) \leq 0 \quad \forall k = 1...N \tag{6.24b}$$

$$g(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{p}_k) = 0 \quad \forall k = 1...N \tag{6.24c}$$

where N denotes the horizon length.

6.4.1 Uncertainty handling

To account for plant-model mismatch / parametric uncertainty, or intrinsic stochasticity of the system, we consider some of the variables (denoted $\mathbf{p}$ in Problem (6.23)) to be stochastic. In particular, it is assumed that the sand production rate SR and the productivity index PI in each of the three wells are stochastic. For simplicity, we assume that the nine uncertain variables are independent and normally distributed, $\mathbf{p}_k \sim \mathcal{N}(\mu_k, \sigma_k)$.

Various approaches for optimization under uncertainty are found in literature. Two of the most popular approaches are worst-case optimization and scenario-based optimization.

Worst-case optimization; stochasticity is acknowledged by substituting in the worst-case realizations in the uncertain parameters. If constraints are satisfied for the worst-case realization, they should also hold for other parameter realizations, for most cases. Though it can be shown that the worst-case solution may be infeasible for other parameter realizations, this approach has been successfully demonstrated for a number of practical applications.

Scenario-based optimization; in which the probability distribution of the uncertain parameters is discretized into a finite number of scenarios and incorporated into the optimization problem in the form of a scenario tree. An illustration of a scenario with four scenarios is shown in Figure 6.2.

By optimizing for all scenarios simultaneously, it is ensured that the obtained solution is not only feasible for the worst-case realization or expected realization, but for all possible realizations in the scenario tree. Furthermore, the degree of sub-optimality of the solution can be reduced by weighing the individual scenarios with their respective probabilities. This leads to a solution that is, on average, less conservative than the worst-case approach. Possibility of future recourse is included in the optimization by design of the scenario tree, which makes this method well suited to RTO problems under uncertainty.

The drawback of the this method compared to the two others is the increased problem size and consequent computation time, due to the need for additional variables for each scenario.

Worst-case and scenario-based RTO maybe classified under the umbrella-term of robust optimization, in which parameter realizations are assumed to occur within bounded uncertainty sets. These approaches are perhaps the easiest to grasp conceptually, but other ways to handle uncertainty (such as chance constrained optimization, dynamic programming, and fuzzy programming) exist in literature. We will however only consider worst-case and scenario optimization in this work.

6.4.2 Worst-case optimization

The worst-case optimization problem can be written as

$$\min_{\mathbf{u}} \sum_{k=1}^{N} \phi\left(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{p}_k^*\right) \tag{6.25a}$$

$$\text{s.t.} \quad \mathbf{p}_k^* = \arg\max_{\mathbf{p}_k} \| f(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{p}_k) \| \quad \forall k = 1...N \tag{6.25b}$$

$$g(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \mathbf{p}_k) = 0 \quad \forall k = 1...N \tag{6.25c}$$

where $\mathbf{p}_k^*$ is the worst-case parameter realization, i.e. the scenario which leads to the largest constraint violation. Due to the two nested optimization problems, this approach is also known as min-max optimization. These problems are generally difficult to solve or even intractable [22]. In general, we must require $\mathbf{p}_k$ to be bounded for the inner problem to have a solution. In some cases, such as the one considered in this work, the worst case parameter realization $\mathbf{p}_k^*$ can be known a priori. This significantly simplifies the problem since the second optimization problem disappears.

6.4.3 Scenario-based optimization

In scenario-based optimization, we discretize the continuous distribution function into a finite number of discrete scenarios, and optimize the following objective

$$\min_{\mathbf{u}} \sum_{i=1}^{S} p_i \sum_{k=1}^{N} \phi\left(\mathbf{x}_{i,k}, \mathbf{z}_{i,k}, \mathbf{u}_{i,k}, \mathbf{p}_{i,k}\right) \tag{6.26a}$$

subject to the following constraints

$$\text{s.t.} \quad f(\mathbf{x}_{i,k}, \mathbf{z}_{i,k}, \mathbf{u}_{i,k}, \mathbf{p}_{i,k}) \leq 0 \quad \forall i = 1...S\,,\ k = 1...N \tag{6.26b}$$

$$g(\mathbf{x}_{i,k}, \mathbf{z}_{i,k}, \mathbf{u}_{i,k}, \mathbf{p}_{i,k}) = 0 \quad \forall i = 1...S\,,\ k = 1...N \tag{6.26c}$$

$$\sum_{i=1}^{S} \mathbf{A}_{i,k} \mathbf{u}_{i,k} = 0 \quad \forall k = 1...N \tag{6.26d}$$

where S is used to denote the number of scenarios, and p_i denotes the probability of realizing scenario i.

Equation (6.26d) are the so-called non-anticipativity constraints, which are needed to enforce non-anticipativity, i.e. making sure that the optimal solution does not depend on yet unrevealed information. The scenario tree shown in Figure 6.2 has $N = 20$ and $S = 4$, for example.

For the kind of scenario trees encountered in RTO problems, each branching represents the different parameter realizations due to uncertainty. One might expect

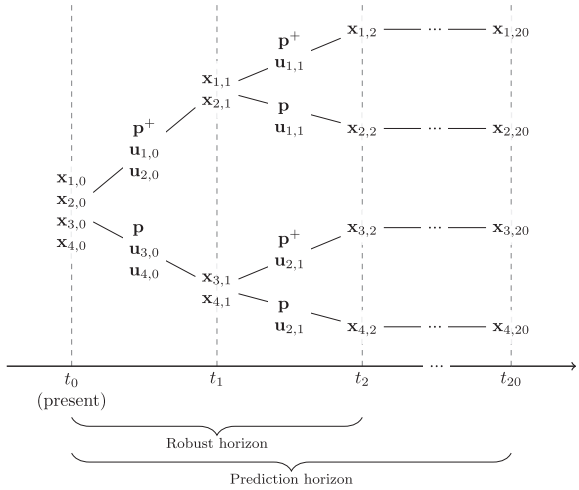

Figure 6.2: Scenario tree with $N = 20$ and $S = 4$.

that the branches from each node should be identical to the branches from its parent node. Alternatively, if the unknown parameters are estimated between each RTO iteration, this information can be included in the scenario tree by propagating the probability distribution into the future from each node and adjusting the parameter realizations of the child nodes according to the propagated probability distribution. In any case, this would lead to exponential growth of the scenario tree, with each scenario tree having n_r^N scenarios, where n_r is the number of discrete realizations of the probability distribution.

To avoid this explosive growth of scenarios, a robust-horizon $N_{robust} < N$, i.e. the stage until which branching occurs, is commonly defined [102]. By choosing a robust horizon shorter than the RTO horizon, we disregard the possibility of future recourse, and are consequently expected to get a sub-optimal solution. However, the loss is expected to be small, since the later stages of the RTO typically do not effect the objective much. A robust horizon longer than $N_{robust} = 1$ or $N_{robust} = 2$ is rarely used, since the marginal improvement of the solution in practice rarely justifies the increased dimensionality of the NLP. In this work, we only branch once, so $N_{robust} = 1$, which yields a two-stage stochastic program.

We further reduce the number of scenarios by limiting n_r, the number of discrete realizations used to approximate the continuous probability function. We generate the scenario tree similarly to what is proposed by [103], i.e. by using all possible combinations of the maximum and minimum uncertain parameter realizations, in addition to a scenario for the expected and nominal uncertain parameter realizations, for a total of $n_r = S = 65$ scenarios. The possible scenario realizations are given in Table 6.1.

Table 6.1: Possible realizations considered for the uncertain parameters in the scenario-based approach.

Variable	PI, wells 1...3	SR wells 1...3
Lower	$6.3 \cdot 10^{-6}$ [-]	$0.80 \cdot 10^{-2}$ [-]
Mean	$6.5 \cdot 10^{-6}$ [-]	$0.85 \cdot 10^{-2}$ [-]
Upper	$6.7 \cdot 10^{-6}$ [-]	$0.90 \cdot 10^{-2}$ [-]

6.4.4 Summary of scenario-based problem formulation

The objective is to maximize the profit, i.e. maximizing the oil and gas production and minimizing the cost of produced water and the cost of gas .

$$\min_{\mathbf{x}_{i,k},\mathbf{z}_{i,k},\mathbf{u}_{i,k}} \sum_{i=1}^{S=65} p_i \sum_{k=1}^{N} NPV\left(\phi\left(\mathbf{x}_{i,k},\mathbf{z}_{i,k},\mathbf{u}_{i,k}\right)\right) \tag{6.27a}$$

where

$$\phi = \sum_{well=1}^{3} c_g \dot{m}_{rg} + c_o \dot{m}_{ro} + c_{lg} \dot{m}_{lg}. \tag{6.27b}$$

Here, NPV is the net present value with a discount factor $r = 0.1$ and c_g, c_o and c_{lg} are the gas price, oil price and gas injection cost, respectively. We assume $c_g =$ 3 USD/MMBtu, $c_o = 44$ USD/bbl, $c_{lg} = 1.3$ USD/MMBtu. Bound constraints for the variables are given in Table 6.2. In addition come the non-anticipativity constraints and the model constraints from Problem (6.26) for the scenario-based approach. For the worst-case method, we have that Problems (6.25) and (6.26) are identical when $S = 1$ and the worst-case scenario $\mathbf{p}_k^*$ is bounded and known a priori. For this particular problem, we have that the worst case scenario occurs when both PI and SR are high in all three wells.

Table 6.2: Bound constraints

Variable	Lower bound	Upper bound	Unit
Choke opening, wells 1...3	0	1	[−]
Gas lift rate, wells 1...3	0	2.5	[kg/s]
Total gas lift rate	0	4.9	[kg/s]
Choke erosion, wells 1...3	0	0.5	[mm]

6.5 Results

We implemented the model in MATLAB using Casadi 3.0.0 [10] and solved the NLP from the discretized problem with IPOPT [185]. Both uncertainty handling strategies, i.e. worst-case RTO and scenario-based RTO, were implemented.

6.5.1 Problem

Problem (6.27) is solved repeatedly in a shrinking horizon fashion, starting with $N = 18$. After finding the optimal solution, only the first input is implemented on the actual plant, before the model is re-optimized. This process is illustrated in Figure 6.3, which shows the open-loop solution of the scenario-based optimization problem at three selected times $t = 0$, $t = 2$ and $t = 4$ years. The predicted states for the 65 scenarios are shown in solid lines, while past states are shown as dashed lines. It can be seen that the first inputs are identical for all scenarios due to the non-anticipativity constraints. Red, blue and yellow color distinguish the first well, second well and third well, respectively.

Figure 6.4 shows the closed-loop solution of the scenario-based method (solid line with circular markers) compared with the worst-case method (dotted line with cross markers). The total profit of the two operational strategies, in terms of Problem (6.27a), is 6.91 bn. USD for the scenario approach and 6.75 bn. USD for the worst-case approach. Although the actual numbers should be taken with a pinch of salt, the relative difference of approx. 2.5% is significant.

6.5.2 Discussion

In this work, we have assumed full state feedback, meaning that the initial values for the states are perfectly known. To simulate plant model-mismatch, the uncertain parameters are perturbed with random noise between each RTO iteration.

Since NPV of production is maximized, we see that early production is higher than late production. Due to depleting reservoir pressure, we also see the need for more lift gas as the field matures. However, due to the decreased density and consequent higher erosion rates, overall production must be throttled down to prevent choke failure, as only so much production can be permitted over the lifetime of the field.

6.6 Conclusion and future work

Our health-aware RTO framework combined diagnostics, prognostics and production optimization of a subsea gas-lifted oil and gas production network subject to sand particle induced choke erosion. We show that by combining a prognostic model for the choke erosion in the production optimization, we can make sure that critical erosion levels are not exceeded during operation, which means that the risk of costly unforeseen maintenance interventions is minimized. We also show that parametric uncertainty in the model should be handled with a scenario-based stochastic optimization approach, as this leads to better economic performance than the conservative min-max formulation that is commonly used.

The objective of the paper is to showcase the efficacy of our framework, rather than providing results which correspond 1:1 with real field data. We have therefore used a simple choke degradation model and reservoir model. It is understood that accurate models must be developed for the specific equipment in question before real-world implementation. These degradation models will have to be devel-

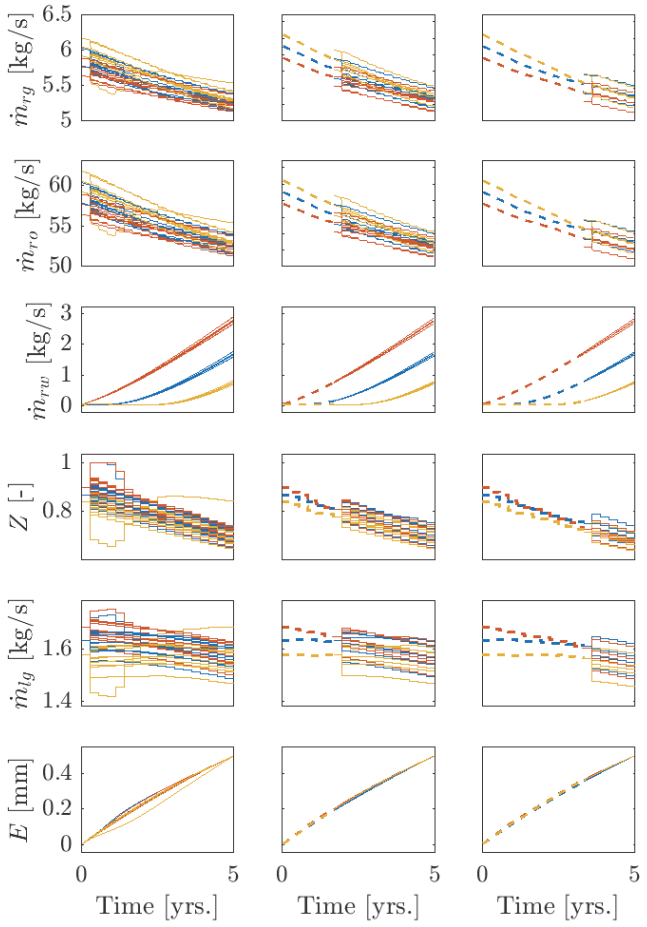

Figure 6.3: Three snapshots of the open-loop solutions of the scenario-based RTO at $t = 0$, $t = 2$ and $t = 4$ years. The red, blue and yellow are for the first, second and third wells, respectively. The dashed lines show the past states, while the solid lines show the predicted states for each of the 65 scenarios.

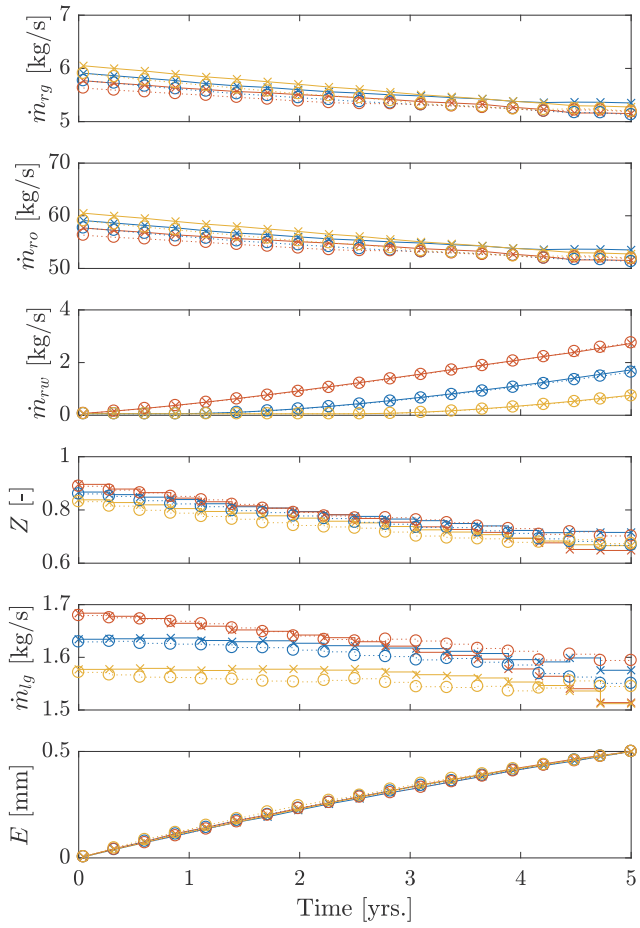

Figure 6.4: Closed-loop solutions for the compared approaches. The solid line with circular markers shows the scenario-based RTO, while the dotted line with cross markers shows the worst-case RTO. The red, blue and yellow scenarios are for the first, second and third wells, respectively.

oped by CFD simulations and/or in collaboration with equipment manufacturers. Furthermore, future work will address the issue of overall plant reliability vs. single component reliability, as correlation between failure modes may significantly impact the overall plant reliability.

Chapter 7

A Unified Approach for Simultaneous Optimization of Production and Maintenance Schedules

This chapter is from the article

- A. Verheyleweghen, H. Srivastav, A. Barros and J. Jäschke. A Unified Approach for Simultaneous Optimization of Production and Maintenance Schedules. *IEEE Transactions on Reliability*. (Under review), 2020

7.1 Abstract

Many systems in process industry experience degradation which is dependent on the usage of the system. At the same time, the systems generate some kind of profit when utilized, which typically increases the more the system is used. Often, this leads to a trade-off between maintenance costs and the profit associated with the utilization of the system. How to optimally operate the system in order to achieve a desired trade-off is a little-explored topic in literature. Common practice in many industries is to optimize operational decisions and maintenance decisions more or less independently of each other, subject to constraints to ensure feasibility of the obtained decisions. This leads to sub-optimal utilization of the system. In this paper, we propose a framework for formulating optimization problems which can be solved to obtain operational strategies which simultaneously optimize the trade-off between production profit and maintenance costs. We show how the discrete decisions pertaining to maintenance times can be approximated by a continuous model and solved using off-the-shelf optimization solvers. Because a global solution cannot be guaranteed using our proposed method, we employ a heuristic multi-start approach to find a near-global solution. We demonstrate the method on a case study inspired by subsea oil and gas production.

7.2 Introduction

In this paper, we propose an unified approach for simultaneous optimization of the production or the system load, and maintenance schedules. Within the process systems engineering community, the development of operational strategies that maximize some economic objective of a chemical production plant, subject to various constraints, is an important research question. How to optimally adjust operational degrees of freedom to achieve higher throughput, while still ensuring that the products are on-spec, is addressed on several time-scales and levels of system complexity, all the way from operation of single components to plant-wide operation [162].

Since plants like these inevitably degrade over time, maintenance interventions need to be performed. The optimization of maintenance schedules is also a much-explored research topic. Especially the field of condition-based maintenance (CBM) has attracted a lot of attention in recent years, both from academia and from industrial practitioners [7]. The main idea behind CBM is to monitor the degradation of the plant, so that maintenance can be performed when necessary according to some safety or economic criterion. This is turn leads to more efficient use of money, time and other resources by avoiding premature (preventive) maintenance. However, in certain industries, continuous process monitoring may not be possible due to lack of technology or prohibitive cost. One such example is subsea oil and gas production, where usually real-time equipment monitoring is currently not available due to lack of qualified instrumentation. Another thing that further complicates the issue is the fact that inspections are very costly, due to the need for special inspection remotely operated vehicles (ROVs). Depending on the location of the subsea installation, inspection might also be impossible to perform for long periods of time due to weather conditions. In such cases, there might be significant economical benefits to finding an optimum inspection and maintenance schedule in a rigorous manner.

In this work, we will therefore focus on such cases where continuous monitoring is not possible, and there is significant benefit to finding an optimal maintenance schedule. For these kind of applications, literature has traditionally focused on finding the optimum frequency of inspections for a periodic (equidistant in time) inspection schedule [84]. However, as we shall discuss later in this paper, periodic maintenance schedules are not optimal in general. One reason for the focus on period schedules in literature is that the tools commonly employed for finding these schedules are not well suited for the aperiodic case.

A major problem associated with maintenance optimization is that the problems tend to be stochastic. Since degradation of process plants is typically not deterministic (either due to inherent stochasticity of the process, or insufficient information), failure times and consequently maintenance costs are usually stochastic in nature. Analytical expressions for the distributions may exist, but could be difficult to express. Consequently, a popular approach to obtain the distributions is by repeated simulation of the system by the use of sampling (Monte Carlo) methods. Once the distribution has been obtained, we can use it to evaluate the maintenance schedule

according to some criteria, such as the expected cost or the value at risk.

During optimization of this criteria, we may need to evaluate the cost function multiple times. Every single one of these function evaluations may be very computationally expensive, due to the need for a large number of Monte Carlo samples when the dimensionality of the problem is large. If the aim is to find optimal non-periodic maintenance schedules, i.e. if maintenance can occur at any point in time and not just at fixed times (maintenance time is a continuous variable), the search space becomes infinitely large and many function evaluations are necessary for the optimization algorithm to converge. Extensions of traditional/textbook Monte Carlo algorithms have been developed to somewhat reduce the amount of samples necessary for each evaluation of the cost function by choosing "smart" sample points. Unfortunately, this only partially mitigates the problem, and evaluations of the cost function through the use of sampling methods tends to be computationally expensive for large problem sizes.

As an alternative to sample-based methods for the evaluation of the cost, we may use numerical integration, which is generally computationally cheap if the functions are sufficiently smooth and convex. An underlying assumption is that we are able to express the evolution of the probabilities as a differential equation. However, it is not necessary that a closed-form solution of the integral exists. The advantage of expressing the cost as a numerical integration is that we can subsequently use theory from the field of dynamic optimization to find the optimal maintenance schedule. Algorithms for numerical optimization can efficiently solve certain large-scale decision making problems, with hundreds of thousands or even millions of variables [67]. Sampling-based methods for evaluating the cost would be infeasible for these large-scale applications due to the sheer amount of cost function evaluations required.

Numerical approaches to combined production optimization and maintenance scheduling were previously addressed in works by e.g. [58, 77, 129, 180, 188]. Typically, the problem is formulated as a lot-allocation problem, meaning that the operation is split into different modes or lots of products, the sequence of which is to be optimized. Consequently, the problems end up being mixed-integer (non-) linear programs (MI(N)LP). Generally, MINLPs are very difficult to solve, with solution times being orders of magnitude larger than for NLP problems of comparable size and complexity. In general, solution time of MINLPs grow exponentially with the number of integer variables [19]. Although algorithms have been developed to deal with integer problems, we would rather avoid the use of integer variables (because this leads to much more difficult optimization problems) by reformulating the problem. In this work, we avoid the use of integer variables by formulating the problem using complementary constraints, which can be solved using off-the-shelf NLP solvers to obtain local solutions in polynomial time [19]. Furthermore, we consider the case where operation of the equipment can be adjusted in a continuous fashion, which is also a largely un-explored topic in reliability literature.

In this paper, we propose a unified approach for simultaneous optimization of the production and maintenance schedules. Note that because the problem is stochastic, we optimize a statistic of the probability density function. Here, we chose to

optimize the expected economical profit of the system by assigning an economic value to each of the discrete degradation states. Note also that the obtained strategies are only optimal in the expected sense, and not optimal for any one given realization of the uncertainties. We start by using a differential model to describe the degradation of the system and resetting the initial conditions each time maintenance is performed. We then show how this piece-wise, discontinuous model can be approximated by a non-smooth, continuous model which is more suitable for optimization.

Several methods for optimization of non-smooth problems exist in literature. Among the most popular methods are subgradients methods and bundle methods. See e.g. [71] for an in-depth discussion on the topic and the different methods. In our case, the non-smoothness stems from the introduction of complementary constraints. Using problem reformulations for mathematical programs with complementarity constraints (MPCCs) from [28], the resulting problem can be solved using of-the-shelf standard non-linear solvers such as IPOPT [185]. One disadvantage of using local solvers is that only locally optimal solutions can be guaranteed. Depending on the non-convexity of the problem, the local solutions may be significantly worse we use a multi-start approach to converge to a solution close to the global optimum in this paper. A subsea case example is used to demonstrate the proposed method.

The remainder of the paper is structured as follows; in Section 7.5, we use a small toy example to show how a non-periodic maintenance schedule can result in higher expected net profit than a periodic maintenance schedule, motivating the rest of the paper. Then in Section 7.3 we show how to derive the general continuous differential model. In Section 7.4.1 we show how to formulate an optimization problem to optimize the net profit of a system by adjusting its inspections and operation. We also show how a multi-stage problem formulation may lead to even better economic performance, at the cost of a larger problem size. The case study illustrating our method is presented in Section 7.7. Finally, concluding remarks are given in Section 7.8.

7.3 Modelling framework for degrading systems

7.3.1 Modelling degrading systems

By "degradation" of process systems, we mean the evolution of a health state or reliability state in time. Both the evolution in time and the evolution in space can be modelled either discretely or continuously. In reliability engineering, continuous-time, discrete-state models are the most common [93, 153]. Consequently, in this work, we limit ourselves to study systems whose degradation cannot be determined with arbitrary fidelity, but whose degree of degradation can be divided into n_y uniquely distinguishable, discrete degradation levels. The transition between these discrete degradation levels is stochastic in nature, and happens at randomly distributed times. Thus we can define a vector of probabilities

$$\boldsymbol{y} \in \left\{ \mathcal{Y} | 0 \leq y_i \leq 1, \sum_i y_i = 1 \right\} \subset \mathbb{R}^{n_y}, \tag{7.1}$$

where the entry y_i is the probability that the system is found in degradation level i at time t. The transition between the degradation levels, i.e. the degradation of the system over time, can be described in terms of the change in the associated probabilities. The change in probability is given by the following ordinary differential equation (ODE) system as

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, t) \tag{7.2a}$$

$$\boldsymbol{y}(t) = g(\boldsymbol{x}, t), \tag{7.2b}$$

where we refer to $f : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ as the state equation and $g : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ as the output equation. The function f is assumed to be piece-wise continuous in time and locally Lipschitz, meaning that at a point $\boldsymbol{x}_0$, there exists a neighborhood around $(\boldsymbol{x}_0, t)$, $N(\boldsymbol{x}_0, r) = \{\boldsymbol{x} \in \mathbb{R}^{n_x} | \|\boldsymbol{x} - \boldsymbol{x}_0\| < r\}$ within which f satisfies the Lipschitz condition.

$$\|f(t, \boldsymbol{x}_1) - f(t, \boldsymbol{x}_2)\| \leq L\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \tag{7.3}$$

where $L > 0$. If there exists a connected open subspace D in which all points are locally Lipschitz, then f is said to be locally Lipschitz in D.

Remark 7.1. Requiring this condition for f guarantees that the solution of the ODE (7.2) does not only exist, but is also unique. Uniqueness and existence of the solution is necessary to guarantee causal determinism. In other words, if we have perfect information about the system and the input-induced loads acting on it, we can predict the future probabilities for degrading. If we were not able to predict the future of the system, we would be unable to optimize decisions influencing the future of the system, as we shall do later.

In the following two examples, we will show what the ODE (7.2) may look like, based on three popular approaches to reliability modelling in literature: reliability block diagrams (RBDs), fault trees (FTs) and Markov chains (MCs). RBDs, FTs and MCs are all somewhat interchangeable modelling techniques, although they do have their strengths and weaknesses [6]. As long as we are able to express the degradation of the system in terms of a differential equation system, it does not matter if the degradation model starts out as an RBD, an FT or an MC.

Example 7.1 (Application to RBDs & FTs)**:**
A common way to represent complex multi-component systems in reliability engineering is through the use of reliability block diagrams (RBDs) [141]. In RBDs, the system reliability is represented as a combination of subsystems / components in parallel or series configuration. In a parallel configuration, all subsystems must fail before the system fails. Conversely, in a series configuration, only a single subsystem failure leads to overall system failure. In RBD analysis, the system reliability is found by checking whether there exists any path from beginning to end of the diagram. If no such path exists, then the system is considered to have failed. A RBD may be converted into a fault-tree (FT), i.e. a set of logic gates that describe the relationship between basic low level events and critical system level events such as failure. An illustration of some simple RBDs and their corresponding FTs is shown in Figure 7.1.

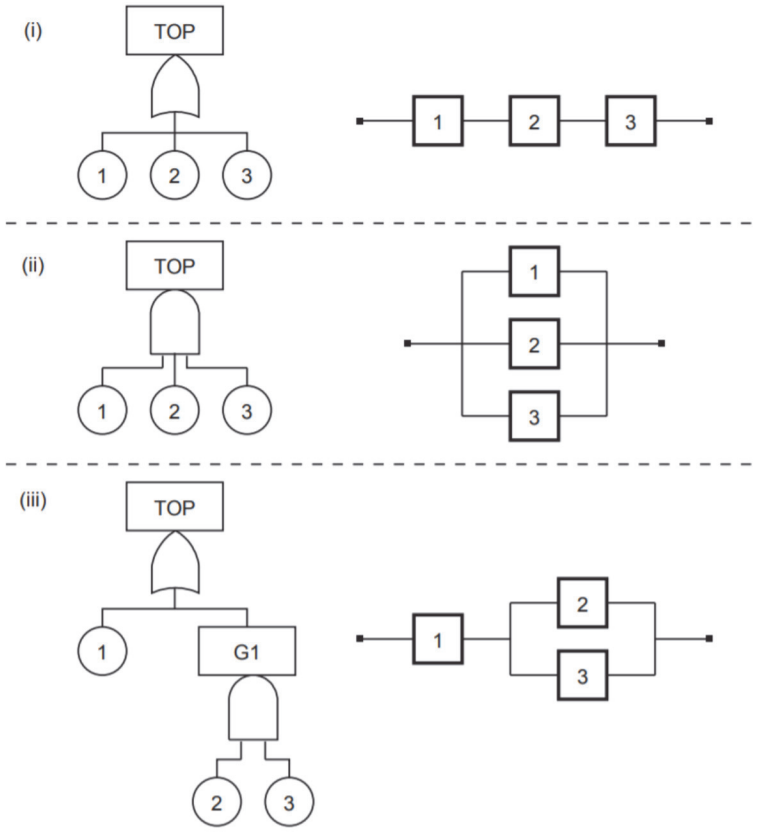

Figure 7.1: Illustration of reliability block diagrams (right) and corresponding fault trees (left). Taken from [141].

Since such a system is compartmentalizable, we split the state vector into k parts, each corresponding to one of the k subsystems.

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_k \end{bmatrix} \tag{7.4}$$

The deterioration of subsystem i is described by

$$\frac{d\boldsymbol{x}_i}{dt} = \tilde{f}_i(\boldsymbol{x}, t). \tag{7.5}$$

Note that under the condition that the deterioration of the k subsystems happens independently from one another, we can write

$$\frac{d\boldsymbol{x}_i}{dt} = \hat{f}_i(\boldsymbol{x}_i, t), \tag{7.6}$$

i.e. the deterioration of component i is only dependent on $\boldsymbol{x}_i$, and not dependent on $\boldsymbol{x}_{j\neq i}$. This independence makes the problem separable, and easier to threat numerically.

Let $\psi : \mathbb{R}^{n_{x_i}} \to \mathbb{R}^k$ be a function that calculates the subsystem reliabilites $\boldsymbol{\chi} \in \mathbb{R}^k$

$$\boldsymbol{\chi}(t) = \psi(\boldsymbol{x}) \tag{7.7}$$

ψ could be a selection function, or a more complex relationship between the component degradation states.

Then the system reliability, χ_s, is

$$\boldsymbol{y}(t) = \chi_s(t) = h(\boldsymbol{\chi}, t). \tag{7.8}$$

$h : \mathbb{R}^k \to \mathbb{R}^1$ is also known as the reliability polynomial *of the system. The reliability polynomial of a system describes the connection between components in a network of components. If we again assume that the k subsystems deteriorate independently, then it can be shown that the reliability polynomial of the system is equal to the* structure function *of the system. How to find this structure function in a systematic and efficient manner is a large research topic in itself, see e.g. [47, 143] for some binary decision diagram (BDD)-based methods that are commonly employed. Finding these structure functions or reliability polynomial functions is outside the scope of this article: we assume that they are known for the systems we are working with.*

For the example RBDs shown in Figure 7.1, the ODE describing the system degradation can be written as

$$\begin{aligned}
\frac{d\boldsymbol{x}}{dt} &= f(\boldsymbol{x}, t) \\
&= \begin{bmatrix} f_1(\boldsymbol{x}_1, t) \\ f_2(\boldsymbol{x}_2, t) \\ f_3(\boldsymbol{x}_3, t) \end{bmatrix} && \text{(7.9a)} \\
\boldsymbol{\chi}(t) &= \psi(\boldsymbol{x}, t) \\
&= \begin{bmatrix} \psi_1(\boldsymbol{x}_1, t) \\ \psi_2(\boldsymbol{x}_2, t) \\ \psi_3(\boldsymbol{x}_3, t) \end{bmatrix} && \text{(7.9b)} \\
\boldsymbol{y}(t) &= g_j(\boldsymbol{\chi}, t) && \text{(7.9c)} \\
&= \chi_s(t), && \text{(7.9d)}
\end{aligned}$$

where

$$\begin{aligned}
g_i &= \chi_1 \cdot \chi_2 \cdot \chi_3 && \text{(7.9e)} \\
g_{ii} &= 1 - (1 - \chi_1) \cdot (1 - \chi_2) \cdot (1 - \chi_3) && \text{(7.9f)} \\
g_{iii} &= \chi_1 \cdot (1 - (1 - \chi_2) \cdot (1 - \chi_3)) && \text{(7.9g)}
\end{aligned}$$

Note again that in this example, the subsystems/components degrade independently of each other, such that f_i and ψ_i are only functions of $\boldsymbol{x}_i$, and not $\boldsymbol{x}_{j\neq i}$.

Example 7.2 (Application to MC)**:**
A third, widespread method for describing the dependability or availability of a system is with Markov chains (MCs). In Markov chains, the system is divided into a set of discrete states. The transition between these states is stochastic, and the transition probabilities are Markovian, meaning that they only depend on the the current state of the system. For the case where the transition rates between the states are constant over time, the evolution of the distribution of states over time in a MC can be written mathematically as

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\boldsymbol{x} \tag{7.10}$$

where $\boldsymbol{A}$ is known as the transition matrix. In this case, the ODE (7.2) is a linear system.

7.3.2 Modelling the input-induced loads

As alluded to in the introduction, there are many examples of processes where the rate of degradation depends on the operation of the system. For example, [156] reports that the degradation of airplane gas turbine engines evolves exponentially, but the rate of degradation is impacted by the operating mode of the airplane, including parameters such as altitude, ambient temperature and throttle lever angle.

In marine vessels, propeller cavitation and fatigue due to cyclic loading are two phenomena which negatively influence the efficiency of a vessel propulsion system [35]. These degradation phenomena are influenced by the operating mode of the vessel, so that higher vessel speeds lead to faster degradation.

Let us denote by $\boldsymbol{u}(t) : \mathbb{R} \to \mathbb{R}^{n_u}$ the input-induced load on (or simply inputs to) the system. We assume now that the transition rates $f = f(\boldsymbol{x}, \boldsymbol{u}(t), t) = f(\boldsymbol{x}, t)$ are a function of these input-induced loads. Consequently, by changing $\boldsymbol{u}(t)$, we can actively steer the evolution of the system degradation in a desired direction.

If f is Lipschitz in $\boldsymbol{u}, t, \boldsymbol{x}$, then the solution $\boldsymbol{x}(t)$ exists and is unique [168].

Example 7.3 (Modelling of wear-out and infant mortality)**:**
If f is time-independent, this results in a constant failure rate for a constant input $\boldsymbol{u}$. Alternatively, a time-dependent function f allows us to model effects such as wear-out (accelerated rate of failure towards the end of a systems lifetime) or infant mortality (decreasing rate of failure at the beginning of a systems lifetime due to burn-in) [164].

A bathtub curve hazard function with infant mortality and wear-out proportional to the cumulative usage of the system can be modelled as

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, t) = \begin{bmatrix} v(t) \\ c \cdot u(t) \end{bmatrix} \tag{7.11a}$$

$$g(\boldsymbol{x}, t) = \chi_s(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \boldsymbol{x}(t) \tag{7.11b}$$

where c is a nonzero constant that determines the rate of degradation, $v(t) = \int c \cdot u(t)dt$ is the cumulative usage of the system, and

$$\boldsymbol{x}(t) = \begin{bmatrix} \chi_1(t) \\ v(t) \end{bmatrix} \tag{7.12}$$

7.3.3 Modelling repairs

In the context of this work, let us by "repairs" refer to the action of resetting the state of the system, i.e. shifting probability from one discrete health state to another. In practice, the probability of being in the degraded or failed state will be reduced and the probability for being in a healthy state will increase correspondingly, i.e. the system will be repaired. Note that since we are dealing with probabilistic systems with multiple discrete health states, the expectancy of being in the fully repaired state is not necessarily 1 after a maintenance intervention, depending on the chosen maintenance strategy. If our maintenance strategy is to only repair failed systems upon inspection, and to leave partially degraded systems as is, then the probability of these partially degraded systems will not be altered by maintenance interventions.

Repairs are assumed to be instantaneous, consequently they cause a discontinuity in the system state. Inspection refers to the action of inspecting the system or a component in order to assess which discrete state it is in.

If the system is found to be failed upon inspection, replacement, i.e. reactive repair, is necessary to restore the system to an operational state. If the system is found to be degraded, preventive maintenance could be performed to restore the system to a less degraded state and reduce probability of a system failure.

We denote by $\boldsymbol{x}^-(T_1)$ the left-handed limit of $\boldsymbol{x}$ when t approaches T_1, meaning $\boldsymbol{x}$ at the moment immediately before the inspection at time T_1. Conversely, $\boldsymbol{x}^+(T_1)$ denotes the resetted state, i.e. after inspections and maintenance. In a similar vein, $\boldsymbol{y}^-(T_1)$ and $\boldsymbol{y}^+(T_1)$ denote the system output before and after inspection, respectively. For inspection and maintenance at time T_1, the system probabilities before and after the inspection and maintenance intervention are

$$\begin{aligned}
\boldsymbol{x}^-(T_1) &= \lim_{t \to T_1} \boldsymbol{x}(t) = \int_{t_0}^{t} f(\boldsymbol{x}, \tau) d\tau \\
\boldsymbol{y}^-(T_1) &= g(\boldsymbol{x}^-(T_1), T_1) \qquad (7.13) \\
\boldsymbol{x}^+(T_1) &= \lim_{t \leftarrow T_1} \boldsymbol{x}(t) = \boldsymbol{M}\boldsymbol{x}^-(T_1) \\
\boldsymbol{y}^+(T_1) &= g(\boldsymbol{x}^+(T_1), T_1) \qquad (7.14)
\end{aligned}$$

Since f is Lipschitz, the limit $\boldsymbol{x}^-(T_1)$ exists and is unique. Here, $\boldsymbol{M} \in \mathbb{R}^{(n_x, n_x)}$ is a square repair (maintenance) matrix.

For k repairs at times $\boldsymbol{T} = \left[T_1, ..., T_k\right]^\intercal$, the system probabilities can be expressed at any time t by the following piecewise model

$$\begin{aligned}
\boldsymbol{x}(t) &= \begin{cases}
\int_{T_0}^{T_1} f(\boldsymbol{x}, t)dt, \\
\boldsymbol{x}(T_0) = \boldsymbol{x}_0 \quad \text{if} \quad T_0 < t < T_1 \\
\int_{T_1}^{T_2} f(\boldsymbol{x}, t)dt, \\
\boldsymbol{x}(T_1) = \boldsymbol{x}^+(T_1) \quad \text{if} \quad T_1 < t < T_2 \\
\vdots \\
\int_{T_k}^{T_f} f(\boldsymbol{x}, t)dt, \\
\boldsymbol{x}(T_k) = \boldsymbol{x}^+(T_k) \quad \text{if} \quad T_k < t < T_f
\end{cases} \\
\boldsymbol{y}(t) &= g(\boldsymbol{x}, t)
\end{aligned} \tag{7.15}$$

where

$$\boldsymbol{x}^-(T_i) = \int_{T_{i-1}}^{T_i} f(x, t)dt \tag{7.16}$$

$$\boldsymbol{x}^+(T_i) = \boldsymbol{M}(T_i) \cdot \boldsymbol{x}^-(T_i). \tag{7.17}$$

Again, because f is Lipschitz, all the pieces of the piecewise $\boldsymbol{x}(t)$ are well defined. $\boldsymbol{M}(t) : \mathbb{R}^1 \to \mathbb{R}^{(n_x, n_x)}$ is a repair function. This allows different maintenance policies to be used at different points in time.

We can differentiate the above model to obtain

$$\frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, t) + \boldsymbol{R}\boldsymbol{r}(t) \tag{7.18}$$

where

$$\boldsymbol{r}(t) = \boldsymbol{S}^\intercal \boldsymbol{x}^-(t) \cdot \left(\sum_{i=1}^{k} \delta(t - T_i) \right) \tag{7.19}$$

$$= \boldsymbol{S}^\intercal \boldsymbol{x}^-(t) \mathbf{e}^\intercal \delta(t - \boldsymbol{T}). \tag{7.20}$$

Here δ is the Dirac delta function, and $\mathbf{e}^\intercal = [1, ..., 1]$ is the unity vector, and

$$\boldsymbol{M} = \left(\boldsymbol{I} + \boldsymbol{R}\boldsymbol{S}(t)^\intercal\right). \tag{7.21}$$

where $\boldsymbol{I} \in \mathbb{R}^{(n_x, n_x)}$ is the identity matrix. $\boldsymbol{R} \in \mathbb{R}^{(n_r, n_x)}$ is the maintenance and repair policy matrix. Finally, $\boldsymbol{S} : \mathbb{R}^1 \to \mathbb{R}^{(n_r, n_x)}$ is a selection function, which at times $t_i, \forall i \in \{1, .., .k\}$ selects the origin states, i.e. which correspond to negative entries in $\boldsymbol{R}$.

Example 7.4 (Structure of $\boldsymbol{R}$ and $\boldsymbol{S}$)**:**
For example, for a four state system where the first state corresponds to a completely new system, and the fourth state corresponds to the failed system, an As-Good-As-New (AGAN) maintenance policy corresponds to the following $\boldsymbol{R}$ matrix

$$\boldsymbol{R}_{\text{AGAN}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}. \tag{7.22}$$

which resets the system state to the first state (as good as new).

The corresponding selection function $\boldsymbol{S}(t)$ that selects which state to reset, is

$$\boldsymbol{S}_{\text{AGAN}}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{7.23}$$

Similarly, if the third state corresponds to a barely functioning system, the As-Bad-As-Old (ABAO) maintenance policy can be expressed as

$$\boldsymbol{R}_{\text{ABAO}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}. \tag{7.24}$$

Since the origin state is the same as for the AGAN policy, the selection functions $\boldsymbol{S}_{\text{AGAN}}(t)$ and $\boldsymbol{S}_{\text{ABAO}}(t)$ are identical. Note that in these two previous examples, the selection function is static and time independent. If our maintenance strategies involves having different maintenance policies applied at different times, this can be expressed as

$$\boldsymbol{R}_{\text{mix}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}. \tag{7.25}$$

In the case where two instances of maintenance are to be performed, one at T_1 and one at $T_2 > T_1$, we may use

$$\boldsymbol{S}_{\text{mix}}(t) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \frac{t-T_1}{T_2-T_1} & \frac{t-T_2}{T_1-T_2} \end{bmatrix} \tag{7.26}$$

to indicate that the AGAN maintenance is performed at T_1 and the ABAO maintenance is performed at T_2. More sophisticated maintenance policies can be obtained by adjusting the $\boldsymbol{R}$ and $\boldsymbol{S}$ matrices continuously.

7.4 Joint optimization of production and maintenance times

7.4.1 General problem description

The aim of this paper is to find a strategy that jointly optimizes the maintenance schedule and the operation of the plant. In terms of the model introduced in the previous section, this may imply finding the optimal system load $\boldsymbol{u}(t)$, and the times $\boldsymbol{T}$ at which to perform maintenance, given repair policies $\boldsymbol{S}$ and $\boldsymbol{R}$. Optimizing the performance of the system is a matter of optimizing the trade-off between the expected monetary benefit of its function, and the expected monetary loss of maintenance, as discussed in the introduction. This trade-off may be described mathematically in terms of a weighted economical objective, which is to be maximized using the available degrees of freedom.

In the following sections, a number of assumptions are made to solve this problem:

1. The system degradation is not monitored in real-time. The system must be inspected to observe in order to determine in which state it currently is.

2. Repairs and replacements are assumed to be instantaneous.

3. The resulting optimization problem is solved initially at the start of the planning horizon to get a production and maintenance schedule which is optimal in the expected sense. The problem is re-optimized regularly to adjust the production and maintenance schedule according to observed plant behavior. This re-optimization should be done as often as possible in order to make sure that the production and maintenance schedule is always up to date and accounting for the current system conditions.

7.4.2 Mathematical problem formulation

A general version of the optimization problem can be formulated as

$$\begin{aligned}\max_{\boldsymbol{u},k,\boldsymbol{T}} \quad & \sum_{i=1}^{k-1} \int_{T_i}^{T_{i+1}} \Big(\phi(t,\boldsymbol{x},\boldsymbol{u})\Big)dt \\ & + \int_{t_0}^{T_1} \Big(\phi(t,\boldsymbol{x},\boldsymbol{u})\Big)dt \\ & + \int_{T_k}^{t_f} \Big(\phi(t,\boldsymbol{x},\boldsymbol{u})\Big)dt \\ & - \phi_I(t,k) - \phi_M(t,\boldsymbol{r}) \end{aligned} \tag{7.27a}$$

$$\text{s.t.} \quad \frac{d\boldsymbol{x}}{dt} = f(t, \boldsymbol{u}, \boldsymbol{x}) + \boldsymbol{R}\boldsymbol{r}(t) \tag{7.27b}$$

$$\boldsymbol{r}(t) = \boldsymbol{S}^{\intercal}\boldsymbol{x}^{-}(t)\mathbf{e}^{\intercal}\delta(t - \boldsymbol{T}) \tag{7.27c}$$

$$\boldsymbol{y}(t) = g(\boldsymbol{x}, t) \tag{7.27d}$$

$$0 \leq \boldsymbol{y} \leq 1 \tag{7.27e}$$

$$\mathbf{e}^{\intercal}\boldsymbol{y} = 1 \tag{7.27f}$$

$$\boldsymbol{u}_{\min} \leq \boldsymbol{u} \leq \boldsymbol{u}_{\max} \tag{7.27g}$$

where $\mathbf{e}$ is the unity vector.

The objective function comprises three terms, as described below.

1. ϕ is the economical profit rate generated by the production, which is to be maximized
2. ϕ_I is the inspection cost, which is proportional to the number of inspections k
$$\phi_I(t, k) \propto k, \tag{7.28}$$
3. ϕ_M is the maintenance cost, which is proportional to the integral of $\boldsymbol{r}(t)$ (maintenance costs only have to be paid if maintenance occurs, i.e. it is dependent on the outcome of the inspection)
$$\phi_M(t, \boldsymbol{r}) \propto \int_{t_0}^{t_f} \boldsymbol{r}(t)dt. \tag{7.29}$$

The constraints of the optimization problem comprise bound constraints on the state variables and the intended input-induced loads, and any additional model equations that describe the system. These model equations could describe aerodynamic properties of an aircraft system, or mass and energy conservation in a chemical production system.

Remark 7.2 (Some special cases of the problem formulation from Problem (7.27)). The problem formulation from Problem (7.27) describes a general case. However, in some circumstances, we may alter the problem formulation slightly to obtain a special problem formulation. Examples of such problem formulations include, but are not limited to:

- **Special case 1: Maintenance optimization for fixed operational strategy**
  If the operational strategy is fixed, as is the case in many industries where a system performs the same task over and over, then the intended input-induced loads $\boldsymbol{u}$ are constant and are thus not subject to optimization. The degrees of freedom are k and $\boldsymbol{T}$.
- **Special case 2: Joint optimization of operational strategy and number of periodic inspections** If the maintenance strategy is fixed, e.g. we have to perform maintenance according to a clock-based or age-based schedule, the inspection times T are no longer subject to optimization, and can

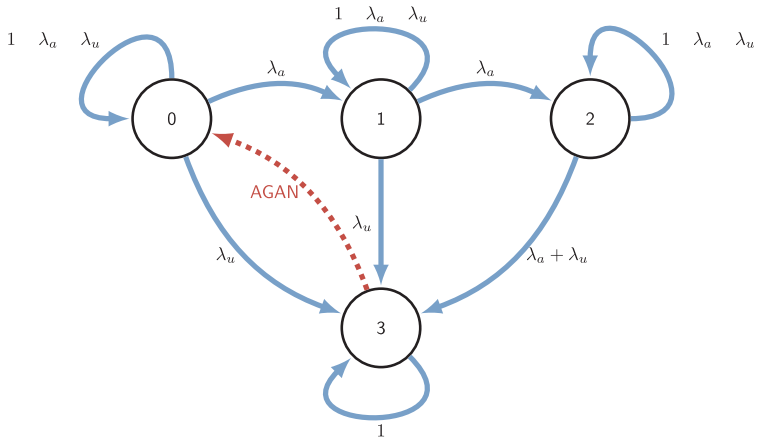

Figure 7.2: Illustration of a typical Markov chain

be reformulated as constraints to the optimization problem. The degrees of freedom are $\boldsymbol{u}$ and k

- **Special case 3: Maintenance optimization for fixed number of inspections**
  If the number of inspections is fixed, k ceases to be a decision variable in the optimization problem. The degrees of freedom are $\boldsymbol{u}$ and $\boldsymbol{T}$

These special cases are numerically easier to solve than the general case. For example, fixing the number of inspections removes the discrete nature of the problem, and we are left with an ordinary NLP instead of an MINLP (in other works) or an MPCC (in this work).

Example 7.5 (What is the benefit of optimizing the inspection times and the number of inspections?)**:**
To illustrate the proposed method, consider the following example. A production system has four degradation states labeled 0, 1, 2 *and* 3, *where state* 3 *is the failed state, and states* 0 - 2 *are progressively degraded states. An illustration of the Markov chain is shown in Figure 7.2 The transition probabilities are assumed constant, meaning that we can use the following model to describe the probability of being in any of the four states at any given time*

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{R}\boldsymbol{r} \tag{7.30}$$

$$\boldsymbol{y} = \boldsymbol{x} \tag{7.31}$$

$\boldsymbol{A}$ *is the constant transition probability matrix*

$$\boldsymbol{A} = \begin{bmatrix} -0.0101 & 0 & 0 & 0 \\ 0.0100 & -0.0101 & 0 & 0 \\ 0 & 0.0100 & -0.0101 & 0 \\ 0.0001 & 0.0001 & 0.0101 & 0 \end{bmatrix} \tag{7.32}$$

and

$$\boldsymbol{R} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \tag{7.33}$$

In this case, we have made the transition rates independent of inputs. The inputs $\boldsymbol{u}$ *do not enter in the problem formulation and are thus omitted (Special case 1).*

The goal is to find the maximum expected net profit by optimizing the k *inspection times* $\boldsymbol{T}$*. The final mission time is* $t_f = 200$ *weeks. The objective terms are*

$$\phi(t, \boldsymbol{x}) = \boldsymbol{c}^\intercal \boldsymbol{x} \tag{7.34}$$

$$\phi_I(t, k) = \boldsymbol{c}_I^\intercal \cdot k \tag{7.35}$$

$$\phi_M(t, \boldsymbol{r}) = \int_{t_0}^{t_f} \boldsymbol{c}_M^\intercal \boldsymbol{r}(t) dt \tag{7.36}$$

with

$$\boldsymbol{c} = \begin{bmatrix} 27.8 \\ 20.8 \\ 13.9 \\ 2.8 \end{bmatrix} \tag{7.37}$$

$$\boldsymbol{c}_I = 30 \tag{7.38}$$

$$\boldsymbol{c}_M = 300. \tag{7.39}$$

Further, fixing the number of inspections to $k = k^*$*, the problem may then be written as*

$$\max_{\boldsymbol{T}} \quad \sum_{i=1}^{k^*-1} \Big(\int_{T_i}^{T_{i+1}} \Big(\boldsymbol{c}^\intercal \boldsymbol{x} \Big) dt \Big) + \int_{t_0}^{T_1} \Big(\boldsymbol{c}^\intercal \boldsymbol{x} \Big) dt + \int_{T_k}^{t_f} \Big(\boldsymbol{c}^\intercal \boldsymbol{x} \Big) dt$$

$$- \boldsymbol{c}_I^\intercal \cdot k^* - \int_{t_0}^{t_f} \boldsymbol{c}_M^\intercal \boldsymbol{r}(t) dt \tag{7.40a}$$

$$\text{s.t.} \quad \frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{R}\boldsymbol{r}(t) \tag{7.40b}$$

$$\boldsymbol{r}(t) = \boldsymbol{S}^\intercal \boldsymbol{x}^-(t) \mathbf{e}^\intercal \delta(t - \boldsymbol{T}) \tag{7.40c}$$

$$\boldsymbol{y}(t) = \boldsymbol{x}(t) \tag{7.40d}$$

$$0 \leq \boldsymbol{y} \leq 1 \tag{7.40e}$$

$$\mathbf{e}^\intercal \boldsymbol{y} = 1 \tag{7.40f}$$

Table 7.1: Optimized maintenance schedule for the motivating example in Section 7.5

Inspections k^*	Net profit [kUSD]	Inspections times [weeks] $\boldsymbol{T}$					
1	3913.9	128.398					
2	3946.0	103.228	149.136				
3	3950.3	88.616	127.258	159.178			
4	3942.9	78.984	112.972	140.713	165.239		
5	3929.4	72.169	102.931	127.776	149.563	169.356	
6	3912.4	67.136	95.549	118.281	138.070	155.928	172.369

Table 7.2: Periodic maintenance schedule for the motivating example in Section 7.5

Inspections k	Net profit [kUSD]	Inspections times [weeks] $\boldsymbol{T}$					
1	3886.1	100.000					
2	3923.1	66.667	133.333				
3	3930.5	50.000	100.000	150.000			
4	3924.9	40.000	80.000	120.000	160.000		
5	3912.5	33.333	66.667	100.000	133.333	166.667	
6	3896.1	28.571	57.143	85.714	114.286	142.857	171.429

Solving the optimization problem Problem (7.27) repeatedly for different k^*. We obtain the optimal maintenance schedules shown in Table 7.1. Note that the decision variables in this optimization problem are only the inspection times $\boldsymbol{T}$.

For reference, Table 7.2 also shows the objective function values for a periodic (equidistant in time) maintenance schedule, which is prevalent in industry today. As expected, the time between inspections becomes shorter and shorter in the optimal maintenance schedule, as opposed to in the periodic maintenance schedule, where the time between inspections is constant.

Figure 7.3 shows the comparison of the net profit from the two approaches, and shows that the optimization approach always yields higher net profits, which motivates the use for an optimization-based method for determining the inspection times, as proposed in this paper.

7.5 Including recourse in the optimization problem

So far we used the expectation of all possible outcomes of an inspection in the problem formulation. The reasoning behind this is that the outcome cannot not be known ahead of time. However, after an inspection has been performed and

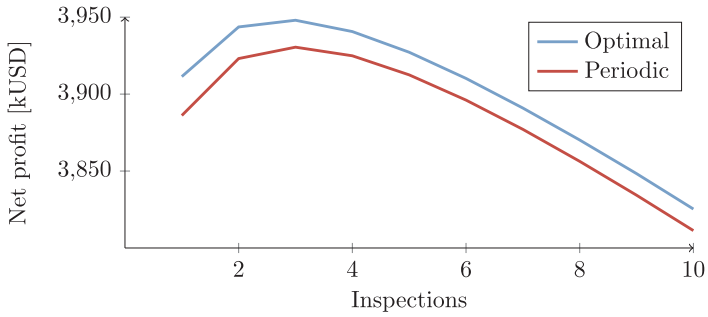

Figure 7.3: Comparison of net profit for optimal and periodic inspection schedules

the true system state has been revealed, we may use this information to revise the future decisions. For instance, if inspection reveals that the system has failed, we perform repairs or replacements to get it working again. However, we may also use this information to delay future inspections. We now have a different system, and the originally scheduled inspections are likely earlier than needed. On the other hand, if we inspect the system and find it to be not failed, we may want to use this information to slightly expedite future inspections, knowing that the system is "overdue".

Mathematically, this can be formulated as a multistage optimization problem. We distinguish between decisions that have to be made now (the next inspection date, and the operational strategy up to that date), and decisions that are made in the future, once more information has been revealed at future inspections (all following inspection dates, and the operational strategies). Due to the different possible outcomes from the inspections, this results in a tree of different scenarios. Each branch represents a unique combination of event realizations. Since current decisions cannot depend on unrevealed information obtained in the future, non-anticipativity constraints are introduced, to enforce that decisions on branches stemming from the same parent node are equal, since these branches share the same history up to that point. See Figure 7.4 for an illustration of this.

Due to the exponential growth of the number of scenarios in multi-stage programming, it is common to prevent branching of the scenario tree beyond a certain point in the future, and consequently solve an approximation of the original problem instead. In scenario-based model predictive control, this point is called the robust horizon [103]. It is argued by [103] that the refinement of the scenario tree by lengthening the robust horizon has negligible effect on the solution of the optimization problem, at the cost of significant increase in the computational burden.

Figure 7.4 shows a scenario tree for the case where k inspections are performed, and each inspection has two possible outcomes. Non-anticipativity constraints ensure that the nearest inspection times are equal on sibling branches, i.e. branches that share a common parent node. The scenario tree shown in the figure has a robust horizon of length two.

Heuristic simplification

Enforcing non-anticipativity for the problem with an arbitrary number of inspections is difficult. Instead, we propose a heuristic approach to address this problem.

1. First, solve the single-scenario problem (Problem (7.27)) corresponding to a robust horizon of length zero to find the optimal number of inspections, k^*, and initial inspections times and initial inputs $\boldsymbol{u}(t)$.

2. Second, once the solution of the single-scenario problem has been found, solve Problem (7.41)

$$
\begin{aligned}
\max_{\boldsymbol{u}_{\text{int}}, \boldsymbol{T}} \quad & \sum_{j=1}^{N_s} p_j \Big(\sum_{i=1}^{k^*-1} \int_{T_{j,i}}^{T_{j,i+1}} \Big(\phi(t, \boldsymbol{x}_j, \boldsymbol{u}_j)\Big) dt \\
& + \int_{t_0}^{T_{j,1}} \Big(\phi(t, \boldsymbol{x}_j, \boldsymbol{u}_j)\Big) dt \\
& + \int_{T_{j,k^*}}^{t_f} \Big(\phi(t, \boldsymbol{x}_j, \boldsymbol{u}_j)\Big) dt \\
& - \phi_I(t, k^*) - \phi_M(t, \boldsymbol{r}_j) \Big) && \text{(7.41a)} \\
\text{s.t.} \quad & \frac{d\boldsymbol{x}_j}{dt} = f(t, \boldsymbol{u}_j, \boldsymbol{x}_j) + \boldsymbol{R}\boldsymbol{r}_j(t) && \text{(7.41b)} \\
& \boldsymbol{r}_j(t) = \boldsymbol{S}^\intercal \boldsymbol{x}_j^-(t) \mathbf{e}^\intercal \delta(t - \boldsymbol{T}_j) && \text{(7.41c)} \\
& \boldsymbol{y}_j(t) = g(\boldsymbol{x}_j, t) && \text{(7.41d)} \\
& 0 \leq \boldsymbol{y}_j \leq 1 && \text{(7.41e)} \\
& \mathbf{e}^\intercal \boldsymbol{y}_j = 1 && \text{(7.41f)} \\
& \boldsymbol{u}_{\min} \leq \boldsymbol{u}_j \leq \boldsymbol{u}_{\max} && \text{(7.41g)} \\
& \boldsymbol{u}_a(\tau) = \boldsymbol{u}_b(\tau) \\
& \forall a, b \in \{1, ..., N_s\} \cap \\
& \boldsymbol{x}_a(t < \tau) = \boldsymbol{x}_b(t < \tau) && \text{(7.41h)}
\end{aligned}
$$

   Note that the optimization problem from Problem (7.41) is similar to the single stage problem in Problem (7.27), except that the number of inspection k is no longer a decision variable, but has instead been substituted by the nominal number of inspections from the single-scenario problem, k^*. Instead, we solve the multi-scenario problem with the inspection times and intended input-induced loads as decision variables (special case 3). Non-anticipativity on the decision variables through the constraint Equation (7.41h)

3. Check that a local optimum has been found by solving Problem (7.41) for $k = k^* + 1$ and $k = k^* - 1$ as well, and confirm that the objective is worse than the solution found for $k = k^*$

Since the solution of the multi-stage problem is often similar to that of the single-scenario problem, we assume that the optimal number of inspections does not change. If that is the case, then this heuristic sequential approach will converge to the true optimal solution.

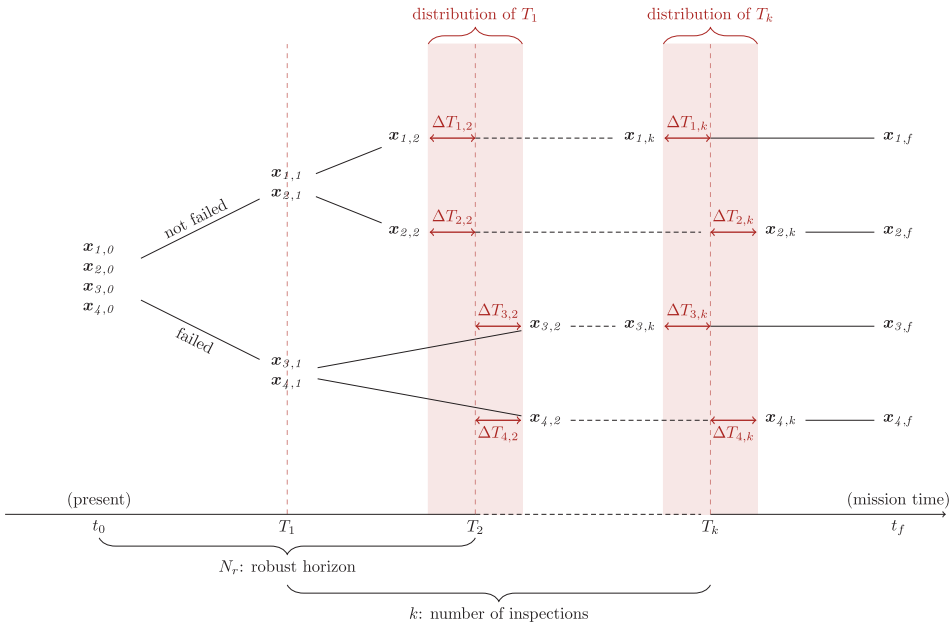

Figure 7.4: Scenario tree for a multi-stage decision making process with k inspections. There are two possible outcomes from each inspection (system has failed / system has not failed), and the length of the robust horizon is two. This results in a total of four scenarios. Note how the inspection times for the different scenarios are not the same after the first inspection, resulting in a discrete distribution of inspection times.

Example 7.6 (Illustrative example of multi-stage formulation)**:**
Let us return to the example from Section 7.5 to illustrate how a multi-stage formulation can be used to obtain a more sophisticated operational strategy and maintenance schedule, and increasing the expected net profit.

We again solve the problem, but this time we solve for all the cases where either $N_s = 2$, $N_s = 3$ or $N_s = 4$ inspections, with varying robust horizons, using the approach outlined in the previous section. The net profits and optimal inspection times are as reported in Table 7.3.

From Table 7.3 it can be seen that the multistage formulation, i.e. when the robust horizon is nonzero, always outperforms the case when the robust horizon is zero. It could be argued that the extra computational burden (owing to the exponential growth of the problem size) is not worth the very small improvement in the objective function value. A robust horizon of length one seems to give a satisfactory

Table 7.3: Optimized multi-stage maintenance schedule for the example in Section 7.6

Inspections k	Robust horizon N_r	Net profit [kUSD]	Branch j	Inspections at times [weeks] $\boldsymbol{T}_j$			
2	0	3946.03	-	103.228	149.140		
2	1	3946.25	F	103.177	162.914		
			H	103.177	148.813		
3	0	3950.29	-	88.6158	127.258	159.178	
3	1	3950.50	F	88.6553	145.733	169.329	
			H	88.6553	126.967	158.923	
3	2	3950.57	FF	88.5761	145.697	175.652	
			FH	88.5761	145.697	169.282	
			HF	88.5761	126.902	169.998	
			HH	88.5761	126.902	158.777	
4	0	3942.87	-	78.9843	112.972	140.713	165.239
4	1	3943.04	F	79.0959	133.404	155.795	173.068
			H	79.0959	112.758	140.443	165.061
4	2	3943.14	FF	78.9957	133.361	165.775	178.858
			FH	78.9957	133.361	155.717	173.013
			HF	78.9957	112.706	156.473	174.400
			HH	78.9957	112.706	140.261	164.901
4	3	3943.17	FFF	78.9631	133.336	165.763	182.661
			FFH	78.9631	133.336	165.763	178.845
			FHF	78.9631	133.336	155.702	178.914
			FHH	78.9631	133.336	155.702	172.989
			HFF	78.9631	112.657	156.453	179.174
			HFH	78.9631	112.657	156.453	174.377
			HHF	78.9631	112.657	140.223	173.972
			HHH	78.9631	112.657	140.223	164.827

trade-off between computational complexity and objective function value. On the other hand, the time-scales for degradation of industrial systems are usually in the range of weeks, months or years, such that there is enough time to perform complex computations.

Observe as well that the different scenarios exhibit the trend hypothesized in Section 7.5: If a system is observed to have failed, inspections are postponed slightly. If the system has not failed, inspections are expedited slightly. For example, for the case with two inspections, we observe that the second inspection is postponed from $T_2 = 149.1$ *weeks to* $T_2 = 162.9$ *weeks if the first inspection reveals failure. Likewise, we observe that the second inspection is preponed from* $T_2 = 149.1$ *weeks to* $T_2 = 148.8$ *weeks if the first inspection reveals that the system is still opera-*

tional. The improvement in the objective function value from 3946.03 *kUSD to* 3946.25 *kUSD is a result of this more sophisticated schedule. The reason why the improvement is so small can be explained by the fact that the failed branch has a very small probability, and thus contributes little to the overall objective.*

7.6 Approximations

7.6.1 Continuous approximation of repairs $\boldsymbol{r}(t)$

Before proceeding, let us recall that the aim of this paper is to find an optimization-based method with which we can determine the optimal operational strategy in terms of production, inspections and maintenance. However, the expression for the evolution of the probabilities in Equation (7.18) contains the Dirac delta function. Due to its discontinuous nature, this is problematic for numerical optimization. A possible solution is to reformulate the problem as a mixed integer problem, as was done in e.g. [9, 15, 63], instead of using Dirac delta functions. Unfortunately, MINLP problems are known to be difficult to solve, despite recent progress in algorithmic development [149].

Instead of using a mixed-integer formulation to solve the problem, we chose to approximate the Dirac delta function by Boxcar functions instead:

$$\boldsymbol{r}(t) \approx \tilde{\boldsymbol{r}}(t) \tag{7.42}$$

$$\begin{aligned} \tilde{\boldsymbol{r}}(t) &= \sum_{i=1}^{k} \text{Boxcar}(t) \\ &= \sum_{i=1}^{k} h_i \Big(\text{H}(t - T_i) - \text{H}(t - T_i - \epsilon_i) \Big). \end{aligned} \tag{7.43}$$

Here, H is the Heaviside function, h_i is the height and ϵ_i is the width of boxcar i. By choosing a nonzero ϵ we ensure that $\tilde{\boldsymbol{r}}(t)$ is bounded and Lipschitz continuous, and consequently that $\boldsymbol{x}(t)$ is continuous (although still non-smooth).

An illustration of the approximation of the Dirac functions with boxcar functions is shown in Figure 7.5. Observe that the approximation for $\boldsymbol{x}(t)$ is good if ϵ is sufficiently small. ϵ should be chosen as small as possible in order to minimize the approximation error, but big enough to avoid numerical problems with the chosen solver. In practice, we have found that choosing $\epsilon \approx \frac{1}{1000}(t_f - t_0)$ yielded satisfactory performance for our applications.

7.6.2 Smooth problem re-formulation for numerical optimization with off-the-shelf NLP solvers

Problem (7.27) is discrete in nature due to the inclusion of the integer variable k, therefore making it difficult to solve without the use of MINLP solvers as explained in Section 7.6.1, or by fixing the number of inspections k, as illustrated in Example 7.5.

In order to avoid both of these solution alternatives, we include the simplifications introduced in the previous sections. As a result, we reformulate optimization problem from (7.27) as

$$\max_{\boldsymbol{u},\tilde{\boldsymbol{r}},\boldsymbol{\epsilon}} \quad \int_{t_0}^{t_f} \Big(\phi(t,\boldsymbol{x},\boldsymbol{u}_{\text{act}})\Big)dt$$

$$-\phi_I(t,\boldsymbol{z}) - \phi_M(t,\tilde{\boldsymbol{r}}) \tag{7.44a}$$

$$\text{s.t.} \quad \frac{d\boldsymbol{x}}{dt} = f(t,\boldsymbol{u}_{\text{act}},\boldsymbol{x}) + \boldsymbol{R}\tilde{\boldsymbol{r}}(t) \tag{7.44b}$$

$$\boldsymbol{y}(t) = g(\boldsymbol{x},t) \tag{7.44c}$$

$$0 \leq \boldsymbol{y} \leq 1 \tag{7.44d}$$

$$\mathbf{e}^{\intercal}\boldsymbol{y} = 1 \tag{7.44e}$$

$$0 \leq \tilde{\boldsymbol{r}} \leq \frac{1}{\epsilon_{\min}} \tag{7.44f}$$

$$\epsilon_{\min} \leq \boldsymbol{\epsilon} \leq \epsilon_{\max} \tag{7.44g}$$

$$\boldsymbol{u}_{\min} \leq \boldsymbol{u} \leq \boldsymbol{u}_{\max} \tag{7.44h}$$

However, we still need to reformulate the inspection and maintenance costs, which contain the integer variable k. In Equation (7.28), the inspection cost is said to be proportional to the number of inspections, k. However, since we reformulated the optimization problem, k is no longer explicitly included. Instead, it is "baked" into $\tilde{\boldsymbol{r}}(t)$.

In order to ensure that the reformulated optimization problem still approximates the original discrete problem, we introduce a variable $\boldsymbol{z}$ which is complementary to $\tilde{\boldsymbol{r}}$

$$0 \leq (1-\boldsymbol{z}) \perp \tilde{\boldsymbol{r}} \geq 0 \tag{7.45a}$$

$$0 \leq \boldsymbol{z} \leq 1 \tag{7.45b}$$

Here, the $\perp$ operator indicates complementary, i.e. we require that at all times either $\tilde{\boldsymbol{r}}$ or $(1-\boldsymbol{z})$ or both to be zero. Now we can formulate the number of inspections as

$$\phi_I(t,\boldsymbol{z}) \propto k \approx \int_{t_0}^{t_f} \frac{\boldsymbol{z}}{\epsilon}, \tag{7.46}$$

where ϵ is the width of the Boxcar in the approximation $\tilde{\boldsymbol{r}}(t)$. A graphical illustration of this is shown in Figure 7.5

Due to the introduction of the complementary constraints into the optimization problem, the resulting problem is classified as a mathematical program with complementary/equilibrium constraints (MPCC / MPEC).

The optimization problem from Equations (7.44a)-(7.44h) is solved using direct collocation [28].

Remark 7.3 (Regarding the complementary constraints). Due to the non-convex nature of the functions in the optimization problem, standard solvers for nonlinear programs, such as IPOPT [185], will only provide locally optimal solutions. To explore more of the feasible space and hopefully converge to a near-global optimum, we solve the problem using a multi-start approach. Here we initiate the optimization algorithm at multiple different initial points. Each optimization run yields a local solution, and the best local solution is chosen. Of course a globally optimal solution cannot be guaranteed, but our experience is that there are many solutions with an objection similar to the the globally optimal value, which are only marginally worse and are acceptable from an engineering perspective.

Another precaution that we have taken is to reformulate the MPEC to make it tractable. Broadly speaking, these reformulations can be categorized as either relaxation methods or penalty methods. In relaxation methods, the complementary conditions are temporarily relaxed. Upon repeated solving of the relaxed problem with gradual tightening of the complementary constraints, the solution of the relaxed problem converges to the true solution. Similarly, penalty methods remove the complementary condition from the constraints and penalize them in the objective instead. Gradual increase of the penalty weight means that the complementary condition is satisfied upon convergence. Using a relaxation and hybrid method, Problem (7.44) with the added complementarity constraints from Equation (7.45) can be reformulated as

$$
\begin{aligned}
\max_{\boldsymbol{u},\tilde{\boldsymbol{r}},\boldsymbol{\epsilon}} \quad & \int_{t_0}^{t_f} \Big(\phi(t,\boldsymbol{x},\boldsymbol{u})\Big)dt \\
& -\phi_I(t,\boldsymbol{z}) - \phi_M(t,\tilde{\boldsymbol{r}}) && \text{(7.47a)}\\
& -\boldsymbol{\gamma}(l)\cdot \operatorname{vec}\left((1-\boldsymbol{z})\otimes\tilde{\boldsymbol{r}}\right) && \text{(7.47b)}\\
\text{s.t.} \quad & \frac{d\boldsymbol{x}}{dt} = f(t,\boldsymbol{u}_{\text{act}},\boldsymbol{x}) + \boldsymbol{R}\tilde{\boldsymbol{r}}(t) && \text{(7.47c)}\\
& \boldsymbol{y}(t) = g(\boldsymbol{x},t) && \text{(7.47d)}\\
& 0 \le \boldsymbol{y} \le 1 && \text{(7.47e)}\\
& \mathbf{e}^{\intercal}\boldsymbol{y} = 1 && \text{(7.47f)}\\
& 0 \le \tilde{\boldsymbol{r}} \le \frac{1}{\epsilon_{\min}} && \text{(7.47g)}\\
& \epsilon_{\min} \le \boldsymbol{\epsilon} \le \epsilon_{\max} && \text{(7.47h)}\\
& \boldsymbol{u}_{\min} \le \boldsymbol{u} \le \boldsymbol{u}_{\max} && \text{(7.47i)}\\
& 0 \le \boldsymbol{z} \le 1 && \text{(7.47j)}\\
& \operatorname{vec}\left(\tilde{\boldsymbol{r}}\otimes(1-\boldsymbol{z})\right) \ge \boldsymbol{\xi}(l) && \text{(7.47k)}
\end{aligned}
$$

Here, $(\cdot)$ denotes the inner product and $(\otimes)$ denotes the outer product. This problem is solved sequentially $\{l\}$ times such that

$$\lim_{l\to\infty}, \boldsymbol{\xi} \to 0, \boldsymbol{\gamma} \to \infty$$

Under assumptions of stationarity, the MPEC solution will be approached asymptotically [19].

7.7 Case example: Subsea oil and gas separation system

To illustrate our proposed method, we apply it to the case study of an subsea oil and gas separation plant. We will solve the problem using the approximations introduced in Section 7.6.2.

7.7.1 Motivation and general description

Despite the trend towards more sustainable energy production, fossil fuel will still play an important role in satisfying the global energy demand in the foreseeable future. Subsea technology is a technology that will increase in prominence in the coming years due to the depletion of easily accessible oil and gas reserves, as it enables production from remote, inaccessible locations. However, due to this inaccessibility and remoteness of subsea oil and gas production facilities, current industrial practice relies on infrequent inspections and conservative operation in order to prevent unplanned maintenance. Another issue preventing real-time monitoring of subsea equipment is the lack of qualified instrumentation. Our method consequently can lead to improved economical performance of the plant by reducing the conservativeness of operation and optimization of the inspection and maintenance times.

An illustration of the plant we are studying in this case example is shown in Figure 7.6.

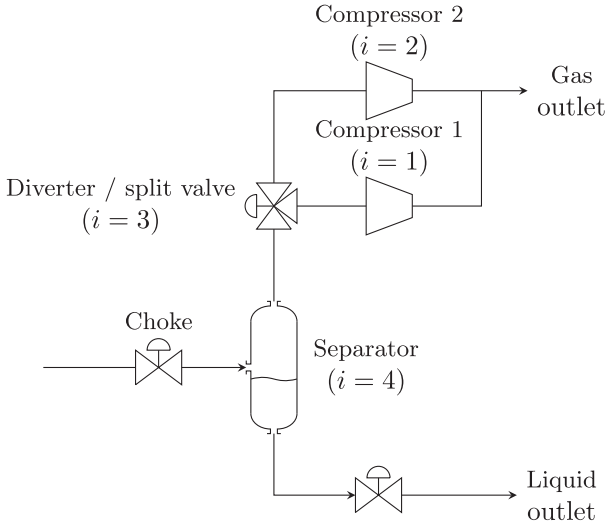

Figure 7.6: Illustration of the subsea compressor station that is studied in this case example, with $i = 1...4$ indicate the degrading components.

It is a subsea gas compression station with two compression trains (referred to as units 1 and 2) in parallel. These units are dry gas compressors, meaning that the wet gas from the reservoir must be cleaned of water and oil droplets in a scrubber-type separator (referred to as unit 4) before entering the compressors. If liquid droplets enter the compressors, this will lead to rapid degradation. The parallel configuration was chosen for the sake of flexibility. Failure of one of the units will only lead to partial failure of the whole system, since the system can continue production. Each compressor alone is able to provide the handle the full load, albeit with the consequence of faster degradation.

A single separator was deemed sufficient, since an additional, second separator unit would significantly increase complexity, cost and footprint of the overall system. Since the separator operates statically without moving parts, its reliability is acceptably high to be in a series configuration with the compressors. A fault tree analysis was performed for the system, resulting in the simplified fault tree shown in Figure 7.7. A comprehensive root cause analysis is outside the scope of this paper, so we instead use the encompassing events "failure from load-induced wear" for the compressors and "actuator failure" for the separator.

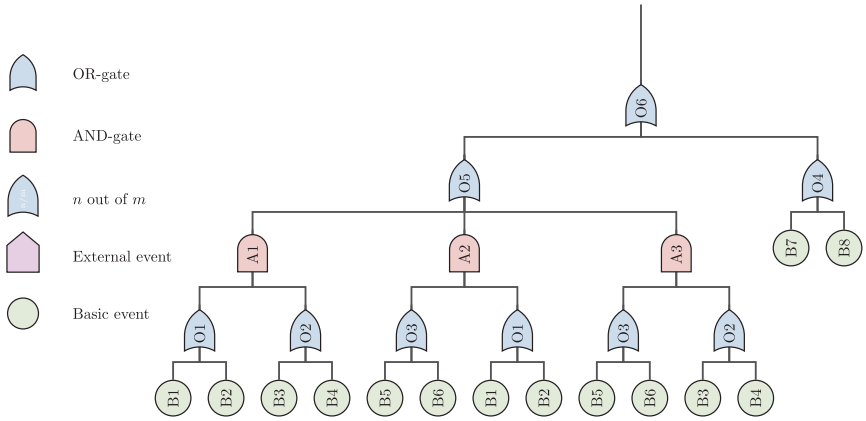

Figure 7.7: Fault tree for the subsea compressor station. The gates and basic events are described in Table 7.4

7.7.2 Modelling the degradation of the subsea separation system

The degradation of each individual component can each be described by four discrete degradation states. The overall system reliability vector $\boldsymbol{x}$ is created by stacking the component reliability state vectors:

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_{comp1} \\ \boldsymbol{x}_{comp2} \\ \boldsymbol{x}_{split} \\ \boldsymbol{x}_{sep} \end{bmatrix}, \tag{7.48}$$

Table 7.4: Description of labels of events and logic gates in fault tree in Figure 7.7

Label	Description
O_6	Failure of the entire system
O_5	Failure of the compression subsystem
O_4	Failure of the separation subsystem
A_1	Failure of compressor 1 and compressor 2
A_2	Failure of compressor 1 and split valve
A_3	Failure of compressor 2 and split valve
O_1	Failure of compressor 1
O_2	Failure of compressor 2
O_3	Failure of split valve
B_1	Sensor failure, loss of communication or other unforseen failure
B_2	Compressor failure due to load induced wear
B_3	Sensor failure, loss of communication or other unforseen failure
B_4	Compressor failure due to load induced wear
B_5	Sensor failure, loss of communication or other unforseen failure
B_6	Insufficient actuation due to load induced wear of split valve
B_7	Sensor failure, loss of communication or other unforseen failure
B_8	Insufficient actuation due to load induced wear of separator valve

and the evolution of the system states can now be expressed as

$$\frac{d\boldsymbol{x}}{dt} = f_a(\boldsymbol{x}, \boldsymbol{u}, t) + f_u(\boldsymbol{x}, t), \tag{7.49}$$

where f_a is the degradation due to aging, which is dependent on the inputs $\boldsymbol{u}$, and f_u is the sudden failure due to shock damage, which is independent from the inputs $\boldsymbol{u}$.

The basic events B_1, B_3, B_5 and B_7 from the fault tree in Figure 7.7 correspond to unexpected shock failure. These events are assumed to be independent of the operating mode, and influence the transition rates through the function f_u, which can be expressed as

$$f_u(\boldsymbol{x}, t) = \begin{bmatrix} f_{u,comp1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & f_{u,comp2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & f_{u,split} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & f_{u,sep} \end{bmatrix} \tag{7.50}$$

where $f_{u,i}$ is

$$f_{u,i}(\boldsymbol{x}_i, t) = \begin{bmatrix} -\lambda_{u,i} & 0 & 0 & 0 \\ 0 & -\lambda_{u,i} & 0 & 0 \\ 0 & 0 & -\lambda_{u,i} & 0 \\ \lambda_{u,i} & \lambda_{u,i} & \lambda_{u,i} & 0 \end{bmatrix} \cdot \begin{bmatrix} x_{i,0} \\ x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix}. \tag{7.51}$$

The basic events B_2, B_4, B_6 and B_8 are linked to aging of the component. The aging of each of the four components is described by the function f_i The Markov chain describing the degradation of each individual component has four discrete degradation states. The final state, i.e. $x_{i,N=3}$, corresponds to the failure of component i. The transition rates are influenced by the input-induced loads, $\boldsymbol{u}$. We model the transition rates as

$$f_a = \begin{bmatrix} f_{a,comp1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & f_{a,comp2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & f_{a,split} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & f_{a,sep} \end{bmatrix} \tag{7.52}$$

where

$$f_{a,i} = \begin{bmatrix} -\lambda_{a,i} & 0 & 0 & 0 \\ \lambda_{a,i} & -\lambda_{a,i} & 0 & 0 \\ 0 & \lambda_{a,i} & -\lambda_{a,i} & 0 \\ 0 & 0 & \lambda_{a,i} & 0 \end{bmatrix} \tag{7.53}$$

where the transition rates of each of the four components is assumed to be proportional to the load applied to it

$$\lambda_{a,comp1} = \tilde{\lambda}_{a,comp1} \cdot (u_1^2 + u_1) \tag{7.54}$$

$$\lambda_{a,comp2} = \tilde{\lambda}_{a,comp2} \cdot (u_2^2 + u_2) \tag{7.55}$$

$$\lambda_{a,split} = \tilde{\lambda}_{a,split} \cdot (u_1 \cdot (1 - x_{comp1,N}) \tag{7.56}$$

$$+ \; u_2 \cdot (1 - x_{comp2,N})) \tag{7.57}$$

$$\lambda_{a,sep} = \tilde{\lambda}_{a,sep} \cdot (u_1 \cdot (1 - x_{comp1,N}) \tag{7.58}$$

$$+ \; u_2 \cdot (1 - x_{comp2,N})), \tag{7.59}$$

Note that the transition rates of the split and the separator depend on the reliability of the two compressors. The reason for this is that if one of the compressors fails, the consequence is that the overall load on the system is reduced since no flow goes through the failed compressor (safety shut-off systems prevent this from happening). In reality, an operator may observe that the overall flow from the system is reduced, and may react to this by increasing the set-points to the compressors in order to keep the total flow constant. However, this manual intervention is not included in the optimization framework.

Furthermore, we note that the transition rates for the compressors are quadratic functions of the flows through the compressor. Due to reduced compressor efficiencies at higher flow rates, compressors have to be run harder to satisfy the desired pressure increase.

At each inspection time, there are two options available for maintenance interventions:

1. If the component is found in the failed state $j = N = 3$, then it is replaced by a new component (as-good-as-new)
2. If the component is found in the degraded state $j = 2$, then preventive maintenance may be performed to restore the component. Due to imperfect repairs, the health component will end up in state $j = 0$ in 80% of the time, and in state $j = 1$ in 20% of the time.

Remark 7.4 (Difference between intended and actual input-induced loads). Note that in the above problem formulation, we distinguish between the **intended** input-induced load, $\boldsymbol{u}_{\text{int}}$, and the **actual**, realized load on the system $\boldsymbol{u}_{\text{act}}$. These two are not the same for several reasons. Firstly, the actual load on the system may be stochastic, and not known exactly. For example, the load may be normally distributed around the intended mean load. Secondly, in a multi-component system, the actual load on the system or on individual components may change due to e.g. the failure of other components. For example, a load-bearing component in parallel configuration might experience a higher load than what was intended upon the failure of one of the parallel components.

7.7.3 Expressing the system states

The condition of the overall system can be formulated as algebraic relationships of the component reliability states. We are interested in the following four states

$$\boldsymbol{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \\ y_4(t) \end{bmatrix} \tag{7.59}$$

where, $y_1, ..., y_4$ denote progressively deteriorated condition of the overall system. Specifically,

1. No failed units, system is working at full capacity

$$y_1 = \mathbb{P}({O_1}^c \cap {O_2}^c \cap {O_3}^c \cap {O_4}^c) \tag{7.60}$$

   Here, $\mathbb{P}$ indicates the probability operator, O_i refers to the corresponding events from Table 7.4, c indicates the complement operator, and $\cap$ is the intersection operator.

2. Either compressor 1 has failed or compressor 2 has failed, but split valve and separator are still operational. The system is running at reduced capacity.

$$y_2 = \mathbb{P}((O_1 \oplus O_2) \cap {O_3}^c \cap {O_4}^c) \tag{7.61}$$

   where $\oplus$ is the exclusive OR operator.

3. The split valve has failed, but the separator and both compressors are still operational. The split ratio can no longer be adjusted. System is running at reduced capacity.

$$y_3 = \mathbb{P}({O_1}^c \cap {O_2}^c \cap O_3 \cap {O_4}^c) \tag{7.62}$$

4. System has failed. No production.

$$y_4 = \mathbb{P}(O_6) \tag{7.63}$$

Note that logically, the assertion

$$\mathbf{e}^\intercal \boldsymbol{y} = 1 \tag{7.64}$$

must hold because the system must be in one of these states. Either the system has no failed components, some non-critical failed components, or the overall system has failed. To get expressions for $\boldsymbol{y}(t)$, we consider at the fault tree and derive the expressions for the probabilities of all four composite events.

First, we define χ_i as the probability of component/subsystem i being in the failed state N, to shorten notation slightly

$$\mathbb{P}(O_1) = \chi_1 = x_{comp1,N}(t) \tag{7.65}$$
$$\mathbb{P}(O_2) = \chi_2 = x_{comp2,N}(t) \tag{7.66}$$
$$\mathbb{P}(O_3) = \chi_3 = x_{split,N}(t) \tag{7.67}$$
$$\mathbb{P}(O_4) = \chi_4 = x_{sep,N}(t) \tag{7.68}$$

Which allows us to express

$$\begin{aligned} \mathbb{P}(A_1) &= \mathbb{P}(O_1 \cap O_2) \\ &= \mathbb{P}(O_1)\mathbb{P}(O_2) \\ &= \chi_1\chi_2 \end{aligned} \tag{7.69}$$
$$\begin{aligned} \mathbb{P}(A_2) &= \mathbb{P}(O_1 \cap O_3) \\ &= \mathbb{P}(O_1)\mathbb{P}(O_3) \\ &= \chi_1\chi_3 \end{aligned} \tag{7.70}$$
$$\begin{aligned} \mathbb{P}(A_3) &= \mathbb{P}(O_2 \cap O_3) \\ &= \mathbb{P}(O_2)\mathbb{P}(O_3) \\ &= \chi_2\chi_3 \end{aligned} \tag{7.71}$$

We can now express the probability of a failure of the compression subsystem as

$$\begin{aligned} \mathbb{P}(O_5) &= \mathbb{P}(A_1 \cup A_2 \cup A_3) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) \\ &- \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3) \\ &= \chi_1\chi_2 + \chi_1\chi_3 + \chi_2\chi_3 \\ &- \chi_1\chi_2\chi_3 - \chi_1\chi_2\chi_3 - \chi_1\chi_2\chi_3 + \chi_1\chi_2\chi_3 \\ &= \chi_1\chi_2 + \chi_1\chi_3 + \chi_2\chi_3 - 2\chi_1\chi_2\chi_3 \end{aligned} \tag{7.72}$$

Consequently, the probability of system failure can be expressed as

$$\begin{aligned}\mathbb{P}(O_6) &= \mathbb{P}(O_4 \cup O_5)\\ &= \mathbb{P}(O_4) + \mathbb{P}(O_5) - \mathbb{P}(O_5 \cap O_4)\\ &= \mathbb{P}(O_4) + \mathbb{P}(O_5) - \mathbb{P}(O_4)\mathbb{P}(O_5)\\ &= \chi_4 + (1-\chi_4)\left(\chi_1\chi_2 + \chi_1\chi_3 + \chi_2\chi_3 - 2\chi_1\chi_2\chi_3\right)\\ &= y_4 \end{aligned} \tag{7.73}$$

Similarly for the other three composite events

$$\begin{aligned} y_1 &= \mathbb{P}({O_1}^c \cap {O_2}^c \cap {O_3}^c \cap {O_4}^c)\\ &= (1-\chi_1)\cdot(1-\chi_2)\cdot(1-\chi_3)\cdot(1-\chi_4) \end{aligned} \tag{7.74}$$

$$\begin{aligned} y_2 &= \mathbb{P}((O_1 \oplus O_2) \cap {O_3}^c \cap {O_4}^c)\\ &= \mathbb{P}(((O_1 \cup O_2) - (O_1 \cap O_2)) \cap {O_3}^c \cap {O_4}^c)\\ &= (\chi_1 + \chi_2 - 2\chi_1\chi_2)\cdot(1-\chi_3)\cdot(1-\chi_4) \end{aligned} \tag{7.75}$$

$$\begin{aligned} y_3 &= \mathbb{P}({O_1}^c \cap {O_2}^c \cap O_3 \cap {O_4}^c)\\ &= (1-\chi_1)\cdot(1-\chi_2)\cdot\chi_3\cdot(1-\chi_4) \end{aligned} \tag{7.76}$$

Finally, we check that the sum of the four composite events is

$$\begin{aligned}\sum_{i_1}^{4} y_i &= y_1 + y_2 + y_3 + y_4\\ &= (1-\chi_1)\cdot(1-\chi_2)\cdot(1-\chi_3)\cdot(1-\chi_1)\\ &+ (\chi_1 + \chi_2 - \chi_1\chi_2)\cdot(1-\chi_3)\cdot(1-\chi_4)\\ &+ (1-\chi_1)\cdot(1-\chi_2)\cdot\chi_3\cdot(1-\chi_4)\\ &+ \chi_4 + (1-\chi_4)\left(\chi_1\chi_2 + \chi_1\chi_3 + \chi_2\chi_3 - 2\chi_1\chi_2\chi_3\right)\\ &= 1 \end{aligned} \tag{7.77}$$

as expected.

Now that we have the expressions for the four system states $\boldsymbol{y}$ as a function of the component states $\boldsymbol{x}$, we may formulate the following optimization problem.

7.7.4 Economic objective

The objective is to maximize the throughput through the compressor station. Simultaneously, we want to minimize the inspection and maintenance costs. The economical objective is the sum of discounted cash flows.

$$\min_{\boldsymbol{u}(t),\tilde{\boldsymbol{r}}(t)} \qquad \int_0^{t_f} \Big(\frac{-\phi + \phi_I + \phi_M}{(1+d)^t}\Big) dt \tag{7.78a}$$

where

- The first term ϕ is expected the production, which is proportional to the system probabilities $\boldsymbol{y}(t)$ via the weighing vector / productivity vector $\boldsymbol{c}_p$.

$$\phi(t, \boldsymbol{y}, \boldsymbol{u}) = \left(\boldsymbol{c}_p^\intercal \cdot \boldsymbol{y}(t)\right) \cdot \left(\mathbf{e}^\intercal \cdot \boldsymbol{u}(t)\right) \tag{7.78b}$$

  $c_{p,i}$ indicates how well the system is able to perform its desired task in the corresponding state y_i.

- The second term ϕ_I is the inspection cost, which is paid every time an inspection is performed. It is proportional to $\boldsymbol{z}(t)$, the variable introduced to count the number of inspections.

$$\phi_I(t, \boldsymbol{z}) = \boldsymbol{c}_I^\intercal \cdot \boldsymbol{z}(t) \tag{7.78c}$$

- The third and last term ϕ_M is the maintenance cost, which is proportional to $\tilde{\boldsymbol{r}}(t)$, i.e. it is proportional to the probability of being in the undesired state at the inspection time.

$$\phi_M(t, \tilde{\boldsymbol{r}}) = \boldsymbol{c}_M^\intercal \cdot \tilde{\boldsymbol{r}}(t) \tag{7.78d}$$

  As mentioned previously, there are two maintenance options:

  1. Corrective replacement
  2. Preventive (imperfect) repairs

$$\boldsymbol{c}_M = \begin{bmatrix} \boldsymbol{c}_{M,corr} \\ \boldsymbol{c}_{M,prev} \end{bmatrix} \tag{7.78e}$$

- The entire objective is discounted by the factor d. The discount factor is chosen according to the economics and the time scale of the optimization.

The decision variables are the maintenance and inspection times, as well as the maintenance type (corrective replacement or preventive repairs)

$$\tilde{\boldsymbol{r}}(t) = \begin{bmatrix} \tilde{\boldsymbol{r}}_{corr}(t) \\ \tilde{\boldsymbol{r}}_{prev}(t) \end{bmatrix} = \begin{bmatrix} \tilde{r}_{corr,comp1}(t) \\ \tilde{r}_{corr,comp2}(t) \\ \tilde{r}_{corr,split}(t) \\ \tilde{r}_{corr,sep}(t) \\ \tilde{r}_{prev,comp1}(t) \\ \tilde{r}_{prev,comp2}(t) \\ \tilde{r}_{prev,split}(t) \\ \tilde{r}_{prev,sep}(t) \end{bmatrix} \tag{7.78f}$$

and the inputs $\boldsymbol{u}$. The maintenance matrix $\boldsymbol{R}$ is

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{7.78g}$$

7.7.5 Constraints

The optimization problem outlined in the previous section is solved subject to the following constraints

$$\begin{aligned} s.t. \quad & \frac{d\boldsymbol{x}}{dt} = f(\boldsymbol{x}, \boldsymbol{u}, t) + \boldsymbol{R}\tilde{\boldsymbol{r}}(t) && \text{(7.78h)} \\ & \boldsymbol{y}(t) = \boldsymbol{g}(\boldsymbol{x}, t) && \text{(7.78i)} \\ & \boldsymbol{x}(0) = \boldsymbol{x}_0 && \text{(7.78j)} \\ & \boldsymbol{u}_{\min} \leq \boldsymbol{u} \leq \boldsymbol{u}_{\max} && \text{(7.78k)} \\ & \mathbf{e}^{\intercal}\boldsymbol{u} \leq \boldsymbol{u}_{\max} && \text{(7.78l)} \\ & 0 \leq \tilde{\boldsymbol{r}} \leq \frac{1}{\epsilon_{\min}} && \text{(7.78m)} \\ & \epsilon_{\min} \leq \boldsymbol{\epsilon} \leq \epsilon_{\max} && \text{(7.78n)} \\ & 0 \leq \boldsymbol{x} \leq 1 && \text{(7.78o)} \\ & 0 \leq \boldsymbol{z} \leq 1 && \text{(7.78p)} \\ & 0 \leq \boldsymbol{y} \leq 1 && \text{(7.78q)} \\ & 0 \leq (1 - \boldsymbol{z}) \perp \tilde{\boldsymbol{r}} \geq 0 && \text{(7.78r)} \end{aligned}$$

Equations (7.78h)-(7.78i) describe the system model. Equation (7.78j) is the initial condition for the states. Equations (7.78k)-(7.78q) are variable bounds. Finally, the last constraint in Equation (7.78r) are the complementarity constraints for the introduced counter variable $\boldsymbol{z}$ that is used to keep track of the number of inspections k.

Table 7.5: Parameters used for the optimization

Parameter	Description	Value
λ_u	Sudden failure transition rate	$10^{-4} \cdot \begin{bmatrix} 2.0 & 2.0 & 1.0 & 0.5 \end{bmatrix}^\intercal$
$\tilde{\lambda}_a$	Base aging transition rate	$10^{-2} \cdot \begin{bmatrix} 1.0 & 1.0 & 0.5 & 0.5 \end{bmatrix}^\intercal$
d	Discount rate	0.001
$\boldsymbol{c}_p$	Productivities in each state	$\begin{bmatrix} 10 & 5 & 5 & 0 \end{bmatrix}^\intercal$
$\boldsymbol{c}_m$	Maintenance cost (replacement)	$10^2 \cdot \begin{bmatrix} 5.0 & 5.0 & 1.0 & 2.0 \end{bmatrix}^\intercal$
$\boldsymbol{c}_m$	Maintenance cost (preventive)	$\frac{10^2}{3} \cdot \begin{bmatrix} 5.0 & 5.0 & 1.0 & 2.0 \end{bmatrix}^\intercal$
$\boldsymbol{c}_i$	Inspection cost	$10^1 \cdot \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix}^\intercal$
t_0	Initial time	0 weeks
t_f	Final time	200 weeks

7.7.6 Simulation results

The model described in Section 7.7 was discretized using direct orthogonal collocation and implemented in MATLAB and CasADi 3.4.1 [11]. IPOPT 3.12.3 [185], an interior point solver for NLPs, was used to solve the problem. The parameter values for the problem are given in Table 7.5.

The optimal system probabilities obtained from the optimization are shown in Figure 7.8. The corresponding decision variables ($\boldsymbol{u}(t)$ and $\tilde{\boldsymbol{r}}(t)$) are shown in Figure 7.9.

We observe that the optimal solution obtained for this case study is very complex. The maintenance times are non-periodically distributed, and not all components are maintained at every inspection, as this would lead to higher maintenance costs. The optimal input profile shows that one compressor is used at maximum capacity the entire time, whereas the other is throttled and run at less than maximum speed for some time to improve the overall economic performance by lowering the probability of system failure and thus the expected maintenance cost. The input-

induced loads increase temporarily at the times when inspections are performed, but these are only numerical artifacts and do not alter the solution significantly.

Initially, the system is inspected and maintained quite often, as the net present value formulation implies a preference for high availability early on. At $T_1 = 32$, $T_2 = 60$, $T_3 = 85$ and $T_4 = 137$ weeks, inspections are performed on all four components. However, various equipment is maintained more often than others. For example, compressor 2 is scheduled to be maintained reactively the first two maintenance interventions, and preventively the first three interventions. The separator, on the other hand, is only scheduled to be repaired reactively on the second and fourth intervention, and preventively at the second inspection. This is due to its overall slower degradation rate and higher reliability. Following the final inspection and maintenance intervention, the system is allowed to degrade and fail, as it is of no economic benefit to keep the availability of the system high when the intended mission time of the system has been reached.

It should be noted that the solution obtained here is only a local solution, as discussed earlier. We performed multi-start optimization until the improvement in the objective function value did not improve significantly for a couple of iterations, at which point it was assumed that a near-global solution was found. At that point, our method had managed to improve the initial guess, a periodic schedule with four inspections and both compressors running at maximum, by about 10%, which we deem to be a significant improvement for engineering applications.

7.7.7 Re-solving the problem using the multi-stage formulation

Now that we have found the optimal inspection, maintenance and operational schedule has been found, we want to investigate whether including the value of future information (i.e. the outcome of inspections) has any significant effect on the solution of the optimization problem. To do so, we lock the number of inspection and maintenance interventions and resolve the optimization problem. The decision variables are now the inspection times and the inputs. By fixing the number of inspections, we can formulate a multi-stage version of the problem, which includes the outcomes of future inspections, as described in Section 7.5. Inspection of a component reveals one of three possible outcomes. Either the component it is as good as new or lightly degraded, i.e. is in one of the discrete degradation states $j = 0$ or $j = 1$. Alternatively, it is found to be severely degraded, i.e. in state $j = 2$. Finally, the component may be found to be broken down, i.e. it is in degradation state $j = N = 3$. Consequently, if all four components are to be inspected, we have $m = 3^4 = 81$ possible outcomes. In other words, the scenario problem has 1 scenario if we choose to use a robust horizon of length $n_r = 0$, 81 scenarios if we use a robust horizon of length $n_r = 1$, and 6561 scenarios if we use a robust horizon of length $n_r = 2$. The optimal objective function values for the three cases where we solve the multi-stage version of the problem with either $n_r = 0$, $n_r = 1$ or $n_r = 2$ are summarized in Table 7.6. For the sake of brevity, the optimal inspection times for the individual scenarios are not shown. As can be seen from the table, the multi-stage formulation with a robust horizon $n_r > 0$ gives better results than the expected value formulation corresponding to $n_r = 0$, although the difference is

Table 7.6: Comparison of different robust horizon lengths

Robust horizon N_r	Number of scenarios N_s	Objective function value [kUSD]
0	1	1473.2
1	81	1474.1
2	6561	1474.2

marginal. The exponential increase in problem size and consequently computation time leads us to conclude that the multi-stage formulation is not necessary for this particular case example.

7.8 Conclusion

We have presented a method for solving the problem of combined maintenance scheduling and production planning for arbitrary degradation models and arbitrary maintenance strategies. The obtained maintenance schedule is non-periodic, and may include multiple different maintenance actions, if included in the model. We derived a general differential model and showed how the resulting optimization problem can be approximated to yield a continuous, although non-smooth, optimization problem that can be solved using standard, off-the-shelf tools for numerical dynamic optimization. Non-convexity of the problem required a multi-start approach to obtain a reasonable, near-global solution. Complementary constraints were handled using a relaxation and penalty method. We also showed how a multi-stage formulation can easily be incorporated in the problem formulation, and may result in better solutions.

Finally, we illustrated the method on a small illustrative example and a more complex case example from the subsea oil and gas industry. For the case study, the method managed to produce a non-periodic and non-intuitive inspection and maintenance schedule, which was found to be better than a simple periodic inspection and production schedule.

Future work within this area should focus on improving the numerical aspects of the problem formulation. Due to non-convex nature of the problem, convergence can be quite slow. A more sophisticated multi-start approach, in combination with a more sophisticated method for handling the complementary constraints, could potentially improve performance. It is known that interior-point solvers such as IPOPT are not well suited to solve MPCCs without reformulations, as the solution has no interior [28]. Active-set methods such as the one in CONOPT [46], which implements a sequential quadratic programming (SQP) method for solving the NLP, may converge quicker due to the efficient detection of active sets and handling of dependent constraints. On the other hand, IPOPT is efficient at solving large-scale optimization problems with many inequality constraints, which is the case when using a direct collocation method to reformulate the dynamic optimization

problem, as we did in this work. A comparison of different NLP solver strategies would be interesting. Furthermore, a comparison with global MINLP solvers such as BARON [149] would also be highly interesting to see the whether the MPCC reformulation is advantageous over solving solving the MINLP directly.

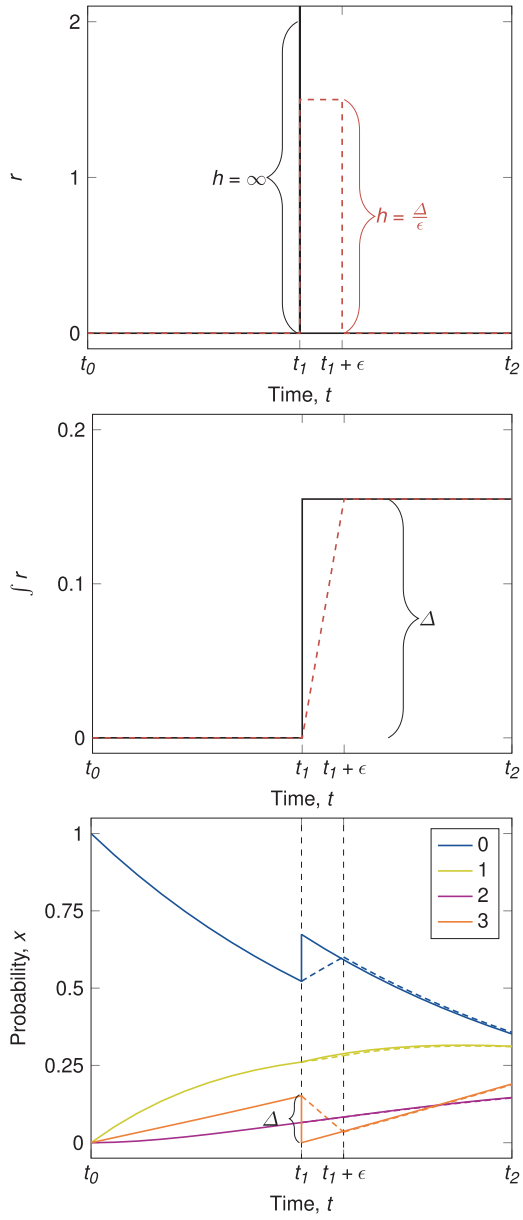

Figure 7.5: Illustration of how $\boldsymbol{r}(t)$ (solid, black line) can be approximated. The top plot shows $\boldsymbol{r}(t)$ and the Boxcar approximation $\tilde{\boldsymbol{r}}(t)$ (dashed, red line). The bottom plot shows the original discontinuous states $\boldsymbol{x}$ (solid lines), and the states due to the approximation (dashed), which are continuous.

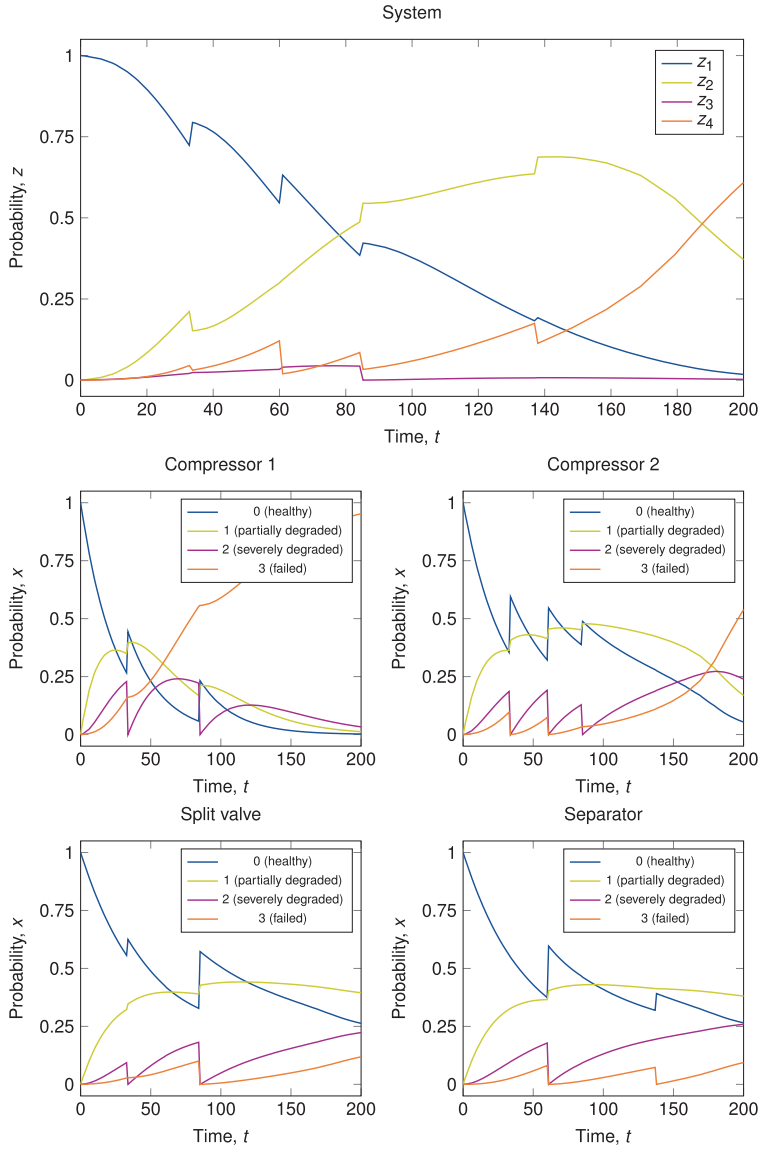

Figure 7.8: Optimal solution obtained from the optimization of the case study. The subplot on top shows the system states, i.e. the four composite events described in Section 7.7.3. The four smaller subplots each show the probabilities x_i of being in the four health states for compressor 1, compressor 2, the split valve and the separator, respectively (counting left to right, top to bottom).

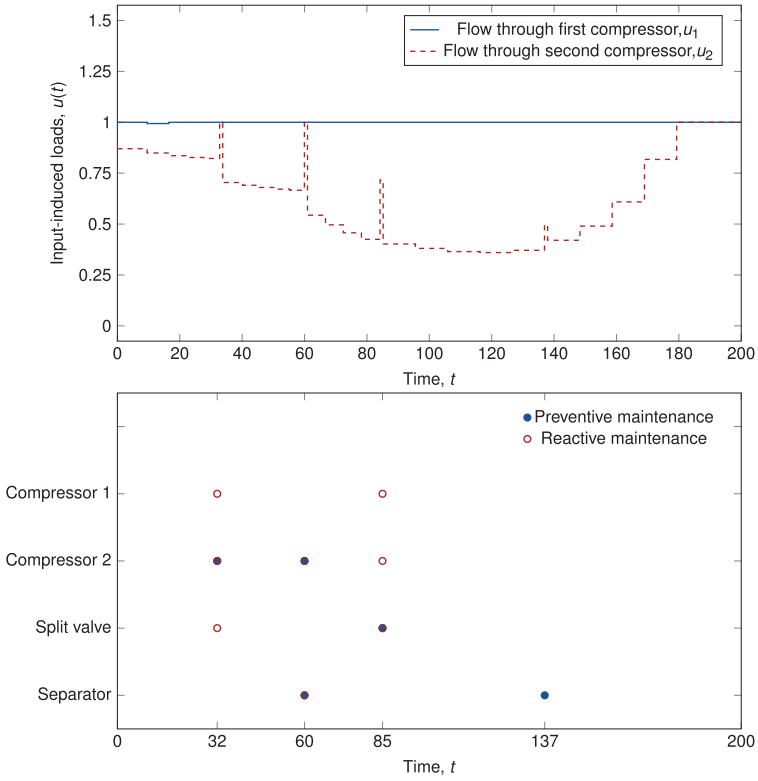

Figure 7.9: Optimal solution from the optimization of the case study. The upper plot shows the two input-induced loads u_1 and u_2, which are the mass flows through the compressors. The middle plot shows the optimal reactive maintenance times, and the bottom plot shows the optimal preventive maintenance times.

Chapter 8

Summary and future work

8.1 Summary

The topic of this thesis is the combination of reliability and process control for subsea production and processing plants. In the introduction, we posed five research questions, which were addressed in this thesis through a series of peer reviewed conference and journal articles. We repeat the research questions and give a short summary of what was found:

Research question 1. *How do we best utilize the novel degrees of freedom introduced by advanced processing equipment such as compressors, pumps and control valves?*

- Several control strategies for various subsea production and processing cases were developed. For new equipment such as compressors, the degrees of freedom were used to operate in a stable region, while ensuring high economic profit without endangering the reliability. A control strategy for choke valves subject to choke erosion was also developed, which ensured that the production of hydrocarbons was maximized while the erosion stayed within acceptable limits.

Research question 2. *What is the current status of subsea PHM technology?*

- The status of subsea PHM technology was discussed in Chapter 2. It was found that monitoring solutions for subsea equipment exist and are used, but that the information from the monitoring solutions is very rarely used to make control decisions in an automated fashion. It was also found that predictive models for degradation of subsea equipment are usually poor and not well documented in literature. The lack of data to construct such models was also found to be an issue. Data is either proprietary, sampled too infrequently or non-existing.

Research question 3. *How can we systematically integrate PHM data into our existing control structures to ensure that we have a control structure that adjusts production such that equipment RUL is not exhausted prematurely?*

- Systematic step-by-step approaches for developing health-aware control structures were presented in Chapters 3 and 4.

Research question 4. *Knowing that equipment will have to be be maintained in the future, what is the optimal way of operating between maintenances, and what is the optimal schedule for maintenance of the equipment?*

- A new method for simultaneously optimizing production and maintenance schedules was developed in Chapter 7 and Appendix B. The problem is formulated as a non-linear problem with complementarity constraints, and solved with IPOPT. Using a multi-start approach, it was found that the method gave near-globally optimal solutions. In comparison to existing periodic maintenance scheduling strategies, which are commonly used in industry, the non-periodic maintenance schedules obtained with this method performed significantly better.

Research question 5. *How can we ensure that our strategy is still optimal in the presence of uncertainty and stochasticity?*

- Uncertainty and stochasticity were addressed by formulating robust formulations of the optimization problems. Several methods were investigated. It was found that scenario-based approaches were robust towards modelled uncertainties, were easy to implement, and had good performance when used with sparsity-exploiting solvers such as IPOPT.

8.2 Future work

Future work should focus on developing better prognostic degradation models for subsea equipment. Very little data exists for how equipment degrades under varying operating conditions, as would be the case for a process with active control. If such models can be developed, the uncertainty of the models will be reduced, leading to less conservative operation.

The method for combined optimization of production and maintenance scheduling should also be improved. The multi-start approach used to ensure near-global optimality is somewhat crude in its implementation, and calculation times can be sped up significantly by using a more advanced method. More sophisticated algorithms for solving complementarity-constrained NLPs should also be used, e.g. branch and bound algorithms or evolutionary algorithms.

Finally, the ideas discussed in this thesis should be applied to a real world case study, either from the area of subsea production and processing, or from other processes in chemical industry.

As a continuation of this project, we have designed and built a small lab rig at the department of chemical engineering at NTNU. The experimental setup allows us to test the algorithms developed for particle erosion from Chapter 6, as well as algorithms developed for gas lift optimization. Preliminary experimental results look very promising.

Appendices

Appendix A

Robust Health-aware operation of a subsea gas compression system

This chapter is from the article

- A. Verheyleweghen and J. Jäschke. Robust Health-aware operation of a subsea gas compression system. *Proceedings of the 2017 Conference of Foundations of Computer Aided Process Operations / Chemical Process Control (FOCAPO/CPC)*. 2017

A.1 Abstract

In this paper we apply health-aware control ideas to the optimal operation of a subsea gas compression plant. Subsea systems operate in harsh environments and under uncertain and varying operation conditions. Because they are very difficult and expensive to access, an optimal operational strategy that tries to maximize hydrocarbon production must ensure that no unplanned shutdowns due to premature equipment failures occur. In this paper we apply two approaches for optimization under uncertainty in order to maximize the economic profit, while ensuring that the subsea compression plant remains operational until the next planned maintenance. We consider a min-max robust optimization and a scenario-based optimization with recourse. Although both methods avoid unplanned shutdowns, the scenario-based method results in a less conservative solution at the cost of a larger problem size.

Introduction

Most oil and gas fields that are easy to develop have been exhausted, forcing the petroleum industry to produce from more difficult fields with larger water depths, longer tie-back distances and harsh climate conditions. Subsea processing technology is an enabling technology for development of such fields, although several new challenges arise when production and processing facilities are put on the seabed [140]. One of the challenges is that the process is not easily accessible for maintenance. Since maintenance interventions require specialized lifting vessels, fair weather conditions and available spare modules, unanticipated breakdowns can

lead to long production halts and large production losses. The lifting vessels sometimes cost several tens to hundreds of millions of dollars to rent, and must be booked several months in advance. For this reason, stringent requirements on safety and reliability are imposed on operation of these processes. This in turn often leads to conservative design and operation strategies and the economic potential of the field is often not fully realized.

In this paper, we propose to combine reliability and operational considerations in an model predictive control-like framework with shrinking horizon. In particular, we present an approach that ensures that the remaining useful life (RUL) of the equipment is not exhausted before the next planned maintenance stop, while at the same time maximizing the expected operational profit.

A few other authors investigated the combination of the prognostics and health monitoring (PHM) with advanced control methods such as model predictive control (MPC). MPC is a control strategy based on repeated optimization of a process model to obtain optimal input trajectories. Due to its ability to deal with multi-variate, constrained problems, MPC has gained popularity in industry in recent years [115]. Health prognostics information is usually not taken explicitly into consideration when calculating the optimal control moves, and this can lead to sub-optimal operation [151]. If a prognostic model is available, the system health state can be included as constraints in the optimization [134, 151], or as terms in the objective function [56]. The term *health-aware control* was introduced by [56] to describe a control structure that pro-actively adjusts the inputs to prevent a fault from occurring. The health-aware control structure thereby distinguishes itself from the more established fault-tolerant control (FTC) structure, which only takes corrective action once a fault has already occurred. Similar ideas are discussed in papers by [134], who include PHM in an MPC framework to redistribute the control efforts among redundant actuators to prevent actuator breakdown, and [151], who model the reliability of pumps in a drinking water network using Bayesian networks and include the system reliability in the MPC formulation.

In this work we present a comparative study of two robust approaches applied to a subsea gas compression system. We model a subsea gas compression station and define the optimal control objective. The reliability of the system is ensured by constraining the health-state of the compressor, which is assumed to be the critical component. The degradation of the compressor health is assumed to be a function of the input usage and uncertain parameters. In particular, we assume parametric uncertainty in the compressor health degradation model and calculate the robust solution using both a scenario-based MPC approach, and a worst-case MPC approach.

Combining Prognostics and Control

To start with, we assume that the health state, h, of the system is observable, and we define a minimum health limit, h_{min}, above which we have to operate. Violation of this constraint corresponds to an unacceptable risk of failure. We assume the health to be monotonously decreasing, i.e. the system is not repaired or maintained

before the final time t_f is reached. Because of the fixed final time, the MPC has a shrinking horizon rather than the more common receding horizon.

Due to the inherent uncertainty in the model, the optimization problem solved at each time step in the MPC is usually stochastic, because most prognostic models are statistics-based. The stochastic optimization problem can be written as

$$\min_{\mathbf{u}} \quad \mathbb{E}(f_0(\mathbf{u},\mathbf{p})) \quad \text{s.t.} \begin{cases} g_i(\mathbf{u},\mathbf{p}) = 0 & i=1,\dots,n_g \\ f_j(\mathbf{u},\mathbf{p}) \geq 0 & j=1,\dots,n_f \\ h_k(\mathbf{u},\mathbf{p}) \geq 0 & k=1,\dots,n_h \end{cases} \tag{A.1}$$

where $\mathbf{p} \in \mathcal{P}$. In the above expression, we use $\mathbf{u}$ to denote the inputs and $\mathbf{p}$ to denote the uncertain parameters, which are contained in the (bounded) set $\mathcal{P}$. f_0 is the objective function, g are the equality constraints, f are the inequality constraints and h are the constraints on the equipment RUL. The operator $\mathbb{E}$ is used to signify the expected value of the objective function. Below, we discuss two approaches for addressing the uncertainty.

Min-Max Model Predictive Control

One way to handle the uncertainty is the "min-max"-approach, in which the objective function is optimized given a worst-case realization of the uncertain parameters. The min-max-approach, sometimes also referred to as the "robust" approach, was implemented in a receding horizon MPC framework in [196].

In the non-linear case, identifying the worst-case realization can usually not be done explicitly. Rather, the worst-case realization is found through maximization of the inequality constraints, subject to bounds on the norm of the random parameters. Consequently, a bi-level optimization problem has to be solved at each stage of the min-max MPC.

$$\min_{\mathbf{u}} \quad \mathbb{E}(f_0(\mathbf{u},\mathbf{p})) \quad \text{s.t.} \begin{cases} \phi_i(\mathbf{u}) \geq 0 & i=1,\dots,n_f+n_h \end{cases} \tag{A.2a}$$

where

$$\phi_i(\mathbf{u}) = \max_{\mathbf{p}} \quad \hat{f}_i(\mathbf{u},\mathbf{p}) \quad \text{s.t.} \begin{cases} g_j(\mathbf{u},\mathbf{p}) = 0 & j=1,\dots,n_g \\ \mathbf{p} \in \mathcal{P} \end{cases} \tag{A.2b}$$

and $\hat{f} = \left[f_1,\dots,f_{n_f},h_1,\dots,h_{n_h}\right]^\intercal$. Bi-level problems are difficult to solve, as they quickly become numerically intractable. [42] propose an approximated robust counterpart of the nonlinear optimization problem, which is numerically efficient. The min-max-approach is often very conservative [159], because the possibility of future information about the realizations, i.e. feedback, and the possibility of other realizations than the worst-case, are ignored when solving the problem.

Scenario-based Model Predictive Control

As a remedy, [159] propose a multi-stage approach with recourse. Scenario-based MPC has its roots in multi-stage stochastic programming. The core idea in scenario-

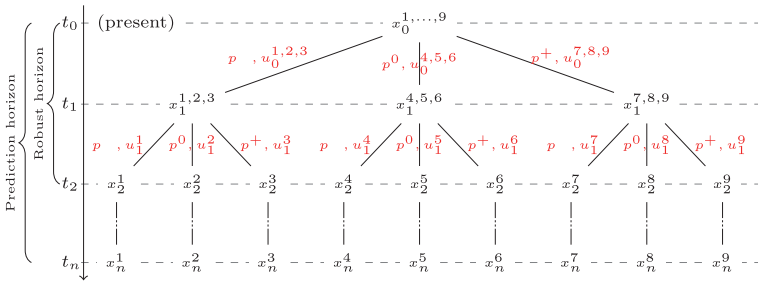

Figure A.1: Illustration of a scenario tree with robust horizon of length $n_{robust} = 2$ and prediction horizon of length n. At each stage there are three possible realizations of the uncertain parameter, p^+, p^0 and p^-.

based optimization is to assume a discrete probability distribution for the uncertain parameters. A finite number of scenarios are then generated to represent how the uncertainty may develop over time. For the resulting scenario tree, the expected objective function value is then minimized subject to non-anticipativity constraints, which require that the decisions only depend on the past realizations of the random parameters and their probability distribution. Future realizations cannot be anticipated, and are therefore not included in the decision making process [48].

Due to the need for additional variables and constraints, the complexity of scenario-based MPC increases with the number of scenarios. In order to keep the problem tractable, the scenario tree only branches up until a certain stage, called the robust horizon [102]. After the robust horizon, the realizations of the uncertain parameters are kept constant.

An illustration of a scenario tree with a robust horizon with length $n_{robust} = 2$ and a prediction horizon with length n is shown in Figure A.1.

A challenging task is the selection of a representative scenario tree. Especially when the dimensionality of the problem becomes large, it is nontrivial to reduce the scenario tree to a manageable size. One way to generate the scenario tree by using combinations of the maximum, minimum and nominal uncertain parameters. A scenario tree generated this way will result in a feasible solution for linear systems, and typically works for non-linear systems in which the degree of non-linearity is not too large [102].

Case study

Process description

A subsea gas compression station (Figure A.2) is used to illustrate the robust health-aware control strategy in this paper. The process is similar to the gas compression stations installed on the Åsgard field and the Ormen Lange pilot.

The plant consists of a single choke, which regulates the flow of oil and gas from the

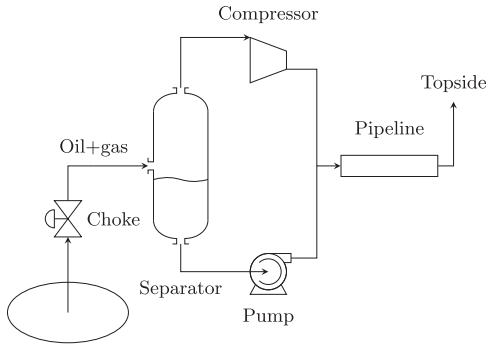

Figure A.2: Process diagram of the subsea gas compression station.

reservoir, a scrubber which separates the gas from the oil, and a wet-gas compressor to achieve sufficient gas pressure for transport through the pipeline. Due to non-perfect separation, some liquid droplets are carried over in the gas stream. The separation efficiency of the scrubber is assumed to be a function of the gas velocity and the fluid density [16]. The compressor in our system is a wet gas compressor which can handle moderate liquid carry-over, with a suction gas-volume-fraction from 0.95 to 1. A full description of the compressor model, including the compressor maps, can be found in [4]. The liquid stream from the separator is boosted before being recombined with the compressed gas stream. Finally, the multiphase flow is sent to the receiving facility through a long subsea pipeline.

We assume that the wet-gas compressor is the critical component in terms of overall system reliability. We therefore make the simplifying assumption that the wet-gas compressor is the only component whose reliability will decrease on the given time horizon. As rotating equipment is prone to wear damage, leakage and signal failure, due to its complexity and many moving parts [99], this assumption seems reasonable. Tests from the Ormen Lange pilot have shown that the RUL of the compressor is strongly linked to the operating conditions [53], which makes it a good choice for showcasing a health-aware operating strategy.

Modelling the compressor degradation

In general it is hard to predict exactly when or how a compressor is going to fail. [99] lists some common causes for compressor failure, including how they can be monitored. Both [53] and [99] report that the magnetic bearings of the compressor are critical components and prone to fatigue failure. [53] also found that the health state of the active magnetic bearings is observable through their power consumption. The reason for this is that damage to the compressor innards causes an imbalance of the driving shaft. Consequently the magnetic bearings require more power to stabilize this imbalance.

We propose to model the health degradation of the compressor over one month of operation, Δh, as a result of wear, which is proportional to the dimensionless

compressor speed N, and shock damage, which is caused by set-point changes in the compressor speed, $|\Delta N|$.

$$\Delta h = -\left(\underbrace{p_N N^n}_{\text{Wear and tear}} + \underbrace{p_\Delta |\Delta N|^{n_\Delta}}_{\text{Shock damage}} \right) \cdot f(\mathbf{y}) \tag{A.3}$$

Here, h is the compressor health, which ranges from 1 to 0 (breakdown). $\mathbf{p} = [p_N p_\Delta]$ denotes the random parameters that affect the degradation. We assume that $\mathbf{p}$ follows a Gaussian distribution. The function f, which is a function of the measurements $\mathbf{y}$, is introduced to take into account the increased rate of wear in multiphase fluids. [53] report that the compressor life scales cubically with the compressor speed. We therefore chose the coefficients $n = n_\Delta = 3$. Furthermore we assume that the compressor degrades exponentially with the liquid content in the gas.

$$\Delta h = -\left(p_N N^3 + p_\Delta |\Delta N|^3\right) \cdot \exp(1 - \text{GVF}) \tag{A.4}$$

where GVF is the gas volume fraction at the inlet of the compressor.

Since p_N, N, p_Δ and $|\Delta N|$are nonnegative, the compressor health is monotonously decreasing, and failure is defined as the event when h goes below a failure threshold value h_{min}. We assume that h is measurable.

Defining Optimal Control Problems

The objective of the plant operation is to maximize the profit of the plant between *planned* maintenance stops. As a simplification, we assume that the variation in the variable operational expenses (in particular the power usage of the compressor) are negligible compared to the income due to gas production. Furthermore, we assume that gas is the only valuable product, and the contribution of oil can be neglected in the objective function. Taken into account that gas that is produced today, is worth more than gas that is produced in the future, we use the net present value (NPV) of the gas in the objective function.

The discharge pressure from the compressor is bounded from below to make sure that the carbohydrate stream has enough pressure to overcome the flow resistance in the transport pipeline. Moreover, we add constraints to prevent compressor surge and compressor choke/Stonewall conditions. Both these phenomena are undesired, so this operating region must be avoided. All bounds are listed in Table A.1.

Deterministic formulation

We formulate the objective function for the optimal control problem as

$$f_0(\dot{m}_{gas}, t_f) = -\int_0^{t_f} \text{NPV}(\dot{m}_{gas}) dt, \tag{A.5}$$

where t_f is the time until the next planned maintenance stop.

Table A.1: Bounds for the variables

Variable	Lower	Upper
Discharge pressure	15 bar	-
Compressor health	0.8	1.0
Compressor surge	0	-
Compressor choke	0	-
Compressor speed	0.6	1.05
Choke opening	0.0	1.0

Table A.2: Values of the uncertain variables p_N and $p_{\Delta N}$ in the scenarios used to generate the scenario tree.

Scenario	p_N	$p_{\Delta N}$
LL	0.006 (μ 2σ)	0.6 (μ 2σ)
LH	0.006 (μ 2σ)	1.8 ($\mu + 2\sigma$)
HL	0.018 ($\mu + 2\sigma$)	0.6 (μ 2σ)
HH	0.018 ($\mu + 2\sigma$)	1.8 ($\mu + 2\sigma$)
mean	0.012 (μ)	1.2 (μ)

The optimization problem is solved using Casadi 3.0.0 [10] in MATLAB R2015a. The problem is discretized using a third order direct collocation scheme and solved with Ipopt 3.12.3 [185].

Stochastic multi-stage approach

Robustness towards parametric uncertainty in the parameters p_N and $p_{\Delta N}$ in the compressor degradation model from Equation (A.3) is achieved by discretizing their probability density function and applying the scenario-based method. Five different scenarios are considered: *HH*, *HL*, *LH*, *LL* and *mean*. These are the combinations of the maximum, minimum and nominal realizations. See Table A.2 for the specific values. All five scenarios are equally probable. An initial prediction horizon of length $n = 20$ and a robust horizon of length $n_{robust} = 1$ is used to speed up the calculation. Higher robust horizons were tested as well, but were not found to improve the solution significantly while resulting in a much higher computational cost.

Min-max approach

Robustness can also be achieved by considering a worst-case scenario in the optimization. For a general, non-linear case, the approximate robust counterpart problem may be solved using the method described in [42]. For the current system, it is not strictly necessary to define the robust counterpart, as it can be determined a priori that the *HH*-scenario from Table A.2 will always be the worst-case scenario.

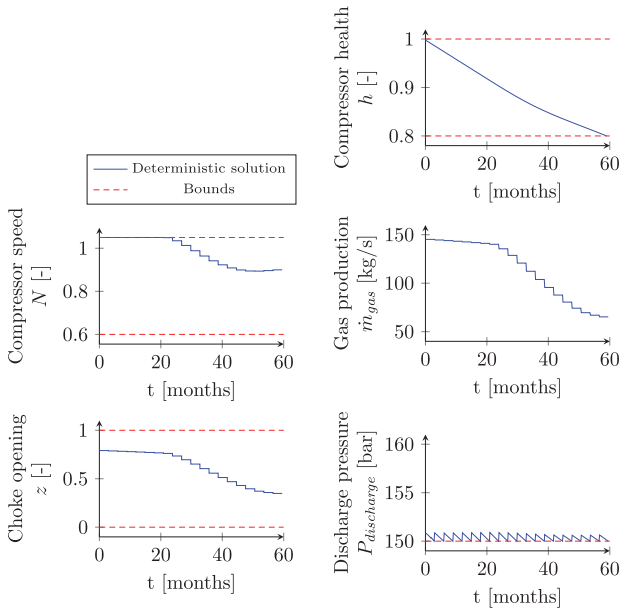

Figure A.3: Deterministic open-loop solution when $p_N = 0.015$ and $p_\Delta = 1.5$.

Results

Deterministic open-loop solution

The deterministic open-loop solution can be seen in Figure A.3. It can be seen that the constraints are satisfied, and that the compressor health constraint is active at the end of the horizon. Since the NPV of the gas production is considered, early production is favored over late production.

Closed-loop results

The closed-loop responses of three control structures are shown in Figure A.4. Firstly, notice that the non-robust approach, in which expected values are considered for the uncertain parameters, leads to repeated violations of the constraints on the discharge pressure and the final health constraint. In contrast, the two robust approaches both satisfy all constraints, as is to be expected. In both cases, there is a back-off from the constraints to account for uncertainty. It can be seen that the scenario-based approach is less conservative than the worst-case approach, since it results in overall higher gas production, $\dot{m}_{gas}$.

The values of the cost function for the three different cases are shown in Table A.3. Note that the scenario-based method yields a higher net present gas production than the worst-case method, but lower than the non-robust method based on expected values. The higher gas production for the non-robust case comes at the cost of constraint violation (i.e. an unplanned maintenance stop). The 2.6% higher net

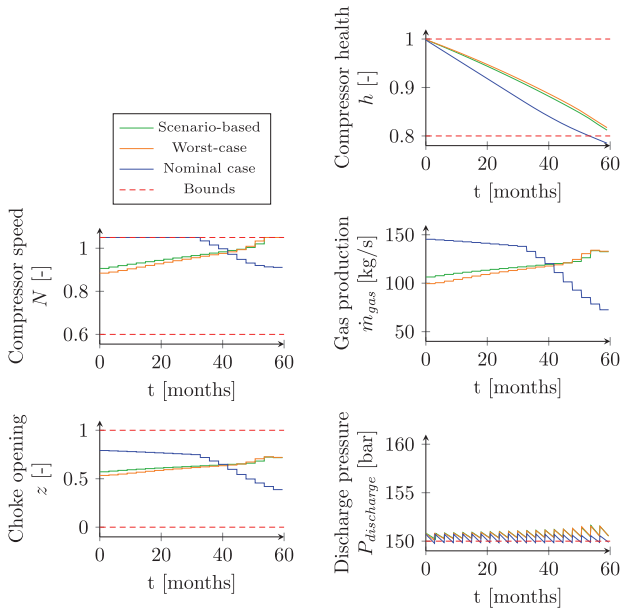

Figure A.4: Comparison of closed-loop performance of three different controllers in the presence of uncertainty. The realizations of the uncertain variables are $p_N = 0.015$ and $p_\Delta = 1.5$.

Table A.3: Normalized profit, i.e. net present gas production, for the three methods (in closed-loop).

Method	**Discounted closed-loop profit**
Scenario-based	1.026
Worst-case	1.000
Nominal case	1.056*

present gas production of the scenario-based method, compared to the worst-case approach, may be a substantial increase in profit.

Conclusion

We have developed a model for a subsea gas compression system and shown how prognostics can be included in the decision-making process to obtain a control structure that gives economical and safe operation. Robustness towards parametric uncertainty is very important in this application, since the health-constraint always will be active. To achieve robustness, we employ a scenario-based optimization method, which is shown to be less conservative than a worst-case approach.

Future work will focus on measurement feedback and health state estimation, more

detailed degradation models and extension to system-wide health-aware operation.

Appendix B

Combined Maintenance Scheduling and Production Optimization

This chapter is from the article

- A. Verheyleweghen, H. Srivastav, A. Barros and J. Jäschke. Combined Maintenance and Production Optimization (European Safety and Reliability Conference (ESREL), Hannover, Germany). *Proceedings of the 29th European Safety and Reliability Conference*. pages 499—506, 2018

B.1 Abstract

Optimal operation of complex production and processing plants is important, but challenging to achieve in practice. The reason for this is that decisions in different disciplines, such as design, operations and maintenance, are made independently of each other. This can lead to a large degree of conservativeness. In this paper, we present a unified approach for maintenance- and production planning, which reduces the conservativeness and leads to more economical operation. We model the system using differential equations and then formulate the problem of optimal operation as a numerical optimization problem. The problem is a mathematical program with equilibrium constraints (MPEC), which we solve using off-the-shelf optimization software. Some model approximations were made to make the system numerically tractable. We demonstrate the method on a subsea-inspired case example.

B.2 Introduction

For certain classes of production systems there exist a trade-off between producing as much as possible, and prematurely wearing the system out. For example, in subsea oil and gas production systems the revenue is directly related with the amount of produced hydrocarbons, while a too high production rate may lead to fast system degradation, with expensive maintenance operations and possible production loss due to downtime.

From an economical point of view, there exists an optimal trade-off between the maintenance cost, the inspection cost and the operational profit. Moreover, when the system has degraded, the operator needs to make a decision relating to what degree the system should be restored. Using commonly employed Monte Carlo methods to obtain the optimal production profile and the optimal maintenance schedule, represents a significant challenge due to the sheer amount of possible scenarios that need to be explored. Numerical optimization seems like an attractive alternative to Monte Carlo methods due to the potential for faster convergence to an optimal solution.

In this paper we propose an integrated approach to operate the system optimally, that is, we propose to integrate the decisions on 1) the system load (how much to produce), 2) when to perform a maintenance operation, and 3) to what degree the system should be restored, in a unified framework.

To demonstrate our approach, we model a subsea oil and gas production system using a four-state Markov chain, where state A represents the new, healthy system, states B and C represent progressively degraded systems, and state D represents the failed, inoperable system. Arrival at the failed state D can be caused by unexpected sudden failure, or due to progressive degradation. The time dependent transition rates are a function of the input usage, thus yielding a multiphase Markov decision process. The production system model is described by a non-linear differential-algebraic equation system (DAE).

Optimal production and operation planning is defined as the case when the sum of the expected value of the revenue minus the inspection cost and the maintenance cost, is maximized. As decision variables in the optimization problem we assume the inspection times and the input profile. By inputs, we here mean the operating mode that the plant is running at. For example, the inputs of a compressor could be its throughput and frequency.

At each inspection, all systems found in the failed state D are restored to state A (if we follow the as-good-as-new (AGAN) policy). If the system is not found in state D, the system does not reveal whether its true state is A, B or C, and no maintenance is performed. Consequently, the model becomes a switched differential algebraic model, and the optimization problem can be formulated as a non-smooth, non-linear program. We solve this problem using state of the art methods for non-smooth optimization.

Authors of previously published work on the topic of combined maintenance and production planning typically formulate the problem as a lot-allocation problem [58, 77, 188]. This often results in a mixed-integer (non-) linear program (MI(N)LP). Our proposed method is different as we do not consider "lots" of products, but rather the case where production can be adjusted in a continuous fashion. We also avoid the use of integer variables by formulating the problem with complementary constraints instead.

The remainder of the paper is structured as follows; in Section B.3 we show how a continuous differential equation can be derived for the case of a degrading system with discrete inspection- and maintenance times. In Section B.4 it is shown how

the derived model can be used in the context of optimization, where the aim is to manipulate inputs and the maintenance times to minimize some objective function. In Section B.5, the method is demonstrated on the subsea case example. Finally, concluding remarks and thoughts on future work are given in Section B.6.

B.3 Modelling the degrading system

Given a four-state Markov process as shown in Figure B.1.

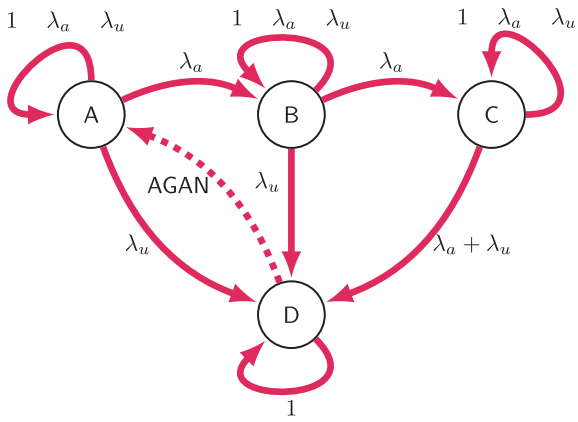

Figure B.1: Markov chain for a system with four discrete degradation states

Let $\boldsymbol{\mu}(t) = \begin{bmatrix}\mu_A(t) & \mu_B(t) & \mu_C(t) & \mu_D(t)\end{bmatrix}^\intercal$ denote the probabilities for being in any state at time t. Assuming time-invariant transition rates λ_a and λ_u between the states, the change in probabilities between two inspections is given by

$$\frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{\Lambda} \cdot \boldsymbol{\mu}(t) \tag{B.1}$$

$$= (\boldsymbol{\Lambda}_a + \boldsymbol{\Lambda}_u) \cdot \boldsymbol{\mu}(t) \tag{B.2}$$

where

$$\boldsymbol{\Lambda}_a = \begin{bmatrix} -\lambda_a & 0 & 0 & 0 \\ \lambda_a & -\lambda_a & 0 & 0 \\ 0 & \lambda_a & -\lambda_a & 0 \\ 0 & 0 & \lambda_a & 0 \end{bmatrix} \tag{B.3}$$

and

$$\boldsymbol{\Lambda}_u = \begin{bmatrix} -\lambda_u & 0 & 0 & 0 \\ 0 & -\lambda_u & 0 & 0 \\ 0 & 0 & -\lambda_u & 0 \\ \lambda_u & \lambda_u & \lambda_u & 0 \end{bmatrix}. \tag{B.4}$$

In the above expressions, $\mathbf{\Lambda}$ is known as the transition matrix. $\mathbf{\Lambda}$ is decomposed into $\mathbf{\Lambda}_a$, which is describing the transitions due to aging, and $\mathbf{\Lambda}_u$, which is describing the transitions due to unforeseen failures.

Integrating Equation (B.1) gives

$$\boldsymbol{\mu}(t) = \exp(\mathbf{\Lambda} \cdot (t - t_0))\boldsymbol{\mu}(t_0) \tag{B.5}$$

B.3.1 Probabilities between two inspections

Upon inspection of the system at time t_1, we reveal if the system is broken down (in state D), or not (either in state A, B or C). This leads to two different cases, depending on the outcome:

Case I:
Upon inspection, we find the system is in state D. Thus, we restore the system by performing maintenance (without time lag). Assuming perfect repairs according to the AGAN policy, the new initial conditions t_1 are

$$\boldsymbol{\mu}^{+}_{\text{Case I}}(t_1) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^{\mathsf{T}} \tag{B.6}$$

and

$$\boldsymbol{\mu}_{\text{Case I}}(t > t_1) = \exp\left(\mathbf{\Lambda} \cdot (t - t_1)\right) \boldsymbol{\mu}^{+}_{\text{Case I}}(t_1) \tag{B.7}$$

Note that we use the notation $\boldsymbol{\mu}^{+}(t_1)$ to indicate the right-handed limit of $\boldsymbol{\mu}(t_1)$, i.e. directly after the inspection at t_1, and $\boldsymbol{\mu}^{-}$ to indicate the left-handed limit of $\boldsymbol{\mu}$, i.e. directly before the inspection at t_1. Because $\boldsymbol{\mu}$ is discontinuous at t_1, these two limits will generally not be equal.

Case II:
Upon inspection, we find that the system is not in state D. However, we do not know if the system is in state A, B or C. The new initial conditions at t_1 are:

$$\boldsymbol{\mu}^{+}_{\text{Case II}}(t_1) = \begin{bmatrix} \frac{\mu^{-}_A(t_1)}{1-\mu^{-}_D(t_1)} & \frac{\mu^{-}_B(t_1)}{1-\mu^{-}_D(t_1)} & \frac{\mu^{-}_C(t_1)}{1-\mu^{-}_D(t_1)} & 0 \end{bmatrix}^{\mathsf{T}} \tag{B.8}$$

and

$$\boldsymbol{\mu}_{\text{Case II}}(t > t_1) = \exp\left(\mathbf{\Lambda} \cdot (t - t_1)\right) \boldsymbol{\mu}^{+}_{\text{Case II}}(t_1) \tag{B.9}$$

Expressing the probabilities from the perspective of t_0
However, we cannot know ahead of time which of the two cases will be observed in the future (non-anticipativity). We must therefore forecast the probabilities into the future by taking the weighted average of both cases.

$$\begin{aligned} \boldsymbol{\mu}^{+}(t_1) &= \boldsymbol{\mu}^{+}_{\text{Case I}}(t_1) \cdot \mu^{-}_D(t_1) && \text{(B.10)} \\ &+ \boldsymbol{\mu}^{+}_{\text{Case II}}(t_1) \cdot \left(1 - \mu^{-}_D(t_1)\right) && \text{(B.11)} \\ &= \begin{bmatrix} \mu^{-}_A(t_1) + \mu^{-}_D(t_1) \\ \mu^{-}_B(t_1) \\ \mu^{-}_C(t_1) \\ 0 \end{bmatrix} && \text{(B.12)} \end{aligned}$$

Or, in matrix notation:

$$\boldsymbol{\mu}^{+}(t_1) = \boldsymbol{M} \cdot \exp\left(\boldsymbol{\Lambda}_0 \left(t_1 - t_0\right)\right) \cdot \boldsymbol{\mu}(t_0) \tag{B.13}$$
$$= \boldsymbol{M} \cdot \boldsymbol{\mu}^{-}(t_1) \tag{B.14}$$

where we further decompose $\boldsymbol{M}$ as

$$\boldsymbol{M} = (\boldsymbol{I} + \boldsymbol{R}\boldsymbol{S}^{\intercal}) \tag{B.15}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{R}$ is the repair matrix, and $\boldsymbol{S}$ is a selection matrix. The selection matrix chooses the failed state D.

$$\boldsymbol{S} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^{\intercal} \tag{B.16}$$

For AGAN repairs, we have

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 & -1 \end{bmatrix}^{\intercal}. \tag{B.17}$$

In Section B.3.2, we will come back to why the decomposition of $\boldsymbol{M}$ into $\boldsymbol{R}$ and $\boldsymbol{S}$ is useful.

The evolution of the probabilities for case 1 and 2, and the weighted average of the two cases is illustrated in Figure B.2.

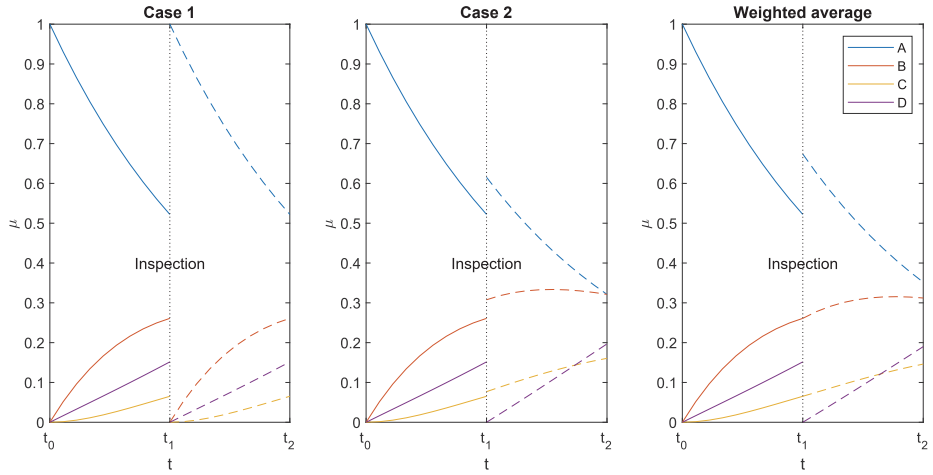

Figure B.2: Illustration of evolution of the probabilities μ before and after inspection. Left: system is found to be in the failed state D upon inspection, and is repaired without time lag. Middle: system is found to be in working order upon inspection, and no repairs are performed. Right: weighted average of the two previous cases.

Generalization

Using the general expression for the probabilities between two maintenance stops

from Equation (B.5), we can express the probabilities at any time t as the piece-wise function

$$\boldsymbol{\mu}(t) = \begin{cases} \exp\left(\boldsymbol{\Lambda} \cdot (t - t_0)\right) \cdot \boldsymbol{\mu}^+(t_0) & \text{if} \quad t < t_1 \\ \exp\left(\boldsymbol{\Lambda} \cdot (t - t_1)\right) \cdot \boldsymbol{\mu}^+(t_1) & \text{if} \quad t > t_1 \end{cases} \tag{B.18}$$

where

$$\boldsymbol{\mu}^+(t_0) = \boldsymbol{\mu}(t_0) = \boldsymbol{\mu}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\intercal \tag{B.19}$$

is the specified initial condition.

Taking it one step further, we can have an arbitrary amount of inspections and maintenances, k, between two times t_0 and t_f and express the probabilities as

$$\boldsymbol{\mu}(t) = \begin{cases} \exp\left(\boldsymbol{\Lambda} \cdot (t - t_0)\right) \cdot \boldsymbol{\mu}^+(t_0) \\ \quad \text{if} \quad t_0 < t < t_1 \\ \exp\left(\boldsymbol{\Lambda} \cdot (t - t_1)\right) \cdot \boldsymbol{\mu}^+(t_1) \\ \quad \text{if} \quad t_1 < t < t_2 \\ \dots \\ \exp\left(\boldsymbol{\Lambda} \cdot (t - t_k)\right) \cdot \boldsymbol{\mu}^+(t_k) \\ \quad \text{if} \quad t_k < t < t_f \end{cases} \tag{B.20}$$

where

$$\boldsymbol{\mu}^+(t_i) = \boldsymbol{M} \cdot \boldsymbol{\mu}^-(t_i) \tag{B.21}$$

$$\boldsymbol{\mu}^-(t_i) = \exp\left(\boldsymbol{\Lambda} \cdot (t_i - t_{i-1})\right) \cdot \boldsymbol{\mu}^+(t_{i-1}) \tag{B.22}$$

B.3.2 Differentiating to get the differential model

Equation (B.1) described the evolution of the state $\boldsymbol{\mu}$ between two inspection times. However, as we have shown, we can express the state at any given time by the piece-wise model from Equation (B.20). If we differentiate Equation (B.20), we get

$$\frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{\Lambda} \cdot \boldsymbol{\mu}(t) + \boldsymbol{R}\boldsymbol{S}^\intercal \boldsymbol{\mu}(t) \cdot \left(\sum_{i=1}^{k} \delta(t - t_i) \right) \tag{B.23}$$

where δ is the Dirac delta function.

Let us now introduce the variable

$$\boldsymbol{r}(t) = \boldsymbol{S}^\intercal \boldsymbol{\mu}(t) \cdot \left(\sum_{i=1}^{k} \delta(t - t_i) \right) \tag{B.24}$$

to obtain the form

$$\frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{\Lambda} \cdot \boldsymbol{\mu}(t) + \boldsymbol{R} \cdot \boldsymbol{r}(t). \tag{B.25}$$

We choose to work with this form of the reliability model as it allows us to distinguish between the degradation of the system due to aging and unforeseen failures (first term of Equation (B.25)), and the maintenance of the system (second term

of Equation (B.25)). Our aim is to do numerical optimization of the maintenance times, which can be achieved by optimizing the breakpoints t_i of the function $\boldsymbol{r}(t)$. More on this in Section B.4.

Furthermore, we can easily change the maintenance strategy from as-good-as-new (AGAN) to as-bad-as-old (ABAO) by changing $\boldsymbol{R}$

$$\boldsymbol{R}_{\text{AGAN}} = \begin{bmatrix} 1 & 0 & 0 & -1 \end{bmatrix}^{\intercal} \tag{B.26}$$

$$\boldsymbol{R}_{\text{ABAO}} = \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix}^{\intercal} \tag{B.27}$$

An illustration of how $\boldsymbol{\mu}(t)$ changes as a function of $\boldsymbol{r}(t)$ is shown in Figure B.3. Note that $\boldsymbol{r}(t)$ is a sum of Dirac functions. In addition, we show the integral $\int_0^{t_f} \boldsymbol{r}(t)$, which is proportional to the maintenance cost.

B.3.3 Modelling the effect of inputs

The second contribution of this paper is the inclusion of the effect of inputs $\boldsymbol{u}(t)$, which influence the degradation rate of the system. Thus by changing $\boldsymbol{u}(t)$, we can actively steer the rate of degradation of the system. This is very useful, as it allows us to optimize the performance of the system by co-optimizing $\boldsymbol{u}(t)$ and $\boldsymbol{r}(t)$. We model this behavior by letting $\boldsymbol{\Lambda}_a$ be a function of the inputs and time, instead of a constant matrix like before. $\boldsymbol{\Lambda}_u$ remains constant, as we assume that unexpected failures cannot be influenced by changing the inputs. The differential model now is

$$\frac{d\boldsymbol{\mu}}{dt} = (\boldsymbol{\Lambda}_a(\boldsymbol{u}, t) + \boldsymbol{\Lambda}_u) \cdot \boldsymbol{\mu}(t) + \boldsymbol{R} \cdot \boldsymbol{r}(t). \tag{B.28}$$

Note that if $\boldsymbol{\Lambda}_a(\boldsymbol{u}, t)$ is a piece-wise constant function, meaning we can write it as

$$\boldsymbol{\Lambda}_a(\boldsymbol{u}, t) = \begin{cases} \boldsymbol{\Lambda}_{a,1} & \text{if} \quad t_0 < t < t_1 \\ \boldsymbol{\Lambda}_{a,2} & \text{if} \quad t_1 < t < t_2 \\ \dots & \\ \boldsymbol{\Lambda}_{a,k} & \text{if} \quad t_k < t < t_f \end{cases}, \tag{B.29}$$

the system becomes a Multiphase Markov process. However, we do not require this assumption, and are free to choose whatever form of $\boldsymbol{\Lambda}_a(\boldsymbol{u}, t)$ we need.

B.4 Formulating the optimization problem

In order to find the optimal combined production and maintenance strategy, we first formulate an optimization problem in terms of an objective function and con-

straints. [1]

$$\min_{\boldsymbol{u},\boldsymbol{r}} \quad \int_{t_0}^{t_f} \Big(- f(t,\boldsymbol{\mu},\boldsymbol{u})\Big)dt$$

$$+f_i(t,\boldsymbol{\mu},\boldsymbol{r}) + f_m(t,\boldsymbol{\mu},\boldsymbol{r}) \tag{B.30a}$$

$$\text{s.t.} \quad \frac{d\boldsymbol{\mu}}{dt} = \boldsymbol{\Lambda}(t,\boldsymbol{u})\cdot\boldsymbol{\mu}(t) + \boldsymbol{R}\cdot\boldsymbol{r}(t) \tag{B.30b}$$

$$\boldsymbol{r}(t) = \boldsymbol{S}^{\intercal}\boldsymbol{\mu}(t)\cdot\left(\sum_{i=1}^{k}\delta(t-t_i)\right) \tag{B.30c}$$

$$0 \leq \boldsymbol{\mu} \leq 1 \tag{B.30d}$$

$$\sum_{i\in\{A,B,C,D\}} \mu_i = 1 \tag{B.30e}$$

$$0 \leq \boldsymbol{r} \leq \infty \tag{B.30f}$$

$$\boldsymbol{u}_{\min} \leq \boldsymbol{u} \leq \boldsymbol{u}_{\max} \tag{B.30g}$$

In the above optimization problem, f denotes some economical objective which is to be maximized (typically profit or production), f_i denotes the inspection cost, which is typically proportional to the number of inspections k,

$$f_i(t,\boldsymbol{\mu},\boldsymbol{u}) \propto k, \tag{B.31}$$

and f_m denotes the maintenance cost, which is assumed to be proportional to the integral of $\boldsymbol{r}(t)$

$$f_m(t,\boldsymbol{\mu},\boldsymbol{u}) \propto \int_{t_0}^{t_f} \boldsymbol{r}(t)dt. \tag{B.32}$$

B.4.1 Problem re-formulation for numerical optimization

The optimization problem from Equations (B.30a)-(B.30g) can be solved in a multitude of ways. A common approach is to approximate the dynamic problem by a static non-linear programming (NLP) problem through the use of so-called direct methods, where direct multiple shooting and direct collocation are common approaches [28]. In this work, we use the direct collocation approach.

One issue with Problem (B.30) is in $\boldsymbol{r}(t)$. Since it is the summation of Dirac functions, it is unbounded as shown in Equation (B.30f). An alternative to using the formulation adopted in this paper is to formulate the problem as a mixed integer problem, as was done in [15]. Due to the nonlinear nature of the problem, we have to solve a mixed-integer non-linear programming (MINLP) problem, as done in e.g. [9, 63]. These MINLP problems are however known to be very difficult to solve in the general case, despite recent progress in algorithmic development.

[1] Note that the constraint in Equation (B.30e) is implied by Equation (B.30b), but we include it for completeness.

Instead, we approximate $\boldsymbol{r}(t)$ using Boxcar functions as

$$\boldsymbol{r}(t) \approx \tilde{\boldsymbol{r}}(t) \tag{B.33}$$

$$\tilde{\boldsymbol{r}}(t) = \sum_{i=1}^{k} \text{Boxcar}(t) \tag{B.34}$$

$$= \sum_{i=1}^{k} h_i \Big(\text{H}(t - t_i) - \text{H}(t - t_i - \epsilon_i) \Big) \tag{B.35}$$

where H is the Heaviside function, h_i is the height and ϵ_i is the width of each "box".

An illustration of this approximation is shown in the middle plot in Figure B.4. Note that the approximation for $\boldsymbol{\mu}$ is good if ϵ is sufficiently small, and that the approximation gives the same cumulative maintenance cost $\int \tilde{\boldsymbol{r}}(t)dt$. Furthermore, we observe that $\boldsymbol{\mu}$ is now continuous (although nonsmooth), which makes the optimization problem easier to solve.

Another numerical issue is posed by the inspection cost from Equation (B.31). In the original formulation with Dirac functions, one might be tempted to find k as

$$k = \int_{t_0}^{t_f} \Big(\sum_{i=1}^{k} \delta(t - t_i) \Big) dt \tag{B.36}$$

$$= \int_{t_0}^{t_f} \Big(\big(\boldsymbol{S}^\intercal \boldsymbol{\mu}(t) \big)^{-1} \cdot \boldsymbol{r}(t) \Big) dt, \tag{B.37}$$

but for this to work, we must assert that

$$\mu_D(t_i) = \mu_D^-(t_i) \tag{B.38}$$

to avoid division by zero. Such a condition might be difficult to impose numerically.

Instead, we propose to solve the problem by introducing the additional variable $\boldsymbol{y}$, which we use to formulate additional constraints:

$$0 \leq (1 - \boldsymbol{y}) \perp \tilde{\boldsymbol{r}} \geq 0 \tag{B.39}$$

$$0 \leq \boldsymbol{y} \leq 1 \tag{B.40}$$

Here, the $\perp$ operator indicates complementary, i.e. we require that at all times either $\tilde{\boldsymbol{r}}$ or $(1 - \boldsymbol{y})$ or both are zero. The inspection cost can then be written as

$$f_i(t, \boldsymbol{\mu}, \boldsymbol{u}) \propto \frac{\boldsymbol{y}}{\epsilon} \tag{B.41}$$

In order to minimize the cumulative inspection cost, $\boldsymbol{y}(t)$ will be a function that is either at its lower bound (zero) when no inspection is performed, or at its upper bound (one) when inspection is performed. By integrating we get

$$k = \int_{t_0}^{t_f} \Big(\sum_{i=1}^{k} \delta(t - t_i) \Big) dt \approx \int_{t_0}^{t_f} \frac{\boldsymbol{y}}{\epsilon} dt \tag{B.42}$$

B.5 Case study

As a case example, we consider a system inspired by subsea oil and gas production. Subsea technology is key to satisfying the energy demands of tomorrow, due to the intermittent nature of renewables and the continued need for petroleum products also in a green society. Reliability is a major issue for subsea installations, as maintenance interventions are very costly. Consequently, it is important to optimize both production from the subsea installation, as well as the maintenance interventions.

Assume that production from the subsea installation actively degrades critical components such as pumps, valves and heat exchangers. The transition rates can therefore be assumed to be proportional to the inputs $\boldsymbol{u}$ that we apply to the system. In our case, $\boldsymbol{u}(t)$ represents the production rate of oil and gas. A higher production rate will give more immediate profit, but also increased degradation.

Poor instrumentation and a lot of measurement uncertainty mean that a system may not be properly diagnosed to have failed without inspection. An example of this could be the failure of a single well going to a manifold with several other wells. The failure of the single well may be masked by the large variability in production of the other wells. Well tests (which can be thought of as inspections) are thus required to reveal the state of the single well.

B.5.1 Objective function

The objective is to maximize the average production from the well over the lifetime of the field, while simultaneously minimizing the inspection and maintenance costs. This economical objective can be written as

$$\min_{\boldsymbol{u}(t),\tilde{\boldsymbol{r}}(t)} \quad \int_0^{t_f} \Big(\frac{-f + f_m + f_i}{(1+d)^t} \Big) dt \tag{B.43a}$$

where

$$f(t) = \boldsymbol{c}_p^\intercal \cdot \boldsymbol{\mu}(t) \cdot \boldsymbol{u}(t) \tag{B.43b}$$
$$f_m(t) = \boldsymbol{c}_m^\intercal \cdot \tilde{\boldsymbol{r}}(t) \tag{B.43c}$$
$$f_i(t) = \boldsymbol{c}_i^\intercal \cdot \boldsymbol{y}(t). \tag{B.43d}$$

Here, $\boldsymbol{c}_p$ is the productivity in each state, $\boldsymbol{c}_m$ is the maintenance cost, $\boldsymbol{c}_i$ is the inspection cost. The entire economic objective is discounted by a factor d to reflect the decreasing value of future income streams compared to present income streams. Note that the objective is non-linear due to the bi-linear term in $f(t)$.

Table B.1: Parameters used for the optimization

Parameter	Description	Value
λ_u	Sudden failure transition rate	10^{-4}
λ_a	Base aging transition rate	10^{-2}
d	Discount rate	.001
$\boldsymbol{c}_p$	Productivities in each state	$\begin{bmatrix}28 & 21 & 14 & 2.8\end{bmatrix}^\intercal$
$\boldsymbol{c}_m$	Maintenance cost	300
$\boldsymbol{c}_i$	Inspection cost	30
t_f	Final time	200 weeks

B.5.2 Constraints

The objective function is optimized subject to the following constraints

$$
\begin{aligned}
s.t. \quad & \frac{d\boldsymbol{\mu}}{dl} = \boldsymbol{\Lambda}(t,\boldsymbol{u})\boldsymbol{\mu}(t) + \boldsymbol{R}\tilde{\boldsymbol{r}}(t) && \text{(B.43e)}\\
& \boldsymbol{\Lambda}(t,\boldsymbol{u}) = \boldsymbol{\Lambda}_{\boldsymbol{a}} \cdot \boldsymbol{u}(t) + \boldsymbol{\Lambda}_u && \text{(B.43f)}\\
& \boldsymbol{\mu}(0) = \boldsymbol{\mu}_0 && \text{(B.43g)}\\
& 0.1 \leq \boldsymbol{u} \leq 1.0 && \text{(B.43h)}\\
& 0 \leq \tilde{\boldsymbol{r}} \leq \frac{1}{\epsilon_{\min}} && \text{(B.43i)}\\
& \epsilon_{\min} \leq \boldsymbol{\epsilon} \leq \epsilon_{\max} && \text{(B.43j)}\\
& 0 \leq \boldsymbol{\mu} \leq 1 && \text{(B.43k)}\\
& 0 \leq (1-\boldsymbol{y}) \perp \tilde{\boldsymbol{r}} \geq 0 && \text{(B.43l)}\\
& 0 \leq \boldsymbol{y} \leq 1 && \text{(B.43m)}
\end{aligned}
$$

Note that the transition matrix $\boldsymbol{\Lambda}(t,\boldsymbol{u})$ is linear in $\boldsymbol{u}$. Since we require $\boldsymbol{u}(t)$ to be a piece-wise constant function, $\boldsymbol{\Lambda}(t,\boldsymbol{u})$ is also a piece-wise function. Consequently we are dealing with a Multiphase Markov process, as discussed in Section B.3.3.

The parameters for the problem are summarized in Table B.1.

B.5.3 NLP formulation

Multiple approaches exist to solve dynamic problems like Problem (B.43), but we choose to use orthogonal collocation on finite elements. The original dynamic problem is reformulated as an NLP, which can be solved using standard non-linear optimization algorithms. We will not go into details about how to discretize the problem, see e.g. [28] for a summary.

The resulting NLP is implemented in MATLAB using Casadi 3.4.1 [11]. The interior-point solver IPOPT 3.12.3 [185] is used to solve the optimization problem.

B.5.4 Solution strategy

Our problem is non-convex and local solvers such as IPOPT will consequently only find local solutions. In other words, we cannot guarantee global optimality of the solution. In order to ensure global optimality, global solvers such as BARON [149] have to be used. Global solvers come with some drawbacks, such as being computationally intractable for large problems.

To remedy this, we use a multi-start approach where the problem is repeatedly re-optimized with different initial guesses. After a certain number of optimizations (1000 in this case), the best local minimum is returned. Fewer than 1000 optimization runs could suffice, but as they are computationally cheap, we choose to run 1000 to ensure that a good solution is obtained.

B.5.5 Solution

The optimized production and maintenance strategy is shown in Figure B.5. As can be seen, the optimal strategy is to operate at maximum $\boldsymbol{u}$ all the time (a typical property of almost-linear optimization problems), while inspections / maintenance is performed at $t = 88$, 127 and 160 weeks. The objective function value is 3924 M$. Note that the first inspection is quite late. The reason for this is that the probabilities of being in the degraded states are initially low. As a result, performing inspections and maintenance at an early stage in the race is sub-optimal.

In an actual implementation, one would re-optimize the problem upon obtaining new information about the system state (such as after an inspection). This is known as model-predictive control (MPC) or rolling horizon optimization [28]. However, the optimization problem itself remains the same with only the initial conditions from Equation (B.43g) changing. We therefore chose to skip the closed-loop results in the interest of time and space.

B.6 Conclusions and future work

In this paper, we have introduced a method for simultaneous production- and maintenance optimization. The problem was motivated by a Markov-chain representation of a degrading production system, which would have been difficult to optimize using the traditional Monte Carlo based approach. We reformulated the problem as a algebraic-differential equation system, which we solved using a non-linear optimization approach. While not all problems can be solved like this, we showed how for the specific problem at hand, the problem could be cast into a form which could be solved using off-the-shelf solvers. The concept of input-dependent transition rates can easily be included in the framework. Some approximations were introduced to make the problem numerically tractable. The method was demonstrated on a case example inspired by subsea oil and gas production.

Possible future research directions include:

- Detailed comparison to Monte Carlo-based methods for optimization.

- Inclusion of more maintenance strategies by modification of $\boldsymbol{R}$ and optimization of the trade-off between the different maintenance strategies.
- A distributionally robust problem formulation to safeguard the solution against uncertainties in the transition rates.
- A multi-step approach to include the value of future information in the open-loop optimization problem.
- Looking at the case where maintenance is not instantaneous, i.e. when there is lag-time.
- Analysis of the losed-loop performance.
- A more complex case study with multiple simultaneously degrading units

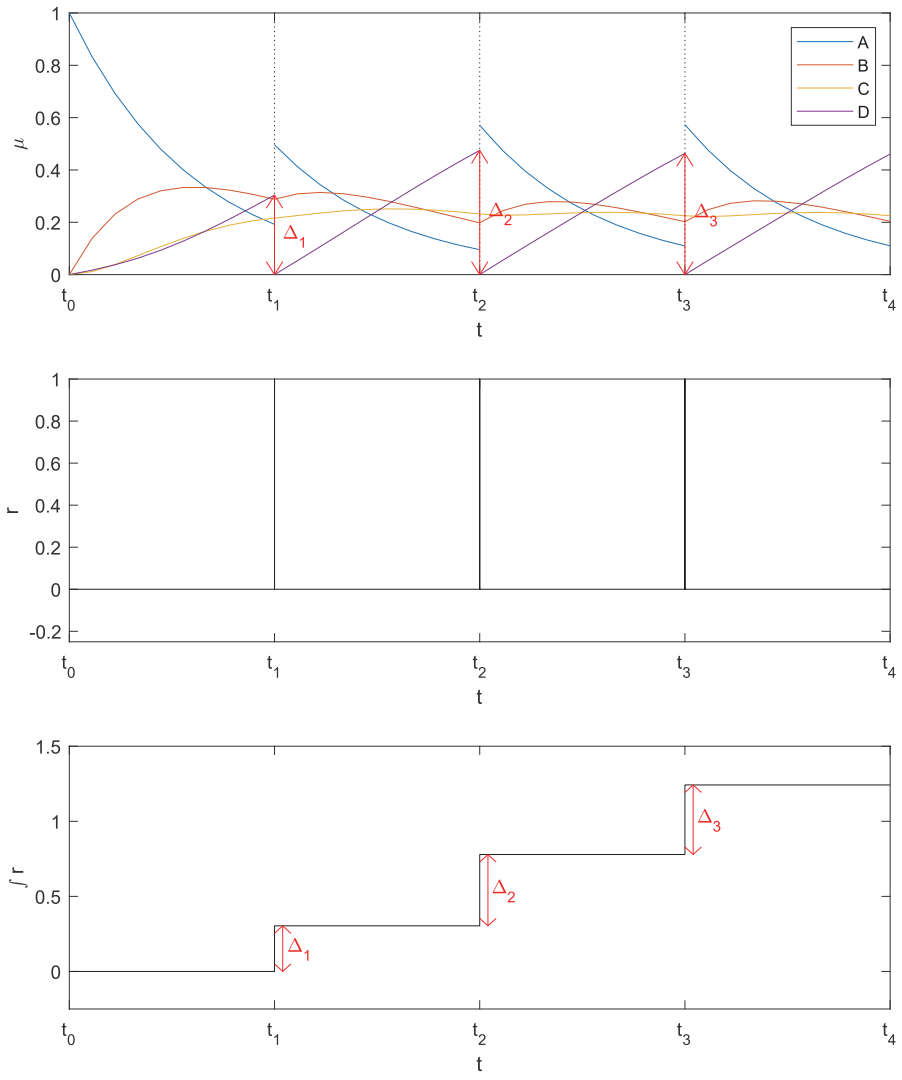

Figure B.3: Illustration of how $\boldsymbol{r}(t)$ influences $\boldsymbol{\mu}(t)$. The cumulative maintenance cost is proportional to the integral of $\boldsymbol{r}$, shown in the bottom plot, while the inspection cost is proportional to number of spikes (three, in this case)

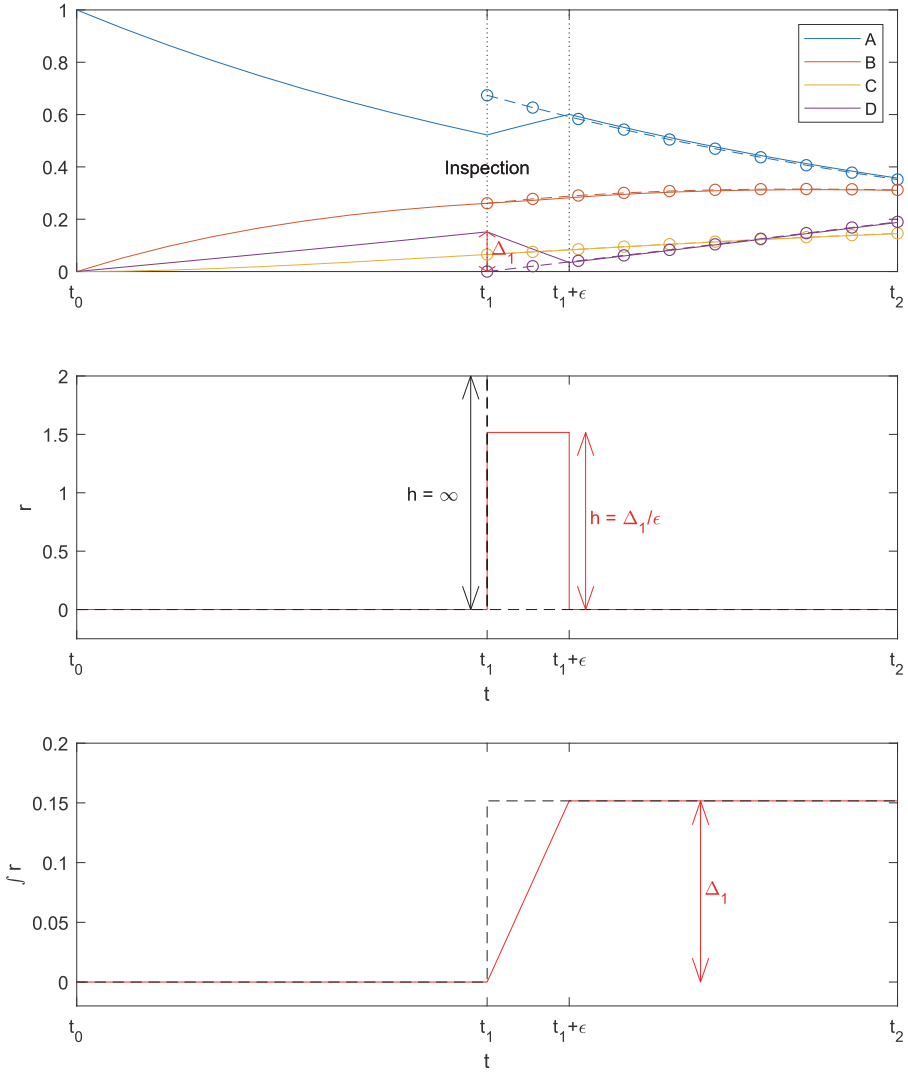

Figure B.4: Illustration of how $\boldsymbol{r}(t)$ (dashed line, circles) can be approximated by $\tilde{\boldsymbol{r}}(t)$ (solid line) to obtain a continuous $\boldsymbol{\mu}$.

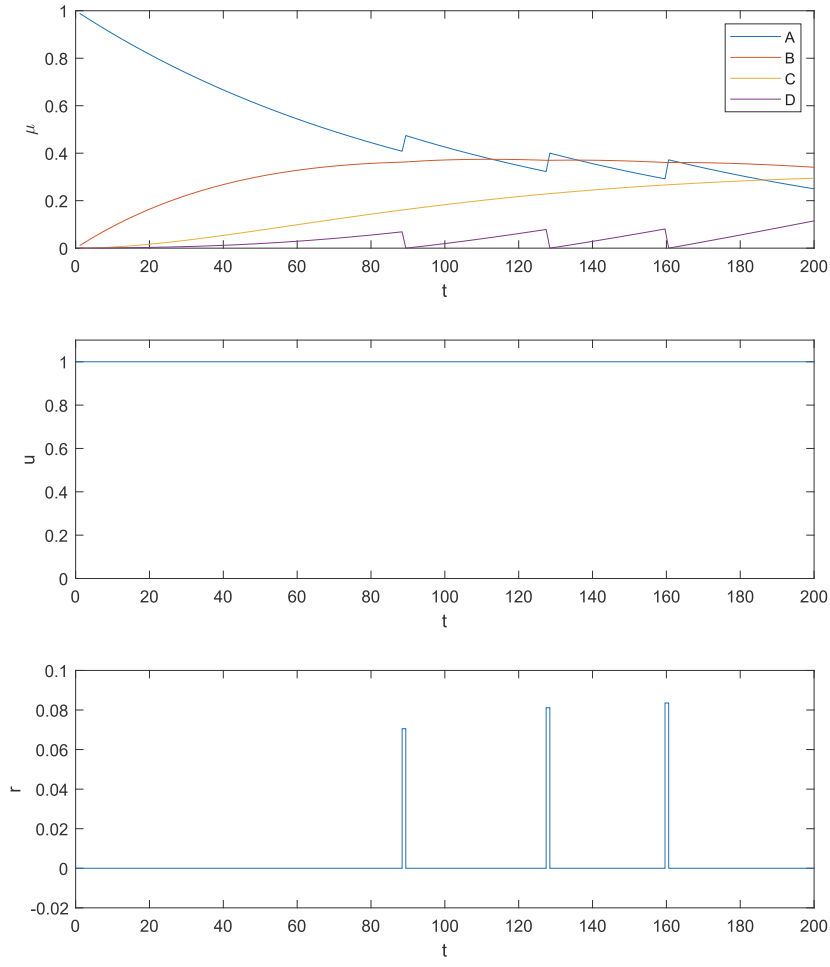

Figure B.5: Optimal solution of the case example

References

[1] ABB. Condition Monitoring and Production Optimization: Insight - Erosion Management System. https://library.e.abb.com/public/a19574f541925d408525771b004ba78b/TP002_Insight_erosion_management_system.pdf. Accessed March 22, 2016.

[2] ABB. Condition Monitoring and Production Optimization Insight - Erosion Management System. https://library.e.abb.com/public/a19574f541925d408525771b004ba78b/TP002_Insight_erosion_management_system.pdf, 2010.

[3] ABB. Summary: SUBPRO Systems Control Workshop on Estimation of unmeasured process variables, February 2016. Presentation slides from SUBPRO Systems Control Workshop at ABB Billingstad.

[4] L. C. P. Aguilera. Subsea Wet Gas Compressor Dynamics. Master's thesis, Norwegian University of Science and Technology, 2013.

[5] A. M. Aguirre and L. G. Papageorgiou. Medium-term optimization-based approach for the integration of production planning, scheduling and maintenance. *Computers & Chemical Engineering*, 116:191–211, 2018.

[6] W. Ahmad, O. Hasan, and S. Tahar. Formal dependability modeling and analysis: A survey. In *International Conference on Intelligent Computer Mathematics*, pages 132–147. Springer, 2016.

[7] S. Alaswad and Y. Xiang. A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157:54–63, 2017.

[8] J. Alexander. Sustaining operator-driven reliability. https://www.efficientplantmag.com/2014/04/sustaining-operator-driven-reliability/, 2014. Accessed: May 20, 2020.

[9] V. J. And and I. E. Grossmann. Cyclic scheduling of continuous parallel-process units with decaying performance. *AIChE Journal*, 44(7):1623–1636, 1998.

[10] J. Andersson. *A General-Purpose Software Framework for Dynamic Optimization*. PhD thesis, Arenberg Doctoral School, KU Leuven, Department of Electrical Engineering (ESAT/SCD) and Optimization in Engineering Center, Kasteelpark Arenberg 10, 3001-Heverlee, Belgium, October 2013.

[11] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, In Press, 2018.

[12] J. S. Andrews, H. Kjorholt, H. Joranson, et al. Production enhancement from sand management philosophy. a case study from statfjord and gullfaks. In *SPE European Formation Damage Conference*. Society of Petroleum Engineers, 2005.

[13] H. Ang, T. Markeset, and T. O. Bang-Steinsvik. Condition monitoring and identification of failure modes of subsea electrical equipments. In *Reliability and Maintainability Symposium (RAMS), 2013 Proceedings-Annual*, pages 1–7. IEEE, 2013.

[14] J. Antoni and R. Randall. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20(2):308–331, 2006.

[15] J. Ashayeri, A. Teelen, and W. Selenj. A production and maintenance planning model for the process industry. *International journal of production research*, 34(12):3311–3326, 1996.

[16] T. Austrheim. *Experimental Characterization of High-pressure Natural Gas Scrubbers*. PhD thesis, University of Bergen, 2006.

[17] C. J. Backi, J. T. Gravdahl, and S. Skogestad. Robust control of a two-state Greitzer compressor model by state-feedback linearization. In *Control Applications (CCA), 2016 IEEE Conference on*, pages 1226–1231. IEEE, 2016.

[18] N. Barton. Erosion in elbows in hydrocarbon production systems: Review document. *Health and Safety Executive*, 2003.

[19] B. Baumrucker and L. Biegler. Mpec strategies for optimization of a class of hybrid dynamic systems. *Journal of Process Control*, 19(8):1248–1256, 2009.

[20] E. Bechhoefer, A. Bernhard, and D. He. Use of paris law for prediction of component remaining life. In *Aerospace Conference, 2008 IEEE*, pages 1–9. IEEE, 2008.

[21] A. Bemporad, F. Borrelli, and M. Morari. Min-max control of constrained uncertain discrete-time linear systems. *IEEE Transactions on automatic control*, 48(9):1600–1606, 2003.

[22] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

[23] M. Benosman. Passive fault tolerant control. In A. Bartoszewicz, editor, *Robust Control, Theory and Applications*, pages 283–308. INTECH Open Access Publisher, 2011.

[24] M. Benosman and K.-Y. Lum. Application of passivity and cascade structure to robust control against loss of actuator effectiveness. *International Journal of Robust and Nonlinear Control*, 20(6):673–693, 2010.

[25] L. F. Bernardino, A. F. de Souza, A. R. Secchi, M. B. de Souza, and A. Barros. Integration of prognostics and control of an oil/co2 subsea separation system. *Processes*, 8(148), 2020.

[26] T. Biagetti and E. Sciubba. Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems. *Energy*, 29(12):2553–2572, 2004.

[27] L. T. Biegler. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers & chemical engineering*, 8(3-4):243–247, 1984.

[28] L. T. Biegler. *Nonlinear programming: concepts, algorithms, and applications to chemical processes*, volume 10. Siam, 2010.

[29] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.

[30] M. Blanke and J. Schröder. *Diagnosis and fault-tolerant control*, volume 2. Springer, 2006.

[31] W. R. Blischke and D. P. Murthy. *Reliability: modeling, prediction, and optimization*, volume 767. John Wiley & Sons, 2011.

[32] K. L. Butler. An expert system based framework for an incipient failure detection and predictive maintenance system. In *Intelligent Systems Applications to Power Systems, 1996. Proceedings, ISAP'96., International Conference on*, pages 321–326. IEEE, 1996.

[33] R. Cagienard, P. Grieder, E. C. Kerrigan, and M. Morari. Move blocking strategies in receding horizon control. *Journal of Process Control*, 17(6): 563–570, 2007.

[34] F. Caleyo, J. Gonzalez, and J. Hallen. A study on the reliability assessment methodology for pipelines with active corrosion defects. *International Journal of Pressure Vessels and Piping*, 79(1):77–86, 2002.

[35] J. Carlton. *Marine propellers and propulsion*. Butterworth-Heinemann, 2018.

[36] F. Cegla and J. Allin. Ultrasonic monitoring of pipeline wall thickness with autonomous, wireless sensor networks. In *Oil and Gas Pipelines: Integrity and Safety Handbook*, pages 571–578. Wiley, 2015.

[37] L. H. Chiang, R. D. Braatz, and E. L. Russell. *Fault Detection and Diagnosis in Industrial Systems*. Springer Science & Business Media, 2001.

[38] H. J. Chizeck and A. S. Willsky. Towards fault-tolerant optimal control. Technical report, DTIC Document, 1977.

[39] L. P. Chze. Importance of Condition and Performance Monitoring to Maximize Subsea Production System Availability. In *SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition*. Society of Petroleum Engineers, 2015.

[40] M. Daigle. Model-Based Prognostics, July 2014. Presentation slides.

[41] M. Diehl and J. Bjornberg. Robust dynamic programming for min-max model predictive control of constrained uncertain systems. *IEEE Transactions on Automatic Control*, 49(12):2253–2257, 2004.

[42] M. Diehl, H. G. Bock, and E. Kostina. An approximation technique for robust nonlinear optimization. *Mathematical Programming*, 107(1-2):213–230, 2006.

[43] DNV-GL. DNVGL-RP-O501 Recommended practice: Managing sand production and erosion, 2015.

[44] DNV-GL. Recommended practice RP-O501: Managing sand production and erosion. `https://rules.dnvgl.com/docs/pdf/dnvgl/RP/2015-08/DNVGL-RP-O501.pdf`, August 2015.

[45] P. Dranchuk, H. Abou-Kassem, et al. Calculation of z factors for natural gases using equations of state. *Journal of Canadian Petroleum Technology*, 14(03), 1975.

[46] A. S. Drud. CONOPT—a large-scale GRG code. *ORSA Journal on computing*, 6(2):207–216, 1994.

[47] J. B. Dugan, K. J. Sullivan, and D. Coppit. Developing a low-cost high-quality software tool for dynamic fault-tree analysis. *IEEE Transactions on reliability*, 49(1):49–59, 2000.

[48] J. Dupačová, G. Consigli, and S. W. Wallace. Scenarios for multistage stochastic programs. *Annals of operations research*, 100(1-4):25–53, 2000.

[49] D. Dwivedi. Process monitoring: Industry example II, February 2016. Presentation slides from SUBPRO Systems Control Workshop at ABB Billingstad.

[50] K. El-Tawil and A. A. Jaoude. Stochastic and nonlinear-based prognostic model. *Systems Science & Control Engineering: An Open Access Journal*, 1 (1):66–81, 2013.

[51] K. El-Tawil, A. Abou Jaoude, S. Kadry, H. Noura, and M. Ouladsine. Prognostic based on analytic laws applied to petrochemical pipelines. In *International Conference on Computer-aided Manufacturing and Design (CMD), November*, pages 1–2, 2010.

[52] K. Eriksson. Control system and condition monitoring for a subsea gas compressor pilot. In *Control 2010, UKACC International Conference on*, pages 1–6. IET, 2010.

[53] K. Eriksson, K. Antonakopoulos, et al. Subsea processing systems: Optimising the maintenance, maximising the production. In *Offshore Technology Conference-Asia*. Offshore Technology Conference, 2014.

[54] K. G. Eriksson. SUBPRO 3.3 Condition and prognostics based maintenance, January 2016. Presentation slides.

[55] K. G. Eriksson, J. Jäeschke, T. Das, and A. Verheyleweghen. Minutes of meeting – Aker, December 2015. Minutes of meeting.

[56] T. Escobet, V. Puig, and F. Nejjari. Health aware control and model-based prognosis. In *Control & Automation (MED), 2012 20th Mediterranean Conference on*, pages 691–696. IEEE, 2012.

[57] J. Eterno, J. Weiss, D. Looze, and A. Willsky. Design issues for fault tolerant-restructurable aircraft control. In *1985 24th IEEE Conference on Decision and Control*, number 24, pages 900–905, 1985.

[58] M.-C. Fitouhi and M. Nourelfath. Integrating noncyclical preventive maintenance scheduling and production planning for a single machine. *International Journal of Production Economics*, 136(2):344–351, 2012.

[59] FMC. Condition and Performance Monitoring. http://www.fmctechnologies.com/en/SubseaSystems/Technologies/AdvancingTechnologies/ProductionMonitoringOpt/CPM.aspx, . Accessed March 22, 2016.

[60] FMC. FlowManager. http://www.fmctechnologies.com/en/SubseaSystems/Technologies/AdvancingTechnologies/ProductionMonitoringOpt/Flowmanagement/FlowManager.aspx, . Accessed March 22, 2016.

[61] J. D. Friedemann, A. Varma, P. Bonissone, N. Iyer, et al. Subsea Condition Monitoring: A Path To Increased Availability And Increased Recovery. In *Intelligent Energy Conference and Exhibition*. Society of Petroleum Engineers, 2008.

[62] J. Ge, M. Roemer, G. Vachtsevanos, et al. An automated contingency management simulation environment for integrated health management and control. In *Aerospace Conference, 2004. Proceedings. 2004 IEEE*, volume 6, pages 3725–3732. IEEE, 2004.

[63] M. Georgiadis and L. Papageorgiou. Optimal energy and cleaning management in heat exchanger networks under fouling. *Chemical Engineering Research and Design*, 78(2):168–179, 2000.

[64] D. Ghiselin. Technology trends seek to reduce the cost of subsea well interventions. *Offshore*, 73(7), July 2013.

[65] G. Gola and B. H. Nystad. From measurement collection to remaining useful life estimation: defining a diagnostic-prognostic frame for optimal maintenance scheduling of choke valves undergoing erosion. In *Proceedings of the 2011 Annual Conference of the Prognostics and Health Management Society*, pages 26–29, 2011.

[66] J. P. P. Gomes, B. P. Leão, W. O. Vianna, R. K. Galvão, and T. Yoneyama. Failure Prognostics of a Hydraulic Pump Using Kalman Filter. In *Annual Conference of the Prognostics and Health Management Society*, 2012.

[67] J. Gondzio. Interior point methods 25 years later. *European Journal of Operational Research*, 218(3):587–601, 2012.

[68] M. J. Goodwin, O. M. Musa, and J. W. Steed. Problems Associated with Sour Gas in the Oilfield Industry and Their Solutions. *Energy & Fuels*, 29 (8):4667–4682, 2015.

[69] R. Gouws. Impact of frequency switching on the efficiency of a fully suspended active magnetic bearing system. 2012.

[70] J. T. Gravdahl and O. Egeland. *Compressor Surge and Rotating Stall: Modeling and Control*. Advances in Industrial Control. Springer Science & Business Media, 1999. ISBN 978-1-4471-0827-6. doi: 10.1007/978-1-4471-0827-6.

[71] J. Haslinger, M. Miettinen, and P. D. Panagiotopoulos. *Finite element method for hemivariational inequalities: theory, methods and applications*, volume 35. Springer Science & Business Media, 2013.

[72] A. Haugan, K. Opel, T. Berg, T. A. Reinen, T. Brukok, and E. Hennie. Condition monitoring of three-phase separators using gamma transmission and passive acoustic measurements. In *Proceedings of the 24th International Congress on Condition Monitoring and Diagnostic Engineering Management*, pages 20–29, 2011.

[73] K. Haugen, O. Kvernvold, A. Ronold, and R. Sandberg. Sand Erosion of Wear-Resistant Materials: Erosion in Choke Valves. *Wear*, 186:179–188, 1995.

[74] A. Heng, S. Zhang, A. C. Tan, and J. Mathew. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical systems and signal processing*, 23(3):724–739, 2009.

[75] J. Houmstuen. Condition Monitoring of Offshore O&G Separator–Cost-Benefit Evaluations and Presentation of Information. Master's thesis, Norwegian University of Science and Technology, 2010.

[76] A. N. S. Institute. *Manual for Determining the Remaining Strength of Corroded Pipelines: A Supplement to ASME B31 Code for Pressure Piping*. American Society of Mechanical Engineers, 1991.

[77] S. M. Iravani and I. Duenyas. Integrated maintenance and production control of a deteriorating production system. *Iie Transactions*, 34(5):423–435, 2002.

[78] M. G. Jacobsen. Process monitoring: Industry example I, February 2016. Presentation slides from SUBPRO Systems Control Workshop at ABB Billingstad.

[79] A. A. Jaoude, H. Noura, K. El-Tawil, S. Kadry, and M. Ouladsine. Lifetime Analytic Prognostic for Petrochemical Pipes Subject to Fatigue. In *8th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes-Safeprocess*, pages 29–31, 2012.

[80] A. K. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006.

[81] J. Jäschke, Y. Cao, and V. Kariwala. Self-Optimizing Control–A Survey. *Annual Reviews in Control*, 43:199–223, 2017.

[82] G. J. Kacprzynski, M. Gumina, M. J. Roemer, D. E. Caguiat, T. R. Galie, and J. J. McGroarty. A prognostic modeling approach for predicting recurring maintenance for shipboard propulsion systems. In *ASME Turbo Expo 2001: Power for Land, Sea, and Air*, pages V001T02A003–V001T02A003. American Society of Mechanical Engineers, 2001.

[83] S. Katipamula and M. R. Brambley. Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part I. *HVAC&R Research*, 11(1):3–25, 2005.

[84] M. C. O. Keizer, S. D. P. Flapper, and R. H. Teunter. Condition-based maintenance policies for systems with multiple dependent components: A review. *European Journal of Operational Research*, 261(2):405–420, 2017.

[85] M. Kermani and A. Morshed. Carbon Dioxide Corrosion in Oil and Gas Production - A Compendium. *Corrosion*, 59(8):659–683, 2003.

[86] F. J. Klever, G. Stewart, and C. A. van der Valk. New developments in burst strength predictions for locally corroded pipelines. Technical report, American Society of Mechanical Engineers, New York, NY (United States), 1995.

[87] R. Kothamasu, S. H. Huang, and W. H. VerDuin. System health monitoring and prognostics–a review of current paradigms and practices. In *Handbook of maintenance management and engineering*, pages 337–362. Springer, 2009.

[88] B. Kouvaritakis, J. A. Rossiter, and J. Schuurmans. Efficient robust predictive control. *IEEE Transactions on automatic control*, 45(8):1545–1549, 2000.

[89] D. Krishnamoorthy, B. Foss, and S. Skogestad. Real-Time Optimization under Uncertainty Applied to a Gas Lifted Well Network. *Processes*, 4(4): 52, 2016.

[90] Y. Langeron, A. Grall, and A. Barros. A modeling framework for deteriorating control system and predictive maintenance of actuators. *Reliability Engineering & System Safety*, 140:22–36, 2015.

[91] H. Langseth and L. Portinale. Bayesian networks in reliability. *Reliability Engineering & System Safety*, 92(1):92–108, 2007.

[92] J. H. Lee and Z. Yu. Worst-case formulations of model predictive control for systems with bounded parameters. *Automatica*, 33(5):763–781, 1997.

[93] F. Lees. *Lees' Loss prevention in the process industries: Hazard identification, assessment and control*. Butterworth-Heinemann, 2012.

[94] N. Lefebvre and A. Barros. Fault detection and isolation, prognosis, 2016. Lecture notes TMR4260 - PK8207.

[95] E. Lemmon, R. Jacobsen, S. Penoncello, and S. Beyerlein. Computer Programs for the Calculation of Thermodynamic Properties of Cryogens and other Fluids. In P. Kittel, editor, *Advances in Cryogenic Engineering*, volume 39 of *Advances in Cryogenic Engineering*, pages 1891–1897. Springer US, 1994.

[96] B. Liang and G. Duan. Robust h-infinity fault-tolerant control for uncertain descriptor systems by dynamical compensators. *Journal of Control Theory and Applications*, 2(3):288–292, 2004.

[97] F. Liao, J. L. Wang, and G.-H. Yang. Reliable robust flight tracking control: an lmi approach. *Control Systems Technology, IEEE Transactions on*, 10(1): 76–89, 2002.

[98] H. Liao and Z. Tian. A framework for predicting the remaining useful life of a single unit under time-varying operating conditions. *IIE Transactions*, 45 (9):964–980, 2013.

[99] F. Liu. Condition Monitoring and Prognosis for Subsea Multiphase Pump. Master's thesis, Norwegian University of Science and Technology, 2015.

[100] R. Lopez-Negrete and L. Biegler. On-line state and parameter estimation for nonlinear dynamic systems: An nlp framework, June 2012.

[101] P. J. Lucas, L. C. van der Gaag, and A. Abu-Hanna. Bayesian networks in biomedicine and health-care. *Artificial intelligence in medicine*, 30(3): 201–214, 2004.

[102] S. Lucia, T. Finkler, and S. Engell. Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty. *Journal of Process Control*, 23(9):1306–1319, 2013.

[103] S. Lucia, S. Subramanian, and S. Engell. Non-conservative robust nonlinear model predictive control via scenario decomposition. In *Control Applications (CCA), 2013 IEEE International Conference on*, pages 586–591. IEEE, 2013.

[104] D. Q. Mayne and W. Langson. Robustifying model predictive control of constrained linear systems. *Electronics Letters*, 37(23):1422–1423, 2001.

[105] D. Q. Mayne, M. M. Seron, and S. Raković. Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, 41(2): 219–224, 2005.

[106] R. McLin. Applying Fault-Tolerant, Hot-Swappable Control Architecture in Subsea Environments. In *Offshore Europe*. Society of Petroleum Engineers, 2011.

[107] G. K. McMillan. *Centrifugal and Axial Compressor Control*. Momentum Press, 1983.

[108] K. R. McNaught and A. Zagorecki. Using dynamic Bayesian networks for prognostic modelling to inform maintenance decision making. In *Industrial Engineering and Engineering Management, 2009. IEEM 2009. IEEE International Conference on*, pages 1155–1159. IEEE, 2009.

[109] A.-R. Merheb, H. Noura, and F. Bateman. Active fault tolerant control of quadrotor uav using sliding mode control. In *Unmanned Aircraft Systems (ICUAS), 2014 International Conference on*, pages 156–166. IEEE, 2014.

[110] S. Miyoshi, D. Zyngier, M. Souza Jr, A. Secchi, A. Teixeira, M. Campos, and E. Lima. Hybrid Monitoring of Offshore Compression Systems. In *IFAC Workshop on Automatic Control in Offshore Oil and Gas Production*, 2012.

[111] A. A. Mokhtar, M. C. Ismail, and M. Muhammad. Comparative study between degradation analysis and first order reliability method for assessing piping reliability for risk-based inspection. *studies*, 4:20, 2009.

[112] Mokveld. Subsea axial control valve. `http://www.mokveld.com/upload/product_document/Mokveld-Qualification_sheet_subsea_axial_control_EN.pdf`, 2016. Accessed April 9, 2016.

[113] R. C. Montgomery and A. K. Caglayan. Failure accommodation in digital flight control systems by bayesian decision theory. *Journal of Aircraft*, 13(2): 69–75, 1976.

[114] R. J. Montoya, W. Howell, W. Bundick, A. Ostroff, R. Hueschen, and C. M. Belcastro. *Restructurable controls*, volume 2277. National Aeronautics and Space Administration, Scientific and Technical Information Branch Washington, DC and Springfield, VA, 1983.

[115] M. Morari and J. H. Lee. Model predictive control: past, present and future. *Computers & Chemical Engineering*, 23(4):667–682, 1999.

[116] J. F. Müller. Subsea processing is making its way up the technology curve, but still several challenges ahead. *Offshore Magazine*, 2015.

[117] NASA. Automated contingency management (acm). http://ti.arc.nasa.gov/tech/dash/pcoe/automated-contingency-management/, 2015. Accessed January 14, 2016.

[118] NAXYS. Subsea Acoustic & Electric Condition Monitoring. http://www.naxys.no/page/281/Condition_Monitoring. Accessed March 22, 2016.

[119] D. N. Nguyen, L. Dieulle, and A. Grall. Remaining Useful Lifetime Prognosis of Controlled Systems: A Case of Stochastically Deteriorating Actuator. *Mathematical Problems in Engineering*, 2015, 2015.

[120] N. Niksefat and N. Sepehri. A qft fault-tolerant control for electrohydraulic positioning systems. *Control Systems Technology, IEEE Transactions on*, 10 (4):626–632, 2002.

[121] B. H. Nystad, G. Gola, J. E. Hulsund, and D. Roverso. Technical condition assessment and remaining useful life estimation of choke valves subject to erosion. In *Proceedings of the 2010 Annual Conference of the Prognostics and Health Management Society*, pages 11–13, 2010.

[122] B. H. Nystad, G. Gola, and J. E. Hulsund. Lifetime models for remaining useful life estimation with randomly distributed failure thresholds. In *First european conference of the prognostics and health management society*, volume 3, 2012.

[123] Y. Oka and T. Yoshida. Practical estimation of erosion damage caused by solid particle impact: part 2: mechanical properties of materials directly associated with erosion damage. *Wear*, 259(1):102–109, 2005.

[124] Y. I. Oka, K. Okamura, and T. Yoshida. Practical estimation of erosion damage caused by solid particle impact: Part 1: Effects of impact parameters on a predictive equation. *Wear*, 259(1):95–101, 2005.

[125] A. Ompusunggu, J.-M. Papy, and S. Vandenplas. Kalman-Filtering-Based Prognostics for Automatic Transmission Clutches. *Transactions on Mechatronics*, 21:419–430, 2015.

[126] OneSubsea. Multiphase compressor - the world's first and only true subsea wet gas compressor. https://www.onesubsea.com/products_and_services/processing_systems/multiphase_compressor.aspx, 2015. Accessed January 12, 2016.

[127] OneSubsea. Electric actuated process control valve. http://cameron.slb.com/onesubsea/technology-and-innovation/control-systems/process-control-valve, 2016. Accessed April 9, 2016.

[128] C. I. Ossai. Pipeline Corrosion Prediction And Reliability Analysis: A Systematic Approach With Monte Carlo Simulation And Degradation Models. *International Journal of Scientific & Technology Research*, 2(3), 2013.

[129] D. Pandey, M. S. Kulkarni, and P. Vrat. Joint consideration of production scheduling, maintenance and quality policies: a review and conceptual framework. *International Journal of Advanced Operations Management*, 2 (1-2):1–24, 2010.

[130] P. Paris and F. Erdogan. A critical analysis of crack propagation laws. *Journal of basic engineering*, 85(4):528–533, 1963.

[131] M. Parsi, K. Najmi, F. Najafifard, S. Hassani, B. S. McLaury, and S. A. Shirazi. A comprehensive review of solid particle erosion modeling for oil and gas wells and pipelines applications. *Journal of Natural Gas Science and Engineering*, 21:850–873, 2014.

[132] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[133] Y. Peng, M. Dong, and M. J. Zuo. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, 2010.

[134] E. B. Pereira, R. K. H. Galvão, and T. Yoneyama. Model Predictive Control using Prognosis and Health Monitoring of actuators. In *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, pages 237–243. IEEE, 2010.

[135] S. Peri, B. M. Rogers, et al. Computational Fluid Dynamics (CFD) Erosion Study for Chokes. In *International Petroleum Technology Conference*. International Petroleum Technology Conference, 2007.

[136] L. T. Popoola, A. S. Grema, G. K. Latinwo, B. Gutti, and A. S. Balogun. Corrosion problems during oil and gas production and its mitigation. *International Journal of Industrial Chemistry*, 4(1):1–15, 2013.

[137] D. E. Powell. Internal corrosion monitoring using coupons and er probes: A practical focus on the most commonly used, cost-effective monitoring techniques. In *Oil and Gas Pipelines: Integrity and Safety Handbook*, pages 495–514. Wiley, 2015.

[138] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[139] S. V. Rakovic, B. Kouvaritakis, M. Cannon, C. Panos, and R. Findeisen. Parameterized tube model predictive control. *IEEE Transactions on Automatic Control*, 57(11):2746–2761, 2012.

[140] R. M. Ramberg, H. Rognoe, O. Oekland, et al. Steps to the Subsea Factory. In *OTC Brasil*. Offshore Technology Conference, 2013.

[141] M. Rausand and A. Høyland. *System reliability theory: models, statistical methods, and applications*, volume 396. John Wiley & Sons, 2003.

[142] M. Rausand and A. Høyland. *System reliability theory: models, statistical methods, and applications*. John Wiley & Sons, 2 edition, 2004.

[143] K. A. Reay and J. D. Andrews. A fault tree analysis strategy using binary decision diagrams. *Reliability engineering & system safety*, 78(1):45–56, 2002.

[144] K. Reber. Eddy current testing in pipeline inspection. In *Oil and Gas Pipelines: Integrity and Safety Handbook*, pages 537–544. Wiley, 2015.

[145] L. C. K. Reuben and D. Mba. Diagnostics and prognostics using switching Kalman filters. *Structural Health Monitoring*, 13(3):296–306, 2014.

[146] D. Rosenthal. Connecting operations personnel to reliability efforts. *Chemical Engineering*, 122(5):59, 2015.

[147] I. Sadeghzadeh, A. Mehta, A. Chamseddine, and Y. Zhang. Active fault tolerant control of a quadrotor uav based on gainscheduled pid control. In *Electrical & Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on*, pages 1–4. IEEE, 2012.

[148] J. H. Sæther. Choke condition and performance monitoring. Master's thesis, Norwegian University of Science and Technology, 2010.

[149] N. V. Sahinidis. *BARON 17.8.9: Global Optimization of Mixed-Integer Nonlinear Programs,* User's Manual, 2017.

[150] M. M. Salama et al. An alternative to API 14E erosional velocity limits for sand laden fluids. *Transactions-American Society of Mechanical Engineers Journal of Energy Resources Technology*, 122(2):71–77, 2000.

[151] J. C. Salazar, P. Weber, F. Nejjari, D. Theilliol, and R. Sarrate. MPC Framework for System Reliability Optimization. In *Advanced and Intelligent Computations in Diagnosis and Control*, pages 161–177. Springer, 2016.

[152] H. Sanchez, T. Escobet, V. Puig, and P. F. Odgaard. Health-aware model predictive control of wind turbines using fatigue prognosis. *IFAC-PapersOnLine*, 48(21):1363–1368, 2015.

[153] G. H. Sandler. *System reliability engineering.* Prentice-Hall, 1963.

[154] S. Sankararaman, M. J. Daigle, and K. Goebel. Uncertainty Quantification in Remaining Useful Life Prediction Using First-Order Reliability Methods. *Reliability, IEEE Transactions on*, 63(2):603–619, 2014.

[155] W. S. Sarle. Neural Networks and Statistical Models. In *Proceedings of the Nineteenth Annual SAS Users Group International Conference*. Citeseer, 1994.

[156] A. Saxena, K. Goebel, D. Simon, and N. Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management*, pages 1–9. IEEE, 2008.

[157] M. Schwabacher. A Survey of Data-Driven Prognostics. In *Proceedings of the AIAA Infotech@ Aerospace Conference*, pages 1–5, 2005.

[158] M. Schwabacher and K. Goebel. A Survey of Artificial Intelligence for Prognostics. In *Aaai fall symposium*, pages 107–114, 2007.

[159] P. Scokaert and D. Mayne. Min-max feedback model predictive control for constrained linear systems. *IEEE Transactions on Automatic control*, 43(8): 1136–1142, 1998.

[160] J. S. Simon. *Feedback Stabilization of Compression Systems*. PhD thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, 77 Massachusetts Ave, Cambridge, MA 02139, USA, February 1993.

[161] S. Skogestad. Plantwide control: The search for the self-optimizing control structure. *Journal of Process Control*, 10(5):487 – 507, 2000.

[162] S. Skogestad. Control structure design for complete chemical plants. *Computers & Chemical Engineering*, 28(1-2):219–234, 2004.

[163] S. Skogestad. Control structure design for complete chemical plants. *Computers & Chemical Engineering*, 28(1-2):219–234, 2004.

[164] D. J. Smith. *Reliability, maintainability and risk: practical methods for engineers*. Butterworth-Heinemann, 2017.

[165] C. Soosaipillai, P. K. Roald, D. Alfstad, T. Aas, G. Smith, J.-Y. Bressand, et al. Condition performance monitoring for subsea: Experience and value documentation from the gjoa field. In *Offshore Technology Conference*. Offshore Technology Conference, 2013.

[166] Statoil. The subsea factory. http://www.statoil.com/en/technologyinnovation/fielddevelopment/aboutsubsea/Pages/Lengre%20dypere%20kaldere.aspx, 2014. Accessed January 2, 2016.

[167] Statoil. The first subsea gas compression plant in the world on line – a step change in subsea technology. http://www.statoil.com/en/NewsAndMedia/News/2015/Pages/17Sep_Aasgard_subsea.aspx, 2015. Accessed January 12, 2016.

[168] P. G. Stechlinski and P. I. Barton. Dependence of solutions of nonsmooth differential-algebraic equations on parameters. *Journal of Differential Equations*, 262(3):2254–2285, 2017.

[169] B. Sun, S. Zeng, R. Kang, and M. G. Pecht. Benefits and Challenges of System Prognostics. *Reliability, IEEE Transactions on*, 61(2):323–335, 2012.

[170] L. Tang, G. J. Kacprzynski, K. Goebel, A. Saxena, B. Saha, and G. Vachtsevanos. Prognostics-enhanced automated contingency management for advanced autonomous systems. In *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pages 1–9. IEEE, 2008.

[171] M. J. Thomas, S. Terpstra, et al. Corrosion monitoring in oil and gas production. In *CORROSION 2003*. NACE International, 2003.

[172] Z. Tian. An artificial neural network approach for remaining useful life prediction of equipments subject to condition monitoring. In *Reliability, Maintainability and Safety, 2009. ICRMS 2009. 8th International Conference on*, pages 143–148. IEEE, 2009.

[173] N. I. Uzelac. In-line inspection (ili)("intelligent pigging"). In *Oil and Gas Pipelines: Integrity and Safety Handbook*, pages 515–536. Wiley, 2015.

[174] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu. *Intelligent fault diagnosis and prognosis for engineering systems*, volume 13. Wiley, 2006.

[175] P. Vaidya. Prognosis-subsea oil and gas industry. In *Annual Conference of the Prognostics and Health Management Society*, 2010.

[176] P. Vaidya and M. Rausand. Remaining useful life, technical health, and life extension. In *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, volume 225, pages 219–231. SAGE Publications, 2011.

[177] A. Verheyleweghen and J. Jäschke. Self-Optimizing Control of a Two-Stage Refrigeration Cycle. *IFAC-PapersOnLine*, 49(7):845–850, 2016.

[178] A. Verheyleweghen and J. Jäschke. Health-Aware Operation of a Subsea Gas Compression System under Uncertainty. In *Foundations of Computer Aided Process Operations / Chemical Process Control 2017. FOCAPO/CPC*, 1 2017.

[179] A. Verheyleweghen and J. Jäschke. Framework for combined diagnostics, prognostics and optimal operation of a subsea gas compression system. In *International Federation of Automatic Control (IFAC) World Congress, 2017.* IFAC, 2017.

[180] A. Verheyleweghen and J. Jäschke. Oil production optimization of several wells subject to choke degradation. *IFAC-PapersOnLine*, 51(8):1–6, 2018.

[181] A. Verheyleweghen, J. Gjøby, and J. Jäschke. *Health-Aware Operation of a Subsea Compression System Subject to Degradation*, pages 1021–1026. 01 2018. ISBN 9780444642356. doi: 10.1016/B978-0-444-64235-6.50179-0.

[182] M. E. Villanueva, R. Quirynen, M. Diehl, B. Chachuat, and B. Houska. Robust mpc via min–max differential inequalities. *Automatica*, 77:311–321, 2017.

[183] G. Vinod, S. Bidhar, H. Kushwaha, A. K. Verma, and A. Srividya. A comprehensive framework for evaluation of piping reliability due to erosion–corrosion for risk-informed inservice inspection. *Reliability Engineering & System Safety*, 82(2):187–193, 2003.

[184] R. R. Viosca. Up the reliability ladder to world class. *Efficient Plant Magazine*, June 1998.

[185] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.

[186] Y.-F. Wang and P. J. Kootsookos. Modeling of low shaft speed bearing faults for condition monitoring. *Mechanical Systems and Signal Processing*, 12(3): 415–426, 1998.

[187] J. Wiebe, I. Cecílio, and R. Misener. Data-driven optimization of processes with degrading equipment. *Industrial & Engineering Chemistry Research*, 57 (50):17177–17191, 2018.

[188] A. Wolter and S. Helber. Simultaneous production and maintenance planning for a single capacitated resource facing both a dynamic demand and intensive wear and tear. *Central European Journal of Operations Research*, 24(3):489–513, 2016.

[189] O. Wu, G. Dalle Ave, I. Harjunkoski, A. Bouaswaig, S. M. Schneider, M. Roth, and L. Imsland. Optimal production and maintenance scheduling for a multiproduct batch plant considering degradation. *Computers & Chemical Engineering*, page 106734, 2020.

[190] R. Yelchuru and S. Skogestad. MIQP formulation for Controlled Variable Selection in Self Optimizing Control. *IFAC Proceedings Volumes*, 43(5):61–66, 2010.

[191] J. Zhang. Hybrid Prognostic Model for Residual Useful Life Estimation of Degraded Equipment. Master's thesis, Norwegian University of Science and Technology, 2015.

[192] J. Zhang and J. Lee. A review on prognostics and health monitoring of Li-ion battery. *Journal of Power Sources*, 196(15):6007–6014, 2011.

[193] X. Zhang and P. Pisu. An unscented kalman filter based approach for the health-monitoring and prognostics of a polymer electrolyte membrane fuel cell. In *Annual Conference of the Prognostics and Health Management Society*, volume 3, page 1, 2012.

[194] Y. Zhang and J. Jiang. Bibliographical review on reconfigurable fault-tolerant control systems. *Annual reviews in control*, 32(2):229–252, 2008.

[195] Y. Zhang, E. Reuterfors, B. McLaury, S. Shirazi, and E. Rybicki. Comparison of computed and measured particle velocities and erosion in water and air flows. *Wear*, 263(1):330–338, 2007.

[196] Z. Q. Zheng and M. Morari. Robust stability of constrained model predictive control. In *American Control Conference, 1993*, pages 379–383. IEEE, 1993.

ISBN 978-82-326-4724-8 (printed ver.)
ISBN 978-82-326-4725-5 (electronic ver.)
ISSN 1503-8181

NTNU

Norwegian University of
Science and Technology