FAULT DIAGNOSIS AND MAINTENANCE OPTIMIZATION FOR INTERCONNECTED SYSTEMS

WITH APPLICATIONS TO RAILWAY AND CLIMATE CONTROL SYSTEMS

Fault diagnosis and maintenance optimization for interconnected systems

WITH APPLICATIONS TO RAILWAY AND CLIMATE CONTROL SYSTEMS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K. C. A. M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 22 november 2016 om 15:00 uur

door

Kim Anna Johanna Verbert

ingenieur in de systeem- en regeltechniek
geboren te Roermond, Nederland.

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. ir. B. De Schutter en Prof. dr. R. Babuška

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. B. De Schutter	Technische Universiteit Delft, promotor
Prof. dr. R. Babuška	Technische Universiteit Delft, promotor

Onafhankelijke leden:

Prof. dr. ir. P.M. Herder	Technische Universiteit Delft
Prof. dr. ir. L. Pintelon	Katholieke Universiteit Leuven
Prof. T. Denoeux	Université de Technologie de Compiègne
Dr. ir. R.J.I. Basten	Technische Universiteit Eindhoven
Dr. ir. J.J.B. Bronswijk	ProRail
Prof. dr. ir. J. Hellendoorn	Technische Universiteit Delft, reservelid

This dissertation has been completed in partial fulfilment of the requirements of the Dutch Institute of Systems and Control (DISC) for graduate studies.

This research is part of the STW/ProRail project "Advanced monitoring of intelligent rail infrastructure (ADMIRE)". This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (project number 12235).

The research leading to these results has received additional funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 324432.

ISBN 978-94-6186-699-8

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

PREFACE

"Difficult roads often lead to beautiful destinations."

-Author unknown-

This thesis is the result of four years of Ph.D. research at the Delft Center for Systems and Control. During these four years, many people helped me to find and prettify my way to the final destination.

First of all my supervisors Bart and Robert, who had the difficult task of keeping me on a right track. "*All roads lead to Rome*" and with three stubborn personalities that does not always make life easy. Bart, I would like to thank you for never giving up, and for your extensive and precise feedback on all aspects of my work. Robert, I would like to thank you for enduring all discussions between Bart and me, for your valuable feedback on the contents of my thesis, and for your suggestions to improve my writing style.

Second, I would like to thank my roommate Renshi. From the beginning till the end, we shared our room as well as our experiences about the Ph.D. trajectory and our supervisor. Renshi, you are an amazing roommate, football player, and person in general. I will never forget your uniqueness, purity, and endless enthusiasm.

During these four years of Ph.D. research, I had the opportunity to work with many students, both in courses and during bachelor and master projects, for which I am grateful. In particular, I would like to thank Tim, Wouter, and Koen for the nice cooperation and their contributions to my Ph.D. research. Moreover, I enjoyed working with Jasper, Paulus, Hajo, Rijk, and Victor on the topic of my master thesis.

To all my (former) colleagues at the DCSC department: I enjoyed your company! Thanks for creating a nice working environment. In particular, I would like to thank:

Noortje, for your friendship and for sharing experience;

Elisabeth and *Laurens*, for the good cooperation in assisting courses;

Edwin, *Ivo*, and *Sachin* for the exciting discussions and football games;

Max and *Bart* for the good company, especially in the early morning;

Marieke, for the nice drives to Zoetermeer and for arranging good food;

Heleen, for your company and for arranging great social events.

Moreover, I would like to thank all members of the Explorail community. A special thanks to Rene and Hans from ProRail for their input and their help with arranging data sets for my research.

In the beginning of 2015, I had the pleasure of visiting the Honeywell Prague Lab. I would like to thank Ondrej, Jan, Henrik, and Karel for the good time and the valuable discussions on my research.

I am grateful to Paulien Herder, Lilianne Pintelon, Thierry Denoeux, Rob Basten, Hans Bronswijk, and Hans Hellendoorn for reading my thesis and participating in the defense committee.

Finally, I thank my mother Miep, sister Denise, and boyfriend Marco for their support and love during these four years of research. Marco, it was mostly you who had to endure my frustrations regarding "annoying" reviewers and supervisors, computers that did not do what I wanted them to do, and tailgating drivers on my daily commute. Thanks for listening and putting things in perspective.

Kim Verbert
Linschoten, October 2016

SUMMARY

For many systems, like medical devices, nuclear reactors, and transportation systems, an adequate maintenance optimization approach is essential to ensure high levels of reliability and safety while keeping operational costs low. A promising approach towards this goal is condition-based maintenance, which plans maintenance only when the system health indicates a need for it. To infer the system health, monitoring devices are installed to collect health-related data. The path from the monitoring data to a maintenance schedule then involves the following steps:

1. fault diagnosis, i.e. detecting abnormal system behavior and identifying its cause;
2. failure prognosis, i.e. predicting future system health;
3. maintenance optimization, i.e. determining the required type of maintenance as well as the optimal time to perform the maintenance task.

Although various methods have been published for all three tasks, discrepancies still exist between the assumptions made in the literature and the conditions encountered in practice. These discrepancies include, e.g., unrealistic assumptions regarding the absence of component interdependencies and regarding the (number of) available monitoring signals. This thesis contributes to resolving these discrepancies by proposing methods for fault diagnosis, failure prognosis, and maintenance optimization, particularly focusing on narrowing the gap between theory and practice. When treating the individual tasks, the dependencies between fault diagnosis, failure prognosis, and maintenance optimization are explicitly taken into account. Below we discuss our contributions to the individual processes in more detail.

EXPLOITING SYSTEM DEPENDENCE FOR FAULT DIAGNOSIS IN INTERCONNECTED SYSTEMS

It is often impracticable to monitor a large number of system variables. This results in a need for fault diagnosis methods that achieve adequate diagnostic performance given a limited number of monitoring signals. We propose such a fault diagnosis method for interconnected systems, i.e. systems consisting of multiple interdependent components. This approach is knowledge-based and uses the temporal, spatial, and spatio-temporal system dependencies as diagnostic features. These features can be derived from the existing monitoring signals; so no additional sensors are required. Moreover, taking spatial dependencies into account makes the approach robust with respect to environmental disturbances.

For a specific railway track circuit fault diagnosis case, we show that, without the temporal, spatial, and spatio-temporal features, it is not possible to identify the cause of a detected fault. Including the additional features allows potential fault causes to be identified. Moreover, we show that the features from the proposed approach are valuable beyond knowledge-based fault diagnosis. More specifically, a good understanding of the correlations between the proposed features and the system health is also helpful for the design of other fault diagnosis approaches. We demonstrate this through two examples: one using a recurrent neural network and another based on a switching Kalman filter model.

MULTIPLE-MODEL APPROACH TO SYSTEM-LEVEL HVAC FAULT DIAGNOSIS

Fault diagnosis in Heating, Ventilation, and Air Conditioning (HVAC) systems is challenging due to interdependencies among system components and the existence of multiple operating modes. Reliable and timely diagnosis can only be ensured when it is performed in all operating modes, and at the system level rather than at the level of the individual components. Nevertheless, almost no HVAC fault diagnosis methods that satisfy these requirements are described in literature. In this thesis, we propose a multiple-model approach to system-level HVAC fault diagnosis that takes component interdependencies and different operating modes into account. For each operating mode, a distinct Bayesian network (diagnostic model) is defined at the system level. The models are constructed based on knowledge regarding component interdependencies and conservation laws, and based on historical data through the use of virtual sensors. We show that component interdependencies provide useful features for fault diagnosis. Incorporating these features results in better diagnostic results, especially when only a few monitoring signals are available. Simulations demonstrate the performance of the proposed method: faults are diagnosed timely and correctly, provided that the faults result in observable behavior.

MULTIPLE-MODEL APPROACH TO SYSTEM RELIABILITY PREDICTION

In recent years, a wide range of prognostic methods have been developed, aiming at predicting future system reliability and remaining useful life with the highest possible accuracy. Almost all of these methods are based on a single degradation measure, and focus on systems with only one degradation and failure mode. In practice, however, multiple degradation measures are often available and needed to adequately predict future system degradation. Moreover, systems may suffer from various kinds of faults, all resulting in different degradation behaviors. To accommodate these properties, we propose a multivariate multiple-model approach to degradation forecasting. A distinct stochastic state-space model is defined for each type of degradation behavior, so as to minimize the modeling error and to manage the uncertainty inherent to degradation forecasting. In addition, we establish a link between failure prognosis and the subsequent maintenance optimization process.

It is concluded that in the presence of multiple degradation modes and provided they are correctly identified, the multiple-model approach outperforms a single-model approach with respect to prediction accuracy. Moreover, in the presence of multiple degradation and failure modes, overall predictions of the remaining useful life as generated by common prognostic approaches are not directly suited for maintenance decision making, as different kinds of system failures and maintenance activities are associated with different costs. In contrast, our approach yields conditional predictions of future system reliability, which much better suit the subsequent maintenance optimization process.

TIMELY MAINTENANCE PLANNING USING DIAGNOSTIC AND PROGNOSTIC INFORMATION

Last-minute maintenance planning is often undesirable, as it may cause downtime during operational hours, may require rescheduling of other activities, and does not allow to optimize the management of spare parts, material, and personnel. Furthermore, it may be beneficial to combine or spread various maintenance activities, which is not possible when maintenance needs are known just in time. In spite of the aforementioned drawbacks of last-minute planning, most existing methods on condition-based maintenance plan maintenance activities at the last minute. In addition, decisions are usually made based on predefined threshold values. As performance degradation varies widely among system components and operating conditions, such methods may generate poor maintenance strategies. To overcome some of the shortcomings of the existing methods, we propose an approach to timely maintenance planning in heterogeneous systems, i.e. systems consisting of multiple types of components. The heterogeneous nature of the systems considered in this thesis makes a top-down approach inappropriate. Hence, we use a bottom-up approach.

The first step is to decide for each system component in need of maintenance independently on the optimal time and type of maintenance. As information regarding the system health is available in real time, it is not obvious when to settle on a decision regarding the time and type of maintenance. On the one hand, it is desirable to determine the maintenance schedule sufficiently in advance. On the other hand, it is desirable to plan the maintenance actions based on accurate predictions of the system health. Since the prediction accuracy increases over time, a trade-off between accuracy and timeliness has to be made. To handle this trade-off, we formulate the problem as a Markov decision process and propose a sequential decision making approach to solve the resulting problem. At each diagnosis instant, we first determine, based on the currently available information, the optimal maintenance strategy (time and type of maintenance). Next, based on the expected costs (including risk) of this strategy, the expected time to failure, and the expected improvements in future predictions, it is decided whether to accept this maintenance strategy or to postpone the decision to a later time. If we accept the strategy, it is forwarded to the system-level optimization.

In the second step, we optimize the system-level maintenance planning for the total costs. By taking account of dependencies between the system components, costs can be reduced by spreading or combining maintenance. Hereby, a trade-off between economies of scale and loss of functionality is made. The resulting optimization problem is a nonlinear combinatorial one. In general, it can be solved using e.g. exhaustive search or generic algorithms.

The applicability of the method is demonstrated on a case study concerning maintenance planning in railway networks. It is analyzed how the different cost functions (e.g., costs of maintenance, downtime, and failure) influence the maintenance decision.

SAMENVATTING

Een adequate onderhoudsoptimalisatiestrategie is voor veel systemen, zoals medische apparatuur, kernreactoren en transportsystemen, van essentieel belang om een hoog niveau van betrouwbaarheid en veiligheid te waarborgen en tegelijkertijd de bedrijfskosten laag te houden. Conditie-gebaseerd onderhoud is een veelbelovende onderhoudsoptimalisatiestrategie om deze doelen te bereiken. Onder deze strategie wordt er enkel onderhoud gepleegd wanneer de gezondheidstoestand van het systeem daar aanleiding tot geeft. De gezondheidstoestand van het systeem wordt bepaald met behulp van sensoren die gezondheidsgerelateerde grootheden meten. Om op basis van deze meetgegevens een onderhoudsschema te genereren, moeten de volgende stappen doorlopen worden:

1. foutdiagnose, d.w.z. het detecteren van afwijkend gedrag en het identificeren van de oorzaak daarvan;
2. faalprognose, d.w.z. het voorspellen van de toekomstige systeemgezondheid;
3. onderhoudsoptimalisatie, d.w.z. het vaststellen van het benodigd type onderhoud en het optimale tijdstip van uitvoering.

Hoewel voor elk van deze taken vele methoden gepubliceerd zijn, bestaan er nog grote verschillen tussen de in de literatuur gemaakte aannames en de daadwerkelijk in de praktijk ondervonden condities. Dit betreft bijvoorbeeld onrealistische aannames over het aantal en de aard van de aanwezige meetsignalen en over de afwezigheid van onderlinge componentafhankelijkheden. Dit proefschrift speelt hierop in middels het ontwikkelen van methoden voor foutdiagnose, faalprognose en onderhoudsoptimalisatie die zich specifiek richten op het verkleinen van de kloof tussen theorie en praktijk. Tijdens het behandelen van de individuele taken houden we expliciet rekening met de afhankelijkheden tussen foutdiagnose, faalprognose en onderhoudsoptimalisatie. Hieronder bespreken we onze bijdragen aan de afzonderlijke processen in meer detail.

BENUTTEN VAN SYSTEEMAFHANKELIJKHEDEN VOOR FOUTDIAGNOSE IN ONDERLING VERBONDEN SYSTEMEN

Het meten van een groot aantal systeemvariabelen is vaak praktisch onhaalbaar. Hierdoor is er een vraag naar foutdiagnosemethoden die goede resultaten leveren op basis van slechts een beperkt aantal meetsignalen. Wij ontwikkelen een dergelijke foutdiagnosemethode voor onderling verbonden systemen. De voorgestelde aanpak is kennisgebaseerd en maakt gebruik van tijd- en ruimtegebaseerde afhankelijkheden in het systeem. Deze afhankelijkheden kunnen afgeleid worden uit reeds gemeten variabelen, waardoor er geen extra sensoren vereist zijn. Het gebruik van de ruimtelijke afhankelijkheden zorgt er verder voor dat de methode robuust is voor verstoringen vanuit de omgeving.

Voor een specifieke *case* betreffende foutdiagnose van spoorsecties laten we zien dat het zonder gebruik te maken van de tijdelijke, ruimtelijke en tijdruimtelijke kenmerken, niet mogelijk is om de oorzaak van een gedetecteerde fout te identificeren. Het toevoegen van de voorgestelde kenmerken stelt ons in staat om potentiële foutoorzaken te identificeren. Verder laten we zien dat de diagnostische kenmerken van de gepresenteerde methode waardevol zijn buiten kennisgebaseerde foutdiagnose. Een goed begrip van de correlaties tussen de voorgestelde diagnostische kenmerken en de systeemgezondheid komt ook van pas voor het ontwerp van andere diagnosestrategieën. We illustreren dit aan de hand van twee voorbeelden: een gebruikmakend van een recurrent neuraal netwerk en een andere gebaseerd op een schakelend Kalman filter model.

MEERDERE-MODELLEN METHODE VOOR SYSTEEMNIVEAU FOUTDIAGNOSE VAN KLIMAATBEHEERSINGSSYSTEMEN

Foutdiagnose van klimaatbeheersingssystemen is uitdagend vanwege onderlinge afhankelijkheden tussen componenten en de aanwezigheid van meerdere werkingsmodi. Betrouwbare en tijdige diagnose kan alleen worden gerealiseerd wanneer de diagnose wordt uitgevoerd in alle werkingsmodi en op systeemniveau in plaats van op het niveau van de individuele componenten. Desondanks zijn er in de literatuur nauwelijks foutdiagnosemethoden voor klimaatbeheersingssystemen beschreven die aan deze eisen voldoen. In dit proefschrift ontwikkelen we een meerdere-modellen aanpak voor systeemniveau-foutdiagnose van klimaatbeheersingssystemen waarin rekening wordt gehouden met onderlinge componentafhankelijkheden en verschillende werkingsmodi. Voor elke werkingsmodus wordt een apart Bayesiaans netwerk (diagnostisch model) gedefinieerd op systeemniveau. Deze modellen zijn naast kennis over componentafhankelijkheden en behoudswetten ook gebaseerd op historische meetgegevens van virtuele sensoren. We laten zien dat componentafhankelijkheden nuttige diagnostische kenmerken leveren, waardoor de kwaliteit van de foutdiagnose verbeterd kan worden, met name wanneer slechts weinig meetsignalen beschikbaar zijn. De prestaties van de gepresenteerde methode worden geëvalueerd aan de hand van simulaties. Hieruit blijkt dat fouten die resulteren in waarneembare gedragsveranderingen tijdig en correct gediagnosticeerd worden.

MEERDERE-MODELLEN BENADERING OM DE BETROUWBAARHEID VAN SYSTEMEN TE VOORSPELLEN

De laatste jaren is er een groot aantal prognostische methodes ontwikkeld. Het doel van deze methodes is het zo nauwkeurig mogelijk voorspellen van de systeembetrouwbaarheid en de resterende levensduur. Bijna al deze methodes maken gebruik van slecht één degradatiesignaal, en zijn gericht op systemen met één degradatie- en faalmodus. In de praktijk zijn er vaak meerdere degradatiesignalen beschikbaar, en vaak is het adequaat voorspellen van de systeemgezondheid alleen mogelijk door gebruik te maken van meerdere degradatiesignalen. Bovendien kunnen systemen last hebben van verschillende soorten fouten, elk met een specifiek degradatiegedrag. In het licht van deze observaties presenteren we een multivariabele meerdere-modellen methode voor degradatievoorspelling. Voor elk type degradatiegedrag definiëren we een apart stochastisch toestandsmodel, om zo de modelleringsfout te minimaliseren en de onzekerheid gerelateerd aan de degradatievoorspelling mee te nemen. Daarnaast bepalen we de relatie tussen de faalprognose en het daarop volgende onderhoudsplanningsproces.

We concluderen dat in de aanwezigheid van meerdere correct geïdentificeerde degradatiemodi, een meerdere-modellen aanpak beter presteert dan een aanpak gebaseerd op een enkelvoudig model op het gebied van nauwkeurigheid van de voorspelling. In de aanwezigheid van meerdere degradatie- en faalmodi zijn voorspellingen van de globale resterende levensduur van standaard diagnostische methoden niet direct geschikt voor onderhoudsbesluitvorming. Verschillende soorten storingen en onderhoudswerkzaamheden gaan immers gepaard met verschillende kosten. Onze aanpak levert voorwaardelijke voorspellingen van de systeembetrouwbaarheid die veel beter aansluiten op het latere onderhoudsoptimalisatieproces.

TIJDIG PLANNEN VAN ONDERHOUD OP BASIS VAN DIAGNOSTISCHE EN PROGNOSTISCHE INFORMATIE

Op het laatste moment plannen van onderhoud is vaak ongewenst omdat dit kan leiden tot een systeemonderbreking tijdens de operationele uren en het moeten herschikken van andere activiteiten. Daarnaast bemoeilijkt het de voorraadoptimalisatie van reserveonderdelen en materiaal, alsmede de inzet van personeel. Tenslotte kan het wenselijk zijn om meerdere onderhoudswerkzaamheden te combineren of juist te spreiden over de tijd, wat niet goed mogelijk is als onderhoudsbehoeften pas op het laatste moment bekend worden. Ondanks deze nadelen plannen de meeste bestaande conditie-gebaseerde onderhoudsoptimalisatiemethoden de onderhoudswerkzaamheden pas op het laatste moment. Beslissingen worden bovendien veelal gemaakt op basis van vooraf bepaalde drempelwaarden. Aangezien systeemdegradatie sterk varieert voor verschillende systeemcomponenten en bedrijfsomstandigheden, leveren zulke methoden vaak slecht presterende onderhoudsstrategieën op. Om een deel van de tekortkomingen van de bestaande onderhoudsoptimalisatiemethoden te verhelpen, ontwikkelen we een aanpak voor tijdige onderhoudsplanning in heterogene systemen, d.w.z. systemen bestaande uit verschillende soorten componenten. Hiervoor maken we gebruik van een *bottom-up* aanpak, omdat het heterogene karakter van de in dit proefschrift beschouwde systemen een *top-down* benadering ongeschikt maakt.

In de eerste stap bepalen we voor elke systeemcomponent afzonderlijk het benodigde type onderhoud en het optimale tijdstip van uitvoering. Doordat informatie over de systeemgezondheid in *real-time* beschikbaar komt, is het niet duidelijk wanneer de beslissing over het tijdstip en de aard van het onderhoud vastgelegd dient te worden. Enerzijds is het wenselijk om het onderhoudsschema ruim op tijd vast te leggen. Anderzijds is het wenselijk om onderhoud te plannen op basis van nauwkeurige voorspellingen van de systeemgezondheid. Aangezien de nauwkeurigheid van de voorspellingen toeneemt in de tijd, moet er een compromis gesloten worden tussen betrouwbaarheid en tijdigheid. Om een goed compromis te sluiten, formuleren we het probleem als een Markov beslissingsproces en stellen we een sequentiële besluitvormingsaanpak voor om het verkregen probleem op te lossen. Op elk diagnosetijdstip bepalen we eerst, op basis van de dan beschikbare informatie, de optimale onderhoudsstrategie (soort onderhoud en tijdstip van uitvoering). Vervolgens besluiten we op basis van de verwachte kosten (inclusief risicokosten) van deze strategie, de verwachte resterende levensduur en verwachte verbeteringen in toekomstige voorspellingen, of we deze onderhoudsstrategie accepteren of dat we de beslissing uitstellen naar een later tijdstip. Als we de strategie accepteren, wordt deze doorgestuurd naar het optimalisatieproces op systeemniveau.

In de tweede stap optimaliseren we het onderhoud op systeemniveau om de totale kosten te minimaliseren. Door rekening te houden met afhankelijkheden tussen systeemcomponenten kunnen de kosten worden verminderd door het spreiden of het combineren van onderhoud. Hierbij moet een compromis worden gesloten tussen schaalvoordelen en functionaliteitsverlies. Dit resulteert in een niet-linear combinatorisch optimalisatieprobleem, dat bijv. kan worden opgelost door middel van uitputtende zoek- of genetische algoritmes.

De toepasbaarheid van de methode wordt gedemonstreerd aan de hand van een *case* over onderhoudsplanning in een spoorwegennet. We analyseren hoe de verschillende kostenfuncties (bijv. kosten van onderhoud, systeemonderbreking en falen) de onderhoudsbeslissing beïnvloeden.

INHOUDSOPGAVE

1

INTRODUCTION

"If we knew what we were doing it would not be called research, would it?"

-Albert Einstein-

This chapter starts with an introduction to condition-based maintenance and its sub-processes: fault diagnosis, failure prognosis, and maintenance optimization. Next, we briefly introduce interconnected systems. Afterwards, we present open challenges in the field of condition-based maintenance. These challenges form the motivation for the research presented in this thesis. Subsequently, we list our specific contributions and provide a roadmap for the rest of this thesis.

1.1. CONDITION-BASED MAINTENANCE

Condition-based maintenance (Yam et al., 2001; Jardine et al., 2006; Ahmad and Kamaruddin, 2012) is an increasingly popular maintenance optimization approach in which maintenance activities are planned based on information collected through real-time condition monitoring. Its promise is twofold (Gebraeel et al., 2005): first, unnecessary maintenance can be eliminated, reducing maintenance costs; second, failures can be avoided, improving safety and reducing unscheduled downtime. As the literature on condition-based maintenance uses different nomenclatures, below we first specify the nomenclature used in this thesis.

First, we make a distinction between a fault and a failure. Following Isermann (2011), we define a *fault* as an unpermitted deviation in the system operation that does not hinder the execution of the system tasks, whereas a *failure* indicates that at least one system task can no longer be executed properly. Following these definitions, condition-based maintenance comprises (see Figure 1.1):

1. data pre-processing;

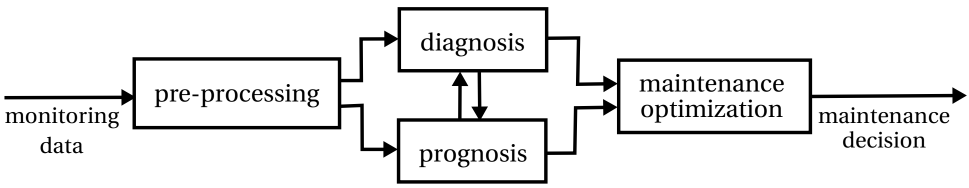
2. fault diagnosis (diagnosis in short);

3. failure prognosis (prognosis in short);
4. maintenance optimization.

Data pre-processing concerns the processing of the collected data to make them suitable for further analysis. It includes, e.g., data cleaning and synchronization of multiple data sets. *Fault diagnosis* concerns the detection of faulty behavior and the determination of its cause(s). *Failure prognosis* refers to the prediction of future degradation behavior and the estimation of the associated failure time. Finally, *maintenance optimization* comprises the determination of the optimal time and type of maintenance based on the diagnostic and prognostic results. Note that, in contrast to Jardine et al. (2006), we do not regard diagnosis as posterior event analysis, nor do we regard prognosis as superior to diagnosis. Diagnosis and prognosis concern different tasks, and both are needed for an adequate implementation of condition-based maintenance.

1

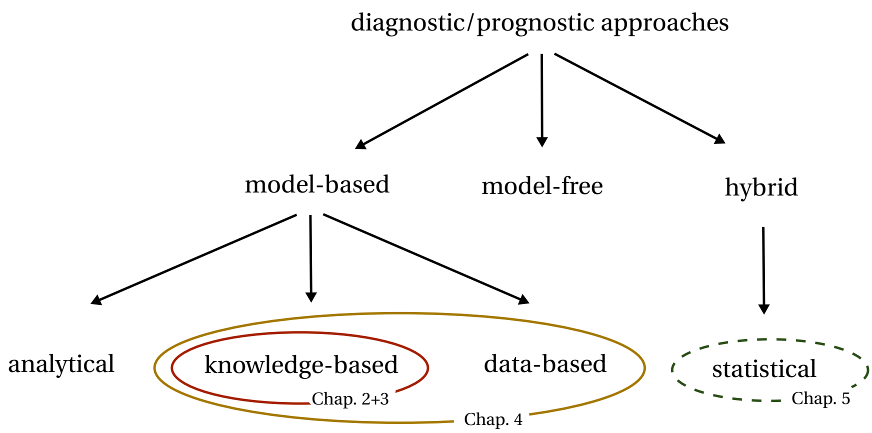
The different processes are interconnected as follows (see Figure 1.1): both the diagnostic and prognostic result serve as an input for the maintenance optimization process. Moreover, we assume that dependencies exist between the diagnosis and the prognosis process. For example, temporal degradation behavior may be influenced, among other things, by the type of fault present. In this thesis, we explicitly account for these dependencies.



Figuur 1.1: The condition-based maintenance process.

1.1.1. FAULT DIAGNOSIS

Fault diagnosis is shifting from a manual process, in which human operators check and evaluate the condition of a system on a regular basis, towards an automated process, in which advanced algorithms are used to detect and identify system faults based on real-time collected system data (Venkatasubramanian et al., 2003). With respect to the algorithms used, a distinction is made between model-based, model-free, and hybrid approaches (see Figure 1.2). *Model-based* approaches (Isermann, 2005; Fekih et al., 2007; Nan et al., 2008; Kukal et al., 2009; Hwang et al., 2010; Chen and Patton, 2012) rely on a qualitative or quantitative description of the relations between the monitoring data and system health, while *model-free* approaches (Oukhellou et al., 2010; Cherfi et al., 2012) use historical data and techniques from machine learning or pattern recognition. Finally, *hybrid* approaches (Chen et al., 2008; Narasimhan et al., 2010; Sandidzadeh and Dehghani, 2013) use a combination of the aforementioned approaches. The difficulty with model-free approaches, and to a lesser extent also with hybrid approaches, is that a representative amount of labeled historical data is required, which is in general difficult to obtain (Cherfi et al., 2012). Furthermore, due to preventive maintenance activities,



Figuur 1.2: Classification of diagnostic and prognostic approaches. The diagnostic (solid) and prognostic (dashed) approaches considered in this thesis are encircled.

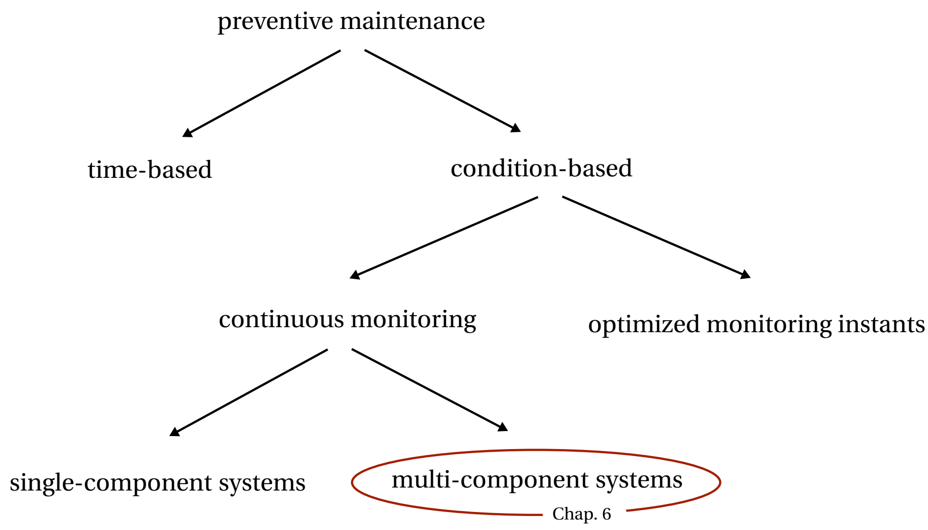
usually only few data samples are available that are characteristic of the natural degradation behavior. For these reasons, this thesis focuses on model-based methods for fault diagnosis.

Model-based approaches can be further divided according to the way the model is created (Frank et al., 2000) (see Figure 1.2). *Analytical* approaches (Isermann, 2005; Hwang et al., 2010; Chen and Patton, 2012) are based on a quantitative model derived from first principles, *knowledge-based* approaches (Nan et al., 2008; Kukal et al., 2009) use expert knowledge to define a qualitative model of the system, while *data-based* approaches (Fekih et al., 2007) use historical data to learn this model. In this thesis, we focus on knowledge-based approaches to fault diagnosis. Whenever possible, we augment our system knowledge with available historical data (see Figure 1.2). We made this choice because for the applications considered in this thesis, namely railway networks and climate control systems, only a limited amount of historical data is available. Moreover, detailed system insight is difficult to obtain because of system complexity and uncertain environmental disturbances.

1.1.2. FAILURE PROGNOSIS

Failure prognosis is a key factor in the success of condition-based maintenance (Engel et al., 2000). Indeed, if we know how the system degrades over time, we can compute the time of a system failure, and accordingly plan the maintenance at a convenient time before the system fails. Point predictions of the remaining life alone, however, do not provide sufficient information to make an informed maintenance decision. Without corresponding measures of uncertainty, predictions of future system health have little practical value (Engel et al., 2000). So, uncertainty handling is an important element of the prognosis task.

Like fault diagnosis, failure prognosis has become an automated process, and the various prognostic approaches can be roughly divided into the three classes indicated in



Figuur 1.3: Classification of maintenance optimization approaches.

Figure 1.2, i.e. model-based, model-free, and hybrid approaches. In this thesis, for prognostic purposes, we consider a subgroup of the hybrid approaches, namely statistical methods. *Statistical methods* rely on available historical data and statistical models (Si et al., 2011), and are considered because of their natural ability to handle the uncertainty inherent to degradation forecasting. Moreover, statistical approaches do not require detailed knowledge of degradation behavior (like model-based approaches) nor the large amount of historical data required by model-free methods.

Note that the choice to use a different approach for the diagnosis and the prognosis task is motivated by the differences in the available diagnostic and prognostic knowledge. In general, quite some (qualitative) diagnostic knowledge is available, whereas prognostic knowledge is often present to a lesser extent and with more uncertainty. Therefore, for the prognosis task we rely more on the (limited amount of) available historical data.

1.1.3. MAINTENANCE OPTIMIZATION

Preventive maintenance optimization approaches, which aim to perform maintenance before a system failure occurs, are gaining ground of the *corrective maintenance approach*, in which maintenance is performed after a system breakdown. Two popular preventive maintenance optimization approaches are time-based maintenance and condition-based maintenance (Jardine et al., 2006; Tinga, 2010; Ahmad and Kamaruddin, 2012) (see Figure 1.3). *Time-based approaches* aim to determine optimal time-based or usage-based intervals to perform maintenance. *Condition-based approaches* use actual health-related system data to determine the optimal time of maintenance. Currently, condition-based maintenance is regarded as the most promising approach, mainly for the following two reasons:

1. Most equipment failures are preceded by certain signs (Yam et al., 2001; Ahmad and Kamaruddin, 2012), indicating the high potential of condition-based maintenance;

2. Systems operate in different environments and under different circumstances. This may lead to highly varying lifetimes and consequently limit the effectiveness of time-based approaches (Mann et al., 1995).

Therefore, this thesis focuses on condition-based maintenance optimization. So, given uncertain estimates of the system health state (diagnostic result) and uncertain predictions of the future system health (prognostic result), we aim to determine an optimal maintenance strategy (time and type of maintenance).

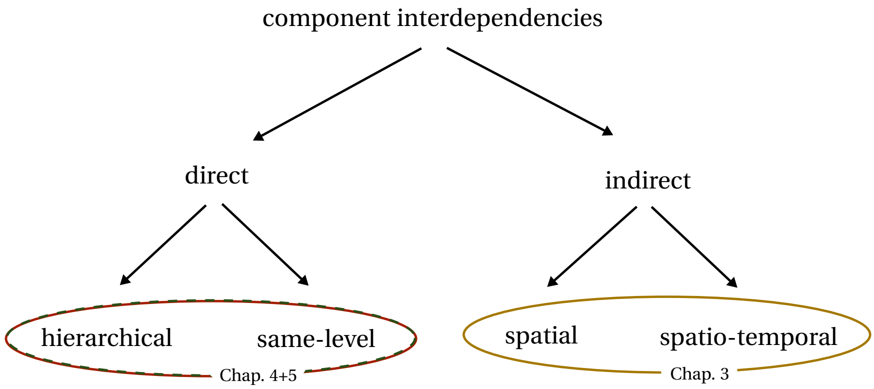
Condition-based maintenance optimization approaches can be classified according to the frequency of monitoring (see Figure 1.3). When the costs of monitoring are high (e.g. when special equipment or personnel has to be sent out for each monitoring round), monitoring is usually carried out at fixed or optimized time instants. In the latter case, the optimization of the monitoring time instants is part of the maintenance optimization process. When the costs of monitoring are low (e.g. when sensors are installed at the system), continuous monitoring is considered. In this thesis we focus on maintenance optimization based on continuous monitoring. The optimization approaches can be further divided into methods that focus on single-component systems and methods tailored to multi-component systems (see Figure 1.3). As practically all systems consist of multiple components, this thesis focuses on multi-component systems.

1.2. INTERCONNECTED SYSTEMS

We define an interconnected system as a multi-component system, the components of which are, directly or indirectly, interconnected with each other. Both for fault diagnosis, failure prognosis, and maintenance optimization, it is necessary and beneficial to account for the interconnections among system components. In this section, we discuss the most important properties of interconnected systems with respect to fault diagnosis and failure prognosis and with respect to maintenance optimization.

1.2.1. FAULT DIAGNOSIS AND FAILURE PROGNOSIS FOR INTERCONNECTED SYSTEMS

The key property of an interconnected system with respect to fault diagnosis and failure prognosis is that the behavior of a system component (and so the associated monitoring signals) may be influenced by or correlated with the behavior of other system components. As a consequence, malfunctioning of a system component may e.g. be caused by a fault in another system component, be masked through compensation by other system components, or be inherently accompanied by malfunctioning of other system components. These interdependencies may have the (negative) consequence that diagnosis and prognosis cannot be performed for each system component independently. On the other side, these dependencies may also have the (positive) effect that additional information about a component's health is contained in the behavior of other system components. The exact implication of component interconnections depends on



Figuur 1.4: Classification of component interdependencies for the purpose of fault diagnosis and failure prognosis.

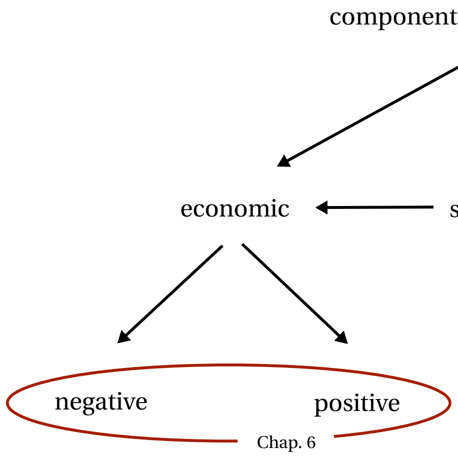
the kind of dependency relation between the respective components. In this thesis, for fault diagnosis and failure prognosis, we make a distinction between the following types of dependency relations (see Figure 1.4):

1. direct component interdependencies, which can be further divided into:
   (a) hierarchical dependencies, implying that the proper functioning of a component relies on the proper functioning of other components;
   (b) same-level dependencies, implying that the effect of malfunctioning of a component depends on the functioning of other components;
2. indirect component interdependencies (i.e. dependencies introduced by external influences), which can be further divided into:
   (a) spatial dependencies, implying that the behavior of some system components may become correlated as a consequence of an external influence simultaneously acting on multiple system components;
   (b) spatio-temporal dependencies, implying that the behavior of some components may become correlated over time as a consequence of an object moving through the system.

Note that next to the spatial and spatio-temporal dependencies, there also exist purely temporal dependencies, which refer to a component's behavior over time. Temporal dependencies are not specific to interconnected systems, but equally included in this thesis.

1.2.2. MAINTENANCE OPTIMIZATION FOR INTERCONNECTED SYSTEMS

Key properties of interconnected systems with respect to maintenance optimization are that different components may require similar maintenance activities at approximately the same time, and that maintenance of one component may cause that not only that



Figuur 1.5: Classification of component interdependencies for the purpose of maintenance optimization.

particular component is out of order, but possibly also other components. As a consequence of these interdependencies, it may be beneficial to combine or spread maintenance activities on the different system components over time. Whether it is optimal to combine maintenance activities or to spread them over time, depends on the type of dependency relation between the respective components.

In the literature on maintenance optimization of multi-component systems, a distinction is made between the following dependencies (Dekker et al., 1997) (see Figure 1.5):

1. economic dependencies, implying that maintenance costs decrease (positive economic dependencies) or increase (negative economic dependencies) when components are jointly maintained instead of separately;

2. structural dependencies, implying that components structurally form a part, so that maintenance of a component implies that other components are out of use as well;

3. stochastic dependencies, implying that the state of a component influences the lifetime distribution of other system components, or if there are causes outside the system that correlate the lifetimes of components (common-cause failures).

For interconnected systems, the economic dependencies depend on the structural dependencies (indicated in Figure 1.5 by an arrow from 'structural' to 'economic'). Indeed, it depends on the structural dependencies whether spreading or combining is more efficient with respect to system downtime. Since system downtime can be expressed in terms of (virtual) monetary units, structural dependencies lead to (positive or negative) economic dependencies. Moreover, note that we plan the maintenance based on the diagnostic and prognostic result. When stochastic dependencies exist, these are reflected in the diagnostic and the prognostic result, and so are accounted for in the maintenance optimization.

1.3. OPEN CHALLENGES & CONTRIBUTIONS

From our literature review on condition-based maintenance (see e.g. Sections 2.1, 3.1, 4.1, 5.1, 5.2, 6.1, 6.2), we conclude that there are plenty of research opportunities left in the field of condition-based maintenance. Many of these opportunities relate to discrepancies that currently exist between the assumptions made in the literature and the conditions encountered in practice. In the light of this observation, the main goal of this thesis is formulated as:

The development of diagnostic, prognostic, and maintenance optimization approaches that connect well with the daily practice of condition-based maintenance.

In this section, we first outline the most important discrepancies and the associated research challenges. Afterwards, we briefly describe how this thesis contributes to these open challenges.

1.3.1. OPEN CHALLENGES

Below, we briefly define the open challenges in the fields of fault diagnosis, failure prognosis, maintenance optimization, and the overall condition-based maintenance process. More details and background information can be found in later chapters (see Section 1.3.2 for an overview of which challenge is handled in which chapter).

In the field of fault diagnosis, we identified the following challenges:

Challenge D_1: Many existing fault diagnosis methods rely on the availability of a large number of monitored variables. Monitoring a large number of variables is however often impracticable. This means that there is a need for diagnosis approaches that achieve adequate diagnostic performance given a limited number of monitoring signals.

Challenge D_2: Fault diagnosis is often considered for individual system components. Ignoring dependencies among system components may however lead to incorrect or incomplete diagnostic results. In addition, by neglecting component interdependencies one discards valuable information. Therefore, the development of fault diagnosis methods that account for dependencies among system components is of importance.

Challenge D_3: Although it is generally known that fault diagnosis is subject to uncertainty, a precise characterization of the uncertainty present is often lacking. Moreover, the uncertainty is often not (completely) accounted for in the diagnostic reasoning. To enhance diagnostic quality as well as the interpretation and usability of the diagnostic result, a better understanding of the underlying uncertainty is required.

In the field of failure prognosis, we identified the following challenges:

Challenge P_1: While in practice often several degradation measures are available and required to characterize degradation behavior, the vast majority of existing prognostic methods solely focus on the availability of a single degradation measure. Since these univariate methods are not straightforwardly applicable to the multivariate case, the development of multivariate prognostic approaches is of practical relevance.

Challenge P_2: Existing prognostic methods are generally based on one overall degradation model. Since degradation may heavily vary for different fault causes, improvement in prediction accuracy is expected when multiple models are considered for degradation modeling.

In the field op maintenance optimization, we identified the following challenges:

Challenge M_1: Existing methods for condition-based maintenance optimization restrict themselves to answering the question whether or not maintenance is needed at a particular decision time instant. However, besides knowing that maintenance is required, it is important to know the required type of maintenance and the optimal time to perform maintenance. Based on this information, one can determine which material and personnel need to be sent to the maintenance location, and with what urgency this needs to be done. Therefore the development of maintenance optimization approaches that include the optimization of type and time of maintenance is of practical importance.

Challenge M_2: Existing methods for condition-based maintenance optimization plan the maintenance based on predefined threshold values. As a consequence, the timing of the maintenance decision cannot be tailored to the individual needs, and maintenance decisions are often taken at the last minute. However, when maintenance needs are known in time, maintenance activities can be planned at a convenient time, users can be informed in advance about the system downtime, and spare parts, material, and personnel required for performing the maintenance activities can be organized in an efficient way. This indicates that there is a need for maintenance optimization approaches that account for the individual needs and, if possible, plan the maintenance in an early stage.

Challenge M_3: Maintenance costs and inconvenience of system downtime can be reduced by combining or spreading maintenance activities in time. So far, relatively little research has been devoted to system-level maintenance optimization, and most of the existing literature solely incorporate positive economic dependencies. Therefore, ample room for improvement is left in the area of system-level maintenance optimization.

Next to the issues regarding the different sub-processes of condition-based maintenance outlined above, the following overall challenge has been identified:

Challenge O_1: Multiple methods have been proposed for fault diagnosis, failure prognosis, and maintenance optimization. The dependencies between the three processes are however often overlooked. As a consequence, the diagnostic and prognostic results are still not used optimally for maintenance optimization.

1.3.2. CONTRIBUTIONS

The main contributions of this thesis are:

1. We thoroughly analyze how a knowledge-based fault diagnosis problem is influenced by uncertainty, and compare how the uncertainty is handled in two well-known frameworks for reasoning under uncertainty, namely the Bayesian and the Dempster-Shafer framework [Challenges D_3, O_1] (Chapter 2).

2. We propose a knowledge-based approach to fault diagnosis in interconnected systems. By using the temporal, spatial, and spatio-temporal dependencies, the approach is robust with respect to environmental disturbances and requires only a small number of variables to be monitored [Challenges D_1, D_2] (Chapter 3).

3. We develop a multiple-model approach to system-level fault diagnosis of climate control systems. By constructing a distinct Bayesian network (diagnostic model) for each operating mode based on both historical data and knowledge regarding component interdependencies and conservation laws, component interdependencies and multiple operating modes are handled effectively [Challenges D_2, D_1] (Chapter 4).

4. We present a multiple-model approach to system reliability prediction in the presence of multiple degradation and failure modes. The multiple models account for different degradation behaviors associated with different fault causes. Provided that faults are correctly identified, a multiple-model approach will in general outperform a single-model approach with respect to prediction accuracy. Moreover, the proposed method generates conditional predictions of future system reliability, instead of commonly used predictions of the remaining useful life. This ensures an adequate connection with the subsequent maintenance optimization process [P_1, P_2, O_1] (Chapter 5).

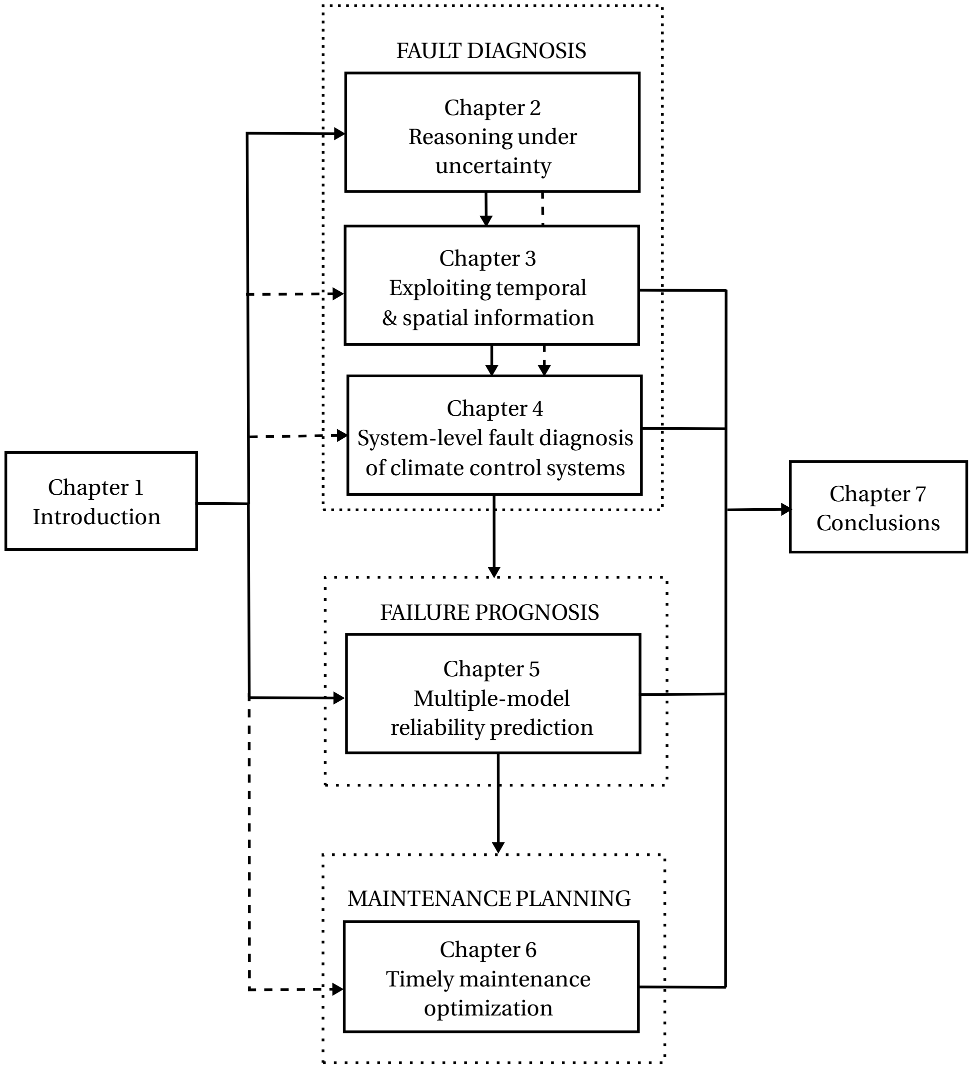
5. We propose an optimization-based approach to timely maintenance planning in multi-component systems. By exploiting diagnostic information, prognostic information, and system dependencies, the costs and inconvenience of maintenance are minimized [M_1, M_2, M_3, O_1] (Chapter 6).

Throughout this thesis we use examples of railway networks and climate control systems to demonstrate our contributions. Although (part of) the choices we made have been inspired by these particular applications, the methods proposed are not limited two these two applications. The methods proposed for fault diagnosis are applicable to all kinds of interconnected systems, where knowledge regarding system dependencies is available, e.g. drinking water distributions networks and highways. The methods proposed for failure prognosis and maintenance optimization are applicable to monitored systems in general.

1.4. THESIS OUTLINE

This thesis consists of seven chapters (including this chapter) and three appendices. Chapters 2 through 4 are about fault diagnosis, Chapter 5 concerns failure prognosis, and Chapter 6 is about maintenance optimization. Finally, in Chapter 7 conclusions and directions for further research are given. More specifically, in Chapter 2 we analyze the uncertain knowledge-based fault diagnosis problem and compare how the problem is handled in the Bayesian and the Dempster-Shafer reasoning framework. Background information regarding reasoning in the Bayesian and the Dempster-Shafer framework is included in Appendix A. In Chapter 3, we propose an approach to fault diagnosis in interconnected systems in the presence of environmental disturbances. The proposed approach is demonstrated on a case concerning fault diagnosis of railway track circuits. Railway track circuits are described in Appendix B. Chapter 4 proposes an approach for system-level fault diagnosis of climate control systems. The diagnostic model is generated based on both historical data and knowledge regarding component interdependencies and conservation laws. Relevant conservation laws are provided in Appendix C. In Chapter 5, we present a multiple-model approach to system reliability prediction. Chapter 6 proposes a condition-based approach to timely maintenance planning in multi-component systems.

Basically, Chapters 2 through 6 can be read separately and independently of each other. An exception is Section 3.5, which is based on the concepts proposed in Chapter 2. Moreover, knowledge of the concepts proposed in Chapters 2 and 3 may be beneficial for the understanding of Chapter 4. Similarly, the concepts introduced in Chapter 5 may improve the readability of Chapter 6. A schematic overview of the structure of this thesis together with the links between the different chapters is given in Figure 1.6.



Figuur 1.6: Structure of this thesis. The arrows indicate possible orders in which the different chapters can be read. The orders indicated by solid lines are preferred over the orders indicated by dashed lines.

2

REASONING UNDER UNCERTAINTY FOR KNOWLEDGE-BASED FAULT DIAGNOSIS

"It is not who is right, but what is right, that is of importance."

-Thomas Huxley-

Even though various frameworks exist for reasoning under uncertainty, a realistic fault diagnosis task does not fit into any of them in a straightforward way. For each framework, only part of the available data and knowledge is in the desired format. Moreover, additional criteria, like clarity of inference and computational efficiency, require trade-offs to be made. Finally, fault diagnosis is usually just a subpart of a larger process, e.g. condition-based maintenance. Consequently, the final goal of fault diagnosis is not (just) decision making, and the outcome of the diagnosis process should be a suitable input for the subsequent reasoning process. In this chapter, we analyze how a knowledge-based diagnosis task is influenced by uncertainty, investigate which additional objectives are of relevance, and compare how these characteristics and objectives are handled in two well-known frameworks, namely the Bayesian and the Dempster-Shafer reasoning framework. In contrast to previous works, which take the reasoning method as the starting point, we start from the application, knowledge-based fault diagnosis, and examine the effectiveness of different reasoning methods for this specific application. It is concluded that the suitability of each reasoning method highly depends on the problem under consideration and on the requirements of the user. The best framework can only be assigned given that the problem (including uncertainty characteristics) and the user requirements are completely known.

Parts of this chapter have been published in (Verbert et al., 2015b).

2

2.1. INTRODUCTION

Fault diagnosis is a challenging task, among other things, due to the presence of uncertainty. Especially for safety-critical systems, like medical devices, railway systems, and nuclear reactors, is it important to deal with the uncertainty in an adequate way (Challenge D_3). In this chapter, we analyze the reasoning problem of knowledge-based fault diagnosis. Knowledge-based fault diagnosis is influenced by uncertainty in various ways. First, the available measurement data may be incomplete, incorrect, or imprecise, e.g. due to sensors with a limited accuracy. Second, knowledge is needed to infer system health from these uncertain data. Also this knowledge is generally uncertain, i.e. (partly) incorrect, subjective, or incomplete.

Despite of the development of various methods for reasoning under uncertainty and the many discussions about the correctness and usefulness of these methods (Cheeseman, 1985; Lindley, 1987; Smets, 1992, 1994; Dubois et al., 1996; Ferson and Ginzburg, 1996; Dubois and Prade, 2001; Cobb and Shenoy, 2003b), no agreement has been reached regarding a consistent and uniform framework to handle problems under uncertainty. In particular the disagreement about the correctness and usefulness of the Bayesian and the Dempster-Shafer framework has led to debates. Bayesian proponents claim that the Bayesian theory is the optimal framework to handle all kinds of "uncertainty" (Cheeseman, 1985; Lindley, 1987). To quote Dennis Lindley, an eminent probabilist (Zadeh, 2008), "probability is the only sensible description of uncertainty and is adequate for all problems involving uncertainty. All other methods are inadequate" and "anything that can be done with fuzzy logic, belief functions, upper and lower probabilities, or any other alternative to probability can better be done with probability." While Bayesian proponents are convinced about their framework, shortcomings are claimed by many researchers (e.g. Shafer (1990); Smets (1992, 1994); Dubois et al. (1996); Ferson and Ginzburg (1996); Dubois and Prade (2001); Cobb and Shenoy (2003b); Haenni (2003)). For example, Shafer (1990); Smets (1992, 1994); Haenni (2003) argue for the need of belief functions and for their added value over probabilities. Especially, they promote belief functions for being superior in representing incomplete and partially reliable knowledge. Dubois et al. (1996) conclude that the Bayesian approach is tailored for decision making, but not necessarily for other kinds of reasoning. Ferson and Ginzburg (1996); Dubois and Prade (2001) consider different sources of uncertainty, all having their own characteristics, and they argue that each of these uncertainty sources requires another reasoning strategy. In contrast, Cobb and Shenoy (2003b) advocate that the Bayesian and Dempster-Shafer frameworks have roughly the same expressive power.

In this chapter, we compare Bayesian and Dempster-Shafer reasoning from an application-oriented point of view. In contrast to previous works, which take the reasoning method as the starting point and use examples to illustrate the effectiveness of the method, we start from the application, namely knowledge-based fault diagnosis, and examine the effectiveness of Bayesian and Dempster-Shafer reasoning for this specific application. More specifically, the contributions of this chapter are:

1. We analyze how the available data and knowledge are influenced by uncertainty;
2. We compare how the fault diagnosis task fits within the Bayesian and Dempster-Shafer reasoning framework;

3. We present additional objectives (e.g. clarity of inference) and analyze how they are accounted for in both reasoning frameworks.

Note that our aim is not to deeply discuss uncertainty methods nor to advocate one of the methods in general. We focus on a specific problem with the related objectives, for which we assess under which circumstances which method is most suitable to reach these objectives.

The remainder of this chapter consists of three parts: The first part (Section 2.2 till Section 2.4) discusses general concepts regarding reasoning under uncertainty. In the second part (Sections 2.5 and 2.6), we analyze the uncertain reasoning problem of knowledge-based fault diagnosis in the Bayesian and Dempster-Shafer framework. Finally, in Section 2.7 we compare the two reasoning frameworks with respect to both reasoning performance and additional objectives. Later in this thesis, in Chapter 3, we compare the two reasoning strategies on an example concerning fault diagnosis of railway track circuits.

2.2. CLASSIFICATION OF UNCERTAINTY

According to Dubois et al. (1996); Ferson and Ginzburg (1996); Dubois and Prade (2001); Zadeh (2008) various sources of uncertainty need to be treated differently. A distinction can be made between the following sources of uncertainty:

1. randomness;
2. incompleteness;
3. imprecision;
4. conflict.

Randomness, also called intrinsic variability, refers to the situation that a future outcome is uncertain, but a probability distribution of the outcome is available, e.g. throwing a known fair die. *Incompleteness* means that an outcome (or probability distribution) is defined, but the information available is not sufficient to identify this outcome (or probability distribution). For example, the evidence that the winner of a competition is a male is only sufficient to identify the winner in the case that there is only one male candidate winner. Otherwise, this evidence only allows to exclude candidate female winners. *Imprecision* refers to the situation that the outcome is known, but with finite precision. For example, we know that the current outside temperature is between 25.5 and 26.5 degrees Celsius. Finally, uncertainty can arise due to (partially) *conflicting* information. For example, two experts give a different answer to a particular question.

For reasoning purposes, uncertainty is often classified into the following two classes (Billinton and Huang, 2008; Kiureghian and Ditlevsen, 2009):

1. aleatory uncertainty;
2. epistemic uncertainty.

Aleatory uncertainty, also called statistical uncertainty, represents intrinsic variability, i.e. the differences that are observed each time the same experiment is repeated. *Epistemic uncertainty*, also called systematic uncertainty, arises due to a lack of knowledge. This is the uncertainty about things that we could in principle know, but in practice we do not know. The two are often distinguished using the fact that epistemic uncertainty can be reduced by gathering more knowledge or more data, whereas aleatory uncertainty cannot be reduced (Ferson and Ginzburg, 1996; Kiureghian and Ditlevsen, 2009). To illustrate this, consider the example of throwing a die. When we throw a die of which we know the underlying model, each time we get a different outcome, but throwing it more often will not provide information to reduce uncertainty about the outcome of a future throw. So, the uncertainty referred to is of the aleatory type. In contrast, when we throw an unknown die and we want to construct a probabilistic model of the outcome of a throw, then the more data we gather, the less uncertainty we have in our model. Here, the uncertainty referred to is of the epistemic type. Ideally, we would like to eliminate all epistemic uncertainty, so that only aleatory uncertainty remains. In practice, which part of the uncertainty actually can be reduced depends on the particular problem, practical constraints, and the assumptions adopted (Kiureghian and Ditlevsen, 2009).

Considering the different uncertainty sources: both imprecision, incompleteness, and conflict refer to a lack of knowledge and they can be regarded as epistemic uncertainty, whereas randomness can be regarded as aleatory uncertainty.

2.3. Methods for reasoning under uncertainty

For completeness and to make a link between the different uncertainty sources and the different reasoning frameworks, in this section we briefly introduce four common frameworks for reasoning under uncertainty, namely the Bayesian framework, the Dempster-Shafer framework, possibility theory, and fuzzy logic. A more extensive discussion of the frameworks compared later in this chapter, i.e. Bayesian and Dempster-Shafer reasoning, can be found in Appendix A.

2.3.1. Notation

We denote a variable by an upper-case letter (e.g. X, Y). A variable X can take values in its domain Θ_X. A particular element of Θ_X is denoted by x_i and a subset of Θ_X is denoted by x. A set of variables is denoted by a bold-face upper-case letter (e.g. **U**, **V**) and the assignment of a value to each variable in the set by the corresponding bold-face lower-case letter (**u**, **v**).

2.3.2. Bayesian probability theory

Probability theory (Laplace, 1814; Savage, 1954) is an established and well-known framework for reasoning under uncertainty. Roughly there are two interpretations of probability (Howie, 2002): the Bayesian and frequentist interpretation. Here, the focus is on the (subjective) Bayesian approach. Whereas frequentists only use data, Bayesians use data to improve their initial belief, i.e. "initial belief" + "data" = "improved belief". The combination of these two is beneficial in situations where relatively little data and a reasonable amount of prior knowledge are available (Goldstein, 2006). Technical details regarding reasoning in Bayesian networks can be found in Section A.1.

2.3.3. Dempster-Shafer framework

The Dempster-Shafer (D-S) framework (Dempster, 1967; Shafer, 1976; Kohlas and Monney, 1995) was developed to handle incomplete information. This is realized by allowing the assignment of belief to *sets of elements in the domain* instead of assigning belief only to *individual elements*, like in the Bayesian framework. Different interpretations of the D-S theory exist, among which are the upper and lower probabilities model and the evidentiary value model (Smets, 1994). In this thesis, we adopt Smets' well-known transferable belief model interpretation (Smets and Kennes, 1994). Technical details regarding reasoning in the transferable belief model can be found in Section A.2.

2.3.4. Possibility theory

Another way to handle incomplete information is using possibility theory (Dubois and Prade, 1988; Zadeh, 1999; Dubois and Prade, 2001). Instead of assigning one probability to each individual element in the domain, like in the Bayesian framework, possibility theory uses two binary values: a possibility value and a necessity value, making it possible to represent incomplete information (Dubois, 2006). The possibility of an event is equal to zero if and only if its negation is known to be true, and is equal to one otherwise. The necessity of an event is equal to one if and only if the event is known to be true. In practice, this binary representation is often not entirely satisfactory and a graded notion of possibility theory is used (see e.g. (Dubois, 2006)).

2

2.3.5. Fuzzy logic

The fuzzy logic framework (Zadeh, 1965, 1975; Klir and Yuan, 1995; Zadeh, 2008) was developed to handle perception-based information. Perception-based information is imprecise and cannot be represented by a single number. In fuzzy logic, everything is, or is allowed to be, graduated (Zadeh, 2006). So in this sense, a proposition can be partially true. Consider for example the proposition "The room temperature is very high". In standard logic, this proposition is true or false. In fuzzy logic, this proposition can be true with a degree between 0 and 1.

2.4. Relation between uncertainty sources and reasoning frameworks

In Section 2.2, we have discussed sources of uncertainty and in Section 2.3 various reasoning frameworks have been mentioned. The question that remains is "How do these relate to each other?". According to Bayesian proponents, probabilities are suited to handle all kinds of uncertainty, which is precisely the advantage of the Bayesian approach (O'Hagan, 2004). According to non-Bayesians, probabilities are suited to handle aleatory uncertainty, but are not suited to handle epistemic uncertainty (Dubois et al., 1996; Ferson and Ginzburg, 1996). Fuzzy set theory (Zadeh, 1965) has been proposed to handle imprecise information, and possibility theory (Dubois and Prade, 1988; Zadeh, 1999) and the theory of belief functions (Dempster, 1967; Shafer, 1976) have been proposed to handle incomplete information. An overview of these relations is given in Table 2.1.

Tabel 2.1: Uncertainty classifications and reasoning frameworks

uncertainty class	uncertainty source	reasoning framework
aleatory	randomness	Bayesian probability theory
epistemic	incompleteness	D-S theory, possibility theory
	imprecision	fuzzy logic
	conflict	D-S theory

2.5. KNOWLEDGE-BASED FAULT DIAGNOSIS

2.5.1. OVERVIEW

2

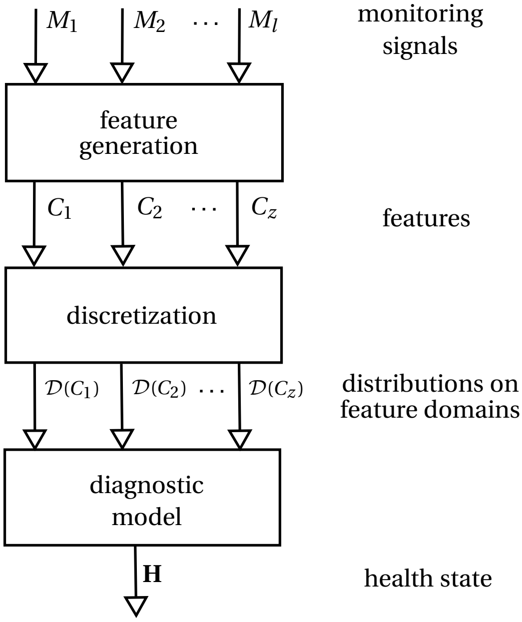
Knowledge-based fault diagnosis (see Section 1.1.1) is a model-based diagnosis strategy that uses knowledge to define the diagnostic model[1] in the form of a qualitative model or a rule-based system (Frank et al., 2000). Figure 2.1 gives an overview of the knowledge-based fault diagnosis process. The monitoring signals M_1 till M_l serve as input for the diagnosis, and the output is the system health represented by a set **H** of variables, indicating whether or not the system is healthy, and if not, what actually causes the faulty behavior. To determine the system health, first, characteristic features C_1 till C_z are extracted from the monitoring signals. Next, the values of features C_1 till C_z are determined and, in the presence of uncertainty, represented by distribution functions over the associated discrete domains $\Theta_{C_1} = \{c_{1,1}, c_{1,2}, \ldots, c_{1,n_1}\}$ till $\Theta_{C_z} = \{c_{z,1}, c_{z,2}, \ldots, c_{z,n_z}\}$. The type of distribution function depends on the reasoning framework used for the fault diagnosis, e.g. in the Bayesian framework, a probability distribution is used, while in the D-S framework, a D-S belief function is used (see Appendix A for more details regarding the different distribution functions). Finally, based on the distributions over the feature domains, the presence and type of faults is inferred by using the diagnostic model.

So, the reasoning task of knowledge-based fault diagnosis concerns the determination of the system health based on the values of the features C_1 till C_z. Therefore, we distinguish between two groups of variables:

1. the set **C** of observable variables $\mathbf{C} = \{C_1, \ldots, C_z\}$;
2. the set **H** of target variables representing the system health.

Assuming that a system has one healthy mode h and ℓ different fault causes f_1 till f_ℓ, the system health is represented by one $(\ell+1)$-valued variable H with $\Theta_H = \{h, f_1, \ldots, f_\ell\}$, or by ℓ two-valued variables F_1 till F_ℓ all taking on values in the set $\{0, 1\}$, indicating the absence (0) or presence (1) of the respective fault causes f_1 till f_ℓ. Generally, the first option is preferred when only single-fault scenarios are considered, while the second option is used when also multiple-fault scenarios are taken into account. A combination of the two can be used when only part of the multiple-fault scenarios are accounted for.

[1]A diagnostic model is a set of static or dynamic relations that link specific input variables – the feature values – to specific output variables – the faults (Isermann, 2011).



Figuur 2.1: Overview knowledge-based diagnosis.

Unless otherwise stated, we allow multiple-fault scenarios and use one binary variable for each possible fault cause.

2.5.2. DIAGNOSIS WITHIN THE CONDITION-BASED MAINTENANCE PROCESS

In general, fault diagnosis is not an isolated task, but it is part of a larger process. As reasoning under uncertainty and decision making in the presence of uncertainty impose different requirements on uncertainty characterization, it is important to consider the purpose(s) of fault diagnosis. We consider fault diagnosis as part of a condition-based maintenance process (see Section 1.1). The final goal of condition-based maintenance is maintenance planning, i.e. deciding on the required maintenance activities. This is done based on the diagnostic and prognostic results. The diagnostic outcome serves as an input for both the prognosis and the maintenance optimization step (see Figure 1.1). So, the final goal of condition-based maintenance is decision making. However, the main goal of fault diagnosis is reasoning about the system health based on monitored variables. Therefore, in this chapter, the main focus is on information fusion and reasoning under uncertainty, and less on decision making.

2.5.3. UNCERTAINTY SOURCES

As already indicated, uncertainty can originate from different sources. For knowledge-based fault diagnosis, we identify the following main sources of uncertainty:

1. uncertainty arising from imperfect sensors;

2. uncertainty regarding the relations between features and faults;
3. uncertainty arising from the conversion from measurement data to the feature space.

More specifically, we characterize the aforementioned uncertainty sources as follows:

Sensors
In general, sensors are imprecise (i.e. they have limited accuracy) and may suffer from structural errors (e.g. off-sets, drift). Due to imperfect sensors, our assumed world differs from reality. Therefore, this type of uncertainty refers to a lack of knowledge. In that sense, the uncertainty can be reduced e.g. by calibrating sensors, implementing better sensors, or using additional sensors. In practice, the available sensors are generally fixed (and cannot be changed) and their precision is approximately known. In this case, the corresponding uncertainty is regarded as intrinsic variability, so it is of the aleatory type.

2

Relations between features and faults
Here, two sources of uncertainty play a role. First uncertainty arises from the relations between features and faults being not completely deterministic due to unmodeled influences. Second, the available knowledge relating faults and features may be incomplete or imprecise. The latter reflects a lack of knowledge (epistemic uncertainty); the former is, for diagnostic purposes, generally regarded as aleatory uncertainty.

Conversion from measurement data to the feature space
Based on the monitoring signals, the features have to be determined. In general, a derived feature C_k does not behave exactly according to one element in its domain Θ_{C_k} (Bayesian framework) or to one element in the power set $2^{\Theta_{C_k}}$ (D-S framework). So, it has to be determined to what extent the observed behavior corresponds to each element of Θ_{C_k} or $2^{\Theta_{C_k}}$. The exact uncertainty characteristics depend on the system behavior and the way the behavior is evaluated, e.g. by subjective human judgment or mathematical (computer) calculations.

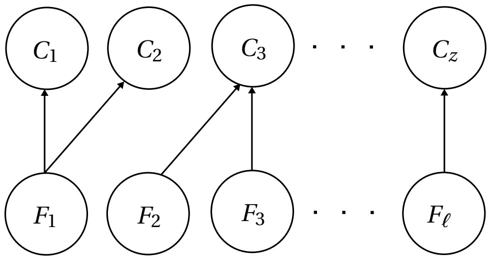
2.6. Reasoning under uncertainty for knowledge-based fault diagnosis

In this section, we discuss how the knowledge-based fault diagnosis problem is handled in the Bayesian and the D-S framework.

2.6.1. Bayesian networks

The considered knowledge-based fault diagnosis problem (see Section 2.5.1) is graphically represented by a Bayesian network such as the one shown in Figure 2.2. The edges indicate that fault f_1 has a direct influence on both feature C_1 and feature C_2, that both fault f_2 and fault f_3 influence feature C_3, and that feature C_z is influenced by fault f_ℓ.

Before the Bayesian network can be used for reasoning, the prior probability distributions of F_1 till F_ℓ (root nodes), and the conditional probability tables of C_1 till C_z need to be determined. The prior probabilities indicate the likelihood of a particular fault f_j,



Figuur 2.2: Bayesian network representation of the knowledge-based fault diagnosis problem. The variables F_1 till F_ℓ indicate the presence or absence of the different system faults f_1 till f_ℓ, and the variables C_1 till C_z represent the diagnostic features.

i.e. $P(F_j = 1)$, before any evidence is collected. The conditional probability table of C_k contains the probabilities of each feature value $c_{k,\iota}$, $\iota = 1,\ldots,n_k$, given the value of each parent of C_k. For the example in Figure 2.2, feature C_3 has parents F_2 and F_3; so, for C_3, the conditional probability table as given in Table 2.2 needs to be defined.

Tabel 2.2: Example of a conditional probability table of C_3

		C_3			
F_2	F_3	$c_{3,1}$	$c_{3,2}$	…	c_{3,n_3}
0	0	…	…	…	…
0	1	…	…	…	…
1	0	…	…	…	…
1	1	$\frac{1}{n_3-1}$	$\frac{1}{n_3-1}$	…	0

Often, the available knowledge is not in probabilistic form, e.g. we are uncertain about the prior probability distribution of F_j, or we are not sure about the conditional probability distribution $P(C_k|\mathbf{U}_{C_k})$ of feature C_k given the values of its parents $\mathbf{U}_{C_k}$. For example, we only know that given that $F_2 = F_3 = 1$, it holds that $P(C_3 = c_{3,n_3}) = 0$. In such case, the remaining probabilities are assigned according to the additivity axiom and the principle of maximum entropy (Jaynes, 1957a,b) (see e.g. the last row of Table 2.2). The *additivity axiom* states that $P(a) + P(\bar{a}) = 1$, and the *principle of maximum entropy* is a strategy in which missing probabilities are assigned such that the distribution is consistent with known constraints, but is otherwise as unbiased as possible. The result of the principle of maximum entropy is a discrete uniform distribution over the missing values.

After the Bayesian network is initialized, i.e. the structure and the set of local probability distributions are defined, it can be used for reasoning. So, we can update the model based on evidences regarding the features C_1 till C_z (the observable variables) and compute the marginal probability distributions of F_1 till F_ℓ (the target variables). When the available evidences are hard evidences[2], they can be easily propagated using standard Bayesian inference algorithms (see Section A.1.3). When the available evidences are uncertain, it first needs to be assured that they are specified by likelihood ratios as required by Pearl's method of virtual evidence (see Section A.1.3). When the evidences are specified as probabilistic evidence[3], standard rules can be used for the conversion (see Section A.1.3). In practice, evidences are often specified by human experts, which do not necessarily follow the Bayesian laws. For example, a (partially) incomplete answer, like "the value of C_k is $c_{k,2}$ or $c_{k,4}$" is also plausible. Probabilistic information is derived from such incomplete information by using the principle of maximum entropy.

To summarize, knowledge-based fault diagnosis in Bayesian networks may require the following pre-processing steps to match the available information with the Bayesian format:

1. Transformation of the uncertain knowledge base (i.e. the relations between features and faults) into a set of conditional probability tables. Usually, the available knowledge is already conditional. Only missing probabilities in the case of incomplete information have to be estimated;

2. Determination of the prior fault probabilities;

3. Transformation of the evidence into the format specified by the virtual evidence method, i.e. likelihood ratios (see Section A.1.3).

2

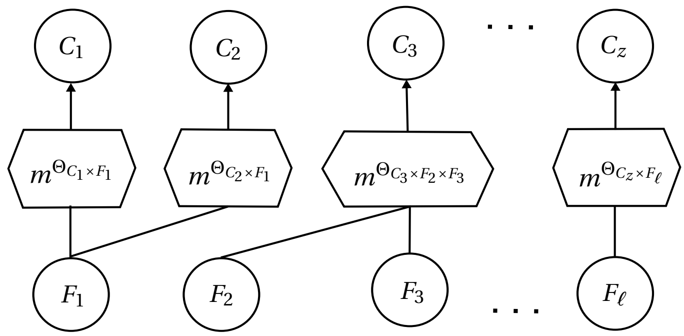
2.6.2. DEMPSTER-SHAFER BELIEF NETWORKS

In the D-S framework, the considered knowledge-based fault diagnosis problem (see Section 2.5.1) is represented by a D-S valuation network (see Figure 2.3 for an example). The notation $\Theta_{A\times B\times C}$ is used as a shorthand for the multidimensional space $\Theta_A \times \Theta_B \times \Theta_C$. Before the valuation network can be used for reasoning, the prior mass distributions of F_1 till F_ℓ and the multivariate mass functions describing the valuations (hexagons in Figure 2.3) need to be defined. The prior mass distributions indicate the likelihood of a particular fault before any evidence is collected. The important difference with the Bayesian analysis is that, in the D-S framework, the prior mass functions[4] of F_1 till F_ℓ can be defined as vacuous mass functions, expressing total ignorance, i.e. $m^{\Theta_{F_j}}(\Theta_{F_j}) = 1$. The relationships (valuations) between variables need to be defined by multivariate mass functions on the product spaces of the domains of the connected variables. For example (see Figure 2.3), the relation between C_3, F_2, and F_3 is characterized by a mass function on the space $\Theta_{C_3} \times \Theta_{F_2} \times \Theta_{F_3}$. A mass needs to be attached to each combination of possible values. For example, to capture the relation between F_1 and C_1, assuming that C_1 can take values in $\Theta_{C_1} = \{c_{1,1}, c_{1,2}\}$, the masses given in Table 2.3 need to determined. When all mass is assigned to the masses in the first column, the information available is complete, but possibly uncertain. The more mass is assigned to the masses in the right columns, the more ignorant one is. Note that even for a two-dimensional mass function

[2]Hard (or certain) evidence for a variable X is evidence that states that X takes a particular value $x_i \in \Theta_X$.

[3]Probabilistic evidence for a variable X is specified by a probability distribution over Θ_X.

[4]Remember that a D-S mass function is a distribution in which belief is assigned to sets of events rather than to single events.



Figuur 2.3: D-S valuation network representation of the knowledge-based fault diagnosis problem. The variables F_1 till F_ℓ indicate the presence or absence of the different system faults f_1 till f_ℓ, the variables C_1 till C_z represent the diagnostic features, and the valuations (multivariate mass functions) $m^{\Theta_{C_1 \times F_1}}, m^{\Theta_{C_2 \times F_1}}, m^{\Theta_{C_3 \times F_2 \times F_3}}$, and $m^{\Theta_{C_z \times F_\ell}}$ express the (uncertain and incomplete) relationships between the respective variables.

2

with small domains, a large number of masses is needed to capture the available (incomplete) knowledge. In the worst case, 15 nonzero masses need to be assigned for the given example. In comparison, the Bayesian model requires only 4 conditional probabilities to be specified. These are the costs that have to be paid for including the possibility of expressing ignorance. For fault diagnosis, the information available is often specified in conditional form, in which case the joint masses are estimated by using the ballooning extension (see Section A.2.1).

Tabel 2.3: masses capturing the relation between F_1 and C_1

$m(0,c_1)$	$m(\{(0,c_1),(0,c_2)\})$	$m(\{(0,c_1),(0,c_2),(1,c_1)\})$	$m(\Theta_{C_1} \times \Theta_{F_1})$
$m(0,c_2)$	$m(\{(0,c_1),(1,c_1)\})$	$m(\{(0,c_2),(1,c_1),(1,c_2)\})$	
$m(1,c_1)$	$m(\{(0,c_1),(1,c_2)\})$	$m(\{(0,c_1),(1,c_1),(1,c_2)\})$	
$m(1,c_2)$	$m(\{(0,c_2),(1,c_1)\})$	$m(\{(0,c_1),(0,c_2),(1,c_2)\})$	
	$m(\{(0,c_2),(1,c_2)\})$		
	$m(\{(1,c_1),(1,c_2)\})$		

A D-S valuation network is used for reasoning as follows: When new evidence becomes available, the network is updated according to Dempster's rule of combination (A.17). These evidences should be represented in the form of a mass function.

To summarize, knowledge-based fault diagnosis in D-S valuation networks may require the following pre-processing steps to match the available information with the D-S demands:

1. Transformation of the uncertain knowledge base into the desired format, i.e. multivariate mass functions on the joint domains. Usually, the knowledge is conditional and the distributions on the joint domains need to be estimated using

the ballooning extension;

2. Transformation of the available evidences into mass functions.

2.7. COMPARISON AND ADDITIONAL CRITERIA

2.7.1. DIAGNOSTIC REASONING PERFORMANCE

From the analysis in Section 2.6, we conclude that the Bayesian model is particularly suited for reasoning about conditional relationships, like the relations between faults and features. In practice, the relationships between faults and features, as well as the available evidences are however not purely probabilistic, and approximations need to be made when using the Bayesian model. In contrast, the D-S model is perfectly suited to handle knowledge that is not purely probabilistic, e.g. incomplete or imprecise. The D-S model is however particularly suited for non-causal reasoning tasks (Cobb and Shenoy, 2003b), e.g. information fusion, and, compared to the Bayesian model, less tailored to diagnostic reasoning. So, when we have to chose for one of the two methods, a trade-off needs to be made. In general, when the problem mainly concerns causal/diagnostic reasoning and the information available is (almost) complete, i.e. probabilistic, the use of the Bayesian model is recommended. When the problem concerns mainly non-causal reasoning and the available information is incomplete, the D-S model is recommended. As the exact reasoning task and the associated uncertainty characteristics are application-specific, this trade-off needs to be made for each specific diagnostic problem individually. Unfortunately, a good insight into the characteristics of all uncertain influences is often missing for practitioners, which complicates the choice of the method.

Table 2.4 gives an overview of the advantages and disadvantages of the Bayesian and the D-S model. The first three properties follow from the previous analysis, the other properties will be discussed in the remainder of this section. Note that in this table, the two methods are compared qualitatively relative to each other, i.e. a minus sign merely indicates that the method is less suited compared to the other method.

Tabel 2.4: Comparison of Bayesian and Dempster-Shafer reasoning

	Bayesian framework	D-S framework
suitability for causal/diagnostic reasoning	+	−
suitability for non-causal reasoning	−	+
handling incomplete information	−	+
computational efficiency	+	−
suitability for decision making	+	+
clarity of inference	+	−
adaptability	+	−

2.7.2. ADDITIONAL CRITERIA

For practical problems, additional criteria like computational efficiency, suitability for decision making, clarity of inference, and adaptability are of importance (see Table 2.4).

COMPUTATIONAL EFFICIENCY

Computationally, D-S networks are more expensive to evaluate than Bayesian networks (Cobb and Shenoy, 2003b; Haenni and Lehmann, 2003). The worst-case complexity of a Bayesian network is $\mathcal{O}(n)$, whereas the worst-case complexity of a D-S network is $\mathcal{O}(2^n)$, with n the dimension of the state space of the largest clique in the join tree[5] (Cobb and Shenoy, 2003b). The size n of the state space of the largest clique depends on the dimensions of the state spaces of variables, the dimensions of state spaces of valuations, and the structure of the graph (Cobb and Shenoy, 2003b). To what extent the higher computational complexity of D-S networks is practically disadvantageous depends on the size of the network and on the available calculation time and power. For online diagnosis this implies that the Bayesian approach has the advantage that the diagnosis can be carried out with a smaller delay due to calculations.

2

SUITABILITY FOR DECISION MAKING

Often, it is argued that only the Bayesian model is appropriate for rational decision making, as probabilities fit within the expected-utility theory (Lindley, 1987). Mass distributions can however be easily transformed to probability distributions at the time decisions have to be made by using the pignistic transformation (see Section A.2.4). Note that in the case of incomplete information, non-probabilistic information is transformed into probabilities without any fundamental reason to do so, except to facilitate decision making. Consider e.g. the extreme case that we have a non-informative mass function m^{Θ_H} regarding variable H:

$$m^{\Theta_H}(\Theta_H) = 1.$$

We can transform this mass function into a probability distribution. However, as we have no knowledge, every probability distribution is equally good (or bad). Is it justified to make decisions based on guessed odds? In addition, if a decision needs to be made, is it justified to ignore that the outcome was just (or partly) based on a guess? Incomplete information indicates that the information collected so far is not sufficient to make a sound decision (Haenni, 2003), so more information should be gathered or the diagnostic setup should be improved. In some situations, decisions need to be made, but even in these cases it seems beneficial to have insight in the underlying mass distributions, e.g. to give feedback about the quality of the monitoring setup. In addition, measures of uncertainty may provide information about the severeness of a fault (Engel et al., 2000). Generally, it holds that the more severe the fault, the lower the ignorance and conflict. This is because for severe faults relatively large amounts of data are available. Moreover severe faults manifests itself more clearly in the data compared to incipient faults. Analyzing and exploiting the uncertainty present require that all computations are done in the D-S framework, which is computationally less attractive. However, applying a technique

[5]A join tree is the moralization of a directed graph into a tree structure that supports efficient inference.

based on probabilities using information that is not probabilistic, may yield erroneous results (Ferson and Ginzburg, 1996).

Based on the considerations presented, we conclude that the Bayesian model naturally fits decision making. Decision making in the D-S framework is slightly more involved compared to decision making in the Bayesian framework. However, mass functions contain more information, so allowing more informed decisions. Therefore, we consider Bayesian and D-S reasoning equally suitable for decision making.

CLARITY OF INFERENCE

Clarity of inference is of importance for most practical applications, as the implementation of a decision support system is much easier when the reasoning is intuitive and understandable. In this sense, Bayesian networks outperform D-S networks, since the causal representation in Bayesian networks is more natural and easier for the user to provide and understand (Yaghlane and Mellouli, 2008).

Although the Bayesian reasoning is considered clearer, the D-S output is clearer, as the D-S framework makes a distinction between probabilistic information and incomplete information. In the D-S framework, two distinct outcomes are obtained in the situation that no information regarding a variable H is available, i.e. $m^{\Theta_H}(\Theta_H) = 1$, and the situation in which we have the information that all elements in Θ_H are equally likely, i.e. $m^{\Theta_H}(h_1) = m^{\Theta_H}(h_2) = \ldots = m^{\Theta_H}(h_n) = 1/n$. In contrast, in the Bayesian framework the two situations are represented by the same probability distribution, $P(H = h_1) = P(H = h_2) = \ldots = P(H = h_n) = 1/n$. The additional information provided by the D-S outcome can be used to reconsider the diagnostic setup (e.g. an incomplete outcome may be a reason to extend the knowledge base, whereas a probabilistic answer may be a reason to implement better sensors) or to assist decision making (e.g. by choosing a conservative decision when the diagnostic result is incomplete).

ADAPTABILITY

Adaptability indicates how easily new knowledge can be incorporated in the network, e.g. when we want to include new faults or features in the model or update the relations between faults and features. This property is mainly important for large networks when it is expected that the model needs to be updated multiple times over time. Both frameworks allow for the incorporation of new knowledge without the need to redefine the whole model. As (new) knowledge relating faults and features is generally in causal form, the incorporation in the Bayesian model is more straightforward.

2.8. CONCLUSIONS

We have compared Bayesian and Dempster-Shafer reasoning for knowledge-based fault diagnosis. The Bayesian model is based on probabilities and is tailored to causal reasoning based on probabilistic knowledge. The Dempster-Shafer model is based on belief functions and is tailored to non-causal reasoning based on both probabilistic and incomplete information. Fault diagnosis comprises causal reasoning, often based on incomplete information. So, none of the two reasoning models fits the diagnostic reasoning task in a straightforward way. In addition, real-life diagnosis problems often include additional criteria, e.g. we want to know how reliable the reasoning results are, or we want to retrieve why a certain conclusion has been reached. For such problems, without an exactly defined performance criterion, it is not possible to unambiguously conclude what the best method is. We conclude that the final choice for a reasoning framework depends on the problem under consideration (including uncertainty characteristics), requirements of the user, and personal preferences. In general, the better the match between the probabilistic description and the real information, the more suitable the Bayesian approach is. The more conflicting and incomplete the available information, the more informative the D-S solution is compared to the Bayesian solution.

3

FAULT DIAGNOSIS USING SPATIAL AND TEMPORAL INFORMATION

"What we see depends mainly on what we look for."

-John Lubbock-

It is often impracticable to monitor a large number of system variables. This results in a need for fault diagnosis methods that achieve adequate diagnostic performance given a limited number of monitoring signals. We propose such a fault diagnosis method for interconnected systems. This approach is knowledge-based and uses the temporal, spatial, and spatio-temporal dependencies as diagnostic features. These features can be derived from the existing monitoring signals; so no additional sensors are required. Moreover, taking spatial dependencies into account makes the approach robust with respect to environmental disturbances. For a railway track circuit example, we show that, without the temporal, spatial, and spatio-temporal features, it is not possible to identify the cause of a detected fault. Including the additional features allows potential causes to be identified.

Parts of this chapter have been published in (Verbert et al., 2015a, 2016).

3.1. INTRODUCTION

In this chapter, we propose an approach to fault diagnosis in interconnected systems, i.e. systems consisting of multiple interdependent components. Key features of this approach are that it relies on the availability of only a limited number of monitored variables (Challenge D_1) and that it is robust with respect to environmental disturbances. To ensure adequate diagnostic performance, the following diagnostic features are taken into account:

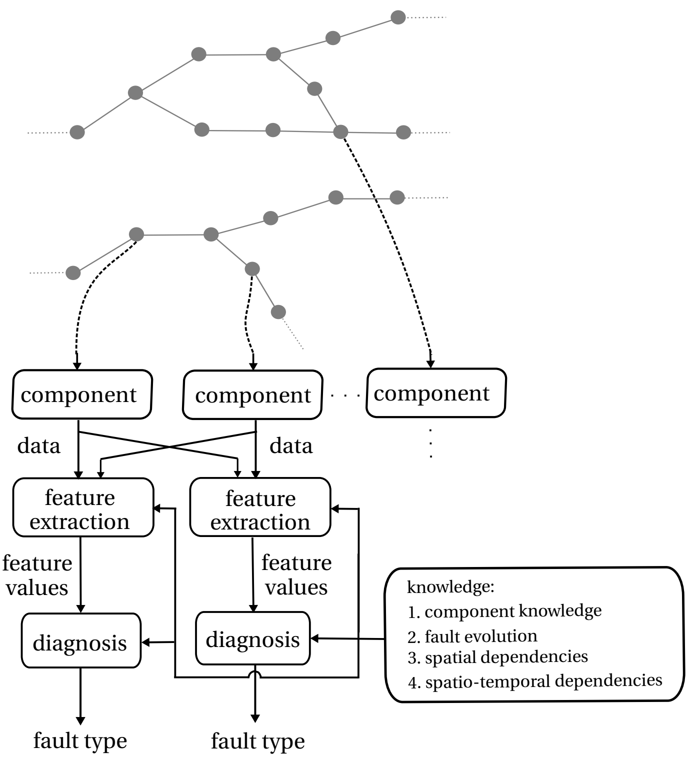
1. temporal dependencies;
2. spatial dependencies;
3. spatio-temporal dependencies.

The *temporal dependencies* are valuable for diagnosis because different faults evolve in different ways. Knowing the temporal degradation behavior provides insight into possible fault causes. Similarly, the *spatial dependencies* are useful because they are different for different types of system faults, i.e. some faults only influence one system component, whereas other faults influence multiple components. Finally, the *spatio-temporal dependencies* become of interest when objects move through the system. In this case, faulty behavior can be caused by the system itself or by a moving object. Since object faults manifest themselves differently in place and time than system faults, spatio-temporal system dependencies are a suitable feature to discriminate between the two fault categories. The temporal, spatial, and spatio-temporal dependencies can be determined from the available monitoring signals, meaning that they do not require the installation of additional monitoring devices.

3

Figure 3.1 gives a schematic overview of the proposed diagnosis approach. The proposed method can be used to monitor all kinds of systems where temporal and spatial knowledge is available, e.g., drinking water distribution networks, building infrastructures, and highways. We illustrate the applicability of the proposed method based on a railway track circuit diagnosis task.

Fault diagnosis for railway track circuits has already been dealt with, e.g. by Chen et al. (2008); Oukhellou et al. (2010); Cherfi et al. (2012); Lin-Hai et al. (2012); Sandidzadeh and Dehghani (2013); Sun and Zhao (2013). A distinction can be made regarding the way the monitoring data are obtained, e.g. using a measurement train (Oukhellou et al., 2010; Cherfi et al., 2012; Lin-Hai et al., 2012; Sun and Zhao, 2013) or using track-side monitoring devices (Chen et al., 2008; Sandidzadeh and Dehghani, 2013). In this thesis, track-side monitoring devices are considered because they continuously monitor the system and are therefore suitable for the early detection and diagnosis of faults. The main difference compared to the approaches by Chen et al. (2008); Sandidzadeh and Dehghani (2013) is that in those works multiple monitoring signals are used, while in this thesis, for each track circuit, only one variable is assumed to be monitored. Although the availability of a wide variety of measured quantities can be beneficial for model-based fault diagnosis (Isermann, 2005), it is not realistic to assume that this will be realized for the whole rail infrastructure, as the related installation and monitoring costs are high. Therefore, we restrict ourselves to one monitoring signal: the current measured at the track circuit receiver.



Figuur 3.1: Overview of proposed diagnosis approach.

This remainder of this chapter consists of four parts: 1. a part regarding fault diagnosis in general interconnected systems (Section 3.2); 2. a part covering fault diagnosis in railway track circuit networks (Section 3.3); 3. two specific track circuit diagnosis examples (Sections 3.4 and 3.5); and 4. a part discussing the extension to other diagnosis strategies (Section 3.6).

3.2. FAULT DIAGNOSIS IN INTERCONNECTED SYSTEMS

In this section, we propose a knowledge-based approach to fault diagnosis in interconnected systems (see Figure 3.2). In brief, we collect monitoring signals from all system components, correct them for the effect of environmental disturbances (Section 3.2.3), and extract diagnostic features from the corrected signals. Based on the extracted features (Section 3.2.2) and knowledge of the component operating states (Section 3.2.1), we infer the health of each system component. In the remainder, the different steps are worked out in more detail.

3.2.1. DIAGNOSIS SETUP

3

Consider a system consisting of n components[1] that can be graphically represented by a, possibly disconnected, graph (see e.g. the graph in Figure 3.1). In this graph, the black dots represent system components and the edges represent connections between components. Here, we consider fault diagnosis of an arbitrary system component i. We assume that each component i has one healthy mode h and ℓ faulty modes f_1 till f_ℓ. For clarity of presentation, in the theory part of this chapter we consider only single-fault scenarios, i.e. the health H_i of each component i takes one value in the set $\Theta_H = \{h, f_1, \ldots, f_\ell\}$. Furthermore, it is assumed that for each system component i a monitoring signal (vector) M_i is available. From these monitoring signals, we extract the following diagnostic features (see Figure 3.2):

- intra-component dependencies K_i;
- temporal dependencies T_i;
- spatial dependencies S_i;
- spatio-temporal dependencies G_i.

We will elaborate on these features in Section 3.2.2.

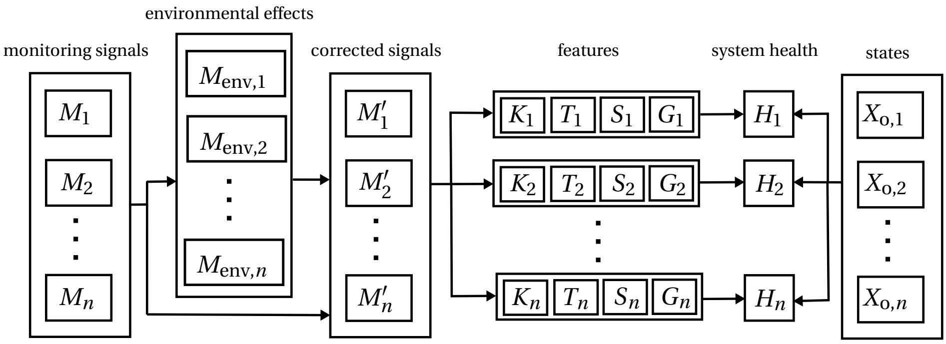
Generally, the operating state $X_{o,i}$ of each component i can take a finite number $m \geq 1$ of possible values $x_{o,1}$ till $x_{o,m}$. For example, for a railway switch, $X_{o,i}$ can take the following values:

$x_{o,1}$: at rest;

$x_{o,2}$: moving from the normal position to the reverse position;

$x_{o,3}$: moving from the reverse position to the normal position.

[1] For clarity of presentation, in this chapter all components are assumed to be identical.



Figuur 3.2: Schematic overview of the proposed knowledge-based fault diagnosis approach.

In general, the monitoring signal (vector) M_i and the extracted features can take different values. The interpretation of these values depends on the operating state. Therefore, for components with more than one state value, i.e. $m > 1$, it is only guaranteed that the health H_i can be inferred from the extracted features if we know the current operating state $X_{o,i}$. In this thesis, the following basic assumption is adopted

Basic assumption A_0: The state $X_{o,i}$ is known for each system component i at all times.

The state $X_{o,i}$ can e.g. be determined from additional analyses or sensor measurements.

3.2.2. DIAGNOSTIC INFERENCE

To determine the health of a system component, a representative set of diagnostic features is extracted from all available monitoring signals (see Figure 3.2). Based on component and system knowledge, these diagnostic features are then linked to the component health. In this section, we first introduce the rule-based system we use to capture the diagnostic model. Next, we propose four diagnostic features and discuss the knowledge required to link these features to the component health.

For sake of brevity, in the sequel we omit the subscript i when the explicit reference to a particular component i is not necessary.

DIAGNOSTIC MODEL

To describe the relations between the features and the component health, we use a rule-based system. Consider that we have z features C_1 till C_z and that each feature C_k is n_k-valued, i.e. C_k takes values in the set $\{c_{k,1}, c_{k,2}, \ldots, c_{k,n_k}\}$. Each feature C_k is linked to the system health H by the following, state-dependent, set of rules:

if $X_o = x_{o,\zeta}$ **then**

$$\textbf{if } C_k = c_{k,1} \textbf{ then } W_{C_k}^{(\zeta)} = w_{k,\zeta,1} \tag{3.1a}$$

$$\textbf{elseif } C_k = c_{k,2} \textbf{ then } W_{C_k}^{(\zeta)} = w_{k,\zeta,2} \tag{3.1b}$$

$$\vdots$$

$$\textbf{elseif } C_k = c_{k,n_k-1} \textbf{ then } W_{C_k}^{(\zeta)} = w_{k,\zeta,n_k-1} \tag{3.1c}$$

$$\textbf{else } W_{C_k}^{(\zeta)} = w_{k,\zeta,n_k} \tag{3.1d}$$

and

$$H \in W_{C_k} \tag{3.2}$$

$$W_{C_k} = W_{C_k}^{(1)} \cap W_{C_k}^{(2)} \cap \ldots \cap W_{C_k}^{(m)} \tag{3.3}$$

with each $w_{k,\zeta,\beta} \subseteq \{h, f_1, \ldots, f_\ell\}$ and W_{C_k} the set of possible fault causes given C_k in all operating modes. When we have m operating states and z features C_k that are n_k-valued, we end up with $m\sum_{k=1}^{z} n_k$ rules. In the ideal case, given the value of each feature C_k in each operating state, we can determine the fault cause, i.e. we know the value of H. This is guaranteed to be the case if each observed combination of feature values corresponds to one possible fault cause, i.e. if:

3

$$|W| = |W_{C_1} \cap W_{C_2} \cap \ldots \cap W_{C_z}| = 1 \tag{3.4}$$

for all possible valid assignments of values for W_{C_1} till W_{C_z}. When $\ell + 1 > |W| > 1$, it is not possible to determine the fault cause, but it is possible to exclude some of the fault causes.

DIAGNOSTIC FEATURES

We propose four diagnostic features for fault diagnosis in interconnected systems. The first two features, intra-component dependencies and temporal dependencies, are applicable to both isolated and interconnected systems. The last two features, spatial and spatio-temporal dependencies, become of interest when the system is part of a larger system.

Intra-component dependencies K_i Component knowledge is generally considered as a first source for feature generation in model-based diagnosis strategies. Component knowledge is used to generate a qualitative description of the nominal, i.e. fault-free, component behavior. Based on insight into component behavior, useful features K_i can be extracted from the monitoring signal vector M_i. Comparing the value of the feature K_i derived from the measurement data with the value of K_i derived from the component model, provides information about the health H_i of component i (Isermann, 2005).

Temporal dependencies T_i Although, in general, a fault may develop in a complex way and an exact quantitative description of the fault evolution cannot be provided, often information is available regarding its qualitative time behavior. For example, it may be known whether the time evolution of a particular fault is intermittent or approximately linear. This information can be used to distinguish between the different faults. Based on the available fault and component knowledge, the expected temporal behavior T_i of the monitoring signal vector M_i as a consequence of a particular fault in component i can be determined. Conversely, based on the observed temporal behavior of M_i, possible underlying fault evolution behaviors can be recovered. So, based on the temporal behavior of M_i, we can infer possible fault evolution behaviors and subsequently the associated fault types.

Spatial dependencies S_i When a component i is part of a larger system, the monitoring signals of other system components may contain valuable information regarding the health of component i. Valuable information is contained in these data if the presence of a system fault introduces dependencies among system components that differ from the dependencies introduced by another type of system fault. In this case, the cross-correlations between the monitoring signals of the different components provide information about the fault causes. So, faults can be classified according to their impact region. For example, a distinction can be made between faults that are specific to one component and faults that affect all interconnected components.

3

Spatio-temporal dependencies G_i In the case that objects move through the system, faulty behavior can be caused by the system itself or by a moving object. To distinguish between a fault in component i and an object fault, we propose to use the spatio-temporal system dependencies G_i. When the faulty behavior is caused by an object o moving through the system, it is expected that this behavior can only be observed in the monitoring signal of component i when the object is in component i, and that time-shifted versions of the faulty behavior are visible in the monitoring signals of each other component lying on the path $\mathcal{P}_o$ of the moving object o. In the case of a fault in component i, the faulty behavior is observed regardless of the objects moving through component i. Note that our main focus is on the fault diagnosis of the system components and not on diagnosing objects passing through the system. Therefore, only a distinction is made between faults that are object-specific and faults that are not object-specific, i.e. G_i is a variable that can take on two values only. This distinction is made to prevent that object faults are incorrectly diagnosed as system faults.

To determine the proposed features, standard techniques from signal analysis or pattern recognition (Papoulis, 1978; Meyer-Baese and Schmid, 2014) can be used. The exact procedure to determine these dependencies is application-specific and a further elaboration is beyond the scope of this thesis. Section 3.4 briefly explains the determination of the feature values for a track circuit diagnosis task.

3.2.3. CORRECTION FOR ENVIRONMENTAL DISTURBANCES

A key property of the proposed approach is that it is robust with respect to environmental disturbances. This is realized by correcting the monitoring signals for environmental disturbances before proceeding with the fault diagnosis. For this, we again use the spatial dependencies, i.e. the correlations between the monitoring signals of the different system components.

Environmental disturbances generally affect all components in a close neighborhood (independent of the system structure) in a similar way. Therefore, if we observe a particular faulty behavior in all nearby components (even in components that are not connected from the system point of view), we can attribute the common part of the faulty behavior to environmental disturbances. So, besides for the diagnosis itself, the spatial dependencies are useful to identify the environmental disturbances.

The contribution of environmental disturbances to M_i, denoted as $M_{\text{env},i}$, can be determined from the monitoring signals of the components in the immediate neighborhood $\mathcal{N}_i$ of component i, assuming that a sufficient number of the components in $\mathcal{N}_i$ is healthy (apart from environmental disturbances). In this thesis, the following basic assumption is adopted:

Basic assumption $\mathbf{A}_1$: In each local neighborhood $\mathcal{N}_i$, the number of healthy components is sufficient to determine $M_{\text{env},i}$.

The optimal size of the neighborhood set $\mathcal{N}_i$ needed to appropriately determine $M_{\text{env},i}$ is application-specific and dependent on the specific environmental disturbances. However, two factors play a role in general:

1. The behavior of the components in neighborhood $\mathcal{N}_i$ should be representative for the behavior of component i. In general, the closer a component is located to component i, the more representative its behavior is. So this requirement asks for a small neighborhood.

2. The diagnostic result should be insensitive to possible faults in the considered nearby components. In general, the more components are considered, the less sensitive the diagnostic result is to possible faults. So, this requirement asks for a large neighborhood.

For each diagnosis task, a trade-off between these two requirements thus needs to be made. Optionally, additional information, e.g. weather reports, can be taken into account to determine the environmental disturbances. Next, based on $M_{\text{env},i}$, monitoring signal M_i is corrected for environmental disturbances. The corrected monitoring signals M'_j for $j \in \mathcal{N}_i \cup \{i\}$ are then used for the diagnosis of component i.

3.2.4. DIAGNOSTIC PROCEDURE

Procedure 1 outlines the proposed approach for online fault diagnosis in interconnected systems. In Procedure 1, the argument τ is used to denote time. In $\mathcal{G}_i(\tau)$ (step 3 of Procedure 1) we collect the behavior corresponding to different objects passing through component i. By analyzing $\mathcal{G}_i(\tau)$, the spatio-temporal dependencies $G_i(\tau)$ can be determined, i.e. it can be inferred whether the problem is object-specific (if the problem is also observed for other components on the path) or component-specific (if the problem is observed independently of the object). The function corr$(\cdot)$ in step 6 corrects the monitoring signal $M_i(\tau)$ for the effect of environmental disturbances, $M_{\text{env},i}(\tau)$, as determined in step 5. An example of how to determine and correct for environmental disturbances can be found in Section 3.4.

Procedure 1 Diagnosis approach at time τ

Input: Graph of the system, neighborhood $\mathcal{N}_i$ and monitoring signal vector M_i for each system component $i = 1,\dots,n$, path $\mathcal{P}_o$ of all objects o passing through the system, time window length δ_w

1: **for** $i = 1,\dots,n$ **do**
  {Selection of components relevant to determine the spatio-temporal dependencies}
2:   **for all** objects o passing through i in $[\tau - \delta_\mathrm{w}, \tau]$ **do**
3:     Add local path $P_{o,i}$ to $\mathcal{G}_i(\tau)$, with $P_{o,i}$ containing all components $j \in \mathcal{N}_i \cap \mathcal{P}_o$
4:   **end for**
  {Determination of and correction for environmental disturbances}
5:   Determine $M_{\mathrm{env},i}(\tau)$ using $M_j(\tau - \delta_\mathrm{w}),\dots,M_j(\tau)$ $\forall j \in \mathcal{N}_i \cup \{i\}$
6:   Correct monitoring signal $M_i(\tau)$ for environmental disturbances:

$$M'_i(\tau) = \mathrm{corr}\big(M_i(\tau), M_{\mathrm{env},i}(\tau)\big)$$

7: **end for**
8: **for** $i = 1,\dots,n$ **do**
  {Feature extraction, fault detection, and diagnosis}
9:   Determine the features $K_i(\tau)$, $T_i(\tau)$, $S_i(\tau)$, and $G_i(\tau)$ in all operating states using $M'_j(\tau)$ for all $j \in \mathcal{N}_i \cup \{i\}$, the graph structure, and the spatio-temporal system knowledge collected in $\mathcal{G}_i(\tau)$
10:   Use (3.1)-(3.3) to determine $W_{K,i}(\tau)$, $W_{T,i}(\tau)$, $W_{S,i}(\tau)$, and $W_{G,i}(\tau)$
11:   $W_i(\tau) = W_{K,i}(\tau) \cap W_{T,i}(\tau) \cap W_{S,i}(\tau) \cap W_{G,i}(\tau)$
12: **end for**

Output: Set $W_i(\tau)$ of possible faults at the current time τ for all components $i = 1,\dots,n$.

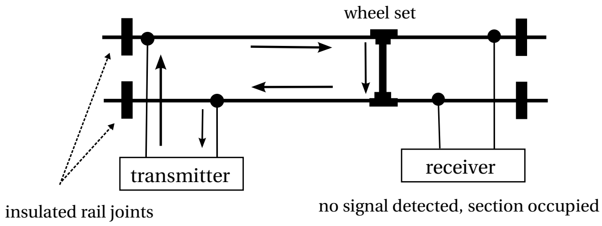
3.3. FAULT DIAGNOSIS OF RAILWAY TRACK CIRCUITS

In this section, we demonstrate the diagnosis approach introduced in Section 3.2 through a case concerning fault diagnosis of railway track circuits.

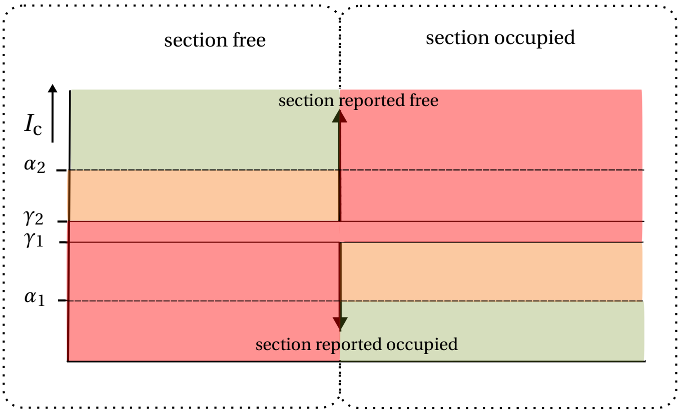
3.3.1. TRACK CIRCUIT WORKING PRINCIPLE

To guarantee the safe operation of a railway network, track circuits are used to detect the absence of a train in a section of railway track. Trains are only allowed to enter sections that are reported free. The track circuit uses the rails as conductors that connect a transmitter at one end of the section to a receiver at the other end. When no train is present in the section, the current will activate a relay in the receiver, which indicates that the section is free. When a train enters the section, the wheels and axles of the train short the circuit (see Figure 3.3). Consequently, the current through the receiver drops, the relay de-energizes, and the section is reported as occupied.

More specifically, the working principle can be described as follows (see Figure 3.4): Under healthy conditions, the current is above a certain threshold α_2 when the section is free and below a threshold α_1 when the section is occupied by a train. The track circuit



Figuur 3.3: Flow of current in a track circuit.



Figuur 3.4: Working principle of a railway track circuit.

is tuned such that even in the case of small current deviations, the presence and absence of a train are correctly reported, i.e.:

if $I_{c,i} > \gamma_2$ **then** section i is reported as free,

if $I_{c,i} < \gamma_1$ **then** section i is reported as occupied,

with[2] $\alpha_2 > \gamma_2 > \gamma_1 > \alpha_1$. So, α_1 and α_2 serve to define system health, whereas γ_1 and γ_2 are settings of the train detection system. For a free section, this means: When $I_{c,i} > \alpha_2$, the track circuit in section i is healthy and is correctly reported as free. When $I_{c,i} < \alpha_2$, the current is too low. However, when $\gamma_2 < I_{c,i} < \alpha_2$ section i is still correctly reported as free and the corresponding system behavior is classified as faulty. Only when $I_{c,i} < \gamma_2$, this fault may result in a *false positive* (FP) train detection result. In this case, we no longer talk about a fault, but about a failure. In the same way, for an occupied section i, it holds that when $I_{c,i} < \alpha_1$, the track circuit is healthy; when $\alpha_1 < I_{c,i} < \gamma_1$, circuit i is faulty (no train detection error); and when $I_{c,i} > \gamma_1$, the circuit fails, i.e. we may have a *false negative* (FN) train detection result.

The proper functioning of a track circuit can be hampered by different fault causes. An overview of the considered track circuit faults can be found in Table 3.1. This table is compiled ourselves based on system manuals, interviews with experts, and available track circuit monitoring data. More details on railway track circuits and the associated faults can be found in Appendix B.

3

3.3.2. DIAGNOSIS SETUP

According to the approach proposed in Section 3.2, a track circuit (i.e. section) can be considered as a system component for which the state $X_{o,i}$ can take two possible values:

$x_{o,1}$: free section;

$x_{o,2}$: occupied section.

Furthermore, $M_i \equiv I_{c,i}$, with $I_{c,i}$ the current measured at the receiver of section i.

For the track circuit case, we only focus on faults and not on failures. This means that the actual operating state $X_{o,i}$ can be inferred from $I_{c,i}$, so $X_{o,i}$ is known at each time, i.e. we assume that basic assumption A_0 (see Section 3.2) is satisfied. Please note that we do not assume that the detection system cannot be broken. We assume that fault diagnosis is preceded by a failure detection mechanism. Failure detection can e.g. be done using redundant measurement equipment, based on train schedules, or by verifying the spatio-temporal dependencies in the railway network.

The following additional assumptions are adopted:

Assumption A_2: Ballast variations (f_{bv}) are caused by environmental disturbances, which are present in all sections;

Assumption A_3: At most one of the faults f_{ij}, f_{co}, f_{rd}, f_{ed}, f_{bd} and one of the faults f_{rc}, f_{lt} are simultaneously present in a section;

[2]When the current is between γ_1 and γ_2, the detection result is not uniquely defined by $I_{c,i}$ and depends on the previous level of the current. In general, $\alpha_1, \alpha_2, \gamma_1$, and γ_2 may vary for different sections.

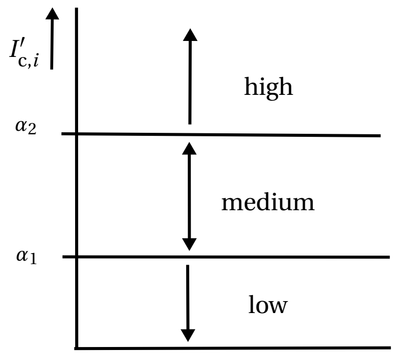
Tabel 3.1: Fault characteristics for track circuits. The relationships between features and faults are defined based on system knowledge and information extracted from historical data. FN = False Negative, FP = False Positive, L = Linear, E = Exponential, A = Abrupt, I = Intermittent, NC = No Correlation, CCS = Correlation with Connected Sections, CAS = Correlation with All Sections, TS = Train-Specific, NTS = Not Train-Specific.

(a) fault characteristics

health (H)	problem	cause	potential error
h	-	healthy state	-
f_{rc}	train shunt imperfection	rail contamination	FN
f_{lt}		lightweight trains	FN
f_{ij}	insulation imperfection	insulated joint defect	FP
f_{co}		conductive objects	FP
f_{rd}	rail conductance impairment	mechanical rail defect	FP
f_{ed}		electrical disturbances	FP
f_{bd}	ballast condition	ballast degradation	FP
f_{bv}		ballast variation	FP

(b) features

intra-component (K)		temporal (T)	spatial (S)	spatio-temporal (G)
section free	section occupied	section occupied	section occupied	
high	low	-	-	-
high	medium or high	-	-	NTS
high	medium or high	-	-	TS
low or medium	low	L or E	NC	NTS
low or medium	low	A	NC	NTS
low or medium	low	E	NC	NTS
low or medium	low	I or A	CCS	TS or NTS
low or medium	low	L or E	CAS or CCS	NTS
low or medium or high	low	A or L or E or I	CAS	NTS



Figuur 3.5: Definition of the feature values of K_i for the track circuit case.

Assumption A$_4$: We have a closed world, i.e. Table 3.1 is complete.

So, ballast variations are considered as environmental disturbances and our aim is to detect and diagnose other faults (f_{rc}, f_{lt}, f_{ij}, f_{co}, f_{rd}, f_{ed}, f_{bd}) in the presence of this natural variation.

3

Since some faults are allowed to be simultaneously present (e.g. faults f_{rc} and f_{co}), we consider two health variables $H_{t,i} \in \{h_t, f_{rc}, f_{lt}\}$ and $H_{f,i} \in \{h_f, f_{ij}, f_{co}, f_{rd}, f_{ed}, f_{bd}\}$. The overall system health H_i equals $(H_{t,i}, H_{f,i})$, with component i being healthy (i.e. $H_i = h$) if $H_{t,i} = h_t$ and $H_{f,i} = h_f$. The set of values $H_{t,i}$ can take given feature C_k is denoted as $W_{t,C_k,i}$ and the set of values $H_{f,i}$ can take given feature C_k is denoted as $W_{f,C_k,i}$. The associated sets $W_{t,i}$ and $W_{f,i}$ (see Section 3.2.2) can be computed according to Procedure 1.

3.3.3. FEATURE EXTRACTION

In this section, we show that component knowledge in combination with the only available monitoring signal $I_{c,i}$ is not sufficient to adequately distinguish between the faults listed in Table 3.1. To improve diagnostic performance, we consider the other diagnostic features proposed in Section 3.2.2, i.e. temporal dependencies, spatial dependencies, and spatio-temporal dependencies.

INTRA-COMPONENT DEPENDENCIES

The actual system knowledge of section i is represented in the form of a single-input single-output system with as (unknown and uncontrollable) input the voltage across the two rails $V_{rail,i}$ and as measured output the current $I_{c,i}$ measured at the receiver. Based on our system knowledge (see Section 3.3.1), we define the feature K_i as the qualitative behavior of $I'_{c,i}$, where K_i takes values in the set {low, medium, high} (see Figure 3.5), with:

low: $I'_{c,i} < \alpha_1$;

medium: $\alpha_1 \le I'_{c,i} \le \alpha_2$;

high: $I'_{c,i} > \alpha_2$.

Note that a finer distinction in current values can be made by using both the thresholds α_1 and α_2 and the thresholds γ_1 and γ_2 (see Figure 3.4). However, since for the purpose of fault diagnosis, a finer discretization does not add additional information, K_i is defined as a three-valued feature. In Table 3.1(b) the value of K_i is given for each of the considered faults. Based on K_i only, it not possible to distinguish between the different faults. Indeed, it is observed that different types of faults have a similar effect on $I'_{c,i}$ (e.g. both ballast degradation, rail conductance impairment, and insulation imperfection cause that $I'_{c,i}$ drops below α_2. Given the system state $X_{o,i}$ (occupied or free), the value of K_i only tells us whether section i is healthy or not, i.e. we can detect faults, but we cannot diagnose them. The system knowledge (see Table 3.1) is represented by the following set of rules:

if K_i = high for $X_{o,i}$ = free **then** $W_{f,K,i} = \{h_f\}$

if $K_i \neq$ high for $X_{o,i}$ = free **then** $W_{f,K,i} = \{f_{ij}, f_{co}, f_{rd}, f_{ed}, f_{bd}, f_{bv}\}$

if K_i = low for $X_{o,i}$ = occupied **then** $W_{t,K,i} = \{h_t\}$

if $K_i \neq$ low for $X_{o,i}$ = occupied **then** $W_{t,K,i} = \{f_{rc}, f_{lt}\}$

3

TEMPORAL DEPENDENCIES

In Table 3.1, a characterization of the time evolution T_i of $I'_{c,i}$ as a consequence of each fault in section i is given. Hereby we have restricted ourselves to the four types of temporal behavior T_i shown in Figure 3.6, namely:

A: abrupt;

L: linear;

E: exponential;

I: intermittent.

The results are included in Table 3.1(b). Note that the temporal dependency T_i is only a relevant feature for the diagnosis of a free section, i.e. $X_{o,i}$ = free. Furthermore, all behaviors that are possible according to the available knowledge are listed[3].

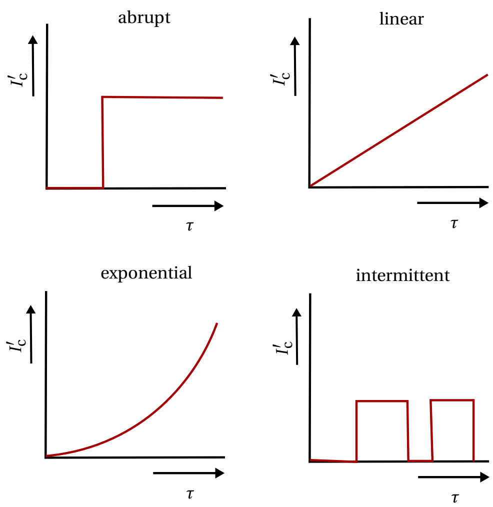
Taking this additional information into account, our knowledge base can be extended with the following rules:

if T_i = L for $X_{o,i}$ = free **then** $W_{f,T,i} = \{h_f, f_{ij}, f_{bd}, f_{bv}\}$

if T_i = E for $X_{o,i}$ = free **then** $W_{f,T,i} = \{h_f, f_{ij}, f_{rd}, f_{bd}, f_{bv}\}$

if T_i = A for $X_{o,i}$ = free **then** $W_{f,T,i} = \{h_f, f_{co}, f_{ed}, f_{bv}\}$

[3] We did not choose one particular type of behavior T if the knowledge to do so was lacking, i.e. for an insulated joint defect, both T = L and T = E are assumed to be possible.



Figuur 3.6: Temporal behaviors.

if T_i = I for $X_{o,i}$ = free **then** $W_{f,T,i} = \{h_f, f_{ed}, f_{bv}\}$

$W_{t,T,i} = \{h_t, f_{rc}, f_{lt}\}$

3

The last rule states that $W_{t,T,i}$ is always set to $\{h_t, f_{rc}, f_{lt}\}$, expressing that the temporal dependencies do not contain information to distinguish between h_t, f_{rc}, and f_{lt}.

SPATIAL DEPENDENCIES

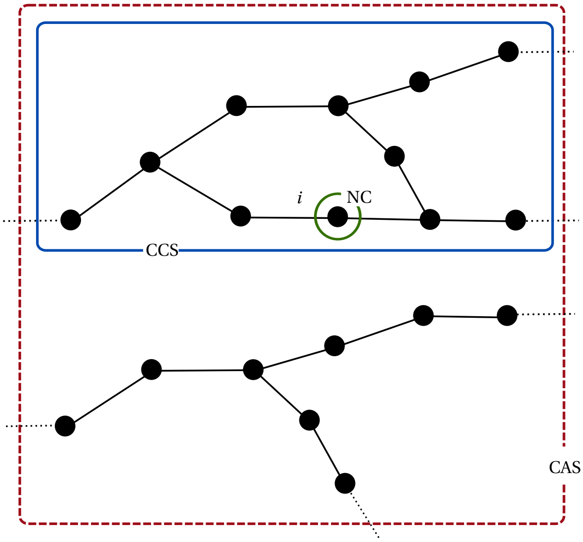
For each section i, only one monitoring signal ($I_{c,i}$) is available. However, this signal is measured for all sections in the network. Additional information is contained in these data thanks to the correlations between the monitoring signals of neighboring sections that vary for different types of faults. Some faults are likely to influence all sections in a small neighborhood (e.g. ballast variation), other faults only influence sections of the same track (e.g. electrical disturbances), while still other faults are specific to one section (e.g. mechanical rail defects). So, the presence of a fault introduces dependencies between (some of) the sections in a local neighborhood. These dependencies introduce correlations between the monitoring signals of the different sections, which can be used for fault diagnosis. An overview of the correlations introduced by the different faults can be found in Table 3.1(b), where the correlations are defined as:

NC: no correlation of $I'_{c,i}$ with the monitoring signals of other sections;

CCS: correlation of $I'_{c,i}$ with the monitoring signals of connected sections, i.e. sections on the same track;

CAS: correlation of $I'_{c,i}$ with the monitoring signals of all nearby sections.

Figure 3.7 gives a graphical overview of the affected sections corresponding to the spatial dependencies NC, CCS, and CAS.



Figuur 3.7: Division of the sections (track circuits) in a railway network for the fault diagnosis of section i.

3

Accordingly we extend our knowledge base with the following set of rules:

if S_i = NC for $X_{o,i}$ = free **then** $W_{f,S,i} = \{h_f, f_{ij}, f_{co}, f_{rd}, f_{bd}\}$

if S_i = CCS for $X_{o,i}$ = free **then** $W_{f,S,i} = \{h_f, f_{ed}, f_{bd}\}$

if S_i = CAS for $X_{o,i}$ = free **then** $W_{f,S,i} = \{h_f, f_{bv}\}$

$W_{t,S,i} = \{h_t, f_{rc}, f_{lt}\}$

SPATIO-TEMPORAL DEPENDENCIES

A distinction can be made between faults that are caused by a train (e.g. train shunt imperfection due to a lightweight train) and faults that are related to the section itself (e.g. rail contamination). Therefore, we make a distinction between two types of faulty spatio-temporal behavior G:

TS: train-specific faulty behavior;

NTS: faulty behavior that is not train related.

Taking this information into account, the following rules can be added to our knowledge base:

if G_i = NTS for $X_{o,i}$ = occupied **then** $W_{t,G,i} = \{h_t, f_{rc}\}$

if G_i = TS for $X_{o,i}$ = occupied **then** $W_{t,G,i} = \{h_t, f_{lt}\}$

if G_i = NTS for $X_{o,i}$ = free **then** $W_{f,G,i} = \{h_f, f_{ij}, f_{co}, f_{rd}, f_{ed}, f_{bd}, f_{bv}\}$

if G_i = TS for $X_{o,i}$ = free **then** $W_{f,G,i} = \{h_f, f_{ed}\}$

CONCLUDING REMARKS

Considering Table 3.1, we conclude that the temporal, spatial, and spatio-temporal dependencies are valuable diagnostic features for fault diagnosis of railway track circuits. Without these features, faults can only be detected. Including these features allows possible fault causes to be identified.

3.3.4. DIAGNOSIS APPROACH

The diagnosis of section i in a monitored track circuit network consists of the following tasks:

1. Select the sections that are relevant for the diagnosis, i.e. determine neighborhood $\mathcal{N}_i$.
2. Infer the system state $X_{o,i}$ from $I_{c,i}$:

   if $I_{c,i} > \gamma_2$ **then** $X_{o,i}$ = free
   if $I_{c,i} < \gamma_1$ **then** $X_{o,i}$ = occupied

3. Determine current fluctuations due to environmental disturbances (ballast variations) based on the behavior of the sections selected in step 1 (step 5 of Procedure 1).
4. If $X_{o,i}$ = free, correct the currents $I_{c,j}$ for all $j \in \mathcal{N}_i \cup \{i\}$ for ballast variations (step 6 of Procedure 1)
5. Check for faulty behavior:

   if $K_i \neq$ "high" for $X_{o,i}$ = free **or** $K_i \neq$ "low" for $X_{o,i}$ = occupied **then** $H_i \neq h$

6. If a fault is detected, determine the spatial dependencies S_i, the temporal dependencies T_i, and the spatio-temporal dependencies G_i and diagnose section i (steps 9 – 11 of Procedure 1).

Below, the determination of the ballast variation over time (tasks 3 and 4) and the fault detection and diagnosis (tasks 5 and 6) are worked out for both free and occupied sections.

DETERMINATION OF THE BALLAST VARIATION OVER TIME

The current fluctuations $I_{bal,i}$ due to ballast variation of section i are determined based on the behavior of healthy sections in $\mathcal{N}_i$. One possible way to compute $I_{bal,i}$ is by taking a filtered (weighted) average of the current fluctuations of the considered sections:

$$I_{bal,i}(\tau) = \text{filter}\left(\sum_{j \in \mathcal{K}_i} \frac{I_{c,j}(\tau) - \bar{I}_{c,j}(\tau)}{|\mathcal{K}_i(\tau)|}\right) \tag{3.5}$$

with $\mathcal{K}_i \subseteq \mathcal{N}_i$ the set of sections in a close neighborhood of section i that are expected to be healthy, i.e. that are not known to be faulty[4], and $\bar{I}_{\mathrm{c},j}$ the nominal value (i.e. long-term average) of $I_{\mathrm{c},j}$. Note that for the determination of these variations, only the measurements corresponding to a free track are considered. When a train is present in the section, the track circuit is generally short-circuited and the current measured at the receiver is approximately zero, independent of the ballast condition.

FAULT DETECTION AND DIAGNOSIS OF A FREE SECTION

The current measurements corresponding to a free section are first corrected for ballast variation based on the previously determined behavior of $I_{\mathrm{bal},i}(\tau)$. The corrected current measurements $I'_{\mathrm{c},i}$ can e.g. be defined as:

$$I'_{\mathrm{c},i}(\tau) = I_{\mathrm{c},i}(\tau) - I_{\mathrm{bal},i}(\tau) \tag{3.6}$$

The corrected current signals $I'_{\mathrm{c},i}$ are then used for the fault diagnosis of section i. When a fault is detected in section i (i.e. $I'_{\mathrm{c},i} < \alpha_2$), the corresponding temporal (T_i), spatial (S_i), and spatio-temporal (G_i) dependencies are determined. Based on T_i, S_i, and G_i the cause (or a set of possible causes) for the faulty behavior can be inferred from Table 3.1.

3

FAULT DETECTION AND DIAGNOSIS OF AN OCCUPIED SECTION

When a section is occupied, ballast variations play no significant role, so we can directly proceed with the detection of faulty behavior. When a fault is detected, i.e. $I'_{\mathrm{c},i} > \alpha_1$, diagnosis is required. Then, it is verified whether the problem is train-specific or not. For this purpose, the monitoring signals of sections lying on the train routes of several passing trains are analyzed. If the problem is train-specific, the faulty behavior is caused by a lightweight train and not due to rail contamination (i.e. fault f_{lt} is present and fault f_{rc} is absent). If the problem is not train-specific, rail contamination (among others) causes the faulty behavior, i.e. fault f_{rc} is present. When rail contamination is present, problems with lightweight trains are no longer guaranteed to be identified in section i. However, defective trains will be detected in any other section on the train path without rail contamination.

3.4. ILLUSTRATIVE EXAMPLE I: WITHOUT UNCERTAINTY

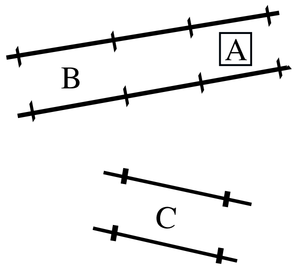
We consider the fault diagnosis of a railway section in a small network. First, we introduce the diagnosis setup together with the adopted assumptions. Next, we consider how to determine and correct for ballast variations and finally, the fault detection and diagnosis is performed.

3.4.1. PROBLEM FORMULATION

Consider that we aim to diagnose section A for which we have the monitoring signals[5] $I_{\mathrm{c,A}}$, $I_{\mathrm{c,B}}$, and $I_{\mathrm{c,C}}$ of the sections A, B, and C as depicted in Figure 3.8 available, with:

[4] Sections that are diagnosed to be faulty, but are still not repaired are excluded from $\mathcal{K}_i$.

[5] The data used in this case study have been provided by Inspectation, VolkerRail.



Figuur 3.8: Sections considered in the diagnosis example.

A: the section to be diagnosed;

B: a nearby preceding section;

C: a nearby section located on another track.

So for this example, we have:

$$\mathcal{N}_A = \{B, C\}$$

3

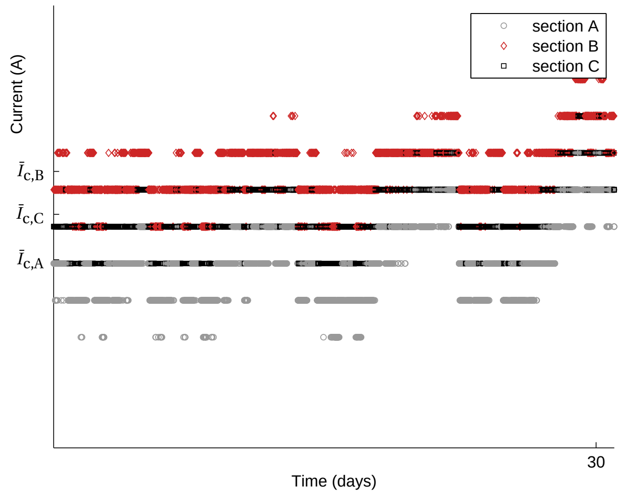
with section B connected to section A, and section C not connected to section A. Furthermore, for this example Basic assumption A_1 is specified as:

Assumption A_1': Section B and C do not suffer from section-specific faults (i.e. faults for which S = NC) and section C does not suffer from track-specific faults (i.e. faults for which S = CCS).

Assumption A_1' is adopted here because (for simplicity) only two neighboring sections are considered. In the case that more sections are considered, the redundant information contained in these signals can be used to detect (and correct for) possible faults in neighboring sections.

3.4.2. DETERMINATION OF AND CORRECTION FOR BALLAST VARIATION

To determine the part of the current signal $I_{c,A}$ that can be attributed to ballast variations, we consider the behavior of the signals $I_{c,A}$, $I_{c,B}$, and $I_{c,C}$ for a free track as shown in Figure 3.9. From this figure, it can be observed that $I_{c,A}, I_{c,B}$, and $I_{c,C}$ exhibit a similar type of variation over time. However, there is also a systematic difference between the current values of the three sections, e.g. the current level at the receiver of section C is generally higher than the current level at the receiver of section A. Furthermore, it can be observed that the measurements are disturbed by noise and quantization. To determine the current fluctuation due to ballast variation $I_{bal,A}$, we first normalize the current signals by subtracting their nominal (i.e. mean) value from the measurements (see (3.5)). In Figure 3.10, the normalized current signals are given. As we know that section C is healthy apart from ballast variation, the resulting signal can basically be attributed to ballast variations. To reduce the effect of the noise and quantization, we first fit a twelfth-degree polynomial model[6] through the data (filter operation in (3.5)) and use the resulting model $I_{\mathrm{bal,A}}$ (black solid line in Figure 3.10) to correct the current signals for the effect of ballast variation. Since only this local neighborhood of three sections is available, we assume $I_{\mathrm{bal,A}} = I_{\mathrm{bal,B}} = I_{\mathrm{bal,C}}$. The corrected current signals are given in Figure 3.11. The remaining variation can mainly be attributed to noise and quantization.



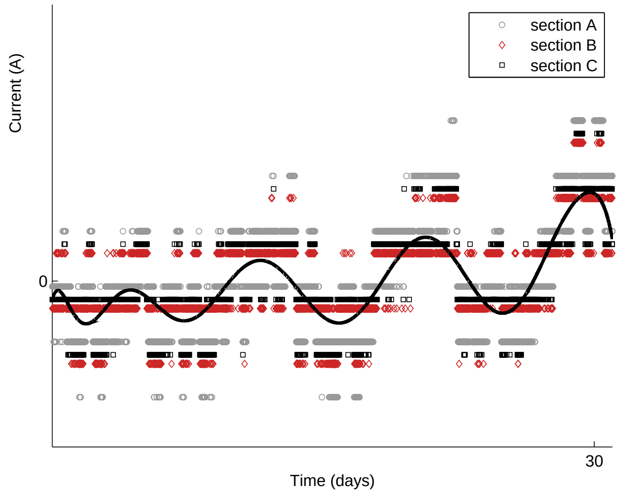
Figuur 3.9: Current measurements of sections A, B, and C when the section is free.

3.4.3. FEATURE EXTRACTION, FAULT DETECTION, AND DIAGNOSIS

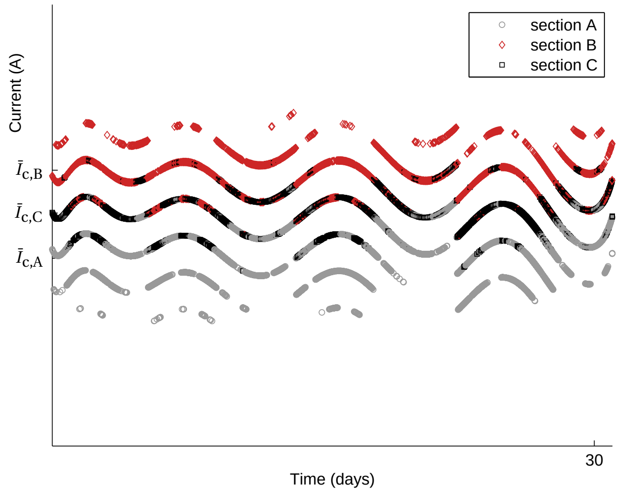
For the fault diagnosis, we focus on a time interval that includes several train passages. The associated corrected monitoring signals are shown in Figure 3.12, with the gray areas indicating the time intervals during which the section is occupied by a train. As expected, after correction for ballast variation (see Section 3.4.2), the free track behavior of section C is as desired; the current $I'_{\mathrm{c,C}}$ is above the threshold α_2 (i.e. K_{C} = "high") when the section is free and below the threshold α_1 (i.e. K_{C} = "low") when the section is occupied. To diagnose section A, first the behavior of K_{A} is analyzed. We conclude that till time $\tau = \tau_1$, K_{A} = "high" when the section is free and K_{A} = "low" when the section is occupied, i.e. the system is healthy (see Section 3.3). At time τ_1, the current level drops as a consequence of a train entering the section, but the current does not decrease below the threshold value α_1 (i.e. $K_{\mathrm{A}} \neq$ "low"), indicating that faults f_{rc} and/or f_{lt} are present (see Table 3.1). To determine which fault is present, feature G_{A} is used[7], i.e. we verify whether the problem is train-specific (see Section 3.3.3). This is done by checking whether the same problem occurred for other train passages. This is not case (G_{A} = TS), indicating

[6] The degree of the polynomial has been tuned manually.

[7] Remember that for an occupied section, it is sufficient to only consider features K_i and G_i. Features T_i and S_i do not put any constraints on the set of possible faults, so $W_{\mathrm{t},K,i} \cap W_{\mathrm{t},G,i} = W_{\mathrm{t},K,i} \cap W_{\mathrm{t},T,i} \cap W_{\mathrm{t},S,i} \cap W_{\mathrm{t},G,i}$.



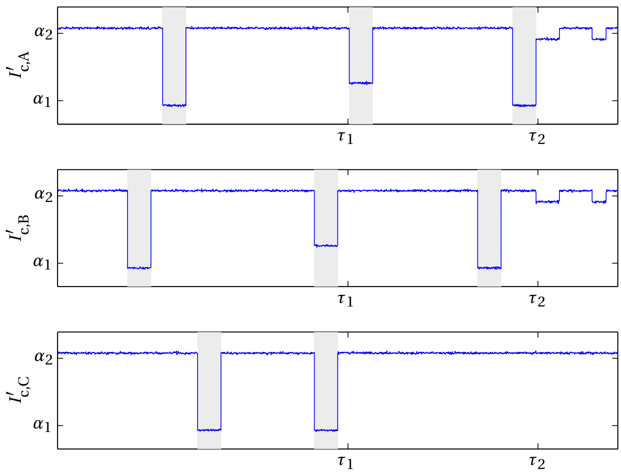
Figuur 3.10: Normalized current measurements of sections A, B, and C together with a polynomial fit (black solid line) through the measurement data of healthy section C.



Figuur 3.11: The current signals corrected for ballast variation.

that the faulty behavior is caused by a train (see Table 3.1). This conclusion is validated by the monitoring signal $I'_{c,B}$ of preceding section B. Also from this monitoring signal, it follows that one particular train suffered from shunt problems.

After the train passage at time $\tau = \tau_1$ the behavior is normal again till $\tau = \tau_2$. From



Figuur 3.12: Monitoring signals of sections A, B, and C.

$\tau = \tau_2$ on, the current sometimes drops below α_2 (i.e. $K_\mathrm{A} \neq$ "high") while the track is free, indicating the presence of one of the faults $f_\mathrm{ij}, f_\mathrm{co}, f_\mathrm{rd}, f_\mathrm{ed}, f_\mathrm{bd}$. To further specify which fault is present, we first consider feature S_A, i.e. we verify whether there is a correlation with neighboring sections. Considering the monitoring signals $I'_\mathrm{c,B}$ and $I'_\mathrm{c,C}$, we observe a similar faulty behavior in section B and healthy behavior in section C, from which we conclude that the disturbance is track-specific, i.e. $S_\mathrm{A} = \mathrm{CCS}$. So far, it can be concluded that $H_\mathrm{t,A} = f_\mathrm{ed}$ or $H_\mathrm{t,A} = f_\mathrm{bd}$. To make a further distinction, the time evolution of $I'_\mathrm{c,A}$ is studied, i.e. we consider feature T_A. Based on the available part of the time signal, we conclude that the time behavior of $I'_\mathrm{c,A}$ is intermittent, i.e. $T_\mathrm{A} = \mathrm{I}$. Then it follows that $H_\mathrm{t,A} = f_\mathrm{ed}$.

In summary, from the signals in Figure 3.12, we conclude that around $\tau = \tau_1$ a "defective" train passes through sections A and B and after $\tau = \tau_2$, sections A and B suffer from electrical disturbances.

3.5. ILLUSTRATIVE EXAMPLE II: WITH UNCERTAINTY

In this example we account for uncertainty and evaluate the techniques presented in Chapter 2 on a track circuit diagnosis task.

3.5.1. PROBLEM FORMULATION

Track circuits only work properly if the conductance properties of the rails are high. Two main causes have been identified that negatively influence rail conductance (see Table 3.1), namely:

1. mechanical rail defects f_rd;

2. electrical disturbances f_{ed}.

The goal is to determine, based on the temporal and spatial dependencies, which fault (f_{rd} or f_{ed}) is present. It is assumed that we already know that the section suffers from a conductance problem.

The Bayesian and D-S graphical representations of the diagnosis problem are given at the top of Table 3.2. Since only single-fault scenarios are allowed, we use one fault variable H with $\Theta_H = \{f_{\text{rd}}, f_{\text{ed}}\}$. Quantitatively, fault variable H is linked to the features S and T as follows:

k_1 : **if** $H = f_{\text{rd}}$ **then** $P(T = \text{E}) = 0.85$

k_2 : **if** $H = f_{\text{ed}}$ **then** $P(T = \text{A} \vee T = \text{I}) = 1$

k_3 : **if** $H = f_{\text{rd}}$ **then** $P(S = \text{NC}) = 1$

k_4 : **if** $H = f_{\text{ed}}$ **then** $P(S = \text{CSS}) = 0.7$

These rules encode that a rail defect f_{rd} likely evolves exponentially over time, whereas an electrical disturbance is characterized by an intermittent or abrupt time behavior. A rail defect only influences the behavior of one particular section, while electrical disturbances likely influence the behavior of sections on the same track (i.e. connected sections). This system knowledge is conditional, uncertain, and incomplete (see Chapter 2).

3

We assume that no prior knowledge about the relative occurrence of the two faults is available and that the following uncertain pieces of evidence are available for diagnosis:

e_1: $P(T = \text{I}) = 0.3, P(T \neq \text{I}) = 0.7$

e_2: $P(T = \text{A} \vee T = \text{I}) = 1$

e_3: $P(S = \text{CCS}) = 0.8$

Evidence e_1 provides information about the temporal dependencies, but can only distinguish between intermittent and non-intermittent behavior. The second evidence indicates that the temporal behavior is not gradual, i.e. not linear or exponential, but cannot discriminate between intermittent and abrupt behavior. Evidence e_3 corresponds to an unreliability information source providing that $S = \text{CCS}$.

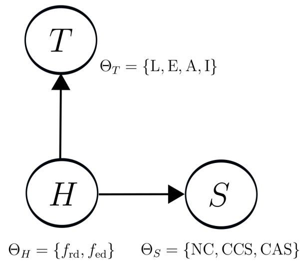
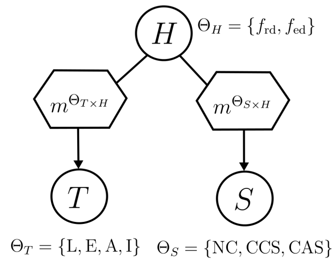
3.5.2. BAYESIAN SOLUTION

INFORMATION PRE-PROCESSING

As indicated in Section 2.6.1, fault diagnosis using Bayesian networks requires some preprocessing steps.

Transformation of the knowledge base The knowledge specified by the rules k_1 till k_4 needs to be represented by two conditional probability tables, one for T and one for S. The knowledge is already in conditional form, so we only have to represent the incomplete knowledge by probabilities. This is done based on the additivity axiom and the principle of maximum entropy. The obtained probability tables are included in Table 3.2.

Tabel 3.2: Summary of the diagnosis example

	Bayesian	Dempster-Shafer
Graph		
Knowledge $T \times H$	T: H | L | E | A | I f_{rd} | 0.05 | 0.85 | 0.05 | 0.05 f_{ed} | 0 | 0 | 0.5 | 0.5	$m^{\Theta_{T\times H}}(\{(E, f_{\mathrm{rd}}), (A, f_{\mathrm{ed}}), (I, f_{\mathrm{ed}})\}) = 0.85$ $m^{\Theta_{T\times H}}(\{(\cdot, f_{\mathrm{rd}}), (A, f_{\mathrm{ed}}), (I, f_{\mathrm{ed}})\}) = 0.15$
Knowledge $S \times H$	S: H | NC | CCS | CAS f_{rd} | 1 | 0 | 0 f_{ed} | 0.15 | 0.7 | 0.15	$m^{\Theta_{S\times H}}(\{(\mathrm{NC}, f_{\mathrm{rd}}), (\mathrm{CCS}, f_{\mathrm{ed}})\}) = 0.7$ $m^{\Theta_{S\times H}}(\{(\mathrm{NC}, f_{\mathrm{rd}}), (\cdot, f_{\mathrm{ed}})\}) = 0.3$
Prior knowledge H	H: f_{rd} | f_{ed} 0.5 | 0.5	$m^{\Theta_H}(\Theta_H) = 1$
Temporal dependencies	T_{obs}: T | $e_1 \wedge e_2$ L | 0 E | 0 A | 7 I | 9	$m^{\Theta_T}(I) = 0.3$ $m^{\Theta_T}(A) = 0.7$
Spatial dependencies	S_{obs}: S | e_3 NC | 1 CCS | 8 CAS | 1	$m^{\Theta_S}(\mathrm{CCS}) = 0.8$ $m^{\Theta_S}(\Theta_S) = 0.2$
Diagnostic result	$P(f_{\mathrm{rd}}) = 0.0167$ $P(f_{\mathrm{ed}}) = 0.9833$	$m^{\Theta_H}(f_{\mathrm{ed}}) = 0.97$ $m^{\Theta_H}(\Theta_F) = 0.03$

Prior probability distribution For the root node H, a prior probability distribution is needed. As we have no prior knowledge regarding the relative occurrence of the two faults, we adopt a uniform prior distribution (principle of maximum entropy).

Temporal evidences Both evidence e_1 and evidence e_2 relate to the temporal dependencies T. In the Bayesian model, e_1 is represented by the following likelihood ratios (principle of maximum entropy):

$$P(e_1|\mathrm{L}) : P(e_1|\mathrm{E}) : P(e_1|\mathrm{A}) : P(e_1|\mathrm{I}) = 0.23 : 0.23 : 0.23 : 0.3 \\ = 7 : 7 : 7 : 9 \tag{3.7}$$

Conditioning this information based on e_2, yields:

$$P(e_1, e_2|\mathrm{L}) : P(e_1, e_2|\mathrm{E}) : P(e_1, e_2|\mathrm{A}) : P(e_1, e_2|\mathrm{I}) = 0 : 0 : 7 : 9 \tag{3.8}$$

These ratios are reflected in the conditional probability of the virtual node T_{obs} (see Table 3.2).

Spatial evidence Evidence e_3 is related to the spatial dependencies. Following the additivity axiom and the principle of maximum entropy, evidence e_3 is represented by the following likelihood ratios:

$$P(e_3|\mathrm{NC}) : P(e_3|\mathrm{CCS}) : P(e_3|\mathrm{CAS}) = 1 : 8 : 1 \tag{3.9}$$

which are reflected in the conditional probability table of the virtual node S_{obs} (see Table 3.2).

3

FAULT DIAGNOSIS

To obtain the posterior probability distribution of H, we propagate the hard evidences on the virtual events T_{obs} and S_{obs} through the augmented Bayesian network. Updating (4.7) with $T_{\mathrm{obs}} = e_1 \wedge e_2$ yields:

$$P(H = f_{\mathrm{rd}}|e_1, e_2) = 0.0909 \tag{3.10}$$

$$P(H = f_{\mathrm{ed}}|e_1, e_2) = 0.9091 \tag{3.11}$$

Subsequently updating (4.7) with $S_{\mathrm{obs}} = e_3$ yields:

$$P(H = f_{\mathrm{rd}}|e_1, e_2, e_3) = 0.0167 \tag{3.12}$$

$$P(H = f_{\mathrm{ed}}|e_1, e_2, e_3) = 0.9833 \tag{3.13}$$

We conclude with a probability of slightly more than 98% that electrical disturbances are responsible for the conductance problem.

3.5.3. DEMPSTER-SHAFER SOLUTION

INFORMATION PRE-PROCESSING

As indicated in Section 2.6.2, fault diagnosis using Dempster-Shafer networks requires some pre-processing steps.

Transformation of the knowledge base To convert the conditional knowledge regarding T and S to a mass function $m^{\Theta_{T\times H}}$ on the space $\Theta_T \times \Theta_H$ and a mass function $m^{\Theta_{S\times H}}$ on the space $\Theta_S \times \Theta_H$, we first use the ballooning extension (A.16) to derive two mass functions on both spaces. Next, we use Dempster's rule of combination (A.17) to combine the two mass functions on each space.

On the space $\Theta_T \times \Theta_H$, the ballooning extension (A.16) of rules k_1 and k_2 yields the following two mass functions:

$$\begin{aligned} m^{\Theta_T}[f_{\text{rd}}]^{\Uparrow\Theta_{T\times H}}(\{(\text{E}, f_{\text{rd}}), (\cdot, f_{\text{ed}})\}) &= 0.85 \\ m^{\Theta_T}[f_{\text{rd}}]^{\Uparrow\Theta_{T\times H}}(\Theta_T \times \Theta_H) &= 0.15 \end{aligned} \tag{3.14}$$

$$m^{\Theta_T}[f_{\text{ed}}]^{\Uparrow\Theta_{T\times H}}(\{(\text{A}, f_{\text{ed}}), (\text{I}, f_{\text{ed}}), (\cdot, f_{\text{rd}})\}) = 1 \tag{3.15}$$

Combining them using (A.17) gives:

$$\begin{aligned} m^{\Theta_{T\times H}}(\{(\text{E}, f_{\text{rd}}), (\text{A}, f_{\text{ed}}), (\text{I}, f_{\text{ed}})\}) &= 0.85 \\ m^{\Theta_{T\times H}}(\{(\text{A}, f_{\text{ed}}), (\text{I}, f_{\text{ed}}), (\cdot, f_{\text{rd}})\}) &= 0.15 \end{aligned} \tag{3.16}$$

On the space $\Theta_{S\times H}$, the ballooning extension (A.16) of rules k_3 and k_4 yields the following two mass functions:

$$m^{\Theta_S}[f_{\text{rd}}]^{\Uparrow\Theta_{S\times H}}(\{(\text{NC}, f_{\text{rd}}), (\cdot, f_{\text{ed}})\}) = 1 \tag{3.17}$$

$$\begin{aligned} m^{\Theta_S}[f_{\text{ed}}]^{\Uparrow\Theta_{S\times H}}(\{(\text{CCS}, f_{\text{ed}}), (\cdot, f_{\text{rd}})\}) &= 0.7 \\ m^{\Theta_S}[f_{\text{ed}}]^{\Uparrow\Theta_{S\times H}}(\Theta_S \times \Theta_H) &= 0.3 \end{aligned} \tag{3.18}$$

Combining them using (A.17) gives:

$$\begin{aligned} m^{\Theta_{S\times H}}(\{(\text{CCS}, f_{\text{ed}}), (\text{NC}, f_{\text{rd}})\}) &= 0.7 \\ m^{\Theta_{S\times H}}(\{(\text{NC}, f_{\text{rd}}), (\cdot, f_{\text{ed}})\}) &= 0.3 \end{aligned} \tag{3.19}$$

Temporal evidences In the D-S framework, evidence e_1 is represented by the following mass function:

$$\begin{aligned} m^{\Theta_T}(\text{I}) &= 0.3 \\ m^{\Theta_T}(\text{L} \vee \text{E} \vee \text{A}) &= 0.7 \end{aligned} \tag{3.20}$$

Conditioning this knowledge based on evidence e_2 yields:

$$\begin{aligned} m^{\Theta_T}(\text{I}) &= 0.3 \\ m^{\Theta_T}(\text{A}) &= 0.7 \end{aligned} \tag{3.21}$$

Spatial evidence In the D-S framework, evidence e_3 is represented as:

$$\begin{aligned} m^{\Theta_S}(\text{CCS}) &= 0.8 \\ m^{\Theta_S}(\Theta_S) &= 0.2 \end{aligned} \tag{3.22}$$

FAULT DIAGNOSIS

To infer the fault cause, we first combine the mass functions m^{Θ_T} and m^{Θ_S} with the corresponding valuation functions $m^{\Theta_{T\times H}}$ and $m^{\Theta_{S\times H}}$. So, m^{Θ_T} is combined with $m^{\Theta_{T\times H}}$ and m^{Θ_S} with $m^{\Theta_{S\times H}}$. Next, we project the mass function on Θ_H. To combine two mass functions on different spaces we use the cylindrical extension (A.14). So, we vacuously extend m^{Θ_T} to the space $\Theta_T \times \Theta_H$ and m^{Θ_S} to the space $\Theta_S \times \Theta_H$.

On the space $\Theta_T \times \Theta_H$ the following results are obtained: The cylindrical extension (A.14) of m^{Θ_T} on $\Theta_T \times \Theta_H$ yields:

$$m^{\Theta_T \uparrow \Theta_{T\times H}}\big(\{(I, f_{\text{rd}}), (I, f_{\text{ed}})\}\big) = 0.3 \tag{3.23}$$

$$m^{\Theta_T \uparrow \Theta_{T\times H}}\big(\{(A, f_{\text{rd}}), (A, f_{\text{ed}})\}\big) = 0.7 \tag{3.24}$$

Combining this mass function with the valuation function $m^{\Theta_{T\times H}}$ according to Dempster's rule of combination (A.17) gives:

$$\begin{aligned} m^{\Theta_{T\times H}}(\text{I}, f_{\text{ed}}) &= 0.3 \cdot 0.85 \\ m^{\Theta_{T\times H}}(\text{A}, f_{\text{ed}}) &= 0.7 \cdot 0.85 \\ m^{\Theta_{T\times H}}(\{(\text{I}, f_{\text{rd}}), (\text{I}, f_{\text{ed}})\}) &= 0.3 \cdot 0.15 \\ m^{\Theta_{T\times H}}(\{(\text{A}, f_{\text{rd}}), (\text{A}, f_{\text{ed}})\}) &= 0.7 \cdot 0.15 \end{aligned} \tag{3.25}$$

Marginalization of $m^{\Theta_{T\times H}}$ on Θ_H according to (A.15) gives:

$$\begin{aligned} m^{\Theta_{T\times H} \downarrow \Theta_H}(f_{\text{ed}}) &= 0.3 \cdot 0.85 + 0.7 \cdot 0.85 = 0.85 \\ m^{\Theta_{T\times H} \downarrow \Theta_H}(\Theta_H) &= 0.3 \cdot 0.15 + 0.7 \cdot 0.15 = 0.15 \end{aligned} \tag{3.26}$$

On the space $\Theta_S \times \Theta_H$, the following results are obtained: The cylindrical extension (A.14) of Θ_S yields:

$$\begin{aligned} m^{\Theta_S \uparrow \Theta_{S\times H}}\big(\{(\text{CCS}, f_{\text{rd}}), (\text{CCS}, f_{\text{ed}})\}\big) &= 0.8 \\ m^{\Theta_S \uparrow \Theta_{S\times H}}\big(\Theta_S \times \Theta_H\big) &= 0.2 \end{aligned} \tag{3.27}$$

Combining (3.27) with the valuation function $m^{\Theta_{S\times H}}$ according to (A.17) gives:

$$\begin{aligned} m^{\Theta_{S\times H}}(\text{CCS}, f_{\text{ed}}) &= 0.7 \cdot 0.8 + 0.3 \cdot 0.8 \\ m^{\Theta_{S\times H}}(\{(\text{CCS}, f_{\text{ed}}), (\text{NC}, f_{\text{rd}})\}) &= 0.7 \cdot 0.2 \\ m^{\Theta_{S\times H}}(\{(\text{NC}, f_{\text{rd}}), (\cdot, f_{\text{ed}})) &= 0.3 \cdot 0.2 \end{aligned} \tag{3.28}$$

Marginalization of $m^{\Theta_{S\times H}}$ on Θ_H according to (A.15) gives:

$$\begin{aligned} m^{\Theta_{S\times H} \downarrow \Theta_H}(f_{\text{ed}}) &= 0.7 \cdot 0.8 + 0.3 \cdot 0.8 = 0.8 \\ m^{\Theta_{S\times H} \downarrow \Theta_H}(\Theta_H) &= 0.7 \cdot 0.2 + 0.3 \cdot 0.2 = 0.2 \end{aligned} \tag{3.29}$$

Combining (3.26) and (3.29) according to the conjunctive rule of combination (A.18) results in the final mass distribution:

$$\begin{aligned} m^{\Theta_H}(f_{\text{ed}}) &= 0.97 \\ m^{\Theta_H}(\Theta_H) &= 0.03 \end{aligned} \tag{3.30}$$

3

In the case that the diagnostic result serves as input for a decision making process, the following pignistic probability distribution is obtained:

$$P_{\text{pig}}(f_{\text{rd}}) = 0.015$$
$$P_{\text{pig}}(f_{\text{ed}}) = 0.985 \tag{3.31}$$

Like in the Bayesian model, it is concluded with a probability of slightly more than 98% that the conductance problem is caused by electrical disturbances.

3.5.4. Modified case: partially conflicting information sources

Consider the case as introduced in Section 3.5.1, but with rule k_2 redefined as:

k_2': **if** $H = f_{\text{ed}}$ **then** $P(T = I) = 1$

The associated conditional probability table of T is given in Table 3.3. The corresponding valuation function $m^{\Theta_{T\times H}}$ is:

$$m^{\Theta_{T\times H}}(\{(\text{E}, f_{\text{rd}}), (\text{I}, f_{\text{ed}})\}) = 0.85$$
$$m^{\Theta_{T\times H}}(\{(\text{I}, f_{\text{ed}}), (\cdot, f_{\text{rd}})\}) = 0.15 \tag{3.32}$$

3

Following the same analysis as before, the following diagnostic results or obtained. According to the Bayesian model:

$$P(f_{\text{rd}}) = 0.015$$
$$P(f_{\text{ed}}) = 0.985 \tag{3.33}$$

According to the D-S model:

$$m(f_{\text{ed}}) = 0.718$$
$$m(f_{\text{rd}}) = 0.052$$
$$m(\Theta_H) = 0.022$$
$$m(\emptyset) = 0.207 \tag{3.34}$$

Both the Bayesian and the D-S solution point towards a conductance problem. The D-S solution encodes more uncertainty about this conclusion compared to the Bayesian solution.

Tabel 3.3: Conditional probability table of T for the modified case

	T			
H	L	E	A	I
f_{rd}	0.05	0.85	0.05	0.05
f_{ed}	0	0	0	1

3.5.5. EVALUATION

We have illustrated how the track circuit diagnosis problem is handled in both the Bayesian and the D-S framework. In the original case, the available information is almost complete and non-conflicting, and both frameworks conclude with a high confidence that electrical disturbances are responsible for the conductance problem. In the modified case, the different evidences are partially conflicting and the results obtained in the two frameworks differ. The Bayesian model, again, concludes with a high confidence that electrical disturbances are responsible for the conductance problem. The D-S model also concludes that the conductance problem is most likely caused by electrical disturbances, but the model is less confident and also indicates that there is some conflicting information. The conflict may e.g. indicate that a fault not included in Θ_H is responsible for the conductance problem, or that one or more of the evidences are unreliable. The different conclusions can partly be explained by the way evidence e_1 and e_2 are interpreted in the two frameworks: According to the Bayesian model, the temporal behavior is most likely intermittent (I). According to the D-S interpretation, the temporal behavior is most likely abrupt (A). The most likely feature values in the D-S model, $T = \mathrm{A}$ and $S = \mathrm{CSS}$, are partially conflicting with respect to the fault cause, which explains the conflict in the D-S solution.

This example confirms that the preferred reasoning framework depends on the way and the extent to which the available knowledge and evidences are disturbed by uncertainty. When the available information is almost complete and non-conflicting, the Bayesian and D-S diagnosis outcome will be close. Considering Table 2.4, in such cases, the Bayesian model seems to be the preferred one since it is computationally less demanding, clearer, and easier to adapt. When the available knowledge is partially incomplete or conflicting, the D-S outcome is more informative and consequently may be preferred over the Bayesian outcome. Whether this advantage outweighs the Bayesian advantages as listed in Table 2.4, depends on the degree to which the information is incomplete and conflicting and on application-specific preferences, e.g. what are the consequences of an incorrect decision, and how important are clarity of inference and adaptability.

3.6. EXTENSION TO OTHER DIAGNOSIS APPROACHES

We motivated our choice for a knowledge-based approach, among other things, by the fact that (labeled) historical data are only scarcely available. However, monitoring devices are increasingly being installed, and it is expected that the amount of available data will increase rapidly over time. When very large amounts of historical data are available, holistic approaches generally tend to outperform approaches that manually incorporate knowledge (Graves and Jaitly, 2014). We anticipate on this trend by investigating how the temporal and spatial system knowledge presented in this chapter can support model-free and hybrid approaches to fault diagnosis. Here, we briefly demonstrate our findings in two examples: a recurrent neural network and a multiple Kalman filter approach. Detailed information regarding the developed methods, as well as details regarding the simulation model used for data generation can be found in (Verbeek, 2015; De Bruin et al., 2016).

3.6.1. RECURRENT NEURAL NETWORK

Neural networks are very expressive models that, in theory and given enough data, can learn any complex non-linear mapping from their inputs to their outputs (Bishop, 1995). Provided the availability of a large and informative set of labeled historical data, neural networks might be able to represent the relationship between monitoring signals and system health more accurately than is currently understood, and so have the potential to outperform model-based diagnosis strategies.

The success of artificial neural networks is partially attributed to a strategy called end-to-end learning (Graves and Jaitly, 2014). This strategy moves away from hand-crafted feature detectors and manually integrating prior knowledge into the network. Instead, neural networks are trained to produce their end results directly from the raw input data. Although neural networks do not explicitly incorporate prior knowledge in the network, they can certainly benefit from the temporal and spatial knowledge presented in this chapter, e.g. in the choice of:

1. the network architecture;
2. the network inputs.

NETWORK ARCHITECTURE

The utilization of temporal dependencies requires a network architecture that is able to learn temporal patterns in data. Standard feed-forward neural networks only capture time dependencies that are within the input size. In order to capture long-term dependencies, the input size has to be large, which can be impractical for multivariate signals or in the case of very long-term dependencies. The solution is to use a model that incorporates temporal coherence, performs temporal pooling, or incorporates memory through recurrent connections (Längkvist et al., 2014). A promising network architecture to handle temporal patterns in data is the Long-Short Term Memory (LSTM) recurrent neural network (Hochreiter and Schmidhuber, 1997). Whereas standard recurrent networks have difficulty to learn long-term dependencies, LSTM networks overcome this problem by introducing special memory cells in the network architecture. These memory cells include gating units to control the flow of information,

De Bruin et al. (2016) have shown that an LSTM network achieves good diagnostic performance on the railway track circuit diagnosis case (see Section 3.3.1). Two drawbacks of the approach are that network training is computationally demanding and quite sensitive to the choice of the hyperparameters[8] (De Bruin et al., 2016). These drawbacks can be avoided by using a convolutional network with a max-pooling operator[9]. However, a convolutional network is in general not able to correctly identify intermittent time behaviors. Since electrical disturbances in a track circuit network are associated with an intermittent temporal behavior, convolutional networks are not appropriate for the considered diagnosis task. For other diagnosis tasks, intermittent time behaviors may be absent and this deficiency may be of less concern. Hence, insight into temporal degradation patterns (e.g., gradual, abrupt, intermittent, long-term versus short-term) is

[8]Hyperparameters are the parameters that are not part of the model that is optimized during the learning phase. The hyperparamaters need to be tuned a priori.

[9]A max-pooling operator introduces a limited invariance to the exact time at which a certain input pattern was detected. This simplifies the learning procedure and improves generalization (De Bruin et al., 2016).

needed to select the most appropriate network architecture (e.g. an LSTM network or a convolutional network).

NETWORK INPUTS

To allow a neural network to learn the spatial dependencies associated with different system faults, the network input should include the monitoring signals of neighboring system components. Which components need to be included depends, among other things, on the set of possible spatial configurations. For example, for the track circuit case, we distinguished between section-specific faults (S_i = NC), track-specific faults (S_i = CCS), and environmental faults (S_i = CAS). In this case, the neural network input should at least contain 1. the monitoring vector M_i of section i to be diagnosed; 2. the monitoring vector M_j of a section j that is connected to section i; and 3. the monitoring vector M_k of a neighboring section k that is not connected to section i. Preferably for the last two groups, signals of multiple components should be included to ensure robustness with respect to possible faults in system components (see Section 3.2.3). Since including too many inputs complicates learning and increases computational burden, a trade-off between robustness and computational costs needs to be made. Hence, insight into spatial system dependencies is needed to carefully select the network inputs.

3.6.2. MULTIPLE KALMAN FILTERS

In a multiple-model approach, multiple (statistical) models are defined, each of them describing the system behavior associated with a particular fault category. The extent to which a particular model describes an observed behavior indicates how likely it is that a fault associated with the considered model is present. The spatial and temporal information presented in this chapter can be explicitly incorporated in the approach by designing a separate model for each combination of qualitative degradation behavior (e.g. linear, exponential, abrupt, intermittent) and spatial configuration (e.g. no correlation with neighboring components, correlation with connected components, correlation with all components in a close neighborhood). Moreover, considering statistical models allows to complement system knowledge with statistical data, and provides a natural framework to handle uncertainty.

Verbeek (2015) proposes a multiple-model approach for fault diagnosis of railway track circuits. Kalman filters (Kalman, 1960) are considered to capture the different models. The various types of degradation behavior are reflected in the different A-matrices, and the different spatial dependencies are reflected in the C-matrices. By solely incorporating the available knowledge (see Table 3.1) it is not possible to discriminate between all fault causes. For example, when we observe an exponential degradation trend in a single component, we can only conclude that either an insulated joint defect or a mechanical rail defect is present. Although both an insulated joint defect and a mechanical rail defect are associated with exponential degradation, the precise degradation rates associated with the two faults likely differ. The exact degradation rates are often unknown, but available statistical data can provide an aid to exploit this information. In (Verbeek, 2015), a Bayesian classifier is included to distinguish among different degradation rates. By comparing the observed degradation rate with previously observed degradation rates for each of the fault types, Bayes' rule can be used to determine the probability that the faulty behavior is caused by a particular fault type.

So, knowledge about the spatial and temporal dependencies allows us to define models that can discriminate between different fault types. By additionally exploiting historical statistical data, a finer distinction in fault causes is possible.

3.7. CONCLUSIONS

In this chapter, a knowledge-based approach to fault diagnosis in interconnected systems has been proposed. Next to intra-component dependencies, the temporal, spatial, and spatio-temporal system dependencies are used as diagnostic features. Two main advantages of this method compared to existing diagnosis methods are that 1. fewer monitoring devices are required and 2. the method is robust with respect to environmental disturbances. The applicability of the method has been demonstrated on a railway track circuit diagnosis case. It has been shown that the proposed method is able to adequately detect and diagnose track circuit faults, even in the presence of environmental disturbances. Compared to the current practice of threshold checking, the proposed approach provides more timely insight into faulty behavior and a characterization of the type of fault present. This additional information is important for creating an effective condition-based maintenance schedule.

Maintenance planning based on diagnostic and prognostic information will be a topic for future research. Other topics for future work include:

1. The development of methods to determine a component's operating state. In this chapter, we assumed that the operating state of each component is known at each time. However, in practice the operating state needs to be determined from the available information. For the railway case, we already mentioned some possible ways to determine the operating state. However, these strategies are application-specific. It would therefore be interesting to examine the possibility of determining the operating state in a more general way.

2. In this chapter, we focused on homogeneous systems, i.e. multi-component systems consisting of only one type of components. Most systems however consist of various types of components. Although the underlying idea behind the proposed method remains valid for heterogeneous systems (i.e. systems consisting of various types of components), the framework needs to be generalized. Therefore, we propose the extension of the diagnostic framework to handle heterogeneous systems as a topic for future research.

3. For the railway track circuit case, we propose the development of systematic methods to determine the feature values as well as the incorporation of extra or more refined features to further improve diagnostic performance.

4

Multiple-model approach to system-level HVAC fault diagnosis

"Everything should be made as simple as possible, but not simpler."

-Albert Einstein-

Interdependencies among system components and the existence of multiple operating modes present a challenge for fault diagnosis of Heating, Ventilation, and Air Conditioning (HVAC) systems. Reliable and timely diagnosis can only be ensured when it is performed in all operating modes, and at the system level, rather than at the level of the individual components. Nevertheless, almost no HVAC fault diagnosis methods that satisfy these requirements are described in literature. In this chapter, we propose a multiple-model approach to system-level HVAC fault diagnosis that takes component interdependencies and multiple operating modes into account. For each operating mode, a distinct Bayesian network (diagnostic model) is defined at the system level. The models are constructed based on knowledge regarding component interdependencies and conservation laws, and based on historical data through the use of virtual sensors. We show that component interdependencies provide useful features for fault diagnosis. Incorporating these features results in better diagnostic results, especially when only a few monitoring signals are available. Simulations demonstrate the performance of the proposed method: faults are timely and correctly diagnosed, provided that the faults result in observable behavior.

4.1. INTRODUCTION

Heating, Ventilation, and Air Conditioning (HVAC) systems, widely used in residential and commercial buildings, are responsible for a large part (20 – 40%) of the worldwide energy consumption (Pérez-Lombard et al., 2008). Malfunction or degradation of HVAC system components causes reduced comfort on the one hand, and approximately 15 – 30% waste of energy on the other hand (Piette et al., 2001; Katipamula and Brambley, 2005). Therefore, the development of effective preventive maintenance strategies for HVAC systems is of major importance.

A promising preventive maintenance strategy is condition-based maintenance, which plans the maintenance according to the needs indicated by the system condition (Yam et al., 2001; Wang, 2014). An important step within the condition-based maintenance process is fault diagnosis (see Chapter 1). Fault diagnosis of HVAC systems is a challenging task for the following reasons:

1. The HVAC system behavior is difficult to model, as it varies from building to building and it is influenced by uncertain factors, like weather and building use;

2. In general, relatively few variables are measured, especially at the component level. For example, air and water flow rates are rarely available for all components, such as radiators and air handling units;

3. The available measurements are often only crude estimates of the underlying variables, e.g. they are collected by single-point air temperature sensors;

4. (Hierarchical) relationships exist among the different system components (Schein and Bushby, 2006). For example, a non-functioning boiler will also affect the working of all radiators and air handling units connected to this boiler. Similarly, the degree to which a radiator fault affects the room temperature depends, among other factors, on the availability and capacity of other radiators in the room;

5. Environmental variations and users settings (e.g. day and night schedules) require that HVAC systems operate in different modes. For example, during the day, both the refresh rate and the supply air temperature are controlled, while during the night only the refresh rate is controlled. Each of the operating modes may require a different diagnostic model.

Although research has been devoted to (HVAC) fault diagnosis (Dexter and Ngo, 2001; Lee et al., 2004; Wang and Xiao, 2004; Schein and Bushby, 2006; Zogg et al., 2006; Liang and Du, 2007; Lo et al., 2007; Namburu et al., 2007; Mulumba et al., 2015), almost no attention has been paid to component interdependencies and to the consequences of multiple operating modes. Most papers focus on specific methods (e.g. principal component analysis (Wang and Xiao, 2004; Namburu et al., 2007), Bayesian networks (Zhao et al., 2013; Xiao et al., 2014) clustering techniques (Zogg et al., 2006), neural networks (Lee et al., 2004), fuzzy systems (Dexter and Ngo, 2001; Lo et al., 2007), or support vector machines (Liang and Du, 2007; Namburu et al., 2007; Mulumba et al., 2015)) for the fault diagnosis of one specific (HVAC) component. For example, Zogg et al. (2006) propose a model-based diagnosis approach for commercial heat pumps; Dexter and Ngo (2001);

Lee et al. (2004); Liang and Du (2007); Lo et al. (2007); Xiao et al. (2014); Mulumba et al. (2015) propose different diagnosis strategies for the fault diagnosis of an air handling unit; Namburu et al. (2007); Zhao et al. (2013) specifically focus on the fault diagnosis of the chiller plant; and Wang and Xiao (2004) present a strategy based on the principal component analysis to detect and diagnose sensor faults in typical air-handling units. Schein and Bushby (2006) consider fault diagnosis at the system level taking component interdependence into account. However, the proposed diagnostic model is captured by a rule-based system, which cannot easily be modified to changing situations and other building configurations and which does not take uncertainty into account.

Fault diagnosis methods that do not take both component interdependencies and changing operating modes into account, will not result in adequate fault diagnosis in practice. To ensure correct and timely diagnosis the problem characteristics should explicitly be taken into account in the formulation of the diagnostic model, and that is what we do. More specifically, we propose a multiple-model approach to system-level fault diagnosis in HVAC systems that:

1. takes the interdependencies among the different HVAC components into account (Challenge D_2);
2. can easily adapt to changing operation conditions and different building configurations.

Each model is captured by a Bayesian network[1]. These Bayesian networks are constructed based on both knowledge regarding *component interdependencies* and *conservation laws*, and based on historical data through the use of *virtual sensors*. This way, advantage is taken of the available knowledge and data, while keeping the reasoning transparent.

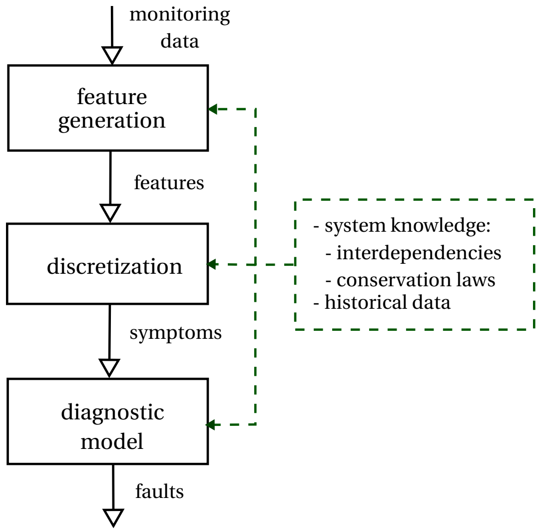
The remainder of this chapter consists of three parts: In the first part (Section 4.2 till Section 4.4) we propose our multiple-model approach to fault diagnosis in HVAC systems. Next in Sections 4.5 and 4.6, we present two case studies involving a simple HVAC-controlled building to demonstrate and evaluate the proposed method. Afterwards, in Section 4.7 we discuss the generalization to other building configurations.

4.2. OVERVIEW OF THE PROPOSED DIAGNOSIS APPROACH

Figure 4.1 gives a schematic overview of the proposed model-based fault diagnosis strategy. First, characteristic features are extracted from the monitoring variables. Next, the continuous-valued features are mapped to discrete-valued symptoms. Finally, based on the symptoms the presence and type of faults are inferred by using the diagnostic model.

We account for component interdependencies by performing diagnosis at the system level (instead of the component level) and by exploiting knowledge regarding component interdependencies in defining the mappings from monitoring data to features, from features to symptoms, and from symptoms to faults (see Figure 4.1). Because the relations between faults and symptoms are uncertain and may differ for different operating modes, an appropriate diagnostic model is defined for each operating mode and

[1] A Bayesian network is an intuitive and transparent model for reasoning under uncertainty that can easily adapt to varying operation conditions and different building configurations (Pearl and Russel, 2001; Boudali and Dugan, 2005; Darwiche, 2009; Wiegerinck et al., 2010).



Figuur 4.1: Overview model-based fault diagnosis.

captured by a Bayesian network. Finally, it is ensured that the method is applicable to a wide range of building configurations by exploiting system knowledge that is applicable to all kinds of building configurations (e.g. conservation laws) in defining the different mappings.

4

4.3. ELABORATION OF THE DIAGNOSIS APPROACH

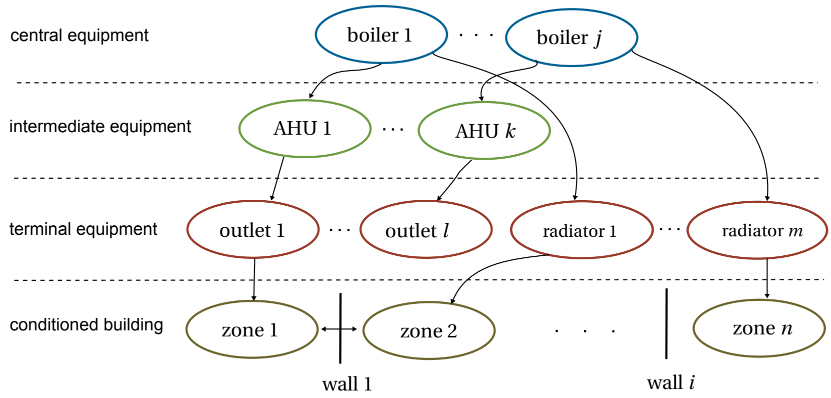
This section elaborates in more detail on the construction of the Bayesian network, i.e. the diagnostic model. A Bayesian network for a set of variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ consists of two components (Heckerman, 1998):

1. network structure G encoding a set of conditional independence assertions about the variables in $\mathbf{X}$;
2. a set D of local probability distributions associated with each variable.

The network structure G is a directed acyclic graph, with the nodes in one-to-one correspondence to the variables in $\mathbf{X}$ and the edges representing direct dependencies. For more background information on Bayesian networks, see e.g. (Pearl, 1988; Charniak, 1991; Cooper and Herskovits, 1992; Heckerman, 1998; Pearl, 2000; Pearl and Russel, 2001; Daly et al., 2011).

The construction of a Bayesian network for fault diagnosis consists of the determination of:

1. the network nodes, which can be divided into:
   (a) observable nodes, representing the symptoms;
   (b) unobservable nodes, representing the faults;



Figuur 4.2: An illustration of the (hierarchical) dependencies among HVAC components in a building.

2. the probabilistic relationships between the nodes.

Because the exact set of symptoms and the relations between symptoms and faults will differ from building to building, we do not propose symptoms here, but we introduce important information sources, namely component interdependencies and conservation laws, and discuss how they can be used for feature extraction and symptom generation. Later, in the case studies in Section 4.6, symptoms are derived based on these information sources.

4.3.1. COMPONENT INTERDEPENDENCIES

In general, an HVAC system can be represented in a hierarchical way as shown in Figure 4.2. At the top there are the boilers, which provide the radiators and Air Handling Units (AHUs) with hot water. These devices in turn transfer the energy of the hot water to the conditioned zones (radiators) and regulate and circulate the zone air (AHUs). The different components interact in various ways with each other. For the purpose of fault diagnosis, we made a distinction between:

1. *Hierarchical dependencies:* The functioning of a component depends on the proper functioning of higher-level components. For example, when a boiler is not able to heat the water to the desired temperature also the connected radiators and AHUs cannot fulfill their function. When an AHU is not able to adequately condition the air, the connected AHU outlets fail to supply the zone with the desired air.

2. *Compensation by same-level components:* The effect of a non-functioning component can be compensated for by another component fulfilling a similar function. For example, a non-functioning radiator can be compensated for by another radiator in the same zone provided that its capacity is sufficient.

Although the presence of these interdependencies complicates the diagnosis in the sense that the diagnosis cannot be carried out for all components individually, the interdependencies are valuable in the sense that they can serve as diagnostic features. Because the interdependencies vary for different faults, their values provide information regarding the fault present. For instance, a boiler fault is probably observed in multiple components or zones, whereas a radiator fault is only locally observed. In this context, an exemplary symptom of a fault in boiler A is "all activate components connected to boiler A are malfunctioning".

4.3.2. CONSERVATION LAWS

Both mass and energy balances apply to an HVAC system. Mass balances can be defined for the water flow in the hot water circuit. Energy balances can be defined for each HVAC component where energy is exchanged, e.g. boilers, radiators, and AHUs, and for each conditioned zone. An overview of applicable energy and mass balances can be found in Appendix C.

Energy and mass balances are a useful source of information for the formulation of diagnostic features. In the case of a fault, the internal relations between variables or between variables and measurements may change. These changes can be detected by verifying internal system relations, including conservation laws. For example, when the measurements do not satisfy the applicable mass balance for the hot water circuit, this could indicate e.g. a leak in the duct work or a sensor fault.

4.3.3. VIRTUAL SENSORS

4

Sometimes, the available knowledge is not sufficiently detailed to define the precise relations between features and faults. Consider e.g. that it is known that, in the absence of a particular system fault f_j (i.e. $F_j = 0$), the variable y can be modeled as an unknown function g_1 of variables x_1 and x_2. However, when fault f_j is present, the variable y no longer depends on both x_1 and x_2, but depends only on x_2, i.e.:

$$y = \begin{cases} g_1(x_1, x_2) & \text{if } F_j = 0 \\ g_2(x_2) & \text{if } F_j = 1 \end{cases} \tag{4.1}$$

From this knowledge, it follows that the symptom "y does not depend on x_1" is characteristic for fault f_j. However, the value of this symptom cannot be assessed based on just this knowledge and instantaneous values of x_1, x_2, and y.

When the available system knowledge is not sufficient to design the diagnostic model, historical data and virtual sensors can be used to complement the available system knowledge, e.g. to find the mapping g_1 in (4.1). *Virtual sensors* (Oosterom and Babuška, 2000; Li and Braun, 2007) estimate system quantities by using mathematical models, which in turn make use of other physical sensor readings to determine the estimate. Virtual sensors can be used in the following situations:

1. Absence of a physical sensor, e.g. because the desired quantity cannot be measured or a physical sensor is too slow or costly.

2. As a backup of a physical sensor, i.e. to introduce analytic redundancy. A significant difference between the real sensor and the virtual sensor indicates that one of the two is faulty.

3. To estimate the behavior of a system variable corresponding to a specific type of system behavior, e.g. healthy behavior. In this case, the virtual sensor is trained using data corresponding to the considered system behavior and a significant difference between the actual sensor reading and the virtual sensor output indicates that the system does not behave according to the considered behavior.

In the case studies in Section 4.6, a virtual sensor covering situation 3 is constructed and in Section 4.7, examples are provided where situation 1 applies.

The design of a virtual sensor essentially consists of three steps:

Step 1: The choice for the quantity to be estimated, i.e. which variables are valuable features for diagnosis.

Step 2: The selection of available sensor measurements that are relevant to estimate these quantities.

Step 3: The choice for the method to capture the relation between the quantity of interest and the relevant sensor measurements, e.g. first-principles or data-based approaches.

In this thesis, the main focus is on the first two steps. For the third step, a standard data-based approach from literature, nearest neighbor regression (Altman, 1992), is used.

4.4. FAULT DIAGNOSIS APPROACH

4.4.1. CONSTRUCTION OF THE DIAGNOSTIC MODEL

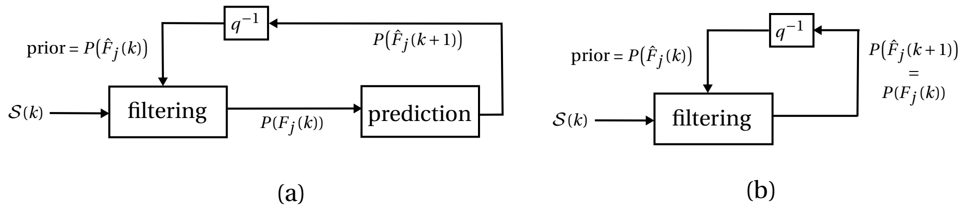
4

Procedure 2 describes the construction of the diagnostic model, in the form of a set of Bayesian networks. In line 1, the system faults f_1 till f_ℓ are determined, e.g. based on expert knowledge. Next, in lines 2 – 4, a binary node F_j is assigned to each system fault f_j. Note that a binary node is used for each of the faults in order to easily handle multiple-fault scenarios (see Chapter 2). Next, in line 5, an appropriate symptom set is determined based on knowledge and data regarding component interdependencies and conservation laws. Subsequently, a node $\mathcal{S}_l$ is assigned to each symptom (lines 6 – 8). Next, the different operating modes are determined (line 9). For each of them, the relationships between the system faults and the symptoms are defined (i.e. the corresponding network is built) (lines 11 – 13).

4.4.2. DIAGNOSTIC INFERENCE

For online fault diagnosis, we use the recursive Bayesian estimation scheme as shown in Figure 4.3, where k denotes a discrete time step and q is the shift operator. In the filtering step, the posterior probability $P(F_j(k))$ of fault f_j is determined based on the evidence $\mathcal{S}(k) = [\mathcal{S}_1(k),\ldots,\mathcal{S}_z(k)]$ and the prior probability $P(\hat{F}_j(k))$. Based on the outcome of the filtering step, a one-step-ahead prediction $P(\hat{F}_j(k+1))$ of the fault probability at the next time step is made, which serves as prior for the next filtering step.

In this work, we assume faults to be binary variables, i.e. a fault is either absent or present. In this case, the fault probability at the next time step can only be estimated based on statistical information regarding fault occurrence rates. Since we do not have



Figuur 4.3: Bayesian fault diagnosis scheme, with evidence $\mathcal{S}(k)$ the observations at time k, $F_j(k)$ the fault variable at time k, and q the shift operator: (a) the full scheme; (b) the simplified scheme adopted in this chapter.

an accurate predictive model, we assume F_j, $j = 1,\dots,\ell$ to be static, i.e. $P(\hat{F}_j(k+1)) = P(F_j(k))$. Now the problem reduces to recursively applying Bayes rule with as prior the previous posterior and as evidence the observations $\mathcal{S}(k)$, i.e. we omit the prediction step (see Figure 4.3(b)). Please note that all concepts proposed in this chapter remain valid when a predictive model is included. Especially, for gradually developing faults, the prediction step is of interest. In this case, prior knowledge of fault evolution can be combined with observed data (see Section 5).

The recursive diagnosis approach is summarized in Procedure 3. As input it uses the set of Bayesian networks defined in Procedure 2. At each diagnosis time instant, first, the actual operating mode is determined (from schedules or measured quantities) (line 3) and next, the corresponding Bayesian network is selected (line 4). Then, based on new evidence e, i.e. observations of the symptoms, the Bayesian network is updated to obtain the posterior fault probabilities (lines 5–7), which serve as prior probabilities at the next diagnosis time instant.

Procedure 2 Model construction

Input: Expert knowledge, historical data

Determine possible system faults f_1 till f_ℓ
for $j = 1,\dots,\ell$ **do**
    Define binary node F_j
end for
Determine symptoms $\mathcal{S}_1$ till $\mathcal{S}_z$ based on expert knowledge and data
for $l = 1,\dots,z$ **do**
    Define discrete-valued node $\mathcal{S}_l$
end for
Determine the system's operating modes 1 till m
for $\zeta = 1,\dots,m$ **do**
    Determine G_ζ, which is the network structure defining the relations between the symptoms $\mathcal{S}_1$ till $\mathcal{S}_z$ and the fault variables F_1 till F_ℓ, in operating mode ζ
    Determine D_ζ, which is the set of local probability functions associated with each node in G_ζ
end for

Output: Bayesian network (G_ζ, D_ζ) for each operating mode ζ

Procedure 3 Fault diagnosis

Input: Bayesian network (G_ζ, D_ζ) for each operating mode ζ, diagnosis instants τ_1 till τ_q

$\mathcal{E} = \{\}$
for $\kappa = 1, \ldots, q$ **do**
  Determine actual operating mode a at τ_κ
  Select corresponding network (G_a, D_a)
  Store new evidences regarding the symptoms in variable e
  $\mathcal{E} \leftarrow \mathcal{E} \cup \{e\}$
  Update probabilities regarding the faults F_1 till F_ℓ

$$\big(P(F_1|\mathcal{E}), \ldots, P(F_\ell|\mathcal{E})\big) = \text{inference}(G_a, D_a, e)$$
$$\big(P(F_1), \ldots, P(F_\ell)\big) \leftarrow P(F_1|\mathcal{E}), \ldots, P(F_\ell|\mathcal{E})$$

  with inference(·) the Bayesian inference algorithm
end for

Output: Conditional probability distributions of F_1 till F_ℓ given $\mathcal{E}$

4.5. HVAC SYSTEM DESCRIPTION

Figure 4.4 gives an overview of the HVAC configuration considered in this work. The main components are:

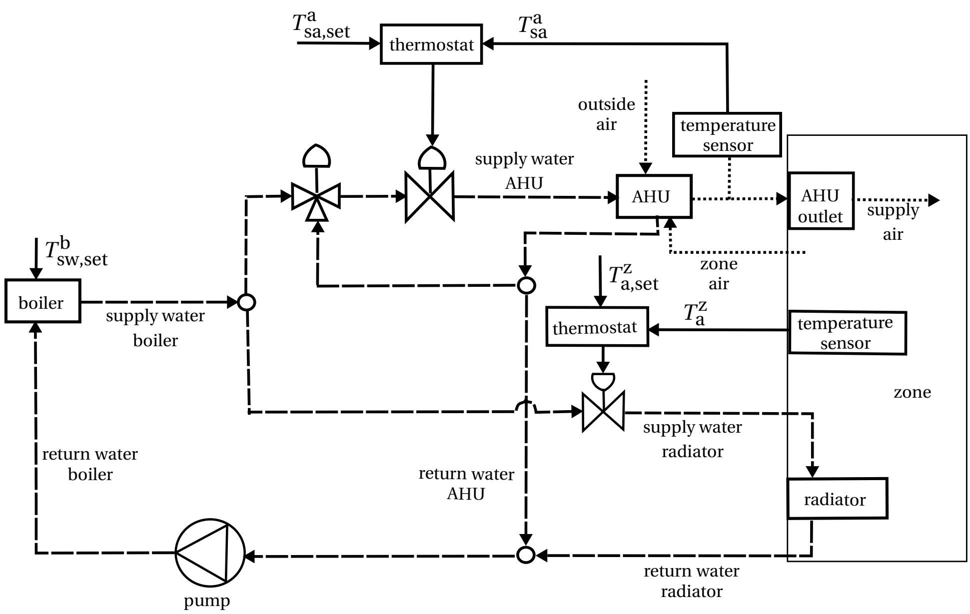
1. the zone to be conditioned.
2. HVAC plants, i.e. the equipment installed to control the zone climate:
   (a) boiler;
   (b) pump;
   (c) radiator;
   (d) air handling unit.

The proper understanding of the case studies requires a basic understanding of the AHU as well as knowledge of the available monitoring signals.

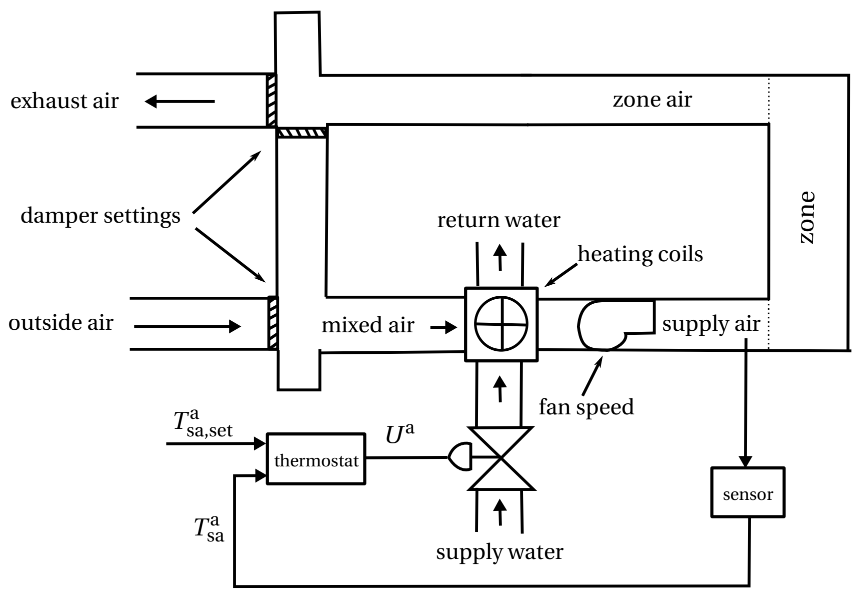
4.5.1. AIR HANDLING UNIT

Figure 4.5 gives an overview of the considered AHU. In the mixing chambers, outdoor air is mixed with air that returns from the zone. The composition of the mixed air is controlled by the positions of three dampers regulating the amount of outdoor air entering the system, the amount of air exhausted from the system, and the amount of return air from the zone to be recirculated. After the mixing, the mixed air passes through the heating coils to condition the air to the desired temperature. The heating in the coils is regulated by the amount and temperature of the water flowing through the coils. The hot water is delivered by the boiler. The temperature of the hot water through the coils is controlled to approximately 40°C using a three-way mixing valve. The amount of water flowing through the coils is determined by the position of a valve, which is controlled by a thermostat based on the differences between the AHU supply air temperature[2] T^{a}_{sa} and

[2] We use the superscript to indicate the location the variable refers to (e.g. AHU, boiler, zone) and the subscript to indicate the particular mass or air flow (e.g. supply air, return water, mixed air).



Figuur 4.4: Overview of the considered HVAC system. Dotted lines represent air flows, dashed lines represent mass flows, and solid lines represent signals.



Figuur 4.5: Schematic overview of an AHU.

its setpoint $T^{a}_{sa,set}$. Finally, a supply fan is present to maintain a pressure in the supply duct to guarantee that the mixed air is pushed through the coil and finally distributed through the duct work to the zone.

4.5.2. MONITORING SIGNALS

The following variables are assumed to be available for fault diagnosis:

- zone air temperature (T^{z}_{a});
- supply air temperature (T^{a}_{sa});
- mixed air temperature (T^{a}_{ma});
- outside air temperature (T^{o}_{a});
- supply water temperature (T^{b}_{sw});
- return water temperature (T^{b}_{rw});
- mass flow through the boiler (w^{b}_{sw});
- control signal to AHU valve (U^{a});
- control signal to the radiator valve (U^{r}).

In addition, the zone air temperature setpoint ($T^{z}_{a,set}$), supply air temperature setpoint ($T^{a}_{sa,set}$), and supply water temperature setpoint ($T^{b}_{sw,set}$) are assumed to be known.

4.6. CASE STUDIES

In this section, we illustrate the proposed method based on two case studies. Case study 1 comprises the fault detection of a stuck AHU heating coil valve and mainly serves to illustrate the problems that occur when neglecting the different operating modes and interdependencies between HVAC components. Case study 2 extends case study 1 in the sense that the possibility of a non-functioning boiler is included. Although this case study is still relatively simple, it clearly illustrates the implications of multiple operation modes and component interdependencies on the fault diagnosis, and how they are handled in the proposed diagnosis approach.

4.6.1. SIMULATION MODEL

SYSTEM MODELING

For the purpose of analysis and validation, experts at Honeywell have developed a simulation model of the considered building (Holub and Macek, 2013). The model has been verified using data obtained from real buildings. The model makes a distinction between two sets of variables: temperatures and mass flows. As the pressure dynamics are much faster than the temperature dynamics, the transient behavior of the mass flow rates is neglected, i.e.:

$$w^{a}_{sw}(\tau) = f_{a}\left(X^{a}(\tau), X^{r}(\tau)\right) \tag{4.2}$$
$$w^{r}_{sw}(\tau) = f_{r}\left(X^{a}(\tau), X^{r}(\tau)\right) \tag{4.3}$$

with w^{a}_{sw} and w^{r}_{sw} the mass flows through the AHU and radiator respectively, and X^{a} and X^{r} the positions of the AHU valve and the radiator valve. For more details on the simulation model, see (Holub and Macek, 2013).

FAULT MODELING

Stuck heating coil valve A stuck valve stays in the position it was before it got stuck, regardless of the control signal U^{a} sent to the valve by the thermostat. This means that the mass flow through the heating coil remains the same. In the simulation model, a stuck valve is modeled by constraining the mass flow to be constant, i.e.:

$$w^{\text{a}}_{\text{sw}}(\tau) = w^{\text{a}}_{\text{sw}}(\tau^{\text{a}}) \quad \forall \tau \geq \tau^{\text{a}} \tag{4.4}$$

with τ^{a} the time that the valve stopped functioning.

Non-functioning boiler When the boiler breaks down, the water returning from the hot water circuit is no longer heated to the supply water temperature setpoint $T^{\text{b}}_{\text{sw,set}}$, i.e. the supply water temperature T^{b}_{sw} becomes equal to the return water temperature T^{b}_{rw}. Therefore, a non-functioning boiler is modeled as follows[3]:

$$T^{\text{b}}_{\text{sw}}(\tau) = T^{\text{b}}_{\text{rw}}(\tau^{\text{b}}) \quad \forall \tau \geq \tau^{\text{b}} \tag{4.5}$$

with τ^{b} the time that the boiler stopped functioning.

SIMULATION SPECIFICATIONS

In the simulation model, the following assumptions are adopted:

1. The daily schedule is defined as:
   - day operation between 04.00 and 18.00 hours;
   - night operation between 18.00 and 04.00 hours.
2. The setpoints of the boiler supply water temperature T^{b}_{sw}, the AHU supply air temperature T^{a}_{sa}, and the zone air temperature T^{z}_{a} are:

$$T^{\text{b}}_{\text{sw,set}} = \begin{cases} 75 & \text{day operation} \\ 65 & \text{night operation} \end{cases}$$

$$T^{\text{a}}_{\text{sa,set}} = \begin{cases} 20 & \text{day operation} \\ \text{-} & \text{night operation} \end{cases}$$

$$T^{\text{z}}_{\text{a,set}} = \begin{cases} 21 & \text{day operation} \\ 18 & \text{night operation} \end{cases}$$

3. Damper positions are fixed, i.e. the ratio between zone air and outside air is constant (1:4 during the day and 3:7 during the night).
4. Fan speed is fixed, i.e. w^{a}_{sa} is constant (0.1 kg/s during the day and 0.001 kg/s during the night).

[3]Note that in practice there is some delay between the time the boiler stops functioning and the time the supply water temperature becomes equal to the temperature of the return water. We assume this delay to be small and neglect it in the remainder.

4

4.6.2. CASE STUDY 1: BASIC EXAMPLE

Consider the building configuration depicted in Figure 4.4 and assume that the system is healthy except for a possibly stuck AHU heating coil valve. Our aim is to determine whether or not the valve is stuck. This is a challenging problem because:

1. the extent to which the fault expresses itself in the measured variables highly depends on the position in which the valve got stuck and on weather conditions;

2. the mass flow through the valve is not measured.

DIAGNOSTIC MODEL

Network structure Given the measurements specified in Section 4.5.2, an obvious way to detect a stuck heating coil valve is to compare the supply air temperature T^{a}_{sa} with its setpoint $T^{a}_{sa,set}$. In the case of a broken valve, a difference between the two temperatures is expected. This knowledge gives rise to define symptom $\mathcal{S}_1$ as:

$$\mathcal{S}_1 = \begin{cases} 1 & \text{if } |T^{a}_{sa} - T^{a}_{sa,set}| > \epsilon_1 \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

with $\epsilon_1 > 0$ a user-defined threshold. Symptom $\mathcal{S}_1$ is related to the system health as follows:

if the system is healthy, i.e. $F^{a} = 0$ **then** likely $\mathcal{S}_1 = 0$

if the valve is broken, i.e. $F^{a} = 1$ **then** likely $\mathcal{S}_1 = 1$

with F^{a} a binary variable indicating whether the AHU valve is healthy ($F^{a} = 0$) or stuck ($F^{a} = 1$). Here, “likely” indicates that due to uncertain influences, we are not completely sure about the relations. The degree of uncertainty is expressed in the conditional probability table of $\mathcal{S}_1$, which we will define later. The relations hold under the assumptions that the system operates in day mode and $T^{a}_{ma} \leq T^{a}_{sa,set}$. Because the supply air temperature T^{a}_{sa} is not controlled during the night, a stuck heating coil valve is only expressed in symptom $\mathcal{S}_1$ during the day. Furthermore, as only heating is present in the considered system, in the summer period when $T^{a}_{ma} > T^{a}_{sa,set}$, too high a value of the supply air temperature can be both due to a stuck valve or due to high outside temperatures.

The proposed diagnostic model is graphically represented by the Bayesian networks in Figure 4.6. Due to the imposed day and night schedule, the system must operate in two modes, which are also reflected in the diagnostic model. As the available simulation data concern the winter season, in which case $T^{a}_{ma} < T^{a}_{sa,set}$, node T^{a}_{ma} is neglected in the remainder.

Local probability distributions To complete the construction of the Bayesian network, the following items need to be determined:

1. the value of ϵ_1;

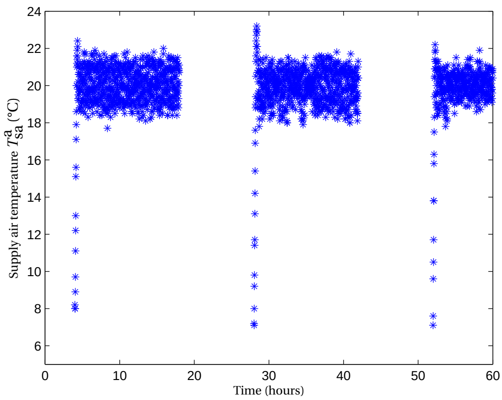
2. the conditional probability table of $\mathcal{S}_1$;

3. the initial prior probability distribution of F^{a}.

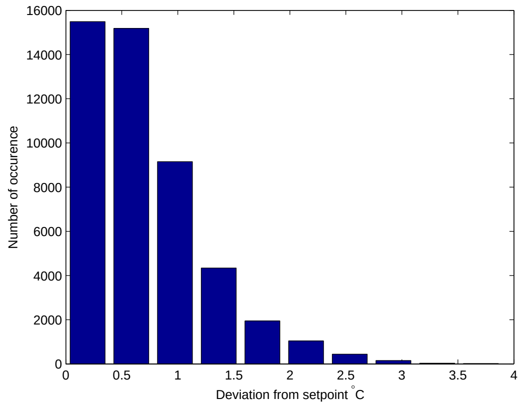
4

Figuur 4.6: Bayesian network representations for case study 1. During day symptom $\mathcal{S}_1$ is influenced by both an AHU fault and by the mixed air temperature. During night, the AHU is switched off and the relations between F^a, T^a_{ma}, and $\mathcal{S}_1$ no longer hold.

4



Figuur 4.7: Typical daily behavior of the supply air temperature with a sample step of one minute. Note that during the night, the AHU supply air temperature is not controlled.



Figuur 4.8: Distribution of $|T^{\mathrm{a}}_{\mathrm{sa}} - T^{\mathrm{a}}_{\mathrm{sa,set}}|$.

Determination of ϵ_1 To determine ϵ_1, the nominal variations in $T^{\mathrm{a}}_{\mathrm{sa}}$ are considered. Figure 4.7 shows the behavior of $T^{\mathrm{a}}_{\mathrm{sa}}$ on three consecutive days. It can be observed that in the morning, when the system switches to day mode, it takes some time (about half an hour) before the supply air temperature has converged to its desired value $T^{\mathrm{a}}_{\mathrm{sa,set}} = 20$°C. After this time, the temperature fluctuates around its desired value. To gain some insight into the degree of fluctuation, in Figure 4.8 the histogram of $|T^{\mathrm{a}}_{\mathrm{sa}} - T^{\mathrm{a}}_{\mathrm{sa,set}}|$ containing data of two consecutive months is shown. We tune the value of ϵ_1 such that 99% of the $T^{\mathrm{a}}_{\mathrm{sa}}$ values between 04.30 and 18.00 hours are within the interval $[T^{\mathrm{a}}_{\mathrm{sa,set}} - \epsilon_1, T^{\mathrm{a}}_{\mathrm{sa,set}} + \epsilon_1]$, resulting in

$$\epsilon_1 = 2.5$$

Conditional probability table of $\mathcal{S}_1$ As ϵ_1 is tuned such that in 99% of the healthy cases it holds that $\mathcal{S}_1 = 0$, the probability that $\mathcal{S}_1 = 1$ given the system is healthy is 1%. To determine the probability that $\mathcal{S}_1 = 1$ given a stuck heating coil valve, simulation data from faulty behavior are considered[4]. Actually, the data set used for this must be representative for faults in all different valve positions and for all relevant weather conditions. Figure 4.9 shows two completely different behaviors of $T^{\mathrm{a}}_{\mathrm{sa}}$ corresponding to a stuck AHU valve. In the first situation, the valve got stuck during night in a cold period, whereas in the second situation, the valve got stuck during day while the outside temperature is increasing. We approximate the probability that $\mathcal{S}_1 = 1$ given an AHU valve fault ($F^{\mathrm{a}} = 1$) based on a finite number of randomly chosen fault scenarios. The results are included in Table 4.1.

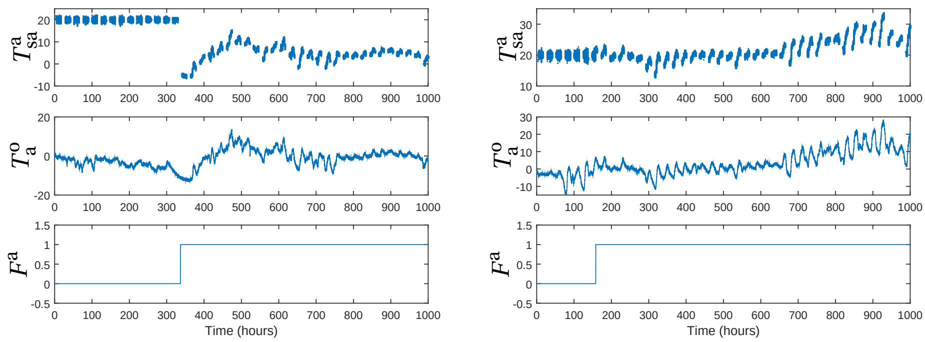
Initial prior probability distribution of F^{a} At the first diagnosis time instant, a user-defined prior $P^0(F^{\mathrm{a}} = 1) = 0.01$ is used. The initial prior probability $P^0(F^{\mathrm{a}} = 1)$

[4]Instead of using (simulation) data, these probabilities can also be directly derived from expert knowledge.

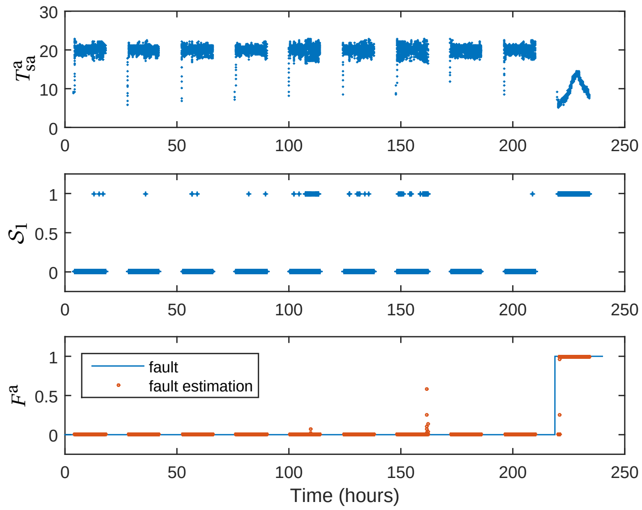
4

Tabel 4.1: Conditional probability table of $\mathcal{S}_1$ for case study 1

	$\mathcal{S}_1$	
F^{a}	0	1
0	0.99	0.01
1	0.24	0.76



Figuur 4.9: Possible behaviors of $T^{\mathrm{a}}_{\mathrm{sa}}$ corresponding to a stuck heating coil valve. Left: the valve got stuck during the night in a cold period. Right: the valve got stuck during the day while the outside temperature is increasing.



Figuur 4.10: AHU fault diagnosis example 1 for case study 1.

indicates how likely we consider the occurrence of an AHU valve fault before observing the monitoring data.

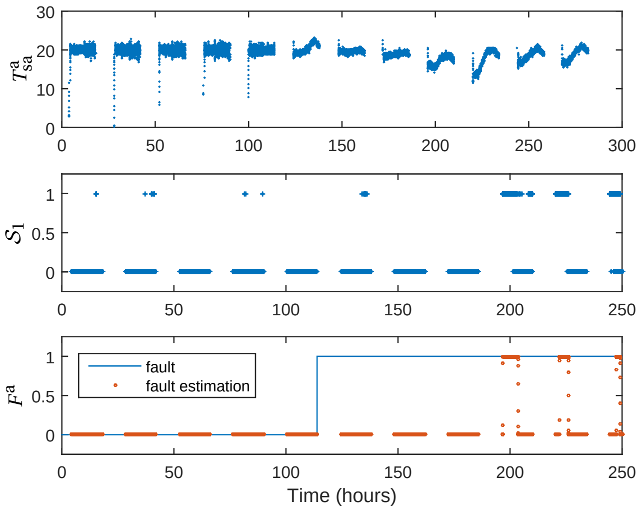
Note that from Bayes' rule, which state that:

$$P(F^a|\mathcal{S}_1) = \frac{P(\mathcal{S}_1|F^a)P(F^a)}{\sum_{y\in\Theta_{F^a}} P\mathcal{S}_1|y)P(y)}, \qquad \text{with } \Theta_{F^a} = \{0,1\} \text{ the domain of } F^a \tag{4.7}$$

it follows that the influence of the initial prior probability on the fault diagnosis is small as the probabilities are recursively updated every minute and the likelihood functions have clearly different values for $F^a = 0$ and $F^a = 1$ (see Table 4.1).

FAULT DIAGNOSIS

The proposed approach is demonstrated by means of two simulations. In the first simulation (see Figure 4.10), the valve got stuck in a cold period during the night (around time $\tau = 220$ hours). As a consequence, the air in the AHU is not sufficiently heated during the subsequent day, symptom $\mathcal{S}_1$ becomes equal to one, and shortly afterwards, an AHU fault is detected, i.e. $P(F^a = 1|\mathcal{E}) \approx 1$, where $\mathcal{E}$ contains all observations of symptom $\mathcal{S}_1$. Besides the correct fault detection around $\tau = 220$ hours, an AHU fault is incorrectly detected around $\tau = 160$ hours. This incorrect detection is of a very short duration and a consequence of the way ϵ_1 is tuned. Recall that ϵ_1 is tuned such that in 1% of the healthy cases symptom $\mathcal{S}_1$ is activated. If this happens at several consecutive time instants, this will lead to a false positive detection. In the second example (see Figure 4.11), the valve got stuck during the day. As the position in which the valve got stuck was quite favorable with respect to the supply air temperature setpoint in the subsequent days, the fault is only detected after four days, i.e. as soon as the effects become observable.



Figuur 4.11: AHU fault diagnosis example 2 for case study 1.

4

CONCLUDING REMARKS

The diagnostic model defined in Section 4.6.2 turned out to be effective in the sense that in the simulations faults are detected as soon as their effects are observable. Nevertheless the proposed model suffers from some shortcomings, namely:

1. Faults cannot be detected during the night;
2. The model is not useful for high mixed-air temperatures;
3. The underlying assumptions are too simplistic, e.g. as only an AHU valve fault is allowed, hierarchical relationships are assumed to be absent.

Therefore, the next section deals with a case study including multiple fault scenarios where the goal is to construct a diagnostic model that it is less sensitive to high values of the mixed air temperature, and that allows for fault diagnosis in all operating modes.

4.6.3. CASE STUDY 2: EXTENDED EXAMPLE

This case study extends the problem discussed in Section 4.6.2 by including the possibility of a non-functioning boiler. In this case, there are four possible fault scenarios:

1. healthy system;
2. stuck heating coil valve:
3. non-functioning boiler;
4. both the valve and the boiler are non-functioning.

DIAGNOSTIC MODEL

Network structure Apart from that the diagnostic model for case study 1 does not support fault diagnosis during the night and is sensitive to high values of the mixed air temperature, the model cannot distinguish between all fault scenarios. If $\mathcal{S}_1 = 1$, all scenarios except for scenario 1 are plausible. To make a further distinction between the different possible fault scenarios, symptom $\mathcal{S}_1$ is extended from a binary valued symptom to a three-valued symptom $\mathcal{S}'_1$:

$$\mathcal{S}'_1 = \begin{cases} -1 & \text{if } (T^{\mathrm{a}}_{\mathrm{sa}} - T^{\mathrm{a}}_{\mathrm{sa,set}}) < -\epsilon_1 \\ 1 & \text{if } (T^{\mathrm{a}}_{\mathrm{sa}} - T^{\mathrm{a}}_{\mathrm{sa,set}}) > \epsilon_1 \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

Symptom $\mathcal{S}'_1$ relates to the system health as follows:

if $F^{\mathrm{a}} = F^{\mathrm{b}} = 0$ **then** likely $\mathcal{S}'_1 = 0$

if $F^{\mathrm{a}} = 1$ and $F^{\mathrm{b}} = 0$ **then** likely $\mathcal{S}'_1 = -1$ or $\mathcal{S}'_1 = 1$

if $F^{\mathrm{b}} = 1$ **then** likely $\mathcal{S}'_1 = -1$

So, $\mathcal{S}'_1 = 0$ characterizes a healthy system and $\mathcal{S}'_1 = 1$ characterizes an AHU valve that got stuck in a too opened position. When $\mathcal{S}'_1 = -1$, scenarios 2, 3, and 4 are all possible. To improve diagnostic power and to allow for diagnosis during both the day and the night, two additional symptoms are proposed: $\mathcal{S}_2$ to the verify the proper functioning of the AHU valve and $\mathcal{S}_3$ to verify the proper functioning of the boiler.

To verify whether or not the valve is stuck, the relationships between the mass flow through the boiler $w^{\mathrm{b}}_{\mathrm{sw}}$ and the control signals U^{a} and U^{r} to the AHU valve and the radiator valve respectively are used:

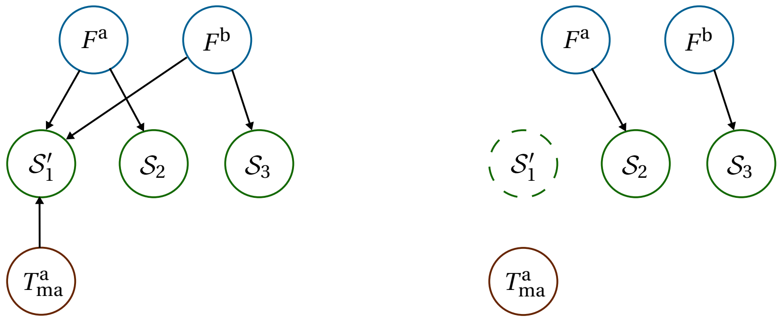
- When $F^{\mathrm{a}} = 0$, the mass flow through the boiler $w^{\mathrm{b}}_{\mathrm{sw}}$ depends both on the control signal to the AHU valve U^{a} and the control signal to the radiator valve U^{r}.
- When $F^{\mathrm{a}} = 1$, the mass flow through the boiler $w^{\mathrm{b}}_{\mathrm{sw}}$ no longer depends on U^{a}, but depends only on U^{r}.

This follows from the applicable mass balance (C.1) and equations (4.2) and (4.3). Since the relationships among $w^{\mathrm{b}}_{\mathrm{sw}}, U^{\mathrm{a}}$, and U^{r} are not known exactly, we construct a virtual sensor that predicts the mass flow through the boiler $w^{\mathrm{b}}_{\mathrm{sw}}$ based on the AHU and radiator valve control signals U^{a} and U^{r}. The virtual sensor is trained based on healthy data. So, the virtual sensor estimate $\hat{w}^{\mathrm{b}}_{\mathrm{sw}}(U^{\mathrm{a}}, U^{\mathrm{r}})$ will be close to its actual value $w^{\mathrm{b}}_{\mathrm{sw}}$ when the AHU valve functions properly. When the AHU valve is broken, the virtual sensor estimate $\hat{w}^{\mathrm{b}}_{\mathrm{sw}}(U^{\mathrm{a}}, U^{\mathrm{r}})$ likely differs from the measured value $w^{\mathrm{b}}_{\mathrm{sw}}$. This gives symptom $\mathcal{S}_2$ as:

$$\mathcal{S}_2 = \begin{cases} 1 & \text{if } |w^{\mathrm{b}}_{\mathrm{sw}} - \hat{w}^{\mathrm{b}}_{\mathrm{sw}}(U^{\mathrm{a}}, U^{\mathrm{r}})| > \epsilon_2 \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

Symptom $\mathcal{S}_2$ is linked to the system health as follows:

if $F^{\mathrm{a}} = 0$ **then** likely $\mathcal{S}_2 = 0$



(a) AHU on (b) AHU off

Figuur 4.12: Bayesian network representations for case study 2. During day, symptom $\mathcal{S}_1'$ is influenced by both F^{a}, F^{b}, and $T^{\mathrm{a}}_{\mathrm{ma}}$, symptom $\mathcal{S}_2$ is influenced by F^{a}, and symptom $\mathcal{S}_3$ is influenced by F^{b}. During night, when the AHU is switched off, only the relations between F^{a} and $\mathcal{S}_2$ and between F^{b} and $\mathcal{S}_3$ still hold.

if $F^{\mathrm{a}} = 1$ **then** likely $\mathcal{S}_2 = 1$

To verify whether the boiler is functioning, a straightforward approach is to compare the boiler supply water temperature $T^{\mathrm{b}}_{\mathrm{sw}}$ with its setpoint $T^{\mathrm{b}}_{\mathrm{sw,set}}$. In case of boiler nonfunctioning these two values will differ significantly. To this end, symptom $\mathcal{S}_3$ is defined as:

4

$$\mathcal{S}_3 = \begin{cases} 1 & \text{if } (T^{\mathrm{b}}_{\mathrm{sw}} - T^{\mathrm{b}}_{\mathrm{sw,set}}) < -\epsilon_3 \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

with $\epsilon_3 > 0$, which links to the system health as follows:

if $F^{\mathrm{b}} = 0$ **then** likely $\mathcal{S}_3 = 0$

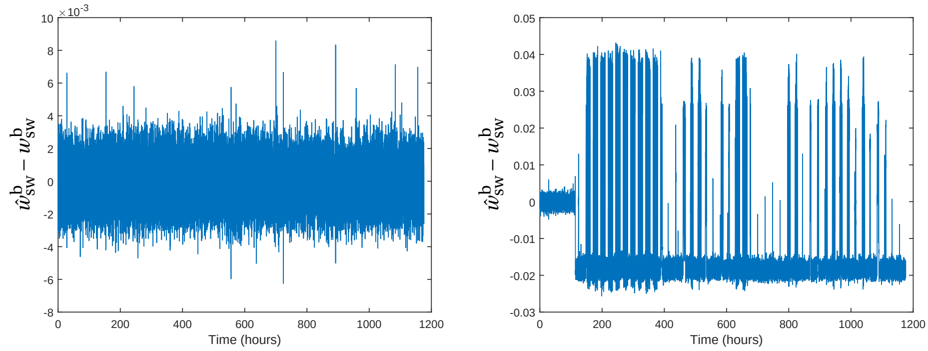
if $F^{\mathrm{b}} = 1$ **then** likely $\mathcal{S}_3 = 1$

Considering the symptoms $\mathcal{S}_1'$, $\mathcal{S}_2$, and $\mathcal{S}_3$, the diagnostic model for this case is represented by the Bayesian network in Figure 4.12. A distinction is made between two operating modes: a day mode (AHU on) and a night mode (AHU off). Fault diagnosis can be carried out in both modes.

Similarly as for case study 1, we restrict ourselves to diagnosis in the cold season, i.e. node $T^{\mathrm{a}}_{\mathrm{ma}}$ is disregarded.

Local probability distributions Before the network can be used for diagnostic inference, the following items need to be determined:

1. the values of ϵ_1, ϵ_2, and ϵ_3;
2. the conditional probability tables of $\mathcal{S}_1'$, $\mathcal{S}_2$, and $\mathcal{S}_3$;
3. the initial prior probability distributions of F^{a} and F^{b}.



Figuur 4.13: Time behaviors of $\hat{w}^{\mathrm{b}}_{\mathrm{sw}} - w^{\mathrm{b}}_{\mathrm{sw}}$: (a) healthy AHU valve; (b) stuck AHU valve.

Determination of ϵ_1, ϵ_2, and ϵ_3 The value of ϵ_1 is chosen similar as in case study 1 (as the variation of $w^{\mathrm{a}}_{\mathrm{sa}}$ is symmetrical around 20°C, there is no need to make a distinction between positive and negative deviations), i.e.

$$\epsilon_1 = 2.5$$

To determine ϵ_2, the variation in $\hat{w}^{\mathrm{b}}_{\mathrm{sw}} - w^{\mathrm{b}}_{\mathrm{sw}}$ is considered. In Figure 4.13, time behaviors of $\hat{w}^{\mathrm{b}}_{\mathrm{sw}} - w^{\mathrm{b}}_{\mathrm{sw}}$ are given for both a healthy and a stuck AHU valve. The value of ϵ_2 is chosen such that $P(S_2 = 0|F^{\mathrm{a}} = 0) = 0.99$. This is the case for

$$\epsilon_2 = 0.003$$

Finally, ϵ_3 is tuned. As the boiler supply water temperature setpoint $T^{\mathrm{b}}_{\mathrm{sw,set}}$ changes at 04.00 hours in the morning and at 18.00 hours in the evening, there is some natural difference between $T^{\mathrm{b}}_{\mathrm{sw}}$ and $T^{\mathrm{b}}_{\mathrm{sw,set}}$ shortly after these times (see Figure 4.14). Therefore, for fault diagnosis and the determination of ϵ_3, only the time intervals 04.30 till 18.00 hours and 18.30 till 04.00 hours are considered. The value of ϵ_3 is chosen such that $P(S_3 = 0|F^{\mathrm{b}} = 0) = 0.99$, i.e.:

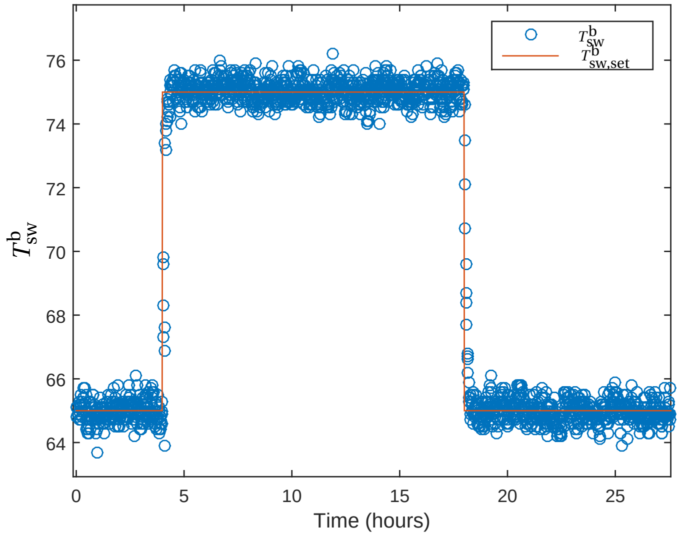
$$\epsilon_3 = 0.8$$

Conditional probability tables of $\mathcal{S}'_1, \mathcal{S}_2$, and $\mathcal{S}_3$ The conditional probability tables are defined similarly as in case study 1. The results are given in Tables 4.2 till 4.4.

Prior probability distributions of F^{a} and F^{b} The initial prior probabilities are defined similarly as for case study 1:

$$\mathrm{Pr}^0(F^{\mathrm{a}} = 1) = \mathrm{Pr}^0(F^{\mathrm{b}} = 1) = 0.01$$

[5] If the boiler is broken the temperature significantly decreases and if the fault holds for some time, this probability converges to one.



Figuur 4.14: Daily behavior of T^{b}_{sw}.

4

Tabel 4.2: Conditional probability table of $\mathcal{S}_1'$ for case study 2

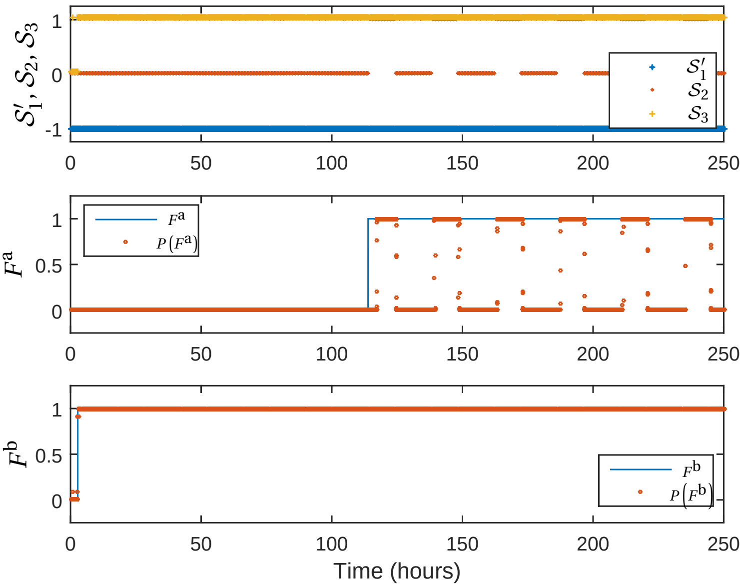
		$\mathcal{S}_1$		
F^a	F^b	-1	0	1
0	0	0.05	0.99	0.05
1	0	0.47	0.24	0.28
0	1	1[5]	0	0
1	1	1	0	0

Tabel 4.3: Conditional probability table of $\mathcal{S}_2$ for case study 2

	$\mathcal{S}_2$	
F^a	0	1
0	0.99	0.01
1	0.11	0.89

Tabel 4.4: Conditional probability table of $\mathcal{S}_3$ for case study 2

	$\mathcal{S}_3$	
F^b	0	1
0	0.99	0.01
1	0	1



Figuur 4.15: Boiler and AHU fault diagnosis example for case study 2.

Again the effect of the initial priors on the fault diagnosis is small as the likelihood functions have clearly different values for the different fault situations (see Tables 4.2, 4.3, and 4.4).

FAULT DIAGNOSIS

Consider the example (see Figure 4.15) in which the boiler breaks down immediately at the start of the simulation. Later, at $\tau = 120$, also the AHU valve gets stuck. From the simulation results, it follows that the boiler breakdown is clearly expressed in symptoms $\mathcal{S}_1'$ and $\mathcal{S}_3$, and that system health is correctly diagnosed till $\tau \approx 120$ hours, i.e., $P(F^a = 1|\mathcal{E}) \approx 0$, $P(F^b = 1|\mathcal{E}) \approx 1$, where $\mathcal{E}$ contains all observations of $\mathcal{S}_1'$, $\mathcal{S}_2$, and $\mathcal{S}_3$. When the AHU gets stuck around $\tau = 120$ hours also symptom $\mathcal{S}_2$ is activated. Because the position in which the valve got stuck is close to the desired position, symptom $\mathcal{S}_2$ is not continuously activated and the stuck valve is not continuously detected. Even though the fault is not continuously detected, the observed behavior clearly indicates the presence of an AHU valve fault.

CONCLUDING REMARKS

The proposed diagnostic model for case study 2 overcomes the limitations of the model proposed for case study 1; diagnosis is possible in all operating modes, multiple fault situations can be handled, and the model is less sensitive to high values of the mixed air temperature. Furthermore, the diagnostic model has shown to be effective in the considered simulation.

4.6.4. ALTERNATIVE SYMPTOMS FOR CASE STUDY 2

Although the diagnostic model for case study 2 results in good performance, situations may exist in which other or additional symptoms are required (e.g. in case of an absent or broken supply water temperature sensor). Therefore, we conclude this section with the proposal of two alternative symptoms for case study 2:

1. Capture and exploit the relationship between the supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$, the mixed air temperature $T^{\mathrm{a}}_{\mathrm{ma}}$, the supply water temperature $T^{\mathrm{b}}_{\mathrm{sw}}$, and the control signal to the AHU valve U^{a}. Depending on the actual system health, the AHU supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$ can be described as a function of:

$$\begin{array}{ll} T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}} & \text{if } F^{\mathrm{a}} = F^{\mathrm{b}} = 0 \\ T^{\mathrm{a}}_{\mathrm{ma}} & \text{if } F^{\mathrm{a}} = 1, F^{\mathrm{b}} = 0 \\ T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}}, T^{\mathrm{b}}_{\mathrm{sw}} & \text{if } F^{\mathrm{a}} = 0, F^{\mathrm{b}} = 1 \\ T^{\mathrm{a}}_{\mathrm{ma}}, T^{\mathrm{b}}_{\mathrm{sw}} & \text{if } F^{\mathrm{a}} = F^{\mathrm{b}} = 1 \end{array} \tag{4.11}$$

These relations follow from the energy balance (C.4), the knowledge that the thermal energy of air/water depends on its temperature and volume, and the fact that, for a healthy valve, the mass flow $w^{\mathrm{a}}_{\mathrm{sw}}$ is directly related to the control signal U^{a}. Since the exact relationships are unknown, we use this knowledge to construct two virtual sensors[6]. Multiple virtual sensors are needed since in this case, a distinction between multiple scenarios has to be made. For example, one virtual sensor $\hat{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}})$ is designed to estimate the AHU supply air temperature $T^{\mathrm{a}}_{\mathrm{sa}}$ corresponding to healthy system behavior ($F^{\mathrm{a}} = F^{\mathrm{b}} = 0$) and another one $\tilde{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}}, T^{\mathrm{b}}_{\mathrm{sw}})$ to estimate the behavior of $T^{\mathrm{a}}_{\mathrm{sa}}$ corresponding to a non-functioning boiler ($F^{\mathrm{a}} = 0$, $F^{\mathrm{b}} = 1$). Accordingly, symptom $\mathcal{S}_{\mathrm{a}_1}$ is defined as:

$$\mathcal{S}_{\mathrm{a}_1} = \begin{cases} -1 & \text{if } |T^{\mathrm{a}}_{\mathrm{sa}} - \tilde{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}}, T^{\mathrm{b}}_{\mathrm{sw}})| < \epsilon_{\mathrm{a}_1} \text{ and } |T^{\mathrm{a}}_{\mathrm{sa}} - \tilde{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}}, T^{\mathrm{b}}_{\mathrm{sw}})| \leq |T^{\mathrm{a}}_{\mathrm{sa}} - \hat{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}})| \\ 0 & \text{if } |T^{\mathrm{a}}_{\mathrm{sa}} - \hat{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}})| < \epsilon_{\mathrm{a}_1} \text{ and } |T^{\mathrm{a}}_{\mathrm{sa}} - \tilde{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}}, T^{\mathrm{b}}_{\mathrm{sw}})| > |T^{\mathrm{a}}_{\mathrm{sa}} - \hat{T}^{\mathrm{a}}_{\mathrm{sa}}(T^{\mathrm{a}}_{\mathrm{ma}}, U^{\mathrm{a}})| \\ 1 & \text{otherwise} \end{cases} \tag{4.12}$$

and linked to the system health as follows:

- **if** $F^{\mathrm{a}} = F^{\mathrm{b}} = 0$ **then** likely $\mathcal{S}_{\mathrm{a}_1} = 0$
- **if** $F^{\mathrm{a}} = 0$ and $F^{\mathrm{b}} = 1$ **then** likely $\mathcal{S}_{\mathrm{a}_1} = -1$

[6] For the sake of clarity, we restrict ourselves to two virtual sensors. In practice, an additional virtual sensor is needed to distinguish between the situation that $F^{\mathrm{a}} = F^{\mathrm{b}} = 1$ and the situation $F^{\mathrm{a}} = 1$ and $F^{\mathrm{b}} = 0$.

- **if** $F^{\text{a}} = 1$ **then** likely $\mathcal{S}_{\text{a}_1} = 1$

A possible drawback of this symptom is that it relies on the availability of historical data of fault situations for designing the virtual sensor (in this case historical data of a non-functioning boiler). However, when a good physical simulator is available, simulated data can also be used to train the virtual sensor.

2. Check whether other AHUs or radiators connected to the same boiler function properly. This strategy can be used provided that multiple systems (e.g. radiators and AHUs) are connected to the same boiler. In case of a boiler fault, also the connected systems will exhibit aberrant behavior (hierarchical dependencies, see Section 4.3.1). In the considered building configuration, one radiator is connected to the same boiler as the considered AHU. If this radiator functions properly this indicates that the boiler cannot be broken (provided that radiator heating is required). This knowledge gives rise to define symptom $\mathcal{S}_{\text{a}_2}$ as:

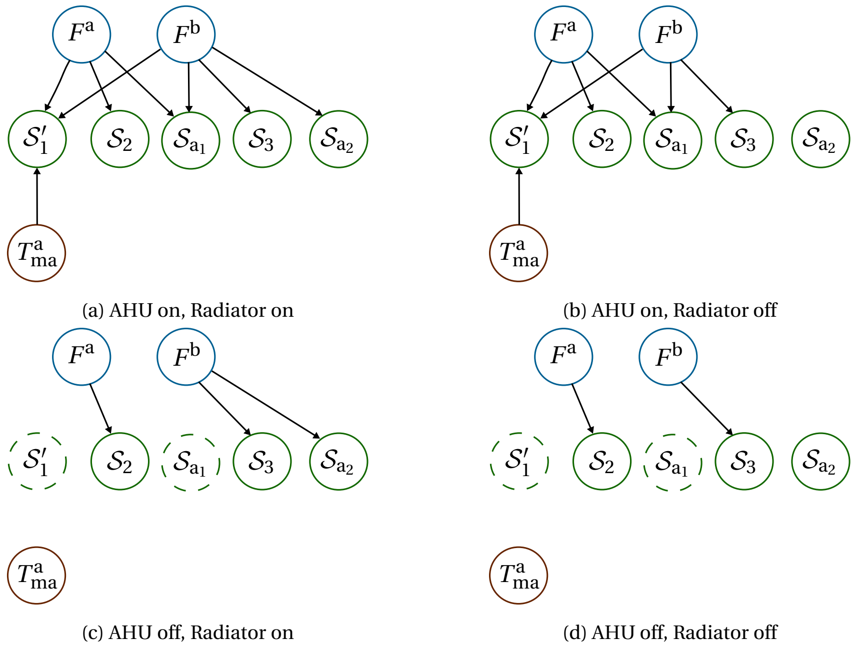
$$\mathcal{S}_{\text{a}_2} = \begin{cases} 1 & \text{if } T^{\text{z}}_{\text{a}} - T^{\text{z}}_{\text{a,set}} < \epsilon_{\text{a}_2} \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

which is linked to the system health as follows:

if $F^{\text{b}} = 0$ **then** likely $\mathcal{S}_{\text{a}_2} = 0$
if $F^{\text{b}} = 1$ **then** likely $\mathcal{S}_{\text{a}_2} = 1$

Note that it is assumed that the radiator functions properly and that this symptom is only useful when radiator heating is required.

Taking the additional symptoms $\mathcal{S}_{\text{a}_1}$ and $\mathcal{S}_{\text{a}_2}$ into account the diagnostic model is represented by the Bayesian network in Figure 4.16. Now, a distinction between four operating modes has to be made. An advantage of this model compared to the original model (see Figure 4.12) is that, due to its redundancy, fault diagnosis is also possible when one of the symptoms is missing. In addition, the redundancy can be used to detect possible sensor faults.



Figuur 4.16: Bayesian network representations for case study 2 including alternative symptoms.

4.7. DISCUSSION ON GENERALIZATION

So far, the focus was on one particular HVAC configuration. In practice, each building is different, e.g. it may have another number of zones, different types of separation between the zones, and different HVAC equipment installed to condition the building. Therefore, it is important to consider how the diagnostic model can be extended to other cases.

4.7.1. DIFFERENT HVAC EQUIPMENT

In general, a building (including HVAC system) can be represented as shown in Figure 4.2. The number of components in each layer and the way the components are connected varies from building to building. These differences influence the diagnostic model. Now, we show that even for two slightly different HVAC configurations the diagnostic model may vary. For this purpose, we assume that an additional radiator has been installed in the building setup considered before (see Figure 4.4). In the original building, a non-functioning radiator, $F^{\mathrm{r}} = 1$, will manifest itself in a too low zone temperature (provided that radiator heating is required). This gives rise to use symptom $\mathcal{S}_{\mathrm{g}_1}$, which is defined as:

$$\mathcal{S}_{\mathrm{g}_1} = \begin{cases} 1 & \text{if } T^{\mathrm{z}}_{\mathrm{a}} - T^{\mathrm{z}}_{\mathrm{a,set}} < -\epsilon_{\mathrm{g}_1} \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

and linked to the system health as:

Tabel 4.5: Variables required by each of the proposed symptoms for case study 2

symptom	required variables
$\mathcal{S}'_1$	T^{a}_{sa}, $T^{a}_{sa,set}$
$\mathcal{S}_2$	w^{b}_{sw}, U^{a}, U^{r}
$\mathcal{S}_3$	T^{b}_{sw}, $T^{b}_{sw,set}$
$\mathcal{S}_{a_1}$	T^{a}_{sa}, T^{a}_{ma}, U^{a}, T^{b}_{sw}
$\mathcal{S}_{a_2}$	U^{r}, T^{z}_{a}, $T^{z}_{a,set}$

if $F^{r} = 0$ **then** likely $\mathcal{S}_{g_1} = 0$

if $F^{r} = 1$ **then** likely $\mathcal{S}_{g_1} = 1$

In the modified building, this relation does not necessarily hold. A non-functioning radiator may be compensated for by the other radiator, provided that its capacity is sufficient. In this case, a non-functioning radiator needs to be identified in an alternative way, e.g. by verifying whether the radiator control signal U^{r} is close to control signal expected based on the outside temperature $\hat{U}^{r}(T^{o}_{a})$. This means that the Bayesian network should be extended with an extra symptom node $\mathcal{S}_{g_2}$ connected to F^{r}, with:

$$\mathcal{S}_{g_2} = \begin{cases} 1 & \text{if } |U^{r} - \hat{U}^{r}(T^{o}_{a})| > \epsilon_{g2} \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

with $\hat{U}^{r}(T^{o}_{a})$ a prediction of U^{r} based on weather information. Symptom $\mathcal{S}_{g_2}$ relates to the system health as:

if $F^{r} = 0$ **then** likely $\mathcal{S}_{g_2} = 0$

if $F^{r} = 1$ **then** likely $\mathcal{S}_{g_2} = 1$

4.7.2. DIFFERENT MONITORING VARIABLES

The symptoms proposed in this work rely on the availability of monitoring data (see Table 4.5 for an overview of the variables required by each of the proposed symptoms). The set of available monitoring signals however varies from building to building. This means that there may exist situations in which part of the monitoring data required to compute the underlying features is missing. In this case, one of the following strategies can be followed:

1. definition of alternative symptoms;
2. use of virtual sensors to estimate missing variables.

The first strategy searches for alternative symptoms that can be determined from the available monitoring signals and that can replace the missing original symptoms. Consider for example that the control signal to the radiator valve, U^{r}, is not measured, meaning that symptom $\mathcal{S}_2$ cannot be defined. In this case, another symptom is needed

4

to identify a stuck AHU heating coil valve. When both the control signal to the AHU valve U^{a}, i.e. the desired position of the valve, and its actual position X^{a} are available, a straightforward alternative symptom $\mathcal{S}_{g_3}$ is:

$$\mathcal{S}_{g_3} = \begin{cases} 1 & \text{if } |U^{\mathrm{a}} - X^{\mathrm{a}}| > \epsilon_{g_3} \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

which relates to the system health as:

if $F^{\mathrm{a}} = 0$ **then** likely $\mathcal{S}_{g_3} = 0$

if $F^{\mathrm{a}} = 1$ **then** likely $\mathcal{S}_{g_3} = 1$

In practice, the definition of adequate alternative symptoms is often not so obvious. In this case, strategy 2 becomes of interest, which aims to estimate the missing variable based on the available variables using a virtual sensor. Considering again that U^{r} is not measured, then symptom $\mathcal{S}_2$ can still be used if U^{r} can be accurately estimated based on the available data, e.g. by estimating U^{r} based on the zone air temperature $T^{\mathrm{z}}_{\mathrm{a}}$ and its setpoint $T^{\mathrm{z}}_{\mathrm{a,set}}$.

4.7.3. DIFFERENT CONTROL STRATEGIES

The way in which the different temperatures and mass flows in the HVAC system are controlled influences the diagnostic model. For example, in the considered building setup, the fan speed and so the air flow $w^{\mathrm{a}}_{\mathrm{sa}}$ through the AHU are fixed. This justifies that for symptom $\mathcal{S}_{a_1}$, only U^{a}, $T^{\mathrm{a}}_{\mathrm{ma}}$, and $T^{\mathrm{b}}_{\mathrm{sw}}$ are used as inputs for the virtual sensor. However, when the fan speed is controlled, a correct implementation of symptom $\mathcal{S}_{a_1}$ requires the mass flow rate $w^{\mathrm{a}}_{\mathrm{sa}}$ to be included as input of the virtual sensor. Indeed, when $w^{\mathrm{a}}_{\mathrm{sa}}$ varies over time, there is no fixed relation between $T^{\mathrm{a}}_{\mathrm{sa}}$ and U^{a} and $T^{\mathrm{a}}_{\mathrm{ma}}$ for a healthy system, and no fixed relation between $T^{\mathrm{a}}_{\mathrm{sa}}$ and U^{a}, $T^{\mathrm{a}}_{\mathrm{ma}}$, and $T^{\mathrm{b}}_{\mathrm{sw}}$ in case of a non-functioning boiler. Similarly, in systems where the supply water temperature $T^{\mathrm{a}}_{\mathrm{sw}}$ to the AHU is not controlled to a fixed value, this variable should be included as an input of the virtual sensor.

4.8. CONCLUSIONS

We have proposed a model-based Bayesian network approach to fault diagnosis in HVAC systems. The diagnostic model was defined using expert knowledge regarding component interdependencies and conservation laws and using historical data by the use of virtual sensors. Important properties of the proposed method are: 1. it adequately handles interdependencies between the different components, 2. diagnosis is carried out continuously in all operating modes, and 3. the method is applicable to all kinds of building setups. The importance of these properties and the applicability of the proposed method have been demonstrated based on various case studies. It is concluded that faults are timely and properly diagnosed, even in the case of multiple faults, provided that the faults result in observable behavior.

Because a different diagnostic model is required for each building and each operation mode, a lot of time and effort will be saved when the diagnostic model can be automatically generated for a class of buildings and operating modes. An interesting topic for further work would therefore be the development of methods to automate the construction of the diagnostic model.

5

MULTIPLE-MODEL APPROACH TO SYSTEM RELIABILITY PREDICTION

"What most experimenters take for granted before they begin their experiments is infinitely more interesting than any results to which their experiments lead."

-Norbert Wiener-

In recent years, a wide range of prognostic methods have been developed, aiming at predicting future system reliability and remaining useful life with the highest possible accuracy. Almost all of these methods are based on a single degradation measure, and focus on systems with only one degradation and failure mode. In practice, however, multiple degradation measures are often available and needed to adequately predict future system degradation. Moreover, systems may suffer from various kinds of faults, all resulting in different degradation behaviors. To accommodate these properties, we propose a multivariate multiple-model approach to system reliability prediction. In addition, we establish a link between failure prognosis and the subsequent maintenance optimization process. We conclude that, in the presence of multiple degradation modes and provided they are correctly identified, a multiple-model approach will outperform a single-model approach with respect to prediction accuracy. Moreover, in the presence of multiple degradation and failure modes, overall predictions of the remaining useful life as generated by common prognostic approaches are not directly suited for maintenance decision making. In contrast, our approach yields conditional predictions of future system reliability, which much better suit the subsequent maintenance optimization process.

5.1. INTRODUCTION

Although in recent years a lot of attention has been devoted to failure prognosis (see Section 5.2), failure prognosis is still an emerging research area with a number of open challenges (Peng et al., 2010; Si et al., 2011). In this chapter, we address two of them. *The first challenge* is the development of methods that can deal with multiple degradation modes and multiple degradation measures (Si et al., 2011) (Challenges P_1, P_2). Most existing methods are based on a single degradation measure and account for just one degradation mode. In practice, multiple degradation measures are often available. Moreover, systems may suffer from various kinds of faults, all resulting in different degradation behaviors. Therefore, improvement in prediction accuracy is expected when considering multiple degradation measures and accounting for variability due to different fault causes. *The second challenge* is to establish a link between failure prognosis and maintenance planning (Peng et al., 2010; Celaya et al., 2012) (Challenge O_1). Most existing prognostic methods have been developed without explicitly considering how the method is going to be used for maintenance planning (Celaya et al., 2012). Accordingly, most existing methods for condition-based maintenance planning base their maintenance decisions solely on diagnostic information, without considering prognostic information (Huynh et al., 2015). Although the link to the subsequent maintenance planning process is often overlooked, it is important in practice (Peng et al., 2010; Celaya et al., 2012).

To address the aforementioned challenges, we propose a multivariate multiple-model approach to degradation forecasting and reliability prediction. Multiple models are considered because degradation can be caused by various system faults. In general, for different system faults, system degradation evolves differently over time. So, to adequately model degradation behavior in the presence of multiple degradation modes, a distinct (multivariate) degradation model is needed for each degradation mode. Based on these models the future system reliability and remaining useful life conditional to each degradation mode can be predicted.

We consider a multivariate approach because in practice often multiple degradation measures are available and required to adequately model the degradation process. In the case of a single degradation measure, system reliability and remaining useful life follow from comparing the predicted degradation signal with a predefined threshold value. In the case of multiple degradation measures, the computations of the system reliability and remaining useful life are less straightforward. Therefore, we consider the computation of the system reliability in the multivariate case for various types of relationships among degradation measures.

For the degradation forecasting, we consider statistical approaches since they are particularly suited to manage and represent the uncertainty inherent to degradation forecasting (see Section 1.1). Good management and representation of uncertainty is of paramount importance for the subsequent decision making process. More specifically, we consider stochastic state space models, which can include most common uncertainty sources inherent to the forecasting process, i.e., temporal uncertainty, case-to-case (or sampling) variability, and measurement uncertainty (Pandey et al., 2009; Si et al., 2014). In addition, Bayesian filtering and prediction are used to estimate and forecast system degradation.

In summary, the contributions of this chapter are:

- We establish a link between failure prognosis and the subsequent maintenance decision making process;
- We propose a multiple-model approach to multivariate degradation forecasting, taking both temporal, sampling, and measurement uncertainty into account;
- For the multivariate case, we provide definitions of the (conditional) system reliability that are in line with the subsequent maintenance optimization process, together with a framework to determine these prognostic measures.

The remainder of this chapter is structured as follows: In Section 5.2, we review the literature on failure prognosis. Section 5.3 introduces the terminology, the adopted assumptions, and the research goals. In Section 5.4, two motivating cases involving reliability prediction for a railway and a climate control system are presented. Section 5.5 describes the position of failure prognosis within a condition-based maintenance scheme, and in particularly, analyzes how the prognostic result should be specified to support maintenance decision making. Next, Section 5.6 presents a multiple-model approach to degradation forecasting. In Section 5.7 we propose a framework to determine the future system reliability based on the predicted degradation measures. Section 5.8 provides a discussion, and finally in Section 5.9 conclusions and possible directions for further research are given.

5.2. RELATED WORK

Over the past few years, various prognostic methods have been proposed, ranging from model-based approaches to artificial neural networks and stochastic filtering approaches. Overviews of the various prognostic methods can be found in the reviews by Schwabacher and Goebel (2007); Peng et al. (2010); Si et al. (2011); Sikorska et al. (2011).

Especially statistical approaches have received a lot of attention in the literature thanks to their ability to handle the uncertainty inherent to the degradation forecasting process. For instance, (hidden) Markov models (Peng and Dong, 2011; Tobon-Mejia et al., 2012; Le et al., 2016) and models based on gamma (Van Noortwijk, 2009; Zhou et al., 2011) and Wiener processes (Wang, 2010; Si et al., 2013, 2014) have frequently been proposed for prognostic purposes. Nevertheless, most of the existing methods take only part of the uncertainty into account. For example, Wang et al. (2011); Si et al. (2013) omit measurement variability, while Xu et al. (2008); Sun et al. (2012) consider measurement variability, but omit the case-to-case or temporal variability. Recently, Si et al. (2014); Zheng et al. (2016) have proposed methods that take both measurement uncertainty, temporal variability, and case-to-case variability into account. In this work, we extend these methods to the multivariate case. Moreover, we address modeling uncertainty by considering a multiple-model approach. In contrast to Reuben and Mba (2014), who consider multiple models to account for different stages in the degradation process, we consider multiple models to account for different qualitative degradation behaviors associated with different fault causes. The main difference compared to the methods proposed by Le et al. (2014, 2015, 2016) is that Le et al. (2014, 2015, 2016) consider multiple hidden Markov models to account for different rates of degradation, while we consider multiple parametric models to describe fault-specific degradation behavior. Variations in degradation rates are accounted for by using stochastic model parameters, which are updated online based on observed monitoring data.

Almost all existing prognostic methods are based on a single degradation measure. An exception is the approach proposed by Lu et al. (2001). In this work, we extend the approach proposed by Lu et al. (2001) to a multiple-model approach, taking into account different degradation modes. In contrast to Lu et al. (2001), we consider a Wiener process-based degradation model accounting for sampling, measurement, and temporal variability. Moreover, we distinguish between different types of possible relationships among the degradation measures (e.g. between redundant and complementary measures).

Characteristic for most methods proposed on prognostics is that the link to the subsequent maintenance optimization process is missing. A small number of papers on maintenance optimization briefly consider the link to the prognosis process (Camci, 2009; Van Horenbeek and Pintelon, 2013; Chen et al., 2015; Huynh et al., 2015). These maintenance methods are however based on a single degradation measure and consider only one degradation mode. In this chapter, we establish a link between diagnosis, prognosis, and maintenance optimization in the case of multivariate degradation measures and multiple degradation and failure modes.

5.3. PROBLEM FORMULATION

5.3.1. TERMINOLOGY

In the sequel, the following terminology is used (see Figure 5.1). First, remember that we made a distinction between three types of system behavior:

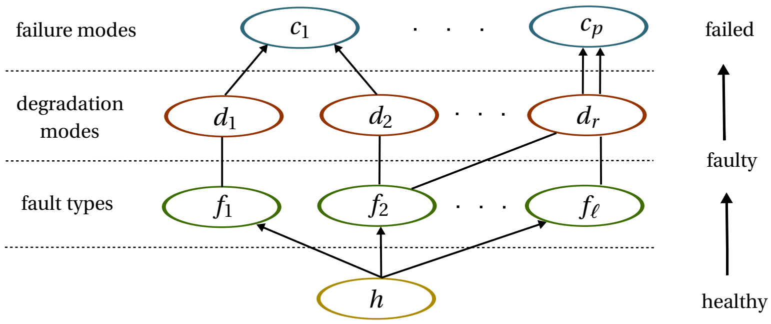
5

1. healthy behavior;
2. faulty behavior;
3. system failure.

Healthy behavior refers to the situation in which the system functions as desired. *Faulty behavior* describes the situation in which the system exhibits some aberrant behavior, but is still functional. When at least one of the system tasks can no longer be executed properly, we talk about a *system failure*. The transition from healthy behavior via faulty behavior to a system failure can take different forms, which we call *degradation modes*. So, the degradation modes d_1 till d_r (see Figure 5.1) describe possible time behaviors of system degradation.

Generally, a system can suffer from different kinds of faults f_1 through f_ℓ. The type of fault present determines to a large extent the temporal degradation behavior. Finally, a distinction is made between different failure modes c_1 till c_p. The failure mode indicates which system function is no longer executed properly.

In summary, the fault type indicates what is causing the faulty behavior, the degradation mode indicates how the system degrades over time, and the failure mode indicates which system function is (going to be) lost.



Figuur 5.1: Relationships between fault types, degradation modes, and failure modes.

5.3.2. ASSUMPTIONS

We assume that the different possible fault types f_1 till f_ℓ, the different possible degradation modes d_1 till d_r, and the different possible failure modes c_1 till c_p are given. Moreover, the temporal degradation behavior corresponding to each fault and with respect to each failure mode is assumed to be known (i.e. the connecting lines in Figure 5.1 are given).

As the focus of this chapter is on failure prognosis, we assume the availability of a diagnostic result in the form of a probability mass function over the current (i.e. for time $\tau = \tau_c$) health state: $P(H(\tau_c))$, where the health state $H(\tau_c)$ takes a value in the set $\{h, f_1, f_2, \ldots, f_\ell\}$.

For failure prognosis, we assume the availability of z degradation measures $X_\xi \in \mathbb{R}$, $\xi = 1, \ldots, z$. A *degradation measure* is a continuous variable that can be computed from sensor information and that is highly correlated with system degradation (Lu et al., 2001; Gebraeel et al., 2005). The set of degradation measures $\mathbf{X} = \{X_1, \ldots, X_z\}$ captures the system's degree of degradation. The degradation measures are linked to the failure modes and fault types as follows: The values of the degradation measures indicate to what extent the system is healthy, faulty, or in a specific failure mode. The evolution of the degradation measures over time is characteristic for the type of fault present.

5

5.3.3. GOALS

The first step of failure prognosis comprises the forecasting of the degradation measure(s) over time. As the degradation behavior varies for different fault types and from case to case (e.g. due to different environmental or operating conditions), the use of a fixed model for degradation forecasting is undesired (Gebraeel et al., 2005; Si et al., 2014). To handle case-to-case variability, methods have been proposed that model the degradation by a parametric model with stochastic parameters (Gebraeel et al., 2005; Si et al., 2014; Ompusunggu et al., 2016). Our *first goal* is to augment these single-model approaches to a multiple-model approach, where a distinct model is defined for each degradation mode. The aim is to explicitly model variability due to different fault types, so as to reduce modeling error.

The second step is to predict, based on the forecasted degradation measures, future system reliability. In univariate, single-model approaches, failure is defined as the degradation measure being larger than or equal to a predefined threshold value. In the case of multiple models and multivariate degradation measures, the definition of a failure and the associated computation of the system reliability are less trivial. Our *second goal* is to extend the threshold-based approaches to the case that we have multiple degradation measures.

As a *third goal*, we aim to explicitly make the link to the subsequent maintenance optimization process. Failure prognosis is not an isolated task, but a task within the condition-based maintenance process. We therefore analyze the dependencies between diagnosis, prognosis, and maintenance optimization. Moreover, we investigate how the prognostic result should be specified to support maintenance optimization in the case of multiple degradation and failure modes.

5.4. MOTIVATING CASES

We motivate the need for multivariate prognostic methods accounting for multiple degradation and failure modes, based on two practical examples: failure prognosis for railway tracks and failure prognosis in buildings.

5.4.1. FAILURE PROGNOSIS FOR RAILWAY TRACKS

A key component of a railway network is the track. Besides that the track provides trains with a dependable surface for their wheels to roll on, it is an essential part of the train detection process using track circuits (see Section 3.3.1 and Appendix B for details regarding railway track circuits).

All together, the railway track serves the following purposes:

1. safe and comfortable guidance of trains;
2. correct detection of a free track;
3. correct detection of an occupied track.

Accordingly, three failure modes are defined (see Figure 5.2).

The proper execution of the aforementioned tasks may be impaired by different faults, four common ones of which are:

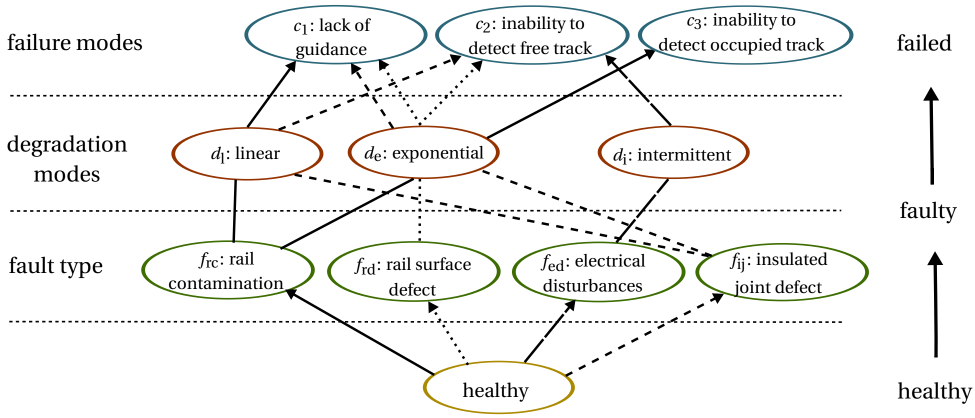
f_{rc} : rail contamination;

f_{rd} : rail surface defect;

f_{ed} : electrical disturbance;

f_{ij} : insulated joint defect.

The faults are related to the failure modes c_1 till c_3 as follows (see Figure 5.2): Contamination between the rail surface and the wheels (e.g. rust films, sand, and leafs) may hamper both the safe and comfortable guidance of trains and the correct detection of an occupied track (because the contamination hinders passing trains to shorten the circuit). Both rail surface defects and insulated joint defects may impair the safe and comfortable guidance of trains, as well as the correct detection of a free track. Finally, electrical disturbances may hamper the correct detection of a free track.



Figuur 5.2: Visualization of the relationships between the faults, degradation modes, and failure modes of a railway track.

Faults f_{rc}, f_{rd}, f_{ed}, and f_{ij} are all associated with different time behaviors of degradation (see Figure 5.2), where a distinction is made between the following three types of qualitative degradation behavior:

d_{l}: linear;

d_{e}: exponential;

d_{i}: intermittent.

From the above description, it follows that adequate degradation modeling for railway tracks requires a multivariate multiple-model approach. First, the railway track is subject to different degradation modes. For example, a false positive train detection (i.e. failure mode c_2) can be caused by both a rail defect, an insulated joint defect, or an electrical disturbance. How the degradation evolves over time depends to a large extent on the type of fault present. Therefore, multiple models are required to forecast degradation behavior. Second, multiple degradation measures are needed to adequately forecast degradation behavior. The system's ability to detect a free track is expressed in the current flowing through the track circuit receiver when the track is vacant (see Chapter 3). The system's ability to detect an occupied track is reflected in the current not flowing through the track circuit receiver when the track is occupied by a train (see Chapter 3). Among other things, the vertical axle box accelerations (Li et al., 2009; Molodova et al., 2011) provide information about the system's ability to safely and comfortably guide vehicles. So, for this application it is not possible to adequately model all degradation behaviors using just one measure, e.g. the guidance abilities cannot be assessed adequately using just electrical information, whereas the detection abilities cannot be assessed adequately based on just mechanical information.

5.4.2. FAILURE PROGNOSIS IN BUILDINGS

Heating, Ventilation, and Air Conditioning (HVAC) systems are another example of systems that fullfill multiple tasks and that are subject to different degradation modes. Without going into detail, an HVAC system serves the following purposes:

1. temperature control;
2. humidity control;
3. ventilation.

Accordingly three failure modes can be defined. Multiple faults can be identified that hinder the proper execution of one or more of these tasks. Just a few examples are (Liang and Du, 2007):

f_{mb}: malfunctioning boiler;

f_{sv}: stuck heating/cooling coil valve;

f_{df}: deteriorating supply fan;

f_{dd}: deteriorating damper (controlling the mixing ratio between outside and recirculation air).

Like for the railway example, multiple degradation measures are needed to model degradation behavior; the system's ability the control zone air temperature is expressed in zone temperature measures, while the system's ability to regulate humidity is reflected in humidity (correlated) measures, and the ventilation quality is reflected in CO_2 (correlated) measures. Because some of the faults affect multiple system goals (e.g. a deteriorating supply fan may affect all goals) the degradation behavior of the different measures may be correlated. Therefore, it is advantageous to consider multivariate degradation modeling at the system level, rather than looking at the individual tasks.

5.5. PROGNOSIS WITHIN THE CONDITION-BASED MAINTENANCE PROCESS

Condition-based maintenance aims to optimize maintenance planning by making use of real-time monitoring data. The path from the monitoring data to an optimal maintenance decision includes data pre-processing, fault diagnosis, failure prognosis, and maintenance optimization (see Section 1.1). Besides the proper implementation of the individual processes, adequate incorporation of the dependencies between the individual processes is crucial for the success of condition-based maintenance. With respect to failure prognosis, the following dependencies are of relevance:

1. dependencies between the diagnosis and prognosis processes;
2. dependencies between failure prognosis and maintenance optimization.

5.5.1. DEPENDENCIES BETWEEN DIAGNOSIS AND PROGNOSIS

Although fault diagnosis and failure prognosis concern different tasks, and are often treated individually, exploiting their mutual dependence is valuable for both diagnosis and prognosis. As outlined before, different fault types are associated with different degradation behaviors. So, information regarding the type of fault present (diagnostic result) provides information about future degradation behavior. Vice versa, information about degradation trends (prognosis result) provides information regarding the type of fault present. We propose to exploit this dependence by using the diagnostic result to determine the likelihood of each prognostic (fault-specific) model.

5.5.2. DEPENDENCIES BETWEEN PROGNOSIS AND MAINTENANCE OPTIMIZATION

The prognostic result serves (together with the diagnostic result) as an input for the maintenance optimization process. It is therefore important to ensure that the prognostic result is specified such that it facilitates maintenance optimization. This requires some insight in the maintenance optimization process.

MAINTENANCE OPTIMIZATION

Maintenance optimization is a typical example of a decision task subject to risk and uncertainty: we have uncertainty regarding the current and future system health, and consequently, we have the risk of making non-optimal maintenance decisions. In the presence of risk and uncertainty, decisions are commonly made based on the expected utility theory (Lindley, 1985), which is a framework for determining the best (maintenance) decision given probabilistic information regarding the actual situation[1] (see Section A.1.4 for more details regarding the expected utility theory).

In contrast to common maintenance optimization methods, which limit the maintenance optimization task to deciding whether or not to perform preventive maintenance at a particular time instant, we augment this task with deciding on:

1. the required type of maintenance;

2. the optimal time to perform maintenance.

So, the possible maintenance decisions are:

$d_{0,\infty}$: do nothing;

$d_{a,t}$: perform maintenance activity $a \in A$ at time $t \in T$,

with a and t, in turn, decision variables, A the finite set of possible maintenance activities, and T the discrete set of available maintenance time instants.

For this maintenance optimization task, and given the following cost functions:

$C_{\mathrm{m}}(\cdot)$: function of a and t, expressing the lifetime-averaged direct costs associated with maintenance action a at time t;

[1] For the maintenance optimization case, the situation is defined by the current and future system health.

5

$C_{\mathrm{i}}(\cdot)$: function of a and t, expressing the lifetime-averaged indirect costs of maintenance (e.g. related to downtime) associated with action a at time t;

C_{c_l}: additional costs of a failure in mode c_l;

$C_{f_j}(\cdot)$: function of a expressing the penalty costs of preparing a (wrong) maintenance activity a in the case of fault type f_j,

the expected utilities are computed as:

$$\mathbb{E}(u|a,t) = -\left(\sum_{l=1}^{p}\sum_{j=1}^{\ell} P(H(t)=f_j)P\big(\mathcal{F}_l(t)=1|H(t)=f_j\big)C_{c_l} + \sum_{j=1}^{\ell} P(H(t)=f_j)C_{f_j}(a) + C_{\mathrm{m}}(a,t) + C_{\mathrm{i}}(a,t)\right) \tag{5.1}$$

The first term expresses the costs related to the risk of maintenance time t being too late to avoid a particular failure, with $\mathcal{F}_l(t)$ a binary variable indicating whether the system fails in mode l at maintenance time t. The second term expresses the costs related to the risk of maintenance action a being not appropriate to repair the system.

To compute $\mathbb{E}(u|a,t)$, next to the cost functions, the probabilities $P(H(t)=f_j)$ and $P(\mathcal{F}_l(t)=1|H(t)=f_j)$ need to be known for $j=1,\ldots,\ell$ and $l=1,\ldots,p$. The probabilities $P(H(t)=f_j)$ reflect the diagnostic result (see Section 5.3.2), the probabilities $P(\mathcal{F}_l(t)=1|H(t)=f_j)$ refer to the prognostic result.

SPECIFICATION OF THE PROGNOSTIC RESULT

From the analysis of the maintenance optimization process, we conclude that, in the case of multiple degradation and failure modes, the prognosis process should output the functions $P^j_{\mathrm{func},l}(\cdot)$ defined by:

$$P^j_{\mathrm{func},l}(\tau) = P\big(\mathcal{F}_l(\tau)=1|H(\tau)=f_j\big), \quad \forall j\in\{1,\ldots,\ell\}, \forall l\in\{1,\ldots,p\} \tag{5.2}$$

where $P^j_{\mathrm{func},l}(\tau)$ indicates the probability of a failure in mode c_l at time τ conditional to degradation mode d_j.

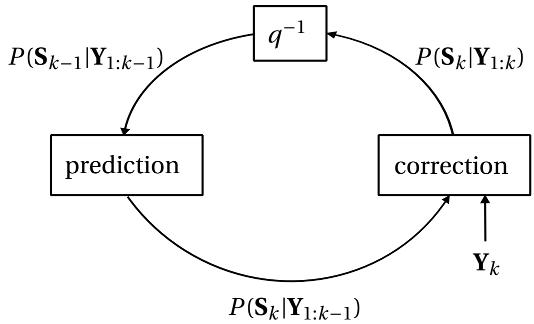
5.6. DEGRADATION MODELING AND FORECASTING

5.6.1. MULTIPLE-MODEL DEGRADATION MODELING

For each fault f_j, $j=1,\ldots,\ell$, the corresponding time behavior of a z-dimensional degradation process $\{\mathbf{X}(\tau)=[X_1(\tau),\ldots,X_z(\tau)]^\top, \tau\geq 0\}$ is described by a Wiener process[2] plus (nonlinear) drift, i.e.

$$\mathbf{X}(\tau) = m_j\big(\tau,\theta_j(\tau)\big) + \sigma_j B(\tau) \tag{5.3}$$

[2]Wiener processes are considered because they can model non-monotonic degradation behavior, which is often encountered in practice (Gebraeel et al., 2005; Crowder and Lawless, 2007; Si et al., 2014). In case of monotonic degradation behavior, gamma or compound Poisson processes can be used instead.



Figuur 5.3: Bayesian filtering. At each time step k, first, the state is estimated based on the model (prediction step). Next, this estimate is updated based on the current measurements $\mathbf{Y}_k$ (correction step).

with $\sigma_j B(\tau) = [\sigma_{j,1},\ldots,\sigma_{j,z}]^\top B(\tau)$ a Wiener process, i.e. $B(\tau)$ represents a standard Brownian motion with $\sigma_{j,\xi} B(\tau) \sim N(0, \sigma^2_{j,\xi}\tau))$. Models m_1 till m_ℓ are z-dimensional vectors the elements of which are (nonlinear) mappings expressing non-decreasing degradation trends (e.g. linear (Si et al., 2014), exponential (Gebraeel et al., 2005), quasi-linear/asymptotic (Ompusunggu et al., 2016)) associated with the corresponding fault mode f_j. The vector $\theta_j(\tau) \in \mathbb{R}^{n_{\theta_j}}$ denotes the model parameters, which might be stochastic. Here we assume $\theta_j(\tau) \sim N(\mu_{\theta_j}, \Sigma_{\theta_j})$. Information regarding the degradation process is obtained through noise-disturbed measurements, i.e.:

$$\mathbf{Y}(\tau) = \mathbf{X}(\tau) + \epsilon(\tau) \tag{5.4}$$

with $\{\mathbf{Y}(\tau) = [Y_1(\tau),\ldots,Y_z(\tau)]^\top, \tau \geq 0\}$ the process describing the time behavior of the measurements, and $\epsilon(\tau) = [\epsilon_1(\tau),\ldots,\epsilon_z(\tau)]^\top$, with $\epsilon_\xi(\tau) \sim N(0,\gamma_\xi^2)$. It is assumed that the random variables ϵ, θ_j, and $B(\tau)$ are mutually statistically independent.

The proposed degradation model (5.3)-(5.4) can describe a wide range of degradation trends, and captures both temporal, sampling, and measurement uncertainty (Si et al., 2014). *Temporal uncertainty*, which is the uncertainty associated with the progression of the degradation over time, is characterized by the dynamics of the Brownian motion $\{B(\tau), \tau \geq 0\}$. *Sampling (or case-to-case) variability* characterizes the heterogeneity among the degradation paths of different systems under different operation conditions, and is represented in (5.3) by the stochastic parameters $\theta_j(\tau)$. Finally *measurement uncertainty* is reflected by the error term $\epsilon(\tau)$ in (5.4), and reflects the fact that the degradation cannot be perfectly measured, i.e. the measurements are disturbed by measurement errors arising e.g. from non-ideal measurement instruments. Moreover, in our modeling framework, *modeling uncertainty* is minimized by considering a separate model for each fault cause.

5.6.2. ONLINE UPDATING AND FORECASTING

Suppose the degradation process is monitored at times $\tau_1 < \tau_2 < \tau_3 < \ldots$ and let $\mathbf{Y}_k = \mathbf{Y}(\tau_k)$ denote the observation vector at time τ_k. The sequence of measurement vectors

5

$\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_k$ is represented by $\mathbf{Y}_{1:k}$ and the corresponding sequence of degradation measures is represented by $\mathbf{X}_{1:k}$, with $\mathbf{X}_k = \mathbf{X}(\tau_k)$. At time τ_k the goal is to estimate the current degradation state $\mathbf{X}(\tau_k)$ and to predict the evolution of the degradation measure $\mathbf{X}(\tau_q)$ for $\tau_q > \tau_k$ based on the model (5.3)-(5.4) and observations $\mathbf{Y}_{1:k}$. For that purpose, we rewrite the model (5.3)-(5.4) as a discrete-time stochastic state space model:

$$\underbrace{\begin{pmatrix} \mathbf{X}_k^j \\ \theta_{j,k} \end{pmatrix}}_{\mathbf{S}_k^j} = \begin{pmatrix} \mathbf{X}_{k-1}^j + m_j(\tau_k, \theta_{j,k-1}) - m_j(\tau_{k-1}, \theta_{j,k-1}) + v_k^j \\ \theta_{j,k-1} \end{pmatrix} \tag{5.5a}$$

$$\mathbf{Y}_k = \mathbf{X}_k^j + \epsilon_k, \quad \text{for system in degradation mode } j \tag{5.5b}$$

with $\mathbf{S}_k^j \in \mathbb{R}^{z+n_{\theta_j}}$ the state vector at τ_k according to model j, which is composed of the degradation measure $\mathbf{X}_k^j \in \mathbb{R}^z$ and the parameter vector $\theta_{j,k} \in \mathbb{R}^{n_{\theta_j}}$, $\mathbf{Y}_k \in \mathbb{R}^z$ the measurements, $v_k^j \sim N(0, \text{diag}(\sigma_{j,1}^2(\tau_k - \tau_{k-1}), \ldots, \sigma_{j,z}^2(\tau_k - \tau_{k-1})))$, and ϵ_k the realization of ϵ at τ_k. Equation (5.5a) is the transition equation, which specifies how each element of the state vector evolves over time according to degradation model j. Equation (5.5b) is the output equation, specifying how the measurements are linked to the system states. In this formulation, degradation forecasting can be considered as a state estimation and prediction problem, where the goal is to estimate and predict the state, so to statistically minimize the state error. This is a common problem that can be solved using Bayesian filtering (Kaipio and Somersalo, 2006). The Bayesian approach to statistics attempts to utilize all available information, i.e. it combines new information with existing knowledge, in order to reduce uncertainty. The formal mechanism to combine new information with existing knowledge is known as Bayes' theorem (Kaipio and Somersalo, 2006). Roughly, this information fusion consists of two steps: a prediction step based on the state transition equation, and a correction step based on new measurements (see Figure 5.3).

Different types of Bayesian filters have been proposed, among which the Kalman filter (Kalman, 1960), its nonlinear extensions, i.e. the extended and unscented Kalman filter, and particle filters (Del Moral, 1996). The choice for a filter depends on the exact form of the model (5.5) and on computational constraints. When the transition and output equation are linear in the state, and the process and measurement noise are additive and Gaussian, the Kalman filter is the optimal filter. When the linearity assumptions are violated (but the noise assumptions are satisfied), an extended or unscented Kalman filter can be used. Another possibility is to use a particle filter, a Monte Carlo methodology, which can also be used when the noise is non-Gaussian or non-additive. The performance of a particle filter depends on the number of particles used. In general, when enough particles are used, a particle filter outperforms the extended and unscented Kalman filter in terms of estimation accuracy and robustness, but at the costs of higher computational demands (Chatzi and Smyth, 2009; Rigatos, 2010; György et al., 2014).

Because of the attractive properties of the Kalman filter (e.g. computational efficiency, analytic solutions), work has been devoted to transform nonlinear degradation data to an approximate linear form. Examples of such transformations are the log transformation (Gebraeel et al., 2005) and the time-scale transformation (Whitmore and Schenkelberg, 1997). This way, analytic solutions can be obtained for the approximate linear degradation process in a computationally efficient way, however, at the cost of modeling error. So, for nonlinear degradation processes a trade-off needs to be made between modeling accuracy and solution accuracy. This trade-off is application-specific and a further elaboration is beyond the scope of this thesis.

For clarity of presentation, in the remainder, we consider the case that the degradation process can be accurately described by a linear stochastic state space model as considered by Zheng et al. (2016), i.e. for all j, model m_j is of the form:

$$m_j(\tau_k, \theta_{j,k}) = \beta_j(\tau_k, \phi_j)\theta_{j,k} \tag{5.6}$$

with $\phi_j \in \mathbb{R}^{n_{\phi_j}}$ a vector of deterministic parameters, $\theta_{j,k} \in \mathbb{R}^{n_{\theta_j}} \sim N(\mu_{\theta_j}, \Sigma_{\theta_j})$ a vector of stochastic parameters, and β_j a $z \times n_{\theta_j}$ matrix. In this case, model (5.5) can be written in a linear form:

$$\mathbf{S}_k^j = \begin{pmatrix} \mathbf{X}_k^j \\ \theta_{j,k} \end{pmatrix} = A_{j,k}\mathbf{S}_{k-1}^j + \eta_k^j \tag{5.7a}$$

$$\mathbf{Y}_k = C\mathbf{S}_k^j + \epsilon_k \tag{5.7b}$$

with:

$$\begin{aligned}
A_{j,k} &= \begin{bmatrix} I & \beta_j(\tau_k,\phi_j) - \beta_j(\tau_{k-1},\phi_j) \\ 0 & I \end{bmatrix} \\
\eta_k^j &= \begin{bmatrix} v_k^j \\ 0 \end{bmatrix} \sim N(0, Q_{j,k}) \\
Q_{j,k} &= \begin{bmatrix} \mathrm{diag}\big(\sigma_{j,1}^2(\tau_k - \tau_{k-1}), \ldots, \sigma_{j,z}^2(\tau_k - \tau_{k-1})\big) & 0 \\ 0 & 0 \end{bmatrix} \\
C &= \begin{bmatrix} I & 0 \end{bmatrix} \\
\epsilon_k &\sim N(0, R) \\
R &= \mathrm{diag}(\gamma_1^2, \ldots, \gamma_z^2)
\end{aligned}$$

Procedure 4 outlines the degradation estimation and forecasting based on the Kalman filter.

Procedure 4 Multiple-model degradation estimation and prediction at time τ_k.

Input: Previous states $\mathbf{S}^j(k-1|k-1)$, previous covariance matrices $P^j(k-1|k-1)$, and matrices $A_{j,k}$ and $Q_{j,k}$, for $j = 1,\dots,\ell$; matrices C and R; failure criteria

Measure $\mathbf{Y}_k$

for $j = 1,\dots,\ell$ **do**

Estimation of current degradation

Prediction step:

$$\mathbf{S}^j(k|k-1) = A_{j,k}\mathbf{S}^j(k-1|k-1)$$
$$P^j(k|k-1) = A_{j,k}P^j(k-1|k-1)A_{j,k}^\top + Q_{j,k}$$

Correction step:

$$K^j(k) = P^j(k|k-1)C^\top\left(CP^j(k|k-1)C^\top + R\right)^{-1}$$
$$\mathbf{S}^j(k|k) = \mathbf{S}^j(k|k-1) + K^j(k)\left(\mathbf{Y}_k - C\mathbf{S}^j(k|k-1)\right)$$
$$P^j(k|k) = \left(I - K^j(k)C\right)P^j(k|k-1)$$

Prediction of future degradation

$n = 0$

while failure criteria are not met **do**

$n \leftarrow n+1$

n-step ahead prediction:

$$\mathbf{S}^j(k+n|k) = \left(A_{j,k}\right)^n\mathbf{S}^j(k|k)$$
$$P^j(k+n|k) = \left(A_{j,k}\right)^n P(k|k)(A_{j,k}^\top)^n + \sum_{l=0}^{n-1}(A_{j,k})^l Q_{j,k}\left(A_{j,k}^\top\right)^l$$

end while

end for

Output: predictions of the degradation measure $\mathbf{X}(\tau_q)$ for $q = k, k+1,\dots,k+n$

5.7. SYSTEM RELIABILITY PREDICTION

5.7.1. MULTIVARIATE DEFINITIONS

Two prognostic measures are (future) system reliability and remaining useful life (Engel et al., 2000; Lu et al., 2001; Si et al., 2011, 2014). Although the remaining useful life is most commonly used, in this thesis we focus on the system reliability. We made this choice because this measure best fits the subsequent maintenance optimization process (see

Section 5.5). Before elaborating on the system reliability, we define system failure in the multivariate case.

In the univariate case, system failure is usually defined as the degradation measure $\mathbf{X}(\cdot)$ being larger than or equal to a predefined threshold λ, i.e.:

$$\mathbf{X}(\tau)\begin{cases} <\lambda & \Longrightarrow \text{system is functional at } \tau \\ \geq \lambda & \Longrightarrow \text{system fails at } \tau \end{cases} \tag{5.8}$$

Failure definition (5.8) can be extended to the multivariate case by defining a failure in mode c_l as $g_l(\mathbf{X}(\cdot))$ being larger than or equal to a predefined threshold λ_l, i.e.:

$$g_l\big(\mathbf{X}(\tau)\big)\begin{cases} <\lambda_l & \Longrightarrow \text{system does not fail in mode } c_l \text{ at } \tau \\ \geq \lambda_l & \Longrightarrow \text{system fails in mode } c_l \text{ at } \tau \end{cases} \tag{5.9}$$

with $g_l(\cdot)$, $l = 1,\ldots,p$, application-specific functions, which we will elaborated on in Section 5.7.2. Accordingly, system failure is defined as:

$$\begin{array}{ll} g_1\big(\mathbf{X}(\tau)\big) < \lambda_1 \text{ and} \ldots \text{and } g_p\big(\mathbf{X}(\tau)\big) < \lambda_p & \Longrightarrow \text{system is functional at } \tau \\ g_1\big(\mathbf{X}(\tau)\big) \geq \lambda_1 \text{ or} \ldots \text{or } g_p\big(\mathbf{X}(\tau)\big) \geq \lambda_p & \Longrightarrow \text{system fails at } \tau \end{array} \tag{5.10}$$

The *system reliability* (P_{func}) is defined as the probability that the system is functional, i.e. does not fail (Lu et al., 2001). Based on (5.9), the system reliability with respect to failure mode c_l at time τ is the probability that $g_l(\mathbf{X}(\tau))$ is smaller than λ_l, i.e.

$$P_{\text{func},l}(\tau) = p\Big(g_l\big(\mathbf{X}(\tau)\big) < \lambda_l\Big) \tag{5.11}$$

The overall system reliability at time τ is defined as the probability that the system is functional, i.e. is not in any of the failure modes c_1 till c_p:

$$P_{\text{func}}(\tau) = p\Big(g_1\big(\mathbf{X}(\tau)\big) < \lambda_1,\ldots,g_p\big(\mathbf{X}(\tau)\big) < \lambda_p\Big) \tag{5.12}$$

5

From these definitions, we conclude that the functions $P^j_{\text{func},l}$ as defined in (5.2) correspond to predictions of the system reliability with respect to failure mode l conditional to degradation mode j. So, in accordance with the above definitions, (5.2) can be written as:

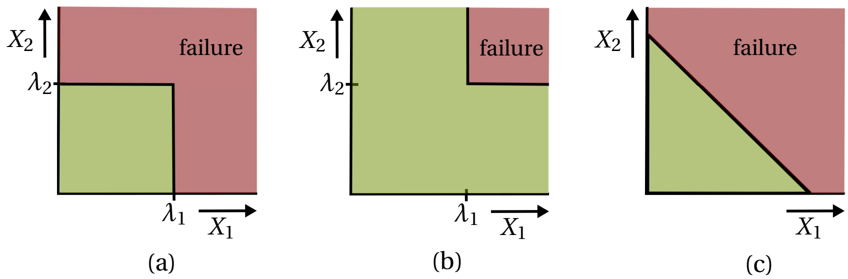
$$P^j_{\text{func},l}(\tau) = p\Big(g_l\big(\mathbf{X}(\tau)\big) < \lambda_l \Big| H(\tau) = f_j\Big) \tag{5.13}$$

5.7.2. DETERMINATION OF FAILURE DEFINITION AND SYSTEM RELIABILITY

The functions $g_l(\cdot)$, $l = 1,\ldots,p$, used to define system failure (5.9) are application-specific and depend on the relationships between the degradation measures X_ξ, $\xi = 1,\ldots,z$. Here, we focus on three common types of relationships (see Figure 5.4 for 2-D example relationships):

1. complementary measures:

   (a) independently assessable;



Figuur 5.4: 2-D illustration of three types of failure definitions. (a) independently-assessable complementary measures, (b) redundant measures, (c) not independently-assessable complementary measures.

   (b) not independently assessable;

2. redundant measures.

Measures X_ξ, $\xi = 1,\ldots,z$, are *complementary* and *independently assessable* if it holds that the system is functional in mode c_l if and only if each measure X_ξ is below an individual threshold $\lambda_{l,\xi}$. In the case of *redundant measures*, only k out of z, $k < z$, of components X_ξ need to be below their threshold $\lambda_{l,\xi}$ for the system to be functioning in mode c_l. For *complementary, not independently-assessable measures*, no relevant individual thresholds exist. For the system to be functioning, all functions $g_l(\cdot)$ of $\mathbf{X}$ should then just be below an overall threshold value λ_l.

For brevity, in the sequel we omit the subscript l whenever the explicit reference to a particular failure mode is not necessary. For the same reason, we omit the time argument τ whenever possible.

5

INDEPENDENTLY-ASSESSABLE COMPLEMENTARY MEASURES

For measures that are complementary and independently assessable, the functions $g_l(\cdot)$, $l = 1,\ldots,p$, can be chosen arbitrarily, as long as they satisfy:

$$g_l(\mathbf{X}) \begin{cases} \geq \lambda_l & \text{if } \max\left(X_1 - \lambda_{l,1},\ldots,X_z - \lambda_{l,z}\right) \geq 0 \\ < \lambda_l & \text{otherwise} \end{cases} \tag{5.14}$$

The system reliability with respect to failure mode c_l is computed as:

$$P_{\text{func},l} = \int_{-\infty}^{\lambda_{l,1}} \int_{-\infty}^{\lambda_{l,2}} \ldots \int_{-\infty}^{\lambda_{l,z}} p(X_1, X_2,\ldots,X_z)\,dX_z \ldots dX_2\,dX_1 \tag{5.15}$$

with $p(\cdot)$ the distribution function of the degradation measure $\mathbf{X}$, which follows from the degradation modeling and forecasting (see Section 5.6).

REDUNDANT MEASURES

Safety-critical systems are often equipped with redundancy, e.g. airplanes having more engines than necessarily for take-off. Systems can be redundant to varying degrees. The redundancy is lowest when $z-1$ out of z components need to be functioning for the whole system to be functioning, and highest when only 1 out of z components needs to be functioning for the whole system to be functioning. If the functioning of each redundant component is reflected by one degradation measure X_ξ, then for a k-out-of-z:G system[3] at least k out of z measures X_ξ, $\xi = 1,\dots,z$, need to be below their threshold $\lambda_{l,\xi}$ for the system to be functioning in mode c_l. So, $g_l(\cdot)$ should be chosen such that:

$$g_l^{(k)}(\mathbf{X}) \begin{cases} < \lambda_l & \text{if } \sum_{\xi=1}^{z} \alpha_{l,\xi} \geq k \\ \geq \lambda_l & \text{otherwise} \end{cases} \tag{5.16}$$

with:

$$\alpha_{l,\xi} = \begin{cases} 1 & \text{if } X_\xi < \lambda_{l,\xi} \\ 0 & \text{otherwise} \end{cases}$$

and the superscript (k) indicating that we consider a k-out-of-z:G system. The system reliability with respect to failure mode c_l is computed as:

$$\begin{aligned} P_{\text{func},l}^{(k)} &= \int\int \dots \int_{\Omega_l^{(k)}} p(X_1, X_2, \dots, X_z)\, dX_z \dots dX_2\, dX_1 \\ &= \underbrace{\sum_{\iota=1}^{\frac{z!}{k!(z-k)!}} \int\int \dots \int_{\Omega_{l,\iota}^{(k)}} p(X_1, X_2, \dots, X_z)\, dX_z \dots dX_2\, dX_1}_{R_l^{(k)}} - kR_l^{(k+1)} \end{aligned} \tag{5.17}$$

with $\Omega_l^{(k)} = \Omega_{l,1}^{(k)} \cup \dots \cup \Omega_{l,\frac{z!}{k!(z-k)!}}^{(k)}$ the integration surface representing the subset of $\mathbf{X} \in \mathbb{R}^z$ for which the system is not in failure mode c_l and $\iota = 1,\dots,\frac{z!}{k!(z-k)!}$ the different configurations for which k degradation measures are in their desired region, with $\Omega_{l,\iota}^{(k)}$ the corresponding surfaces. The last term $kR_l^{(k+1)}$ corrects for the overlap between the integration surfaces associated with the different configurations $\iota = 1,\dots,\frac{z!}{k!(z-k)!}$. To illustrate, Figure 5.5 gives the integration surfaces $\Omega^{(3)}$, $\Omega^{(2)}$, and $\Omega^{(1)}$ for a three-dimensional case, which are defined as:

$$\Omega^{(3)} = \{\mathbf{X} \in \mathbb{R}^3 : X_1 < \lambda_1 \text{ and } X_2 < \lambda_2 \text{ and } X_3 < \lambda_3\}$$

$$\Omega^{(2)} = \{\mathbf{X} \in \mathbb{R}^3 : (X_1 < \lambda_1 \text{ and } X_2 < \lambda_2) \text{ or } (X_1 < \lambda_1 \text{ and } X_3 < \lambda_3) \text{ or } (X_2 < \lambda_2 \text{ and } X_3 < \lambda_3)\}$$

$$\Omega^{(1)} = \{\mathbf{X} \in \mathbb{R}^3 : X_1 < \lambda_1 \text{ or } X_2 < \lambda_2 \text{ or } X_3 < \lambda_3\}$$

NOT INDIVIDUALLY-ASSESSABLE COMPLEMENTARY MEASURES

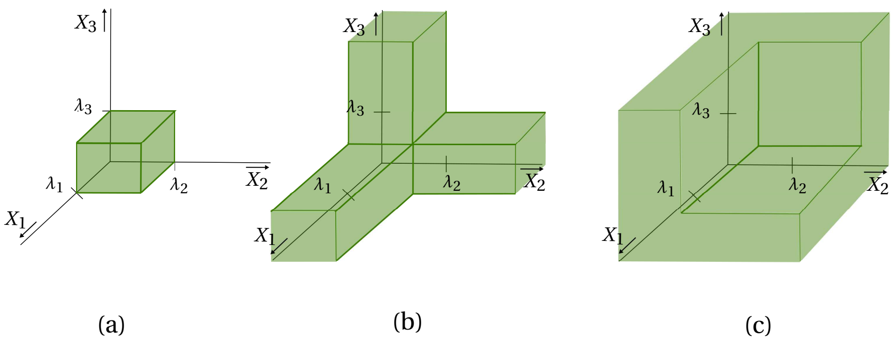
In practice, it is common that the functioning of a system is defined as a combination of the degradation measures satisfying a certain criterion, e.g. the sum or product of measures X_ξ, $\xi = 1,\dots,z$, should be below a threshold. In such situations, system reliability cannot be assessed based on individual threshold values. However, in such cases, the critical surface is generally smooth and can be written in the form:

$$s_{\text{cr}}(X_1, X_2, \dots, X_z) = 0 \tag{5.18}$$

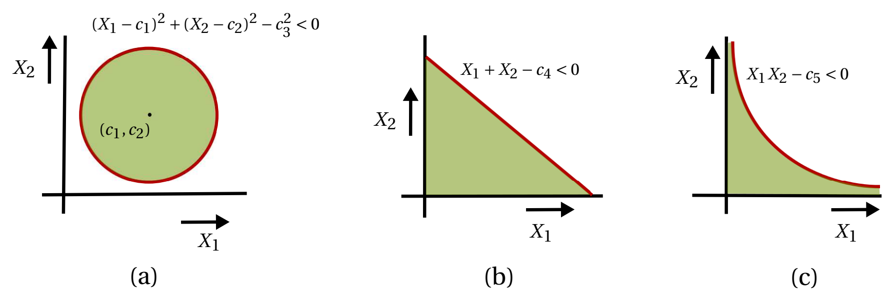
with $s_{\text{cr}}(\cdot)$ a continuous function (see Figure 5.6 for some two-dimensional example surfaces and the associated functions $s_{\text{cr}}(\cdot)$). In this case, the integration bounds directly

[3] A k-out-of-z:G system is a system that works well if at least k-out-of-z components work well.

5



Figuur 5.5: 3-D visualization of the integration surfaces indicating the subsets of $\mathbf{X} \in \mathbb{R}^3$ for which the system is functional: (a) 3-out-of-3: G system; (b) 2-out-of-3: G system, (c) 1-out-of-3: G system.



Figuur 5.6: 2-D example surfaces indicating the subset of instances of $\mathbf{X} = [X_1\, X_2]^\top$ for which the system is functional.

follow from $s_{\mathrm{cr}}(\cdot)$.

CONCLUDING REMARK

In the multivariate case, the failure definition and the associated computation of the system reliability depend on the relationships among the degradation measures. Three common relationships have been discussed. In general, the dependencies among degradation measures do not always fall within one category. Consider for example a system with four degradation measures X_1 till X_4 and failure defined as:

$$g(X_1, X_2) > \lambda_1 \text{ and } g'(X_3, X_4) > \lambda_2 \implies \text{system failure} \tag{5.19}$$

For this system, we have to deal with both redundant and not individually-assessable complementary measures. In such cases the strategies discussed before can be combined.

5.8. EVALUATION AND DISCUSSION

A realistic and thorough evaluation of the proposed approach is only possible for a particular application and in combination with a fault diagnosis and maintenance optimization approach. Such an evaluation goes beyond the scope of this thesis. In this section, we will reflect on two main attributes of the proposed approach, namely:

1. its position within the condition-based maintenance process;
2. the added value of using multiple models over using a single model on the prediction quality, and its dependence on the diagnostic result.

5.8.1. POSITION WITHIN THE CONDITION-BASED MAINTENANCE PROCESS

Procedure 5 summarizes the proposed prognosis strategy within a condition-based maintenance process. Although we ensure that the different processes are compatible with each other in the sense that the diagnostic and prognostic results support maintenance optimization, we do not impose further requirements on the diagnosis and maintenance optimization process. In particular, we only assume that the diagnosis process outputs a probability distribution over the system health state, and that decision making is done based on expected utilities. Even when the diagnostic result is specified using another uncertainty formalism (e.g. possibilities, fuzzy measures, mass functions) the proposed strategy is of use. In this case the alternative uncertainty distribution first has to be transformed into a probability distribution. For this task, transformation rules are available in literature (Dubois et al., 1993; Cobb and Shenoy, 2003a). Moreover, if desired, another multivariate multiple-model forecasting algorithm can be used instead of the forecasting strategy outlined in Procedure 4. For example a multiple-model method based on gamma processes (Van Noortwijk, 2009; Zhou et al., 2011) in the case that degradation behavior is best described by a monotonic process. We regard the freedom to independently select an optimal diagnosis and forecasting algorithm as a practical advantage. Indeed, problem characteristics vary widely among applications, and so the best combination of diagnosis and prognosis approach is highly application-specific. As another advantage, we regard the fact that the maintenance optimization model we rely on is based on cost functions that are easily assessed by practitioners (e.g., costs of maintenance, costs associated with a failure, costs associated with downtime).

Procedure 5 Prognosis within condition-based maintenance at time τ_k.

Input: Failure functions $g_l(\cdot)$ and thresholds λ_l for $l = 1, \ldots, p$, set T of possible maintenance time instants

Fault diagnosis

generates $P\big(H(\tau_k)\big)$

Prognosis

for $t \in T, t > \tau_k$ **do**
  for $j = 1, \ldots, \ell$ **do**
    Determine $\mathbf{X}^j(t)$ using Procedure 4
    for $l = 1, \ldots, p$ **do**
      Determine $P^j_{\text{func},l}(t)$:

$$P^j_{\text{func},l}(t) = \int\int \ldots \int_{\Omega_l} p\big(\mathbf{X}^j(t)\big) dX_z \ldots dX_2 dX_1$$

      with $\Omega_l \in \mathbf{X}$ the surface for which $g_l(\mathbf{X}) < \lambda_l$
    end for
  end for
end for

Maintenance optimization

Based on $P(H(\tau_k))$ and $P^j_{\text{func},l}(t)$ $\forall t \in T, t > \tau_k$, $\forall j \in \{1, \ldots, \ell\}$, $\forall l \in \{1, \ldots, p\}$, determine the optimal maintenance decision, e.g. according to (5.1)-(5.2)

Output: Maintenance decision for τ_k

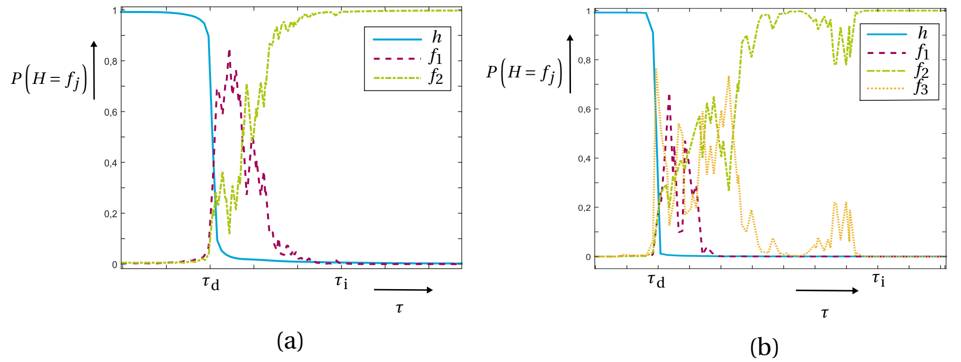
5.8.2. MULTIPLE MODELS

We motivated our choice for a multiple-model approach by the fact that system degradation may evolve differently over time for different types of faults. For example, for the railway case (see Figure 5.2) the ability to detect a free track decreases approximately linearly over time in the case of an electrical insulation problem, while the temporal degradation behavior is best described by an exponential model when a rail surface defect is present. Therefore, a natural choice is to use a linear model to forecast degradation resulting from an insulation problem, while using an exponential model to forecast degradation behavior as a consequence of a rail surface defect.

We conclude that a multiple-model approach has the potential to outperform a single-model approach with respect to prediction accuracy. We say "has the potential to" because the actual prediction performance of a multiple-model approach heavily relies on knowledge of the underlying degradation mode. In practice, we do not know with certainty which fault is present, and so which model best describes degradation behavior. This means that for online degradation forecasting the potential improvement in prediction accuracy cannot be fully utilized. Whether and to what extent a multiple-model approach will outperform a single-model approach with respect to prediction accuracy depends, among other things, on the accuracy of the diagnostic result. Although promising fault diagnosis methods have been proposed in the literature (see e.g. (Chen and Patton, 2012; Cherfi et al., 2012; Yin et al., 2012)), the achievable diagnostic accuracy is rather application-specific. Moreover, in general, diagnostic quality improves with the severity of the fault; so for incipient faults, diagnostic quality may be low.

Figure 5.7 shows two typical temporal behaviors of the diagnostic result for gradually evolving faults taken from (Verbeek, 2015). These behaviors relate to a track circuit diagnosis example similar to the example described in Section 3.4. For the fault diagnosis a Kalman filter approach has been used (see Section 3.6.2). In both Figures 5.7(a) and (b), the system is healthy till τ_{d}, i.e. $H(\tau) = h$ for $\tau < \tau_{\mathrm{d}}$. Afterwards, the system suffers from fault f_2, i.e. $H(\tau) = f_2$ for $\tau \geq \tau_{\mathrm{d}}$. From the diagnostic results in Figure 5.7, we conclude that in both cases the presence of a fault is almost instantly detected, i.e. $P(H(\tau) = h) \approx 0$ for $\tau > \tau_{\mathrm{d}}$. However, only from $\tau = \tau_{\mathrm{i}}$ on the system behavior is adequately diagnosed. In Figure 5.7(a), the fault is initially misdiagnosed, i.e. we conclude with a probability of approximately 80% that the system suffers from fault f_1. Slightly later, when more data have been collected, fault f_2 is correctly identified. In Figure 5.7(b) for τ between τ_{d} and τ_{i} there is (much) uncertainty about the cause of the faulty behavior. Initially all faults are plausible, i.e. both $P(H(\tau) = f_1)$, $P(H(\tau) = f_2)$, and $P(H(\tau) = f_3)$ are significantly larger than zero. Afterwards, doubt remains between faults f_2 and f_3 only. From time τ_{i}, the actual fault cause is identified with high accuracy, i.e. $P(H(\tau) = f_2) \approx 1$.

In general, the longer the fault is present, the more data of the faulty behavior are available, and the more accurate and reliable the diagnostic result is. How long it takes before adequate diagnostic results are obtained is however rather application-specific. Since in general both diagnostic and degradation forecasting performance improve over time, it is important to account for this time behavior in the subsequent decision making process. Therefore, in the next chapter, which concerns maintenance optimization, we give particular attention to the timing of maintenance decisions.



Figuur 5.7: Examples of commonly observed time behaviors of the diagnostic result (Verbeek, 2015). The monitored system is healthy till $\tau = \tau_d$, and suffers from fault f_2 afterwards. From the diagnostic results, we conclude that: (a) fault f_2 is incorrectly diagnosed in its incipient phase; (b) in the incipient phase, there is doubt between the different faults.

5.9. CONCLUSIONS

We have proposed a multiple-model approach to degradation modeling and forecasting for systems with multiple degradation and failure modes. A stochastic filtering approach is considered to handle the different sources of uncertainty inherent to degradation forecasting. Moreover, the links with the other tasks of the condition-based maintenance process, i.e. diagnosis and maintenance planning, have been established. We conclude that conditional predictions of future system reliability best support the subsequent decision making process. Accordingly, a framework has been proposed to determine the (future) system reliability in the multivariate case for different types of relationships among degradation measures.

We conclude that by using multiple models to forecast degradation behavior, the modeling error can be reduced. However, since the applicable model is selected based on the diagnostic result, the benefit of using multiple models over using a single model highly depends on the accuracy of the diagnostic result. Given the current quality of diagnosis methods, we do not expect this to be a serious drawback. However, caution is needed when faults are in their incipient phase. In this phase, the diagnostic results are often less accurate. A thorough analysis of the accuracy of diagnosis and prognosis results over time, and its implications on the subsequent maintenance optimization process is therefore an interesting topic for further research. As another topic for further research, we propose the thorough evaluation of the proposed approach within a condition-based maintenance process.

6

TIMELY CONDITION-BASED MAINTENANCE PLANNING

"Success is going from failure to failure without loss of enthusiasm."

-Winston Churchill-

Last-minute maintenance planning is often undesirable, as it may cause downtime during operational hours, may require rescheduling of other activities, and does not allow to optimize the management of spare parts, material, and personnel. In spite of the aforementioned drawbacks of last-minute planning, most existing methods on condition-based maintenance plan maintenance activities at the last minute. In this chapter, we propose an approach to timely maintenance planning in heterogeneous systems. As a first step, we determine for each system component independently the most appropriate maintenance planning strategy. This way, the maintenance decisions can be tailored to the specific situations. For example, conservative maintenance decisions can be taken when the risk tolerance is low, whereas more optimistic decisions can be taken when the risk tolerance is high; maintenance decisions can be made timely in situations in which we can accurately predict the future degradation behavior, while last-minute maintenance decisions can be made when the degradation behavior is less predictable. As a second step, we optimize the maintenance plan at the system level. Here, we account for economic and structural dependence with the aim to profit from spreading or combining various maintenance activities. The applicability of the method is demonstrated on a railway case. It is shown how the different cost functions (e.g., costs of maintenance, downtime, and failure) influence the maintenance decisions.

6.1. INTRODUCTION

For many systems, like manufacturing and transportation systems, maintenance activities have a major influence on the availability, safety, and operational costs of the system. The ideal maintenance strategy prevents failures without resorting to over-maintenance. Such a strategy depends on the current and the future health of the system, which are never completely known in practice. However, if the right system variables are measured and processed adequately, good estimates and predictions of the system health can be obtained, based on which the maintenance can be planned. This is the motivation behind *condition-based maintenance* (Yam et al., 2001; Wang, 2014).

Although much research has been devoted to maintenance planning based on real-time condition monitoring, most existing methods use only diagnostic information, consider planning of individual components, and focus on last-minute planning. Such approaches are sufficient for systems for which it is convenient to perform maintenance shortly after the decision to do so has been made. Last-minute maintenance planning is however often undesirable as it may cause downtime during operational hours, may require rescheduling of other activities, and does not allow to optimize the management of spare parts, material, and personnel. Furthermore, in multi-component systems, like road and railway networks and wind farms, it may be beneficial to combine or spread maintenance activities. This is not possible when maintenance needs are known just in time.

Motivated by the shortcomings of last-minute maintenance planning, we propose a two-stage bottom-up approach[1] for timely maintenance decision making based on real-time condition monitoring. A bottom-up approach is preferred over a top-down (aggregate) approach, because of its applicability to heterogeneous systems, i.e. systems consisting of multiple types of components (Medury and Madanat, 2011; Sathaye and Madanat, 2011; Furuya and Madanat, 2012; Yeo et al., 2013).

The *first stage* consists of determining the need for maintenance on each of the individual system components. If maintenance is required, based on the nature and urgency of the problem, the most appropriate type and time of maintenance have to be decided. As information regarding the system health is available in real time, it is not obvious when to settle on the decision regarding the time and type of maintenance. Generally, the more data are available, the better we can estimate the current and future system health, allowing a better decision on the time and type of maintenance. However, it also holds that the more data are available, the further the system fault has already developed, reducing the freedom to efficiently plan the maintenance. This trade-off implies that the timing of the final maintenance decision is an important decision variable. To the author's best knowledge, the timing of maintenance decisions, i.e. trading off accuracy and timeliness, has not been previously considered in the context of condition-based maintenance decision making. However, similar kinds of problems have been studied in economics as the *intertemporal choice problem* (Loewenstein et al., 2003) and in probability theory as the *optimal stopping problem* (Fisher, 1965; Agrawal et al., 2011). In this work, we extend the intertemporal choice problem to the maintenance domain.

In the *second stage*, we determine, based on the previously determined maintenance

[1] In a bottom-up approach, optimal maintenance strategies for the individual system components are determined first.

strategies for the individual components, the system-level maintenance strategy that minimizes total costs. So far, most work on system-level maintenance optimization has focused on incorporating budget constraints (Robelin and Madanat, 2008; Yeo et al., 2013). In this work, we incorporate economic and structural dependencies with the aim to minimize costs by spreading or combining maintenance activities. As already pointed out by Medury and Madanat (2011); Furuya and Madanat (2012) in the context of infrastructure management, the benefits of combining maintenance highly depend on the (spatial) relationships among the assets. Combining maintenance activities on multiple system components in road and railway networks is usually advantageous when the components form a series configuration, as the particular section is not available anyway. For system components arranged in a parallel structure, simultaneous maintenance is often undesirable, as it further reduces network capacity. An analogous reasoning holds for multi-component systems in general: in deciding whether to combine or spread maintenance, a trade-off between economies of scale and loss of functionality has to be made. The exact form of this trade-off depends on economic and structural dependencies between the considered components. In this work, we propose a systematic way for incorporating these dependencies in the system-level optimization.

The contributions of this chapter can be summarized as follows:

- We propose a component-level approach to maintenance planning in heterogeneous systems that trades timeliness for accuracy (Section 6.5);
- We propose a system-level maintenance optimization method that allows trading loss of functionality for cost reduction by incorporating economic and structural dependencies in a fundamental way (Section 6.6).

To demonstrate both approaches, we apply them in a case study concerning maintenance planning for a railway network (Section 6.7), as a railway network is a typical example of a system subject to heterogeneity and interdependence for which last-minute maintenance planning is undesirable.

6.2. RELATED WORK

Over the past years, various papers have been published on condition-based maintenance planning. Most of these works (e.g. (Wang, 2000; Grall et al., 2002; Amari et al., 2006; Wang et al., 2008; Elwany et al., 2011; Knowles et al., 2011; Wu et al., 2013; Lam and Banjevic, 2015; Wang and Wang, 2015)) base their maintenance decisions on just diagnostic information. Currently, only a few papers have been published that consider prognostic information for maintenance optimization (Camci, 2009; Van Horenbeek and Pintelon, 2013; Chen et al., 2015; Huynh et al., 2015; Tang et al., 2015). The advantage of including prognostic information is demonstrated by Camci (2009). Chen et al. (2015) propose an optimal condition-based replacement policy for systems for which degradation conforms to an inverse Gaussian process. Van Horenbeek and Pintelon (2013); Huynh et al. (2015) optimize maintenance planning for complex multi-component systems, assuming that component degradation follows a gamma distribution. Tang et al. (2015) consider maintenance optimization for systems whose degradation behavior is described by a random coefficient auto-regressive model.

Two drawbacks of the aforementioned methods are: 1. they only address the question whether or not to perform maintenance at a particular decision time instant; 2. except for (Camci, 2009), they rely on a specific degradation model. In contrast, the method proposed in this work is not restricted to a specific degradation model, but uses the diagnostic and prognostic result to predict both the *required type* and *optimal time* of maintenance.

Maintenance planning for *multi-component systems* has been treated e.g. by Castanier et al. (2005); Camci (2009); Bouvard et al. (2011); Tian et al. (2011); Van Horenbeek and Pintelon (2013); Huynh et al. (2015); Nguyen et al. (2015); Vu et al. (2015); Zhu (2015); Keizer et al. (2016). Because, in general, interconnections exist among system components (see Chapter 1), the optimal maintenance strategy for a multi-component system is not simply the set of optimal component-level solutions (Castanier et al., 2005). Consequently, recently few works have been proposed that account for (some of) these interactions, most of them focusing on (positive) economic dependencies (Castanier et al., 2005; Bouvard et al., 2011; Tian et al., 2011; Vu et al., 2015; Zhu, 2015; Keizer et al., 2016). Stochastic dependencies are e.g. considered by Camci (2009), and Van Horenbeek and Pintelon (2013) propose a method that incorporates both economic, stochastic, and structural dependencies. In this work, we incorporate both economic and structural dependencies. Like Van Horenbeek and Pintelon (2013), we model structural dependencies as economic dependencies. We do not explicitly consider stochastic dependencies as we assume that these dependencies are accounted for in the diagnosis and prognosis process. The main difference compared to the method proposed by Van Horenbeek and Pintelon (2013) is that they update the maintenance schedule each time new information becomes available without considering when to finalize the maintenance decision. In contrast, our method optimizes the time instant at which to settle on the maintenance decision. We aim to settle on the maintenance decision in an early stage, so as to avoid last minute re-scheduling of other activities and to allow for the optimization of spare parts, material, and personnel. Especially for intensively-used systems, like railway networks, it is desirable to plan the maintenance in time. Moreover, in contrast to Van Horenbeek and Pintelon (2013), we account for different types of maintenance activities.

Although the two-stage bottom-up approach that we consider for maintenance optimization is based on the approach proposed by Yeo et al. (2013), the two strategies are clearly different: while Yeo et al. (2013) focus on infrastructure management, we use it for condition-based maintenance planning, resulting in other objective functions. Moreover, Yeo et al. (2013) only account for budget constraints in the system-level optimization, while we account for economic and structural dependencies.

In summary, compared to existing methods on condition-based maintenance, the proposed method adds the following:

1. Next to deciding whether maintenance is needed, we optimize the required type of maintenance and the time to perform the maintenance (Challenge M_1);

2. We optimize the time to settle on the aforementioned maintenance decisions, hereby trading accuracy for timeliness (Challenge M_2);

3. We decouple the maintenance optimization from the diagnosis and prognosis process. This way, we are able to exploit both diagnostic and prognostic information

for maintenance optimization without restricting ourselves to a particular degradation model;

4. We provide a systematic framework for incorporating economic and structural dependencies among system components in the system-level optimization (Challenge M3).

6.3. ASSUMPTIONS

For clarity and to limit the scope of this chapter, the following assumptions are adopted:

A_1 We have fixed monitoring time instants at which diagnosis, prognosis, and maintenance planning are carried out.

A_2 Only major maintenance is considered. After maintenance is carried out, the component is in an as-good-as new state.

A_3 The different, application-specific, cost functions (e.g., costs of maintenance, downtime, and system failure) are assumed to be known.

A_4 The costs of performing a maintenance action are independent of the actual system health.

Moreover, as the focus of this chapter is on maintenance optimization, we assume that adequate diagnostic and prognostic results are available (see Chapters 3 till 5 for details and reference works on diagnosis and prognosis). Below, we specify how we assume the diagnostic and prognostic result to be specified.

DIAGNOSTIC RESULT

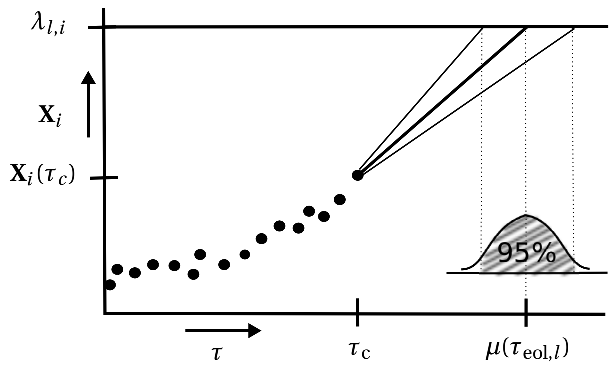
The health state of component i at time τ is captured by a discrete variable $H_i(\tau) \in \{h, f_{i,1}, f_{i,2}, \ldots, f_{i,\ell_i}\}$, where h represents the healthy state and $f_{i,1}$ through f_{i,ℓ_i} denote the possible fault types for component i. Since it is generally not possible to determine the health state with complete certainty, the diagnostic result is a probability mass function over the current (i.e. for time $\tau = \tau_c$) health state: $P(H_i(\tau_c))$.

PROGNOSTIC RESULT

The prognostic result (see Figure 6.1) of component i includes the current value of the degradation measure[2], $\mathbf{X}_i(\tau_c)$, as well as its predicted evolution. We assume $\mathbf{X}_i(\cdot)$ to be a continuous-time stochastic process. Since the prognostic result is a probability distribution over this process, we denote it as $p(\mathbf{X}_i|\mathcal{I}_c)$, where $\mathcal{I}_c$ represents all data available at time τ_c. The prognostic result can be used to predict the value $p(\mathbf{X}_i(\tau)|\mathcal{I}_c)$ of the degradation measure at a given time $\tau > \tau_c$.

Since different fault types generally result in different time behaviors of $\mathbf{X}_i(\cdot)$ (see Chapter 5), we require a distinct prognostic model to be available for every possible fault type $f_{i,1}, \ldots, f_{i,\ell_i}$. In the sequel, we assume that the prognostic models are captured by parametric models (see Chapter 5). So, for each component i we have ℓ_i parametric models: $m_{i,1}(\cdot|\theta_{i,1}), \ldots, m_{i,\ell_i}(\cdot|\theta_{i,\ell_i})$, each characterizing the expected temporal behavior of

[2]For sake of clarity and without loss of generality, in this chapter, $\mathbf{X}_i$ is assumed to be univariate.



Figuur 6.1: Illustration of the prognostic result at $\tau = \tau_c$. For $\tau \leq \tau_c$, the degradation measure $\mathbf{X}_i(\tau)$ is known. For $\tau > \tau_c$ we predict the behavior of the degradation measure based on its previous values and the fault-specific parametric model (linear here). From this prediction and the failure threshold $\lambda_{l,i}$, the distribution of the estimated failure time in mode l, $p(\tau_{eol,l})$, can be determined.

the degradation measure as a consequence of the corresponding fault type. Vector $\theta_{i,j}$ denotes the model parameters, which might be stochastic. The applicable parametric model is indicated by the diagnostic result, and its parameters are estimated based on the available data (see Chapter 5).

6.4. PROBLEM DEFINITION

In this chapter, condition-based maintenance planning for a heterogeneous system consisting of n independently and continuously monitored components is considered. In defining the problem, we make a distinction between decision making at the component level and decision making at the system level (see Figure 6.2).

COMPONENT-LEVEL OPTIMIZATION

6

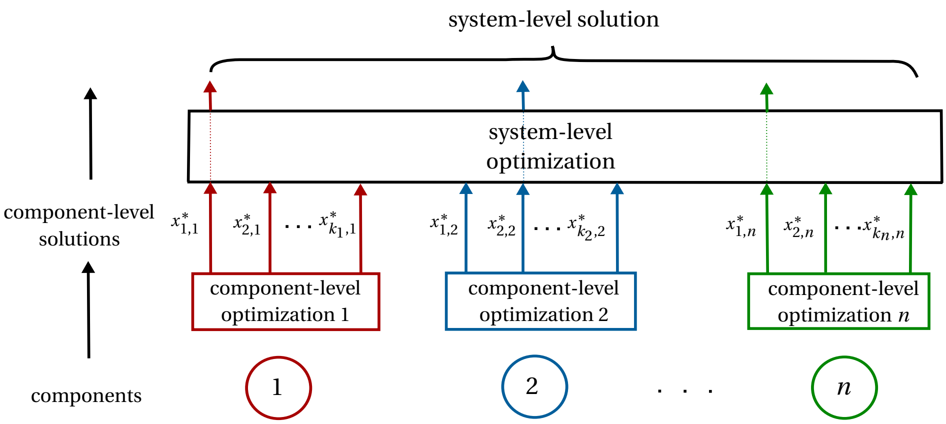
At this level, we aim to find a set of optimal and near-optimal maintenance strategies for each system component in need of maintenance. The optimal maintenance strategy is defined by:

1. the required type of maintenance;
2. the optimal time of performing maintenance.

The *required type of maintenance* refers to the maintenance action that brings the system to an as-good-as-new condition. As only major maintenance is considered, the optimal maintenance action depends only on the system health state H_i.

The *optimal time of performing maintenance* refers to the maintenance time that minimizes total costs, i.e. the time of maintenance is chosen such that the component's lifetime is maximized, while accounting for the following additional objectives:

- prevention of failure;



Figuur 6.2: Two-stage bottom-up approach (adapted from (Yeo et al., 2013)). First stage: optimal and near-optimal maintenance strategies $x_{1,i}$ till $x_{k_i,i}$ are determined for each component i in need of maintenance. Second stage: the optimal combination of component-level maintenance strategies is determined at the system level.

- minimization of downtime during operational hours.

The optimal time depends on the expected degradation over time and on the applicable cost functions and risk tolerances.

As information regarding the system health is available in real time, it is not obvious when to settle on a decision regarding the time and type of maintenance. On the one hand, it is desirable to make the maintenance decisions as early as possible, as the creation of an effective maintenance schedule requires that the maintenance needs are known in time, e.g.:

- To prevent a failure, the right decision should be made at least z_i time units before component i fails, where z_i refers to the time needed to get the personnel and material at the maintenance location and to maintain component i.

- Moreover, to prevent system downtime during operational hours, the right decision should be made at least $z_i + q_i$ time units before component i fails, with q_i the maximum possible time gap between any two consecutive out-off-service periods of component i. Indeed, when the system will fail just before an out-of-service period, maintenance has to be performed in a previous out-of-service period to avoid failure and system downtime during operational hours.

- To optimize (system-level) maintenance planning: The earlier the maintenance requirements are known, the more freedom there is in maintenance scheduling, and the more cost efficient the resulting maintenance schedule will be.

On the other hand, it is desirable to plan the maintenance based on reliable and accurate predictions of the system health. This is done to avoid that maintenance is performed too late, resulting in sudden failures, or that maintenance is done too early, leading to

6

over-maintenance. Since the prediction accuracy increases over time, a trade-off between accuracy and timeliness needs to be made. Clearly, early predictions that are not accurate at all are of no use. Perfect predictions less than z_i time units before a functional failure of component i are of no use either. In such a case, a failure could not be prevented.

In summary, next to determining the required type of maintenance and optimal time to perform the maintenance, the component-level optimization comprises the timing of the maintenance decision, i.e. trading off accuracy and timeliness.

System-level optimization

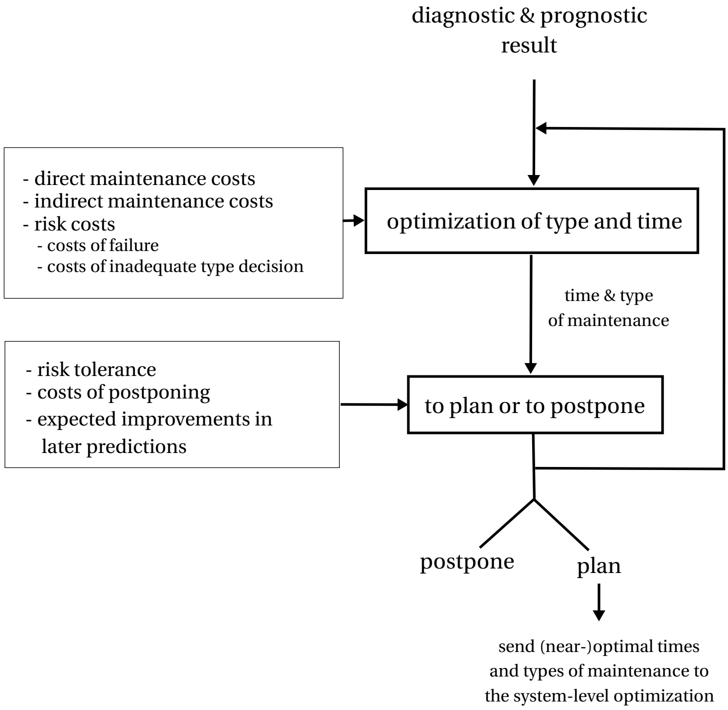
At this level, we search for the optimal maintenance strategy from the system-level point of view. More specifically, for each system component in need of maintenance we search for a maintenance strategy that minimizes the maintenance costs for the system as a whole. In the simplest case, the system-level solution coincides with the set containing the optimal strategy for each component. However, if budget or resource constraints are binding or system dependence applies, the system-level solution may differ from the set of optimal component-level solutions.

As only major maintenance is considered, the only optimization variable at the system level is the time of performing maintenance. At the system-level, we aim to maximize the cost benefits resulting from combining or spreading maintenance activities. In general, direct maintenance costs can be reduced by combining maintenance on nearby system components. This way, part of the costs, e.g. setup work and transportation costs, can be shared between the simultaneously maintained components. Although the direct maintenance costs generally decrease when maintenance on nearby components is combined, the indirect costs (e.g. costs related to downtime) do not necessarily decrease. The effect of combining maintenance on the indirect costs depends on the extent to which the additional maintenance of a component influences the functionality of the whole system. When the whole system is out of service during the maintenance of component A, simultaneously maintaining an arbitrary component B has no negative impact on the functionality of the system, and so on the indirect costs. However, when the system is still (partly) functional when only component A is maintained, but no longer (or less) functional when component A and B are maintained simultaneously, combining maintenance on components A and B may have a negative effect on the indirect maintenance costs. In this case, the potential reduction in direct costs (economies of scale) must be traded against the potential increase of indirect costs (loss of functionality). To handle this trade-off, the influences of combining maintenance on both the direct costs and the indirect costs need to be clear. Here, we assume that the potential reduction in direct costs depends on:

1. the number of simultaneously maintained components;
2. the similarity between the maintenance activities;

and the potential reduction in indirect costs depends on:

1. the reduction of downtime when combining maintenance;
2. the structural dependencies between the components.



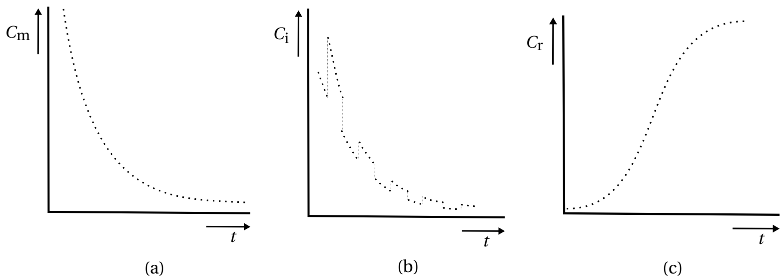
Figuur 6.3: Component-level maintenance optimization. First, based on the currently available information, the optimal type and time of maintenance are determined. Next, it is decided whether to plan the maintenance according to the previously determined strategy or to postpone the decision to a later time (when more data are available).

6.5. DECISION MAKING AT THE COMPONENT LEVEL

At the component-level, for each component i in need of maintenance, the optimal and near-optimal[3] maintenance strategies are determined. The focus is on the individual components and dependencies among system components are not yet taken into account. The determination of the (near-)optimal maintenance strategies is done in two steps (see Figure 6.3). At each monitoring time instant, we first determine, based on the currently available information, optimal and near optimal maintenance strategies. Next, we determine whether we want to plan the maintenance according to the determined maintenance strategy or to postpone the maintenance decision to a later time, when more data are available.

For the sake of brevity, in the sequel we omit the subscript i when the explicit reference to a particular component i is not necessary. For the same reason, we omit the time argument τ whenever possible.

[3]The (near-)optimal strategies $x^*_{1,i}, x^*_{2,i}, \ldots, x^*_{k_i,i}$ serve as an input for the system-level maintenance optimization (see Figure 6.2).



Figuur 6.4: Example relations between the different cost components and the time of maintenance. (a) life-time averaged direct maintenance costs; (b) lifetime-averaged indirect maintenance costs; (c) costs related to risk.

6.5.1. OPTIMIZATION OF TYPE AND TIME OF MAINTENANCE

In this step we search for the optimal maintenance action a^*, maintenance time t^*, and the associated costs C^*:

$$\left(a^*, t^*\right) = \arg \min_{a \in A, t \in T} C_m(a,t) + C_i(a,t) + C_r(a,t) \tag{6.1}$$

$$C^* = \min_{a \in A, t \in T} C_m(a,t) + C_i(a,t) + C_r(a,t) \tag{6.2}$$

with A the discrete set of possible maintenance activities, T the discrete set of available maintenance time instants, $C_m(a,t)$ the lifetime-averaged direct costs of performing maintenance activity a at time t, $C_i(a,t)$ the lifetime-averaged indirect costs of maintenance activity a at time t, and $C_r(a,t)$ the costs associated with the risk of action a being inadequate or time t being too late, where all costs are expressed in terms of (virtual) monetary units. More specifically, we define the lifetime-averaged direct costs of maintenance $C_m(a,t)$ as:

$$C_m(a,t) = \frac{c_m(a)}{t - t_{mnt}} \tag{6.3}$$

with t_{mnt} the previous maintenance time and $c_m(a)$ the direct costs of performing maintenance activity a (e.g. material, personnel). So, the lifetime-average direct costs of maintenance $C_m(a,t)$ correspond to the costs $c_m(a)$ averaged over the lifetime $t - t_{mnt}$; the larger t is chosen, the lower are the lifetime-averaged direct costs of maintenance (see Figure 6.4(a)).

The lifetime-averaged indirect costs of maintenance $C_i(a,t)$ are defined as:

$$C_i(a,t) = \frac{c_i(a,t)}{t - t_{mnt}} \tag{6.4}$$

with $c_i(a,t)$ the indirect costs (e.g. the costs related to downtime) of maintenance activity a at time t. In contrast to the direct costs, the indirect costs depend on the time of maintenance. Indeed, for most systems, the cost of downtime depends on the time of the downtime. For example, for road or railway networks, the inconvenience of downtime is less during night than during day. So, given a particular action a, the life-time averaged indirect costs $C_{\mathrm{i}}(a,t)$ intermittently decrease for increasing t (see Figure 6.4(b)).

Finally, we define the costs $C_{\mathrm{r}}(a,t)$ related to risk as:

$$C_{\mathrm{r}}(a,t) = \sum_{l=1}^{p}\sum_{j=1}^{\ell} P(H(t)=f_j)P(\mathcal{F}_l(t)=1|H(t)=f_j)C_{c_l} - \sum_{j=1}^{\ell} P(H(t)=f_j)C_{f_j}(a) \quad (6.5)$$

The first term expresses the costs related to the risk of maintenance time t being too late to avoid a failure. The second term expresses the costs related to the risk of maintenance action a being inappropriate to repair the system. In (6.5), $\mathcal{F}_l(t)$ is a binary variable indicating whether the system fails in mode c_l at time t, C_{c_l} represents the costs of a failure in mode c_l, and $C_{f_j}(a)$ represents the penalty cost of an (inadequate) maintenance type decision a in case of fault f_j. So, given a particular maintenance action, and as long as no maintenance is done, the costs $C_{\mathrm{r}}(a,t)$ related to risk increase for increasing t (see Figure 6.4(c)).

We conclude that the first two terms in (6.1), i.e. $C_{\mathrm{m}}(\cdot)$ and $C_{\mathrm{i}}(\cdot)$, are minimized for t chosen as large as possible, while the last term, $C_{\mathrm{r}}(\cdot)$, is minimized for t chosen as small as possible. The overall optimum depends, besides on the diagnostic result ($P(H(t)=f_j)$) and the prognostic result ($P(\mathcal{F}_l(t)=1|H(t)=f_j)$), on the cost functions $c_{\mathrm{m}}(\cdot)$, $c_{\mathrm{i}}(\cdot)$, C_{c_l}, and $C_{f_j}(\cdot)$.

Besides the optimal maintenance strategy, alternative, near-optimal, strategies are determined. These alternative strategies are required for the system-level optimization and can be found by excluding the optimal strategy from the search space (Yeo et al., 2013):

$$x^*_{k+1} = \left(a^*_{k+1}, t^*_{k+1}\right) = \arg\min_{(a,t)\in A\times T\setminus\{(a^*_l,t^*_l)|l=1,\dots,k\}} C_{\mathrm{m}}(a,t) + C_{\mathrm{i}}(a,t) + C_{\mathrm{r}}(a,t) \quad (6.6)$$

$$C^*_{k+1} = \min_{(a,t)\in A\times T\setminus\{(a^*_l,t^*_l)|l=1,\dots,k\}} C_{\mathrm{m}}(a,t) + C_{\mathrm{i}}(a,t) + C_{\mathrm{r}}(a,t) \quad (6.7)$$

The resulting optimization problem can be solved using standard non-linear optimization techniques (see e.g. (Fletcher, 2013)). Which of the available algorithms is the most suitable depends e.g. on characteristics of the specific cost function (e.g. unimodal versus multimodal), the size of the system, the frequency at which diagnosis, prognosis, and maintenance planning are carried out, and the number of times the problem has to be solved each time (i.e. the number of required near-optimal solutions).

6.5.2. TO PLAN OR TO POSTPONE

In this step, it is decided whether to accept the previously found component-level maintenance strategy (i.e. the combination of maintenance time t^* and type a^*) or to wait for a potential better maintenance strategy (i.e. a strategy with lower costs C^*) at a later time. So we have to trade between potential cost savings and the risk and inconvenience of postponing maintenance planning. The problem can thus be considered as a sequential decision problem. At each monitoring instant, based on the outcome of the previously determined optimal maintenance strategy, it has to be decided whether to plan maintenance or to postpone the maintenance decision. If the decision is postponed, we face the decision again at the next monitoring instant. The problem repeats itself until the maintenance is planned or the system fails, in which case corrective maintenance is needed.

Inspired by the approaches proposed by Etzioni et al. (2003); Agrawal et al. (2011); Groves and Gini (2015), we propose to solve the sequential decision problem by formulating it as a Markov decision process, and to use dynamic programming or reinforcement learning to solve the obtained Markov decision problem. The choice to take advantage of the methods proposed by Etzioni et al. (2003); Agrawal et al. (2011); Groves and Gini (2015) is motivated by the resemblance between the considered maintenance optimization task and the task of buying durable goods or airline tickets: First, the risk that an airline ticket is sold out before buying is comparable to the risk of failure before maintenance is performed. Second, both buying durable goods and planning of maintenance activities are associated with a cost for postponing; when buying durable goods, postponing means that you cannot immediately use the product: when planning maintenance, postponing means that less freedom remains in scheduling. Third, for both buying durable goods or airline tickets and planning maintenance, there is uncertainty about the future "costs", while having some (qualitative) knowledge about their further evolution.

BACKGROUND ON MARKOV DECISION PROCESSES

A (finite) Markov decision process is defined by the tuple $(S_{\mathrm{m}}, A_{\mathrm{m}}, P_{\mathrm{t}}, R)$, where S_{m} represents a finite set of states, A_{m} a finite set of actions, P_{t} a probabilistic transition function, and R a reward function (Powell, 2011; Feinberg and Shwartz, 2012). The transition function indicates how the state changes as a result of action a_{m}. Assuming a probabilistic setting, the transition function $P_{\mathrm{t}}(a_{\mathrm{m}}, s_{\mathrm{m}}, s'_{\mathrm{m}})$ outputs the probability that the system moves to state s'_{m} given that it is currently in state s_{m} and action a_{m} is taken. The reward function $R(a_{\mathrm{m}}, s_{\mathrm{m}}, s'_{\mathrm{m}})$ evaluates the immediate effect of moving from state s_{m} to state s'_{m} under action a_{m}. Note that both the transition function and the reward function satisfy the Markov property. The goal is to design an optimal policy π that defines which action $a_{\mathrm{m}} \in A_{\mathrm{m}}$ to take when the system is in state $s_{\mathrm{m}} \in S_{\mathrm{m}}$ such that the expected long-term reward is maximized. When the reward function R and the state transition function P_{t} are known, the problem can be solved by dynamic programming (Bertsekas, 1995) by recursively solving the following equations:

6

$$\pi(s_{\mathrm{m}}) = \arg\max_{a_{\mathrm{m}}} \Big(\sum_{s'_{\mathrm{m}}} P_{\mathrm{t}}(a_{\mathrm{m}}, s_{\mathrm{m}}, s'_{\mathrm{m}}) \big(R(a_{\mathrm{m}}, s_{\mathrm{m}}, s'_{\mathrm{m}}) + \gamma V(s'_{\mathrm{m}}) \big) \Big) \tag{6.8}$$

$$V(s_{\mathrm{m}}) = \sum_{s'_{\mathrm{m}}} P_{\mathrm{t}}\big(\pi(s_{\mathrm{m}}), s_{\mathrm{m}}, s'_{\mathrm{m}}\big) \Big(R\big(\pi(s_{\mathrm{m}}), s_{\mathrm{m}}, s'_{\mathrm{m}}\big) + \gamma V(s'_{\mathrm{m}}) \Big) \tag{6.9}$$

with $\gamma \in (0,1]$ a discount factor. When the reward function or the transition probabilities are unknown, reinforcement learning (Van Otterlo and Wiering, 2012) can be used to learn the optimal policy from data/experience, which can be done offline or online. Online learning is not recommended for maintenance decision making, as we want to avoid safety-critical errors. Advantages of learning in sequential decision making are that it relieves the designer of the system from deciding upon everything in the design phase and that it can cope with uncertainty and changing situations (Van Otterlo and Wiering, 2012). The final choice for dynamic programming or learning is problem-specific and depends on the available domain knowledge and data.

REFORMULATION AS A MARKOV DECISION PROCESS

To reformulate the considered decision task, i.e. deciding whether to plan maintenance or to postpone the maintenance decision, as a Markov decision problem, we first define the set of states S_m as the set of all possible instances of the following state vector:

$$s_\mathrm{m} = \left[C^* \ P(H) \ \mu(\theta) \ \sigma(\theta) \ \tau - \tau_0 \ \mathcal{D} \ \mathcal{F}\right]^\top \tag{6.10}$$

with C^* the costs associated with the maintenance strategy under consideration (see Section 6.5.1), $P(H)$ the probability mass function over the health state H, $\mu(\theta)$ and $\sigma(\theta)$ the mean and standard deviation of the parameter vector $\theta = [\theta_1, \ldots, \theta_\ell]^\top$ of the degradation model, τ_0 the initial time of the decision, $\mathcal{D}$ a binary variable indicating whether maintenance is already planned or not, and $\mathcal{F}$ a binary variable indicating whether the system fails. The state vector is chosen this way as the decision whether or not to plan depends on the current costs C^* and the expected costs of future strategies, which can be predicted based on C^*, $P(H)$, and the forecasted degradation measure (characterized by $[\mu(\theta_1), \ldots, \mu(\theta_\ell)]^\top$ and $[\sigma(\theta_1), \ldots, \sigma(\theta_\ell)]^\top$). Since timely maintenance planning is preferred over last-minute planning, the benefits of planning decrease over time. This motivates to include a decreasing term in the immediate reward function for planning. To ensure that the Markov property is satisfied, $\tau - \tau_0$ is included in the state. The last two elements, $\mathcal{D}$ and $\mathcal{F}$, are included to define the terminal states. When one of the two is true, the decision process ends, i.e. all states for which holds that $\mathcal{D} = 1$ or $\mathcal{F} = 1$ are terminal states.

The set of actions A_m is:

$$A_\mathrm{m} = \{\text{plan, postpone}\} \tag{6.11}$$

and the transition probabilities can be defined as:

$$P\big(C^*(\tau+\Delta)\big) = f_\mathrm{C}\Big(C^*(\tau), P\big(H(\tau)\big), \mu(\theta,\tau), \sigma(\theta,\tau)\Big) \tag{6.12}$$

$$P\big(H(\tau+\Delta)\big) = P\big(H(\tau)\big) \tag{6.13}$$

$$\mu(\theta,\tau+\Delta) = \mu(\theta,\tau) \tag{6.14}$$

$$\sigma(\theta,\tau+\Delta) = f_\sigma\Big(P\big(H(\tau)\big), \mu(\theta,\tau), \sigma(\theta,\tau)\Big) \tag{6.15}$$

$$\mathcal{D}(\tau+\Delta) = \begin{cases} 0 & \text{if } a_\mathrm{m} = \text{postpone} \\ 1 & \text{otherwise} \end{cases} \tag{6.16}$$

$$P\big(\mathcal{F}(\tau+\Delta) = 1\big) = P\big(\mathcal{F}(\tau+\Delta) = 1 | \mathcal{I}_\tau\big) \tag{6.17}$$

with Δ the time interval between two monitoring instants. Equation (6.12) predicts the costs at the next time instant given the current costs, the diagnostic result, and the prognostic result. The function $f_\mathrm{C}(\cdot)$, which specifies the relation between $P(C^*(\tau+\Delta))$ and $C^*(\tau)$, $P(H(\tau))$, $\mu(\theta,\tau)$, and $\sigma(\theta,\tau)\big)$, is application-specific and needs to be learned from historical data, possibly in combination with expert knowledge. Note that an important

6

condition is that the diagnosis, prognosis, and maintenance optimization method are in use when collecting the data. Equation (6.13) indicates that the prediction of the diagnostic result (i.e. the probability of presence of a fault) at time $\tau+\Delta$ equals the diagnostic result at time τ. Similarly, (6.14) specifies that the prediction of the mean of parameter vector θ at time $\tau+\Delta$ equals the mean at time τ. The variance of the parameter vector at time $\tau+\Delta$ is expressed as a (decreasing) function of $P(H(\tau)), \mu(\theta,\tau)$, and $\sigma(\theta,\tau)$. Like $f_{\mathrm{C}}(\cdot)$, the function $f_\sigma(\cdot)$ is application-specific and needs to be learned from data or defined by domain experts. Equation (6.16) specifies that maintenance is not planned as long as a_{m} = "postpone". Finally, (6.17) specifies the probability that the system fails at the next monitoring instant.

With respect to the reward function, we model the loss of utility due to waiting by means of discounting (Agrawal et al., 2011). If we decide to plan the maintenance at time τ, this gives us a utility of $u_{\max}\delta_{\mathrm{d}}^{\tau-\tau_0}$, with τ_0 the initial time maintenance planning is considered, δ_{d} the discounting rate, and $u_{\max}$ the maximum utility for planning. The net reward for planning maintenance equals the obtained utility minus the costs:

$$R(a_{\mathrm{m}} = \text{plan}, s_{\mathrm{m},\mathcal{D}} = 0, s'_{\mathrm{m},\mathcal{D}} = 1) = u_{\max}\delta_{\mathrm{d}}^{\tau-\tau_0} - C^*$$

with $s_{\mathrm{m},\mathcal{D}}$ the $\mathcal{D}$-component of the state (6.10). The immediate reward associated with postponing the decision is zero as long as the system does not fail in the next time step. When the system fails, this is penalized with a negative reward $-\alpha$. All together, we define our reward function as follows:

$$R(a_{\mathrm{m}}, s_{\mathrm{m}}, s'_{\mathrm{m}}) = \begin{cases} u_{\max}\delta_{\mathrm{d}}^{\tau-\tau_0} - C^* & \text{for } s'_{\mathrm{m},\mathcal{D}} = 1 \\ -\alpha & \text{for } s'_{\mathrm{m},\mathcal{D}} = 0 \text{ and } s'_{\mathrm{m},\mathcal{F}} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.18}$$

with $s_{\mathrm{m},\mathcal{F}}$ the $\mathcal{F}$-component of the state (6.10). We propose the use of a finite decision horizon, where the horizon corresponds to the predicted failure time $\mu(\tau_{\mathrm{eol}})$ minus $(w\sigma(\tau_{\mathrm{eol}}) + z)$, with w an application-specific and user-defined parameter defining the latest time one wants to schedule maintenance. The larger w, the more failure-avoidant one is. Planning is forced when the end of the horizon is reached. Because we consider a finite horizon, we do not discount future rewards, i.e. $\gamma = 1$.

6.6. SYSTEM-LEVEL MAINTENANCE OPTIMIZATION

In the system-level optimization we search for the optimal system-level maintenance strategy accounting for economic and structural dependencies among system components.

6.6.1. PROBLEM FORMULATION

As we consider maintenance planning based on real-time condition monitoring, we face both newly entered maintenance needs and already scheduled[4] (but not yet carried

[4]The order of scheduling at the system-level is determined by the order in which the maintenance needs are set at the component level.

out) maintenance activities in the system-level optimization. Consider that η components need to be maintained and that for η_{ns} of them maintenance has not yet been planned. Without loss of generality they are renumbered such that $l = 1,\ldots,\eta_{\text{ns}}$ represent the components for which maintenance is not yet planned. For these components, we have to select the optimal system-level strategy x_l from the set A_l^* of optimal and near-optimal component-level maintenance strategies, with $A_l^* = \{x_{1,l}^*, x_{2,l}^*, \ldots\} = \{(a_{1,l}^*, t_{1,l}^*), (a_{2,l}^*, t_{2,l}^*), \ldots\}$ (see Section 6.5.1). For each other component $l = \eta_{\text{ns}}+1,\ldots,\eta$, the maintenance time and type are already defined, i.e. $x_l = x_{\text{o},l}^*$. Now the aim is to find the optimal system-level maintenance strategy X_{free}^*, i.e. a strategy $x_l \in A_l^*$ for each component $l = 1,\ldots,\eta_{\text{ns}}$ such that the system-level criterion $C_{\text{SL}}(X_{\text{free}}, X_{\text{fixed}})$ is minimized:

$$X_{\text{free}}^* = \arg\min_{X_{\text{free}}} C_{\text{SL}}(\underbrace{X_{\text{free}}, X_{\text{fixed}}}_{X}) \tag{6.19}$$

with

$$\begin{aligned} X_{\text{free}} &= \left(x_1,\ldots,x_{\eta_{\text{ns}}}\right) \text{ with } x_l \in A_l^* \\ X_{\text{fixed}} &= \left(x_{\eta_{\text{ns}}+1},\ldots,x_\eta\right) \text{ with } x_l = x_{\text{o},l}^* \end{aligned}$$

The subscript “free” refers to the variables for which the maintenance strategy is not yet set, and the subscript “fixed” refers to the components for which the maintenance strategy is already set in a previous system-level optimization.

6.6.2. OPTIMIZATION CRITERION

The optimization criterion $C_{\text{SL}}(X)$ expresses the total costs of system-level strategy X. The total costs equal the sum of the individual maintenance costs corrected for the cost benefits/drawbacks obtained from combining or spreading maintenance.

COST OF INDIVIDUAL MAINTENANCE ACTIVITIES

The first part of the optimization criterion expresses the total costs of system-level strategy X in the absence of economic and structural dependence, i.e.:

$$C_0(X) = \sum_{l=1}^{\eta} C_{x_l}, \quad x_l \in X \tag{6.20}$$

with C_{x_l} the component-level costs of maintenance strategy x_l.

SYSTEM DEPENDENCE

We make a distinction between:

1. economies of scale;
2. loss of utility.

To incorporate economies of scale, we split the direct costs of maintenance $c_{\text{m}}(a)$ into three components:

$$c_{\text{m}}(a) = c_{\text{m},1}(a) + c_{\text{m},2}(a) + c_{\text{m},3} \tag{6.21}$$

with $c_{m,1}(a)$ the fixed costs that do no depend on economies of scale, $c_{m,2}(a)$ the costs that can be shared between components that simultaneously undergo maintenance action a, and $c_{m,3}$ the costs that can be shared between all simultaneously maintained components, regardless of the type of maintenance. This way, the cost savings resulting from economies of scale can be computed as:

$$C_{EOS}(X) = \sum_{t\in T}\sum_{a\in A}\beta_{1,(a,t)}(X)c_{m,2}(a) + \sum_{t\in T}\beta_{2,t}(X)c_{m,3} \tag{6.22}$$

with:

$$\begin{aligned} \beta_{1,(a,t)}(X) &= \max(0, n_{1,(a,t)}(X)-1)\\ \beta_{2,t}(X) &= \max(0, n_{2,t}(X)-1)\\ n_{1,(a,t)}(X) &: \text{number of components that undergo action } a\\ &\quad\ \text{at time } t \text{ under strategy } X\\ n_{2,t}(X) &: \text{number of maintained components at time } t\\ &\quad\ \text{under strategy } X \end{aligned}$$

Indeed, if we divide some costs c between $n > 1$ components, the cost reduction at the system-level equals $(n-1)c$.

To incorporate loss of utility, we consider both the reduction in downtime and the reduction in system functionality as a consequence of combining. To assess the cost savings resulting from downtime reduction, we split the indirect maintenance costs $c_i(a,t)$ into two parts:

$$c_i(a,t) = c_{i,1}(t) + c_{i,2}(a,t) \tag{6.23}$$

with $c_{i,1}(t)$ the costs that are directly related to system downtime and $c_{i,2}(a,t)$ all other indirect costs. The reduction in downtime when $n_{2,t} \geq 1$ components are maintained at time t is $(n_{2,t}-1)c_{i,1}(t)$. Accordingly we define the reduction in downtime costs of strategy X as:

6

$$C_{DT}(X) = \sum_{t\in T}\beta_{2,t}(X)c_{i,1}(t) \tag{6.24}$$

To incorporate the additional loss of functionality when multiple components are maintained simultaneously, we divide the components into ν disjoint groups $g_1,\ldots,g_\nu$, such that if only components of one group are maintained at time t, no additional loss of functionality is induced. When components of different groups are maintained simultaneously, the functionality of the system may reduce, and an additional cost term $C_{LF}(X)$ is added to the optimization criterion, with $C_{LF}(X)$ defined as:

$$C_{LF}(X) = \sum_{t\in T} f_{LF}(\mathcal{X}_t(X))$$

with $\mathcal{X}_t(X)$ the set containing all groups g_ζ of which a component is maintained at time t under strategy X, and $f_{LF}(\cdot)$ a function assigning a penalty cost to each possible set $\mathcal{X}_t$.

6.6.3. SYSTEM-LEVEL OPTIMIZATION FORMULATION

Taking all together, the system-level optimization criterion C_{SL} is defined as follows:

$$C_{SL}(X) = C_0(X) - C_{EOS}(X) - C_{DT}(X) + C_{LF}(X) \tag{6.25}$$

Together with (6.19) this defines the system-level optimization problem.

System-level optimization problems of small size can be solved by brute-force approaches. For computationally demanding problems, approximate algorithms, like pattern search heuristics or evolutionary algorithms can be used (Yeo et al., 2013).

6.7. CASE STUDY: MAINTENANCE PLANNING IN RAILWAY NETWORKS

We illustrate the proposed approach on a case study concerning maintenance planning for a railway network. A railway network consists of different types of components (e.g., tracks, switches, bridges) that are located in areas with different environmental conditions, meaning that the system should be considered as heterogeneous with respect to deterioration processes and costs. Furthermore, railway networks are subject to economic and structural dependencies, meaning that costs and downtime can be reduced when maintenance activities are combined or spread in time. Note that this case study is fictitious and has an illustrative purpose. A full evaluation of the proposed approach is beyond the scope of this thesis. In the case study, all optimization problems are solved using an exhaustive search algorithm.

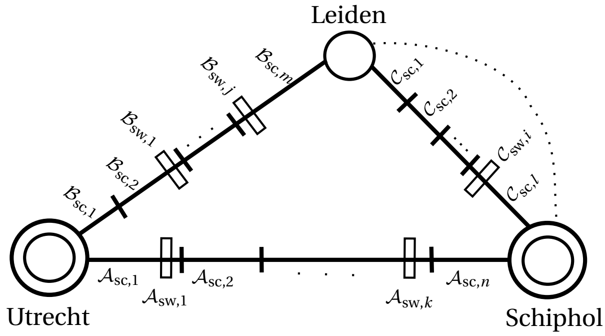
6.7.1. PROBLEM SPECIFICATION

SYSTEM DESCRIPTION

Consider the network depicted in Figure 6.5, representing a part of the Dutch railway network. Utrecht and Schiphol are two important and busy railway stations in the Netherlands. Besides the direct line between the two cities, there is an indirect connection between the two cities via Leiden. Furthermore, there is a bus connection between Leiden and Schiphol.

Assume that the railway network consists of two types of components:

6



Figuur 6.5: Railway network considered in the case study.

1. sections (parts of the track);

2. switches.

A section can suffer from two types of faults: rail defects f_{rd} and rail contamination f_{rc}. A switch can suffer from one fault f_{sw}. Accordingly we have three maintenance actions: a_{rd} to repair a rail defect, a_{rc} to remove rail contamination, and a_{sw} to repair a switch. We assume that the temporal behavior of the degradation measure $\mathbf{X}(\cdot)$ as a consequence of each fault can be described by the following parametric models:

$$X^{\text{rd}}(\tau) = \theta_{\text{rd},1} + \theta_{\text{rd},2} e^{\theta_{\text{rd},3}(\tau-\tau_{\text{c}})} \tag{6.26}$$

$$X^{\text{rc}}(\tau) = \theta_{\text{rc},1} + \theta_{\text{rc},2}(\tau-\tau_{\text{c}}) \tag{6.27}$$

$$X^{\text{sw}}(\tau) = \theta_{\text{sw},1} + \theta_{\text{sw},2}(\tau-\tau_{\text{c}}) \tag{6.28}$$

with $\theta_{\text{rd},3}$, $\theta_{\text{rc},2}$, and $\theta_{\text{sw},2}$ normally distributed random variables and the other parameters deterministic. For all components and failure modes, the failure threshold is set to 100.

COMPONENT-LEVEL COST FUNCTIONS

Type and time of maintenance We define the cost C_{c_l} of a failure in mode l as follows[5]: For section faults, f_{rd} and f_{rc}, we make a distinction depending on the location of the section. Sections close to switches or level crossings influence the proper functioning of these switches and crossings. Therefore, a failure of such a section is more disastrous than a failure of another section. Sections that influence the proper functioning of other assets are said to be of type II, all other sections are of type I. Accordingly, the additional costs of a failure C_{c_l} are defined as:

$$C_{c_{\text{rd}}} = \begin{cases} 500 & \text{if the section is of type } I \\ 2000 & \text{if the section is of type } II \end{cases} \tag{6.29}$$

$$C_{c_{\text{rc}}} = \begin{cases} 750 & \text{if the section is of type } I \\ 2500 & \text{if the section is of type } II \end{cases} \tag{6.30}$$

$$C_{c_{\text{sw}}} = 1000 \tag{6.31}$$

Let the costs of a wrong maintenance decision $C_{f_j}(a)$ be given by:

$$C_{f_{\text{rc}}}(a_{\text{rd}}) = 350 \tag{6.32}$$

$$C_{f_{\text{rd}}}(a_{\text{rc}}) = 500 \tag{6.33}$$

Note that we know the type of monitored component (section or switch). Hence, we will not schedule a section maintenance action (a_{rd} or a_{rc}) for a switch. Vice versa, we will not schedule a switch maintenance action (a_{sw}) for a section.

[5]For the sake of clarity, we assume that each fault f_j is associated with one failure mode c_j.

Let the different components of the direct maintenance costs $c_{\mathrm{m}}(a)$ be given by:

$$c_{\mathrm{m},1}(a) = \begin{cases} 87.5 & \text{if } a = a_{\mathrm{rd}} \\ 72.5 & \text{if } a = a_{\mathrm{rc}} \\ 130 & \text{if } a = a_{\mathrm{sw}} \end{cases} \tag{6.34}$$

$$c_{\mathrm{m},2}(a) = \begin{cases} 7.5 & \text{if } a = a_{\mathrm{rd}} \\ 10 & \text{if } a = a_{\mathrm{rc}} \\ 15 & \text{if } a = a_{\mathrm{sw}} \end{cases} \tag{6.35}$$

$$c_{\mathrm{m},3} = 5 \tag{6.36}$$

The different components of the indirect maintenance costs $c_{\mathrm{i}}(a,t)$ are defined as follows:

$$c_{\mathrm{i},1}(t) = \begin{cases} 70 & \text{if } t \text{ is during the day} \\ 20 & \text{if } t \text{ is during the night} \end{cases} \tag{6.37}$$

$$c_{\mathrm{i},2}(a,t) = 55 \tag{6.38}$$

To plan or to postpone We define the parameters defining the reward function for timely scheduling (see Section 6.5.2) as:

$$u_{\mathrm{max}} = 100 \tag{6.39}$$

$$\delta_{\mathrm{d}} = 0.99 \tag{6.40}$$

$$\alpha = 5000 \tag{6.41}$$

SYSTEM-LEVEL COST FUNCTIONS

To define the cost function $f_{\mathrm{LF}}(\cdot)$ expressing the additional loss of functionality when multiple components are maintained simultaneously, we divide all components into three groups:

g_A: components of line A "Utrecht-Schiphol", i.e. $\mathcal{A}_{\mathrm{sc},1},\ldots,\mathcal{A}_{\mathrm{sc},n},\mathcal{A}_{\mathrm{sw},1},\ldots,\mathcal{A}_{\mathrm{sw},k}$

g_B: components of line B "Utrecht-Leiden", i.e. $\mathcal{B}_{\mathrm{sc},1},\ldots,\mathcal{B}_{\mathrm{sc},m},\mathcal{B}_{\mathrm{sw},1},\ldots,\mathcal{B}_{\mathrm{sw},j}$

g_C: components of line C "Leiden-Schiphol", i.e. $\mathcal{C}_{\mathrm{sc},1},\ldots,\mathcal{C}_{\mathrm{sc},l},\mathcal{C}_{\mathrm{sw},1},\ldots,\mathcal{C}_{\mathrm{sw},i}$

and accordingly define the cost function $f_{\mathrm{LF}}(\cdot)$ as:

$$f_{\mathrm{LF}}(\mathcal{X}_t) = \begin{cases} 35 & \text{if } g_A \in \mathcal{X}_t \text{ and } g_B \in \mathcal{X}_t \\ 20 & \text{if } g_A \in \mathcal{X}_t \text{ and } g_C \in \mathcal{X}_t \text{ and } g_B \notin \mathcal{X}_t \\ 0 & \text{otherwise} \end{cases} \tag{6.42}$$

Simultaneously maintaining components on lines A and B or components on lines A and C means that there is no train connection between Utrecht and Schiphol. Hence combining maintenance activities on lines A and B and on lines A and C is penalized. Because there are other public transport options between Leiden and Schiphol, which are absent between Utrecht and Leiden, simultaneously maintaining lines A and B is penalized more severely.

6.7.2. COMPONENT-LEVEL OPTIMIZATION

Consider that maintenance is planned once a day, i.e.:

$$\Delta = 1 \text{ day}$$

Assume that at time $\tau_0 = 0$, section $\mathcal{A}_{\text{sc},1}$, which is of type II indicates a need for maintenance and that component $\mathcal{A}_{\text{sc},1}$ was last maintained 150 days ago, i.e. $t_{\text{mnt}} = \tau_{-150}$. At time τ_0, the diagnostic result of component $\mathcal{A}_{\text{sc},1}$ is specified as:

$$P\big(H(\tau_0) = f_{\text{rd}}\big) = 0.1$$
$$P\big(H(\tau_0) = f_{\text{rc}}\big) = 0.9$$

and the prognostic result as:

$$\begin{bmatrix} \theta_{\text{rd},1}(\tau_0) \\ \theta_{\text{rd},2}(\tau_0) \\ \theta_{\text{rd},3}(\tau_0) \end{bmatrix} = \begin{bmatrix} 2.5 \\ 1 \\ 0.15 \pm 0.1 \end{bmatrix}$$
$$\begin{bmatrix} \theta_{\text{rc},1}(\tau_0) \\ \theta_{\text{rc},2}(\tau_0) \end{bmatrix} = \begin{bmatrix} 3.5 \\ 0.4 \pm 0.2 \end{bmatrix}$$

The prognostic result is shown in Figure 6.6.

From the diagnostic and prognostic results, we conclude that if rail contamination is present, the system degrades slowly and the 95% confidence interval of the expected time to failure is [161, 483]. However, there is also a small possibility of a rail defect, in which case the system degrades much faster and the 95% interval of the expected time to failure is [18, 92].

Assume that at time τ_0 the following times are available for maintenance:

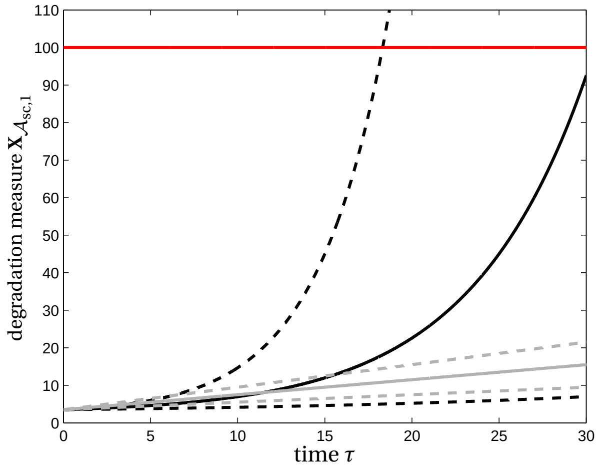
$$t_1 = \tau_0 + 0.2$$
$$t_p = \tau_{p-1}, \; p = 2, \ldots, 500$$

There is one immediate possibility (t_1) during the day in the case of an urgent fault. When the problem is not urgent, the maintenance will be scheduled at the most convenient time slot during the night.

The optimal and suboptimal maintenance strategies are found based on (6.1)-(6.6), with the cost functions as defined in (6.29)-(6.38). We found the optimal maintenance strategy with associated costs:

$$(a^*, t^*) = (a_{\text{rc}}, t_{12})$$
$$C^* = 418.2$$

while the first ten alternative strategies are given in Table 6.1.



Figuur 6.6: Prognostic result at τ_0. The solid black line correspond to the expected time behavior of the degradation measure in case of a rail defect and the solid gray line represents the time behavior in case of rail contamination. The dashed lines represent the 2.5% and 97.5% bounds of the distribution.

Tabel 6.1: Alternative maintenance strategies and associated costs at τ_0

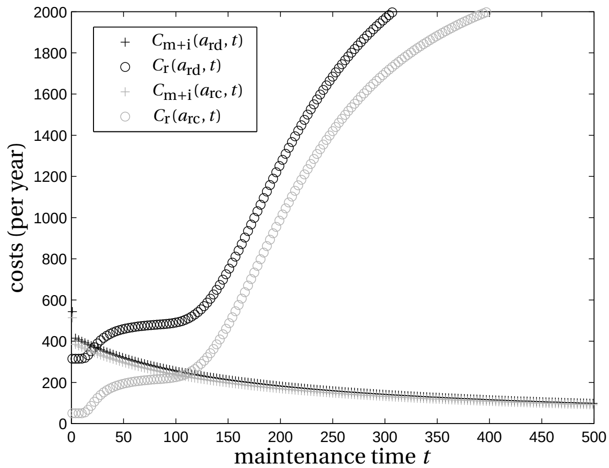
k	a_k^*	t_k^*	C_k^*	k	a_k^*	t_k^*	C_k^*
2	a_{rc}	t_{13}	418.2	7	a_{rc}	t_9	423.1
3	a_{rc}	t_{11}	419.2	8	a_{rc}	t_{16}	424.6
4	a_{rc}	t_{14}	419.3	9	a_{rc}	t_8	425.4
5	a_{rc}	t_{10}	420.9	10	a_{rc}	t_7	427.8
6	a_{rc}	t_{15}	421.5	11	a_{rc}	t_{17}	428.4

6

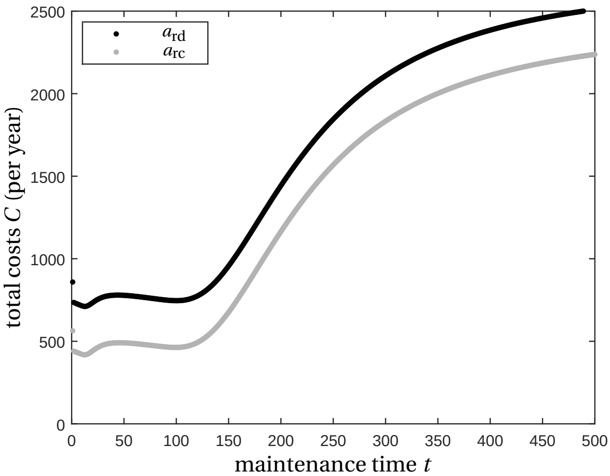
Figure 6.7 shows the cost components[6] $C_{\mathrm{m+i}}(a,t)$ and $C_{\mathrm{r}}(a,t)$ for the different maintenance strategies. For both types of maintenance a, the larger we chose t, the lower the (direct plus indirect) maintenance costs $C_{\mathrm{m+i}}(a,t)$, but the higher the risk costs $C_{\mathrm{r}}(a,t)$. The total costs C for the different strategies are shown in Figure 6.8.

The next step is to decide whether to accept this maintenance strategy or to postpone the maintenance decision to a later time (see Figure 6.3). The decision is postponed if it is expected that scheduling at a later monitoring time instant results in a higher reward. So, we postpone if 1. it is expected that at least at one later monitoring instant the costs of the optimal maintenance strategy have reduced more than the associated penalty costs for postponing have increased; and 2. according to our sequential decision

[6] $C_{\mathrm{m+i}}(a,t) = C_{\mathrm{m}}(a,t) + C_{\mathrm{i}}(a,t)$



Figuur 6.7: Cost components for the different maintenance strategies ($\tau = \tau_0$).



Figuur 6.8: Expected costs for the different maintenance strategies ($\tau = \tau_0$).

making strategy, we will actually decide to plan maintenance at one of these monitoring instants. Assuming that we have a certain prediction of the future costs and considering reward function (6.18), with the parameters as defined in (6.39)-(6.41), the decision is postponed if there exists $\tau_{\mathrm{h}} > \tau_0$ for which:

$$\Big(\underbrace{100}_{u_{\max}} \cdot \underbrace{0.99}_{\delta_{\mathrm{d}}}{}^{\tau_{\mathrm{h}}-\tau_0} - \hat{C}^*(\tau_{\mathrm{h}}|\mathcal{I}_0)\Big)\Big(1 - P\big(\mathcal{F}(\tau_h) = 1\big)\Big)$$
$$- \underbrace{5000}_{\alpha} P\big(\mathcal{F}(\tau_{\mathrm{h}}) = 1\big) \geq \underbrace{100}_{u_{\max}} - C^*(\tau_0)$$

with $\hat{C}^*(\tau_{\mathrm{h}}|\mathcal{I}_0)$ the expected costs of the optimal maintenance strategy determined at τ_{h} given the available diagnostic and prognostic information up to τ_0. At τ_0, the probability of failure is negligible in the first few days. Because of the low prediction accuracy, we expect a cost reduction that is larger than the penalty cost of postponing. Therefore, we postpone the decision to τ_1. Similarly we postpone at the next 149 decision instants τ_1 till τ_{149}, meaning that for all $\tau_{\mathrm{i}} \in \{\tau_1, \ldots, \tau_{149}\}$ there exists a $\tau_{\mathrm{h}} > \tau_{\mathrm{i}}$ for which:

$$\big(100 \cdot 0.99^{\tau_{\mathrm{h}}-\tau_{\mathrm{i}}} - \hat{C}^*(\tau_{\mathrm{h}}|\mathcal{I}_{\mathrm{i}})\big)\big(1 - P\big(\mathcal{F}(\tau_{\mathrm{h}}) = 1\big)\big)$$
$$-5000 P\big(\mathcal{F}(\tau_{\mathrm{h}}) = 1\big) \geq 100 - C^*(\tau_{\mathrm{i}})$$

We assume that at time τ_{150}, the diagnostic result of component $\mathcal{A}_{\mathrm{sc},1}$ is specified as:

$$P\big(H(\tau_{150}) = f_{\mathrm{rd}}\big) = 0.0025$$
$$P\big(H(\tau_{150}) = f_{\mathrm{rc}}\big) = 0.9975$$

and the prognostic result as:

$$\begin{bmatrix} \theta_{\mathrm{rd},1}(\tau_{150}) \\ \theta_{\mathrm{rd},2}(\tau_{150}) \\ \theta_{\mathrm{rd},3}(\tau_{150}) \end{bmatrix} = \begin{bmatrix} 64 \\ 1 \\ 0.10 \pm 0.05 \end{bmatrix}$$
$$\begin{bmatrix} \theta_{\mathrm{rc},1}(\tau_{150}) \\ \theta_{\mathrm{rc},2}(\tau_{150}) \end{bmatrix} = \begin{bmatrix} 65 \\ 0.41 \pm 0.075 \end{bmatrix}$$

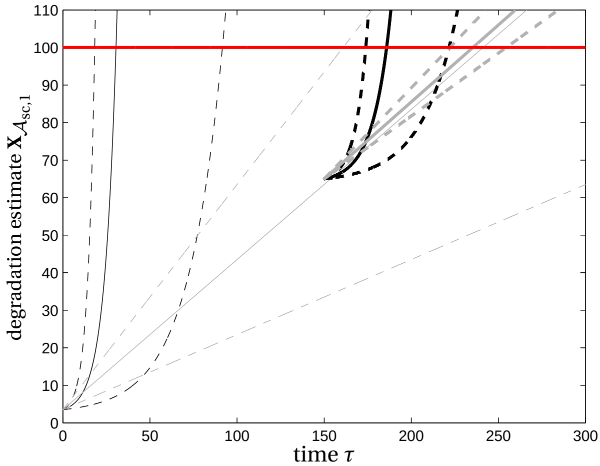
The prognostic result is shown in Figure 6.9.

We conclude that the actual value of the degradation measure $\mathbf{X}_{\mathcal{A}_{\mathrm{sc},1}}(\tau_{150})$ is close to the value predicted by the parametrized model of rail contamination defined at τ_0. The newly obtained parametrized models are however more accurate. Given that rail contamination is present, the 95% confidence interval of the time-to-failure distribution has reduced to $[222 - 150, 254 - 150]$. In case of a rail defect, the probability of which has become really small, the 95% confidence interval is $[174 - 150, 222 - 150]$.

Assume that at time τ_{150} the following maintenance time slots are available:

$$t_{151} = \tau_{150} + 0.2$$
$$t_p = \tau_{p-1}, \; p = 152, \ldots, 500$$

6



Figuur 6.9: Prognostic results at τ_0 and τ_{150}. The black lines correspond to the expected time behavior of the degradation measure in case of a rail defect and the gray lines represent the time behavior in case of rail contamination. The dashed lines represent the 2.5% and 97.5% bounds of the distribution.

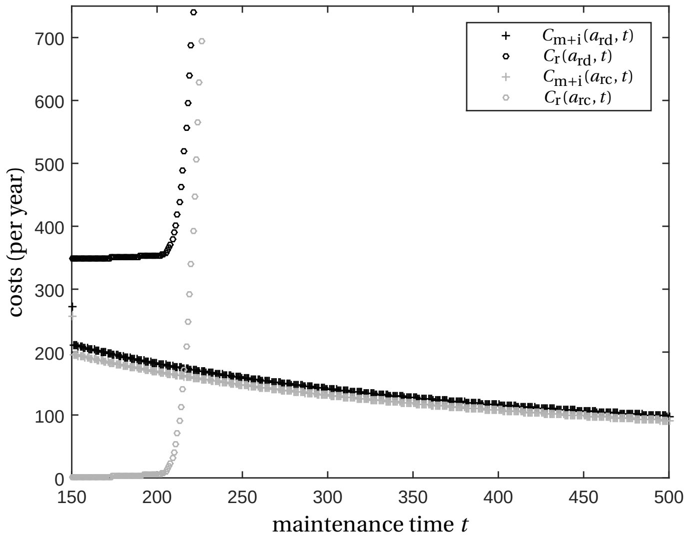
The optimal and suboptimal maintenance strategies are found according to (6.1)-(6.6), with the cost functions as defined in (6.29)-(6.38):

$$(a^*, t^*) = (a_{\text{rc}}, t_{201})$$
$$C^* = 174.0$$

6

The first ten alternative strategies and their associated costs are given in Table 6.2; Figures 6.10 and 6.11 show the costs for the different maintenance strategies.

We conclude that the costs of the optimal strategy obtained at τ_{150} are significantly lower compared to the costs of the optimal strategy obtained at τ_0. By postponing the decision we have limited the scheduling possibilities. However, the penalty cost for the delay in planning (maximum 100) is lower than the reduction in maintenance costs $(418.2 - 174.0)$. We assume that the expected cost reduction with later policies is small. Therefore, we accept this maintenance strategy.



Figuur 6.10: Cost components for the different maintenance strategies ($\tau = \tau_{150}$).

Tabel 6.2: Alternative maintenance strategies and associated costs at τ_{150}

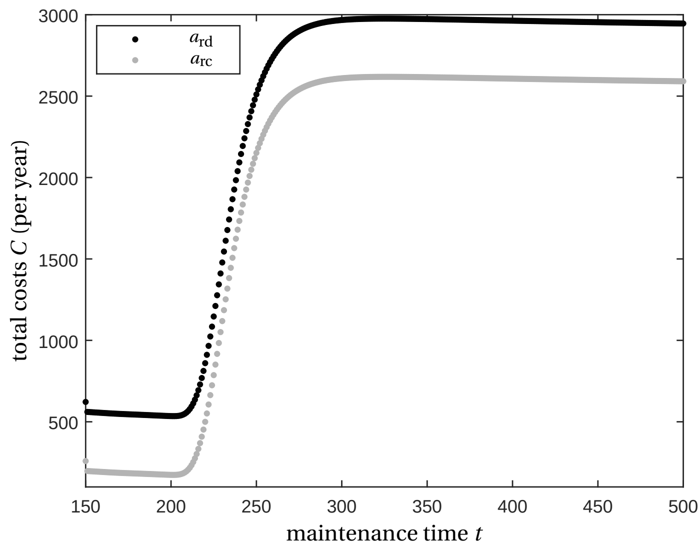
k	a_k^*	t_k^*	C_k^*	k	a_k^*	t_k^*	C_k^*
2	a_{rc}	t_{202}	174.0	7	a_{rc}	t_{197}	175.2
3	a_{rc}	t_{200}	174.1	8	a_{rc}	t_{204}	175.2
4	a_{rc}	t_{203}	174.4	9	a_{rc}	t_{196}	175.6
5	a_{rc}	t_{199}	174.4	10	a_{rc}	t_{195}	176.0
6	a_{rc}	t_{198}	174.8	11	a_{rc}	t_{194}	176.5

6

6.7.3. SYSTEM-LEVEL OPTIMIZATION

Consider that 7 system components, namely $\mathcal{A}_{sc,1}, \mathcal{A}_{sc,2}, \mathcal{A}_{sw,1}, \mathcal{B}_{sc,1}, \mathcal{B}_{sw,1}, \mathcal{C}_{sc,1}$, and $\mathcal{C}_{sw,1}$, are in need of maintenance. For section $\mathcal{A}_{sc,1}$, maintenance is already scheduled[7] (see Table 6.3). For each other component, the optimal and two near-optimal component-level strategies are given in Table 6.3.

[7] Remember that the time of optimization at the system-level is determined by the time the maintenance needs are finalized at the component-level.



Figuur 6.11: Expected costs for the different maintenance strategies ($\tau = \tau_{150}$).

Tabel 6.3: Inputs system-level optimization, with the optimal system-level strategy marked in gray

component	action a	time t	costs C
$\mathcal{A}_{sc,1}$	a_{rc}	t_{202}	181.6
$\mathcal{A}_{sc,2}$	a_{rc}	t_{155}	186.2
	a_{rc}	t_{180}	190.0
	a_{rc}	t_{202}	242.6
$\mathcal{A}_{sw,1}$	a_{sw}	t_{155}	195.0
	a_{sw}	t_{180}	191.7
	a_{sw}	t_{202}	198.3
$\mathcal{B}_{sc,1}$	a_{rc}	t_{155}	201.0
	a_{rc}	t_{180}	191.2
	a_{rc}	t_{202}	179.8
$\mathcal{B}_{sw,1}$	a_{sw}	t_{155}	181.0
	a_{sw}	t_{180}	211.2
	a_{sw}	t_{202}	233.0
$\mathcal{C}_{sc,1}$	a_{rc}	t_{155}	203.6
	a_{rc}	t_{180}	191.2
	a_{rc}	t_{202}	169.8
$\mathcal{C}_{sw,1}$	a_{sw}	t_{155}	201.5
	a_{sw}	t_{180}	189.3
	a_{sw}	t_{202}	165.4

Minimizing the system-level optimization criterion (6.25) with the costs as defined in Section 6.7.1 results in the following system-level maintenance schedule:

- at t_{155}: $\mathcal{B}_{\text{sw},1}$
- at t_{180}: $\mathcal{A}_{\text{sc},2}$
- at t_{220}: $\mathcal{A}_{\text{sc},1}, \mathcal{A}_{\text{sw},1}, \mathcal{B}_{\text{sc},1}, \mathcal{C}_{\text{sc},1}, \mathcal{C}_{\text{sw},1}$

So, except for switch $\mathcal{B}_{\text{sw},1}$ and section $\mathcal{A}_{\text{sc},2}$, maintenance is scheduled simultaneously with the maintenance on section $\mathcal{A}_{\text{sc},1}$ at t_{220}. For components $\mathcal{B}_{\text{sc},1}$, $\mathcal{C}_{\text{sc},1}$, and $\mathcal{C}_{\text{sw},1}$, t_{202} coincides with the optimal component-level time. For switch $\mathcal{A}_{\text{sw},1}$, t_{202} is not optimal. However, the cost benefit of combining maintenance on two switches is larger than the increase of the component-level maintenance costs when scheduling the maintenance at a near-optimal time. At t_{155} maintenance on line B is scheduled, meaning that scheduling maintenance on $\mathcal{A}_{\text{sw},1}$ at t_{155} results in an additional penalty of 35. Therefore, the maintenance of switch $\mathcal{A}_{\text{sw},1}$ is scheduled at t_{202}.

Maintenance for section $\mathcal{A}_{\text{sc},2}$ and switch $\mathcal{B}_{\text{sw},1}$ is scheduled to be performed at another time. For these components, the costs of performing maintenance at t_{202} are too high compared to the costs at their optimal component-level time t_{155}, i.e. the reduction as a consequence of economies of scale does not outweigh the increase in component-level strategy costs. To avoid penalty costs due to performing maintenance on line A and B simultaneously, maintenance on $\mathcal{A}_{\text{sc},2}$ is scheduled at the first alternative maintenance time t_{180}.

Note that the optimal system-level strategy is a direct result of the adopted cost functions. In the case the cost reductions from economies of scale are extremely high, all components will be maintained at t_{220}. When loss of functionality is severely penalized, all components on line A will be maintained at maintenance time t_{155} and t_{202}, while all components on line B and C will be maintained at t_{180}.

6.8. CONCLUSIONS

We have proposed a two-stage optimization approach to timely maintenance planning in heterogeneous systems. In the first stage, the maintenance needs of the individual system components are determined. In the second stage we optimize the maintenance schedule at the system level. More specifically:

1. We optimize both the required type of maintenance and the time to perform the maintenance;
2. We optimize the time to settle on the aforementioned maintenance decisions, hereby trading accuracy with timeliness;
3. We decouple the maintenance optimization from the diagnosis and prognosis process. In this way, we are able to exploit both diagnostic and prognostic information for maintenance optimization without restricting ourselves to a particular degradation model;

4. We provide a systematic framework for incorporating economic and structural dependencies among system components.

Advantages of the proposed approach are: 1. at the component level, maintenance is planned timely when this does not lead to violations of cost and safety constraints. Timely planning allows to perform maintenance at a convenient time, to inform users regarding system downtime, and to optimize the management of spare parts, material, and personnel; 2. at the system level, maintenance costs are significantly reduced by adequately combining or spreading maintenance activities. The applicability of the method is demonstrated through a case study concerning maintenance planning in a railway network.

The proposed method can be extended and improved in various ways. Some possible directions for future research are:

1. In this work, we assumed that monitoring is done at fixed time intervals. This is a realistic assumption for systems for which high-frequency monitoring is relatively cheap. When each monitoring round is associated with high costs (e.g. when the monitoring data are collected by a measurement train) it is more realistic to perform the monitoring at optimized time instants. Therefore, the optional inclusion of the next monitoring time instant as a decision variable would be an interesting extension of the proposed approach.

2. The approach can be extended with the possibility to perform additional monitoring before taking a maintenance decision. In the context of the railway case, this could e.g. refer to additional measurements by a measurement train when the track-side monitoring data do not provide enough information to make an informed decision.

3. In this work, we considered only the possibility of major maintenance. Depending on the type of system, it would be beneficial to include additional maintenance options, e.g. minimal maintenance, major maintenance, and repair.

4. In this work, we assumed that the cost savings due to economies of scale do not depend on the locations of the components. However, for large-scale systems the cost savings may depend on the exact locations of the different components. Therefore, an interesting extension of the proposed method is to make the cost savings resulting from economies of scale dependent on the locations of the components.

5. The approach can be extended by including the possibility to reschedule maintenance.

6. A further optimization of the order and frequency at which maintenance activities are planned in the system-level optimization.

7

CONCLUSIONS

"Success comes from knowing what you do not know,
more than coming from what you do know."

-Ray Dalio-

In this thesis, methods have been proposed for the different tasks of the condition-based maintenance process; fault diagnosis, failure prognosis, and maintenance optimization. Compared to existing methods, the assumptions underlying the proposed approaches better match the conditions encountered in practice, and so the presented approaches support the actual implementation of condition-based maintenance. In the remainder of this chapter, we summarize our main contributions and conclusions. Additionally, we present open challenges and give recommendations for further research.

7.1. SUMMARY OF CONTRIBUTIONS AND CONCLUSIONS

Here, we outline our main contributions and conclusions, which concern fault diagnosis, failure prognosis, and maintenance optimization as well as the overall condition-based maintenance process.

REASONING UNDER UNCERTAINTY FOR KNOWLEDGE-BASED FAULT DIAGNOSIS

We have analyzed the knowledge-based diagnosis problem and, in particular, investigated how the diagnosis task is influenced by uncertainty. It is concluded that the diagnosis task is influenced by uncertainty in various ways. The exact uncertainty characteristics depend e.g. on the measurement equipment used and on the available diagnostic knowledge. We have compared how the diagnosis problem is handled in two well-known reasoning frameworks, namely the Bayesian framework and the Dempster-Shafer framework. The Bayesian model is tailored to causal reasoning based on probabilistic information, while the Dempster-Shafer model is tailored to non-causal reasoning based on both probabilistic and incomplete information. Since fault diagnosis comprises causal reasoning, often based on incomplete information, none of the two reasoning approaches fits the diagnostic reasoning task in a straightforward way. Moreover, additional objectives, like limited computational power and clarity of inference, may influence the final choice for a method. The suitability of a particular method therefore depends, among other things, on the specific diagnosis task and on user requirements. Consequently, we cannot conclude the superiority of one of the two methods for fault diagnosis in general. Instead, guidelines have been provided to support the choice for one of the two reasoning frameworks. In general, the better the match between the probabilistic description and the real information, the more suitable the Bayesian approach is. The more conflicting and incomplete the available information, the more informative the Dempster-Shafer solution is compared to the Bayesian solution.

EXPLOITING SYSTEM DEPENDENCE FOR FAULT DIAGNOSIS OF INTERCONNECTED SYSTEMS
In this thesis, two approaches to fault diagnosis have been proposed. These approaches explicitly account for the following, often overlooked, practicalities:

1. Only a limited set of monitoring signals is generally available;
2. Most systems are subject to environmental disturbances and multiple operating modes;
3. Dependencies may exist among system components.

The first approach concerns fault diagnosis for general interconnected systems. This approach is knowledge-based and uses the temporal, spatial, and spatio-temporal system dependencies as diagnostic features. These features can in general be derived from existing monitoring signals; so, no additional monitoring equipment is required. In addition, by using the spatial and spatio-temporal dependencies as diagnostic features, component interdependencies are automatically accounted for. Finally, thanks to the use of the spatial dependencies, the method is robust with respect to environmental disturbances. The applicability of the method has been demonstrated on a railway track circuit diagnosis case. It has been shown that the proposed method is able to adequately detect and diagnose track circuit faults, even in the presence of environmental disturbances. Compared to the current practice of threshold checking, the proposed approach provides more timely insight into faulty behavior and a characterization of the type of fault present. This additional information is important for creating an effective condition-based maintenance schedule.

Second, we have proposed a multiple-model approach to fault diagnosis of heating, ventilation, and air conditioning (HVAC) systems. Next to prior knowledge, historical data are incorporated in the diagnostic model through the use of virtual sensors. This way, diagnostic power improves, while the reasoning remains transparent. To handle multiple operating modes and uncertainty, a separate diagnostic model is defined for each operating mode and captured by a Bayesian network. Component interdependencies are handled by performing diagnosis at the system level, rather than at the level of the individual components, and by incorporating knowledge regarding component interdependencies in the diagnostic model. The need for and the applicability of the proposed method have been demonstrated based on various case studies. It is concluded that faults are timely and properly diagnosed, even in the case of multiple faults, provided that the faults result in observable behavior.

A MULTIPLE-MODEL APPROACH TO SYSTEM RELIABILITY PREDICTION

We have presented a multiple-model approach to system reliability prediction. The approach consists of two parts: the first part includes a multiple-model approach to multivariate degradation forecasting. The second part includes a framework to determine, based on the predicted system degradation, (future) system reliability.

Compared to existing methods for failure prognostics, the proposed approach has the following features:

1. The method fits well with the subsequent maintenance optimization process;
2. Multiple degradation models are considered to minimize the modeling error;
3. The method effectively combines the information from multiple degradation measures.

We conclude that, in the presence of multiple degradation modes and provided they are correctly identified, a multiple-model approach outperforms a single-model approach with respect to prediction accuracy. However, since the applicable model is selected based on the diagnostic result, the benefit of using multiple models over using a single model highly depends on the accuracy of the diagnostic result.

TIMELY MAINTENANCE PLANNING USING DIAGNOSTIC AND PROGNOSTIC INFORMATION

We have proposed a two-stage optimization approach to timely maintenance planning in heterogeneous systems. In the first stage, the maintenance needs of the individual system components are determined. In the second stage we optimize the maintenance schedule at the system level. Compared to existing methods on condition-based maintenance planning, our method adds the following:

1. We optimize both the required type of maintenance and the time to perform the maintenance;
2. We optimize the time to settle on the aforementioned maintenance decisions, hereby trading accuracy with timeliness;
3. We decouple the maintenance optimization from the diagnosis and prognosis process. In this way, we are able to exploit both diagnostic and prognostic information for maintenance optimization without restricting ourselves to a particular degradation model;
4. We provide a systematic framework for incorporating economic and structural dependencies among system components.

Main features of the proposed approach are: 1. At the component level, maintenance is planned timely when this does not lead to violations of cost and safety constraints. Timely planning allows to perform maintenance at a convenient time, to inform users regarding system downtime, and to optimize the management of spare parts, material, and personnel; 2. At the system level, maintenance costs are significantly reduced by adequately combining or spreading maintenance activities. The applicability of the method is demonstrated through a case study concerning maintenance planning in a railway network.

THE OVERALL CONDITION-BASED MAINTENANCE PROCESS

Fault diagnosis, failure prognosis, and maintenance optimization serve a common goal; improving system safety while minimizing maintenance costs and system downtime. To reach this goal, the different processes should connect to each other. The methods proposed in this thesis account for the, often overlooked, dependencies between fault diagnosis, failure prognosis, and maintenance optimization. More specifically:

- We ensure that the diagnostic and prognostic result are suitable inputs for the maintenance optimization process;
- We account for interactions between the diagnosis and prognosis process;
- We use both the diagnostic and the prognostic result for maintenance optimization without restricting ourselves to a specific diagnostic and prognostic approach.

It was a deliberate choice to treat the different sub-tasks individually instead of considering the condition-based maintenance task as a whole. Since every problem is different and the available knowledge and data may vary from case to case and over time, the optimal combination of a diagnosis and prognosis strategy is situation-specific and possibly time-varying. Individually optimizing the different sub-tasks, provides the freedom to accommodate the individual needs. As an example we mention the differences in available historical data for the track circuit case and the HVAC case. Accordingly a knowledge-based approach is considered for the railway case, while a combined knowledge and data-based approach is used for the HVAC case.

7.2. RECOMMENDATIONS FOR FUTURE RESEARCH

The methods proposed in this thesis address the challenges presented in Chapter 1. Nevertheless, various opportunities remain left for future research. First, the methods proposed in thesis need to be further developed and validated. Second, during our research we faced some new challenges. In the individual chapters, we already pointed out some possible directions for future research. In this section, we discuss the most important directions.

FAULT DIAGNOSIS

- For each system and each operation mode a different diagnostic model needs to be constructed. Much time and effort is saved when the diagnostic model can be generated automatically for a class of systems (e.g. buildings, railway networks). An interesting topic for further research would therefore be the development of methods to automate the construction of the diagnostic model.

FAILURE PROGNOSIS

- We conclude that by using multiple models to forecast degradation behavior, the modeling error can be reduced. However, since the applicable model is selected based on the diagnostic result, the benefit of using multiple models over using a single model highly depends on the accuracy of the diagnostic result. Given the current high quality of diagnosis methods, we do not expect this to be a serious drawback. However, caution is needed when faults are in their incipient phase. In this phase, the diagnostic results are often less accurate. A thorough analysis of the accuracy of diagnostic and prognostic results over time, and its implications on the subsequent maintenance optimization process would therefore be an interesting topic for further research.

MAINTENANCE OPTIMIZATION

- The maintenance optimization approach proposed in this thesis excludes the possibility to reschedule maintenance. Preferably maintenance is planned correctly at once. However, it may happen that at some time we find out that a previously made decision is far from optimal. In this case, the benefits of rescheduling (e.g. reduction of maintenance costs) may outweigh its drawbacks (e.g. last minute adjustments, non-optimal management of spare parts and personnel). Therefore an interesting extension of the proposed maintenance optimization approach is to include the possibility of maintenance rescheduling.

- At the system level, maintenance is scheduled in the first optimization round after the maintenance strategy is finalized at the component level. More efficient maintenance schedules are however expected when the optimization frequency and order are optimized. Consider for instance that for some components maintenance strategies are set at the component-level, but the associated optimal times to perform the maintenance activities are in the far future. At the same time, we expect that component-level maintenance strategies of other components become available soon. In this case, it might be advantageous to wait and to optimize the maintenance for all components together. This would require communication between the component-level and system-level optimization processes, which is currently lacking. A further optimization at the system level would therefore be an interesting direction for future research.

OVERALL CONDITION-BASED MAINTENANCE PROCESS

- As indicated before, extensive evaluations of the proposed approaches are still required. The individual approaches can only be evaluated in an adequate way when they are implemented within a condition-based maintenance scheme. For example, to predict future system reliability, we rely on the diagnostic result. Moreover, we are not directly interested how adequate our estimates of the current system health and predictions of future system reliability are, but we are interested in how these results contribute to the maintenance planning. Therefore, an important topic for future research includes the thorough evaluation of the proposed approaches within a condition-based maintenance scheme.

- In this thesis, we focused on continuous condition monitoring. Moreover, diagnosis, prognosis, and maintenance optimization are a performed frequently in time. When continuous monitoring is not possible, or when the costs of diagnosis, prognosis, or maintenance optimization are high, the time instants of monitoring, diagnosis, prognosis, and maintenance optimization need to be optimized. This can for example be done by determining the next monitoring instant based on the current diagnostic and prognostic results, or by considering an event-triggered approach. For fault diagnosis in a railway network this could, e.g., mean that the track side monitoring signals are used to trigger train-based or manual inspections. As a topic for future work, we propose the development of approaches to find the optimal trade-off between monitoring costs and maintenance optimization performance.

A

Bayesian and Dempster-Shafer reasoning

This appendix provides background information regarding reasoning in Bayesian and Dempster-Shafer networks. This background information is in support of the concepts discussed in Chapter 2.

A.1. Reasoning in Bayesian networks

A.1.1. Uncertainty representation

In the Bayesian framework uncertainty is represented by (conditional) probabilities. At each time and for each variable[1] X, a conditional probability $P(x_i|\mathcal{E})$ between zero and one is assigned to *each individual element* x_i in the domain Θ_X of X such that (Darwiche, 2009):

$$\sum_{x_i \in \Theta_X} P(x_i|\mathcal{E}) = 1 \tag{A.1}$$

with $\mathcal{E}$ the collection of the currently available information.

A.1.2. Bayesian networks

A Bayesian network is a graphical model for probabilistic relationships among a set of variables that provides a powerful way to embed knowledge and to update one's beliefs about target variables given new information about other variables (Heckerman, 1998; Mrad et al., 2015). Formally, a Bayesian network for a set of variables $\mathbf{V}$ is a pair (G, D) (Heckerman, 1998; Mrad et al., 2015), with:

1. $G = (\mathbf{V}, \mathbf{E})$ a directed acyclic graph with nodes $\mathbf{V}$ and directed edges $\mathbf{E}$ that encodes a set of conditional independence assertions about the variables in $\mathbf{V}$;

[1] This section is elaborated for variables X for which the domain Θ_X is finite.

2. D a set of local probability distributions associated with each variable in $\mathbf{V}$.

In a Bayesian network, a directed edge from a variable X to a variable Y indicates that X has a direct influence on variable Y. Variable X is then called a *parent* of variable Y and variable Y is called a *child* of variable X. The absence of an edge in G encodes conditional independence (Heckerman, 1998). Bayesian networks satisfy the Markov condition, meaning that any node is conditionally independent of its non-descendants given its parents. Thanks to the Markov assumption, the joint distribution of the complete system can be obtained in an efficient way by combining the conditional distributions of each variable given its parents (Pearl, 1988; Darwiche, 2009): Given the network structure (G, D), the joint probability distribution for $\mathbf{V}$ is given by:

A

$$P(\mathbf{V} = \mathbf{v}) = \prod_{x_i \in \mathbf{v}} P(x_i | \mathbf{u}_X) \tag{A.2}$$

with $\mathbf{U}_X \subset \mathbf{V}$ the parents (immediate predecessors) of $X \in \mathbf{V}$ and $P(x_i|\mathbf{u}_X)$ the local probabilities associated with variable X, which are collected in D. Consequently, the pair (G, D) uniquely defines the joint probability distribution of $\mathbf{V}$.

A.1.3. REASONING UNDER UNCERTAINTY

Once the Bayesian network has been constructed (from prior knowledge, data, or a combination of both), we can use it to determine the probabilities of interest. This process is known as *probabilistic inference* (Heckerman, 1998). In explaining probabilistic inference, we make a distinction between inference with hard evidences and inference with uncertain evidences.

INFERENCE WITH HARD EVIDENCES

Probabilistic inference with hard evidences can be regarded as a mechanism for automatically applying Bayes' rule:

$$P(y_i | x_i) = \frac{P(x_i | y_i) P(y_i)}{\sum_{y_j \in \Theta_Y} P(x_i | y_j) P(y_j)} \tag{A.3}$$

with:

$P(y_i)$: prior probability that $Y = y_i$

$P(y_i|x_i)$: posterior probability, i.e. the probability that $Y = y_i$ after observing $X = x_i$

$P(x_i|y_i)$: likelihood function, i.e. the probability of observing $X = x_i$ given $Y = y_i$

The importance of Bayes' rule is that it expresses a quantity $P(y_i|x_i)$, which is often difficult to assess, in terms of quantities that often can be drawn directly from expert knowledge (Pearl, 1988). For a more thorough discussion on inference algorithms in Bayesian networks, we refer the interested reader to e.g. (Pearl, 2000).

INFERENCE WITH UNCERTAIN EVIDENCES

In practice, the available evidences are often uncertain, in which case Bayes' rule is not directly applicable. With respect to uncertain evidences, a distinction can be made between (Mrad et al., 2015):

1. likelihood (or virtual) evidence;
2. probabilistic evidence:
   (a) fixed;
   (b) non-fixed.

A likelihood evidence on a variable $X \in \mathbf{V}$ is specified by likelihood ratios $L(X)$:

$$L(X) = P(\eta|x_1) : \ldots : P(\eta|x_n) \tag{A.4}$$

with $P(\eta|x_i)$ the probability of the observation η given $X = x_i$. Likelihood evidence concerns evidence with uncertainty, i.e. the uncertainty refers to the meaning of the input (Dubois et al., 1998); the existence of the input itself is uncertain due to e.g. the unreliability of the source that supplies the input (Mrad et al., 2014). Note that likelihood evidence is specified "without a prior". As a consequence, its correct propagation requires both the evidence and the current belief in X to be taken into account.

A probabilistic evidence on a variable $X \in \mathbf{V}$ is specified by a local probability distribution $R(X)$ that defines a constraint on the beliefs on the variable X after the evidence has been propagated, i.e. $R(X)$ is an absolute constraint on the posterior probability distribution of X. Probabilistic evidence concerns evidence of uncertainty, i.e. the uncertainty is part of the input (Dubois et al., 1998). Fixed probabilistic evidence cannot be altered by any further information, while non-fixed probabilistic information can be modified based on later evidences (Mrad et al., 2015).

Two main methods exist for revising probabilistic belief in the case of uncertain evidence (Chan and Darwiche, 2005):

1. Jeffrey's rule of probability kinematics;
2. Pearl's method of virtual evidence.

Likelihood evidence is propagated by Pearl's method of virtual evidence, while probabilistic evidence is propagated following Jeffrey's rule. Note that for the propagation of multiple fixed probabilistic evidences, specific iterative algorithms, such as big clique or BN-IPFP, are needed to ensure that all constraints imposed by the different evidences are satisfied (Mrad et al., 2015). Although the various belief revision principles seem to be different, they are all based on the principle of probability kinematics (Chan and Darwiche, 2005), which can be viewed as a principle for minimizing belief change while satisfying the (absolute or relative) constraints imposed by the evidence. In addition, it has been shown that one can translate an evidential constraint used by Jeffrey's rule into one used by Pearl's method and vice versa (Chan and Darwiche, 2005). Furthermore, as Pearl's method is directly applicable to Bayesian networks, while Jeffrey's rule is not, we will only elaborate on Pearl's method here. For a more thorough discussion on belief propagation based on uncertain evidences, we refer the interested reader to (Chan and Darwiche, 2005; Mrad et al., 2015) and the references therein.

Pearl's method of virtual evidence (Chan and Darwiche, 2005): Given an original distribution $P(\mathbf{V})$ and some uncertain evidence η regarding variable $X \in \mathbf{V}$, a likelihood evidence is specified by $\lambda_1, ..., \lambda_n$ as:

$$P(\eta|x_1) : ... : P(\eta|x_n) = \lambda_1 : ... : \lambda_n \tag{A.5}$$

meaning that:

$$\frac{P(\eta|x_i)}{P(\eta|x_j)} = \frac{\lambda_i}{\lambda_j} \text{ for } i = 1,\dots,n,\ j = 1,\dots,n$$

A

The method assumes that the observation η depends only on variable X and is independent of any other variable $Y \in \mathbf{V}$ given X:

$$P(\eta|x_i, y_i) = P(\eta|x_i), \text{ for } i = 1, ..., n \tag{A.6}$$

This results in the following expression for the revised distribution:

$$P(y_i|\eta) = \frac{\sum_{j=1}^{n} \lambda_i P(y_i, x_j)}{\sum_{j=1}^{n} P(x_j)} \tag{A.7}$$

In a Bayesian network, the virtual evidence is represented by adding an auxiliary variable Z and a directed edge $X \to Z$, where one value of Z, say z_i, corresponds to the virtual event η. This ensures assumption (A.6): the virtual event η is independent of every variable Y given X. The uncertainty of the evidence is quantified by the likelihood ratios $\lambda_1, ..., \lambda_n$ and the conditional probability table of variable Z is assigned such that $P(z_i|x_1), ..., (z_i|x_n) = \lambda_1 : ... : \lambda_n$. The Bayesian network (augmented with variable Z) is updated in the standard way with observation $Z = z_i$, which is a hard evidence.

A.1.4. DECISION MAKING

Often, decisions have to be made given uncertain information regarding the situation you are in. When the uncertain information is represented by a probability distribution, expected-utility theory is generally used for decision making. *Expected-utility theory* (Lindley, 1985) provides a framework for determining the optimal action given probabilistic information regarding the situation you are in. Its two main ingredients are:

1. Utilities, which indicate the desirability of a particular action in a particular situation, i.e. utilities express preferences among the available choices.
2. Probabilities, which indicate how likely a particular situation is.

The expected utility $\mathbb{E}(u|d)$ of a decision $d \in \Theta_D$ is computed as:

$$\mathbb{E}(u|d) = \sum_{v \in \Theta_V} P(v) u(d, v) \tag{A.8}$$

with Θ_D the discrete set of possible decisions, Θ_V the set of possible situations, $u(d, v)$ the utility of decision d given situation v, and $P(v)$ the probability of v. Then, an optimal decision d^* is:

$$d^* = \arg\max_{d \in \Theta_D} \mathbb{E}(u|d) \tag{A.9}$$

A.2. REASONING IN DEMPSTER-SHAFER NETWORKS

We adopt Smets' Transferable Belief Model (TBM) interpretation of the Dempster-Shafer theory (Smets and Kennes, 1994). In the TBM, uncertainty is managed at two levels: the *credal level* where beliefs are entertained and the *pignistic level* where beliefs are used to make decisions (Smets and Kennes, 1994). At the credal level, the model does not rely on a probabilistic quantification, but on a more general system based on belief functions (Smets, 1994). In contrast to Bayesian probabilities, belief functions can express states of ignorance.

The theory presented in the remainder of this section is based on (Shafer, 1976; Smets, 1978; Yager, 1987; Shenoy, 1989; Smets, 1990; Shenoy, 1992a,b; Smets, 1994; Smets and Kennes, 1994; Cobb and Shenoy, 2003b; Yaghlane et al., 2003; Yaghlane and Mellouli, 2008).

A

A.2.1. UNCERTAINTY REPRESENTATION

To enable the expression of (partial) ignorance, in the D-S framework, belief is assigned to each subset of the domain Θ_Y of a variable Y. The power set of Θ_Y, denoted as 2^{Θ_Y}, is a set containing all the possible subsets of Θ_Y. The mapping $\text{bel}: 2^{\Theta_Y} \to [0,1]$ is a belief function if and only if there exists a basic belief assignment (bba) $m^{\Theta_Y}: 2^{\Theta_Y} \to [0,1]$ such that (Yaghlane and Mellouli, 2008):

$$\sum_{y \subseteq \Theta_Y} m^{\Theta_Y}(y) = 1 \tag{A.10}$$

$$\text{bel}(y) = \sum_{\emptyset \neq x \subseteq y} m^{\Theta_Y}(x), \text{ and } \text{bel}(\emptyset) = 0 \tag{A.11}$$

$$\text{pl}(y) = \sum_{x \cap y \neq \emptyset} m^{\Theta_Y}(x), \text{ and } \text{pl}(\emptyset) = 0 \tag{A.12}$$

The mass $m^{\Theta_Y}(y)$ allocated to $y \subseteq \Theta_Y$ is the degree of belief that is exactly committed to y and that cannot be allocated to a more specific subset. The value $\text{bel}(y)$ quantifies the strength of the belief that the event y occurs. The value $\text{pl}(y)$ quantifies the maximum amount of potential specific support that could be given to y. It can be interpreted as the degree to which the evidence is not contradictory with y, i.e. $\text{pl}(y) = 1 - \text{bel}(\bar{y})$, where $\bar{y}$ is the complement of y with respect to the domain Θ_Y. When mass is only assigned to singleton elements, the mass distribution reduces to a probability distribution.

For illustration consider the uncertain variable H, with $\Theta_H = \{h_1, h_2, h_3\}$ and belief assignment:

$$\begin{aligned} m^{\Theta_H}(\{h_1\}) &= 0.1 \\ m^{\Theta_H}(\{h_2, h_3\}) &= 0.9 \\ m^{\Theta_H}(\{h_2\}) = m^{\Theta_H}(\{h_3\}) = m^{\Theta_H}(\{h_1, h_2\}) &= 0 \\ m^{\Theta_H}(\{h_1, h_3\}) = m^{\Theta_H}(\Theta_H) = m^{\Theta_H}(\emptyset) &= 0 \end{aligned} \tag{A.13}$$

Mass distribution (A.13) indicates that no information is available to discriminate between the outcomes h_2 and h_3. Note that the Bayesian model cannot represent such incomplete information. In the Bayesian model, the mass assigned to $\{h_2, h_3\}$ would typically be equally divided between the two elements (principle of maximum entropy). The situation in which one knows that h_2 and h_3 are equally likely and the situation in which one does not know anything about the individual probabilities $P(h_2)$ and $P(h_3)$ result in the same probability distribution. This is precisely what D-S proponents claim as the main shortcoming of the Bayesian framework (Cobb and Shenoy, 2003b).

When we model aspects of the real world, we often have to deal with multivariate situations (Yaghlane et al., 2003), where the state space is a product space and information may be available in a conditional form. Multivariate belief function theory is well suited to handle real-world problems. A multivariate mass function $m^{\Theta_X \times \Theta_Y}$ on $\Theta_X \times \Theta_Y$ can be seen as an uncertain relation between variables X and Y. To extend the theory discussed so far to multivariate problems, the following operations are defined (Cobb and Shenoy, 2003b; Yaghlane et al., 2003):

A

1. cylindrical extension to convert a mass function to a mass function on a larger space;
2. marginalization to convert a mass function to a mass function on a smaller space;
3. ballooning extension to convert conditional information to a mass function on the joint space.

The different extensions are specified as follows:

Cylindrical extension (Cobb and Shenoy, 2003b): Let m^{Θ_X} be a mass function on Θ_X. To extend this information to the space $\Theta_X \times \Theta_Y$, we use the cylindrical extension defined as:

$$m^{\Theta_X \uparrow \Theta_X \times \Theta_Y}(z) = \begin{cases} m^{\Theta_X}(x) & \text{if } z = x \times \Theta_Y \\ 0 & \text{otherwise} \end{cases} \qquad \forall z \subseteq \Theta_X \times \Theta_Y \tag{A.14}$$

Marginalization (Cobb and Shenoy, 2003b): Let $m^{\Theta_X \times \Theta_Y}$ be a mass function on $\Theta_X \times \Theta_Y$. The marginal mass function $m^{\Theta_X \times \Theta_Y \downarrow \Theta_Y}$ on Θ_Y is defined as:

$$m^{\Theta_X \times \Theta_Y \downarrow \Theta_Y}(y) = \sum_{z \subseteq \Theta_X \times \Theta_Y | \mathrm{Proj}(z \downarrow \Theta_Y) = y} m^{\Theta_X \times \Theta_Y}(z) \qquad \forall y \subseteq \Theta_Y \tag{A.15}$$

with $\mathrm{Proj}(z \downarrow \Theta_Y) = \{y \in \Theta_Y | \exists x \in \Theta_X, (y, x) \in z\}$

Ballooning extension (Smets, 1978): Let $m^{\Theta_Y}[x]$ denote the conditional mass function on Θ_Y given $x \subseteq \Theta_X$. The ballooning extension of $m^{\Theta_Y}[x]$ on the space $\Theta_X \times \Theta_Y$ is the least committed mass function whose conditioning on x yields $m^{\Theta_Y}[x]$. It is obtained as:

$$m^{\Theta_Y}[x]^{\Uparrow \Theta_X \times \Theta_Y}(z) = \mathbb{1}_Z \cdot m^{\Theta_Y}[x](y), \qquad \forall z \subseteq \Theta_X \times \Theta_Y, \tag{A.16}$$

with:

$$\mathbb{1}_Z = \begin{cases} 1 & \text{if } z = (x \times y) \cup (\bar{x} \times \Theta_Y), \\ 0 & \text{otherwise.} \end{cases}$$

Tabel A.1: Comparison of Bayesian networks and valuation networks (Yaghlane et al., 2003)

	Bayesian network	Valuation network
Graphical structure		
1. type of graph	directed acyclic graph	hypergraph
2. relations	conditional independence relations	joint form
3. nodes	random variables	variables & valuations
Inference procedure		
4. type of uncertainty	probabilistic	several
5. inference	quantitative based on probability propagation	quantitative based on fusion algorithm

A

A.2.2. VALUATION NETWORKS

Valuation networks are a graphical tool to represent uncertain knowledge in the form of belief functions (Shenoy, 1989, 1992a,b; Yaghlane et al., 2003). In contrast to Bayesian networks, which emphasize conditional independent relations, valuation networks emphasize factorizations of the joint distribution function. Formally, a valuation network can be regarded as a 3-tuple $(\mathbf{V}, \{\Theta_X\}_{X\in\mathbf{V}}, \{W_1, ..., W_m\})$ with operators $\{\oplus, \downarrow\}$ (Yaghlane et al., 2003), where:

1. $\mathbf{V}$ is the set of variables representing the universe of discourse
2. $\{\Theta_X\}$ is the set of frames associated with each variable $X \in \mathbf{V}$
3. $\{W_1, ..., W_m\}$ is a collection of valuations[2] defined on the subsets of variables
4. $\oplus$ is the combination operation. Intuitively, combination corresponds to the aggregation of knowledge
5. $\downarrow$ is the marginalization operation. Intuitively, marginalization corresponds to the coarsening of knowledge.

When the uncertainty is represented by belief functions, the valuations are multivariate basic belief assignments, and the combination operator corresponds to the conjunctive rule of combination (see Section A.2.3).

There are two types of vertices in a valuation network. One set of vertices represents variables, indicated by circles, and the other set represents valuations, indicated by diamonds. In a valuation network, there are edges only between variables and valuations. There is an edge between a variables and a valuation if and only if the variable is in the domain of the valuation.

A comparison of the Bayesian networks and the valuation networks is given in Table A.1.

[2]A valuation is a function representing the relationship among the variables in its domain.

A.2.3. REASONING UNDER UNCERTAINTY

When new evidences become available, this is incorporated by combining the existing mass function with the mass function describing the new evidence. Consider two distinct mass functions $m_1^{\Theta_\mathbf{X}}$ and $m_2^{\Theta_\mathbf{X}}$ on $\Theta_\mathbf{X}$. The belief function $m^{\Theta_\mathbf{X}}$ that quantifies the combined impact of the two mass functions according to Dempster's original rule of combination is defined ad follows: for any $\mathbf{x} \subseteq \Theta_\mathbf{X}$

$$m^{\Theta_\mathbf{X}}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} = \emptyset, \\ k_\mathrm{n} \sum\limits_{\mathbf{x}' \cap \mathbf{x}'' = \mathbf{x}} m_1^{\Theta_\mathbf{X}}(\mathbf{x}') m_2^{\Theta_\mathbf{X}}(\mathbf{x}'') & \text{otherwise,} \end{cases} \tag{A.17}$$

with $m_1^{\Theta_\mathbf{X}}$ and $m_2^{\Theta_\mathbf{X}}$ two mass functions on the same (multivariate) space $\Theta_\mathbf{X}$, $m^{\Theta_\mathbf{X}}$ the combined mass function, and k_n a normalization constant.

A

In the TBM, an open world assumption is allowed, i.e. the truth value of a variable X may be not included in Θ_X. In this case, two pieces of evidence are combined using the conjunctive rule of combination, which is an unnormalized form of Dempster's original rule of combination (Yager, 1987; Smets, 1990):

$$m^{\Theta_\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{x}' \cap \mathbf{x}'' = \mathbf{x}} m_1^{\Theta_\mathbf{X}}(\mathbf{x}') m_2^{\Theta_\mathbf{X}}(\mathbf{x}'') \tag{A.18}$$

The mass assigned to the empty set can be regarded as a measure of conflict between the different information sources.

Combination rules (A.17) and (A.18) assume that the two sources $m_1^{\Theta_\mathbf{X}}$ and $m_2^{\Theta_\mathbf{X}}$ are both reliable and independent. Alternative combination rules, e.g. the disjunctive rule of combination and the cautious rule of combination, have been proposed to handle dependent and unreliable sources of evidence (Destercke and Dubois, 2011). A detailed discussion about combination rules is beyond the scope of this thesis. For a more thorough discussion, we refer the interested reader to (Zadeh, 1984; Yager, 1987; Smets, 1990).

A.2.4. DECISION MAKING

In the TBM, decisions are made by transforming the mass distribution to a probability distribution and then applying the expected-utility theory (see Section A.1.4). Belief masses are transformed to probabilities using the pignistic transformation (Smets, 2005):

$$P_\mathrm{pig}(x_i) = \sum_{x \subseteq \Theta_X} \frac{|x_i \cap x|}{x} \frac{m^{\Theta_X}(x)}{1 - m^{\Theta_X}(\emptyset)} \tag{A.19}$$

with $|x|$ the cardinality of the set x. So, the mass allocated to a non-singleton set x is proportionally divided among the singleton elements in x, and the mass allocated to the empty set is proportionally distributed among all focal sets[3]. Although other transformation rules have been proposed, of which the plausibility transformation (Cobb and Shenoy, 2006) is the most well-known, we adopt Smets' view (Smets, 2002, 2005) and regard the pignistic transformation as the most sensible transformation rule in the context

[3]The focal sets of a bba m are all subsets $A \subseteq \Theta$ for which $m(A) > 0$ (Smets, 1998).

Tabel A.2: Probabilities according to the pignistic and plausibility transformations of the mass distribution in Example A.2.1

hypothesis	P_{pig}	P_{pl}
h_1	0.30	0.13
h_2	0.23	0.29
h_3	0.23	0.29
h_4	0.23	0.29

of decision making with incomplete information. In our view, a main drawback of the plausibility transformation is that it may result in probability distributions that contradict the information available in belief form, which we illustrate with an example.

A

Example A.2.1. Consider the basic belief assignment m^{Θ_H} for a variable H with domain $\Theta_H = \{h_1, h_2, h_3, h_4\}$:

$$\begin{aligned} m^{\Theta_H}(\{h_1\}) &= 0.3 \\ m^{\Theta_H}(\{h_2, h_3, h_4\}) &= 0.7 \end{aligned} \tag{A.20}$$

The associated probabilities resulting from the pignistic transformation P_{pig} and plausibility transformation P_{pl} are given in Table A.2. According to the pignistic transformation, hypothesis h_1 is most likely, whereas according to the plausibility transformation, the hypotheses h_2, h_3, and h_4 are most likely.

The probability distribution P_{pl} for H obtained from the plausibility transformation indicates that every $h_i \neq h_1$ is more likely than hypothesis h_1. However, this is not in agreement with the information contained in the basic belief assignment m^{Θ_H} (A.20). Given the information contained in m^{Θ_H} the following probabilities cannot be all satisfied:

$$\begin{aligned} P(h_1) &< P(h_2) \\ P(h_1) &< P(h_3) \\ P(h_1) &< P(h_4) \end{aligned}$$

B

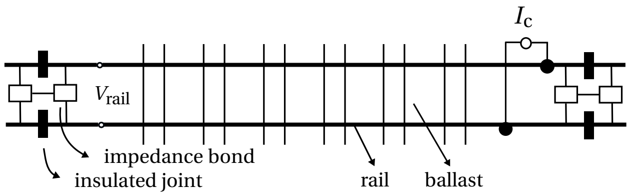
RAILWAY TRACK CIRCUITS

In this appendix, track circuits are described and modeled and possible system faults are discussed. In particular, double-rail, 75 Hz AC track circuits, as used in the Netherlands, are considered.

B.1. WORKING PRINCIPLE

Throughout the world, track circuits are the most commonly used devices for train detection (Chen et al., 2008). For the purpose of train detection, the railway track is then divided into electrically separated sections, each having its own track circuit, see Figure B.1. In this figure, V_{rail} represents the voltage applied between the two rails at the side of the transmitter and I_{c} represents the signaling current measured at the receiver. The insulated joints prevent current flow via the rails to the neighboring sections. The impedance bonds allow direct traction currents to flow to adjacent sections, while blocking the alternating currents used for train detection.

Track circuits operate by transmitting an electric current to a receiver via the two rails. When a section is free, the transmitted signal reaches the far end of the section. When the section is occupied by a train, the circuit is short-circuited by the wheel sets and the current does not reach the receiver (see Figure 3.3).



Figuur B.1: Overview of a railway section and the corresponding track circuit.

B.2. SYSTEM MODELING

To get insight into the system behavior and possible fault causes, a track circuit model will be derived hereafter. To model the relation between the input voltage V_{rail} and the output current I_c, a good understanding of the electrical properties of the rails, ballast, and train shunts is required.

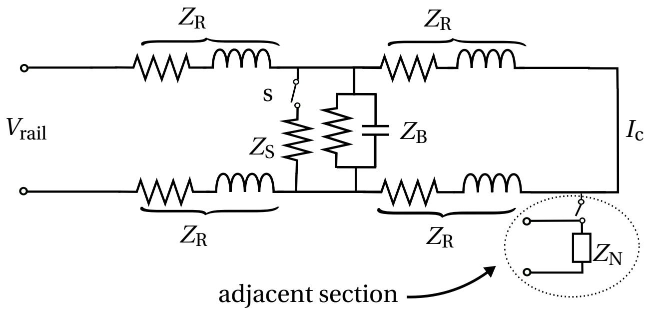
B.2.1. RAIL AND BALLAST IMPEDANCE

The rail bars are made of iron, having a low resistance for DC currents and an increasing resistance for AC currents as frequency increases. Here, we are interested in the resistance (Ω/km) that a 75 Hz current encounters when flowing in the longitudinal direction of the rail bars. The ballast impedance is a measure of how easily current can flow between the two rails of a track circuit and it consists of the leakage between the rail fixings, sleepers, and earth (Spoors, 1998).

To model the rail impedance Z_R and ballast impedance Z_B, the two-line transmission line model (Chen et al., 2008) is often used. This model assumes that the rail and ballast impedance are evenly distributed over the length of the track. For practical purposes, lumped parameter models, consisting of a finite number of (identical) cascaded subsections, are often considered to approximate the transmission line behavior. The number of subsections determines the accuracy of the model considered. In Figure B.2, a model with one subdivision is shown. A connection to an adjacent section is also included in the figure to describe insulated joint defects (see Section B.3.2).

B.2.2. TRAIN SHUNT

When a train is present in a section, the wheels and axles create low-impedance connections between the two rails. Such a connection can be modeled by the shunt impedance Z_S between the two rails, parallel to the ballast impedance Z_B. Resistor Z_S is only connected when there is a train in the section (i.e. switch s is closed).



Figuur B.2: Model of a track circuit.

Tabel B.1: Overview of considered track circuit faults.

health (H)	problem	cause	potential error
h	-	healthy state	-
f_1	train shunt imperfection	rail contamination	FN
f_2		lightweight trains	FN
f_3	insulation imperfection	insulated joint defect	FP
f_4		conductive objects	FP
f_5	rail conductance impairment	mechanical defect	FP
f_6		electrical disturbances	FP
f_7	ballast condition	ballast degradation	FP
f_8		ballast variation	FP

B.3. FAULT CAUSES

Due to several causes, a track circuit can behave in an undesired way. For instance, due to an increased resistance of the rails (e.g. as a consequence of a broken rail), the current level at the receiver may be too low. In the worst case, this hinders the execution of the system task (train detection), resulting in a functional failure. To prevent functional failures, it is important to recognize system faults as early as possible. Therefore, in the sequel, different types of system faults, the related causes, and their effect on the system behavior are investigated. Table B.1(a) gives an overview of the faults considered.

B.3.1. TRAIN SHUNT IMPERFECTION

The proper functioning of a track circuit requires that every train short-circuits the section, meaning that the path “rail-wheels-axles-wheels-rail” should have a sufficiently low resistance for 75 Hz AC currents. A good train shunt can be hampered by different causes; the two most important ones are: 1. contamination between the rail surface and the wheels, and 2. lightweight trains. Contamination between the rails and the wheels (e.g., rust films, sand, and leaf residue) acts as a semi-conductor, in the sense that it exhibits a high resistance until the voltage exceeds a threshold (Spoors, 1998). When the contamination level is too high, the voltage between the rails and the train is too low to realize a good train shunt. In addition, lightweight trains may suffer from shunting problems because they can be too light to make good contact and to clean the rails. In the case of a bad train shunt, the resistance of Z_S is relatively high, meaning that the path via the train is electrically less attractive and more current flows to the receiver.

B.3.2. INSULATION IMPERFECTION

Insulated joints are used to prevent that 75 Hz AC currents leak to neighboring sections. Problems occur when insulated joints degrade or when conductive objects lie over the joints. Insulated joints are implemented in a way that they are fail-safe. This is achieved by using phase-shifted currents in adjacent sections, so that a current signal of one section cannot energize the relay of an adjacent section. Insulated joint defects can be modeled by a connection to another circuit (see Figure B.2). The impedance of this circuit determines the amount of current flowing to the adjacent section. In the case of an insulation problem, the circuit leaks current and consequently, the current I_c is too low.

B.3.3. RAIL CONDUCTANCE IMPAIRMENT

The proper functioning of a track circuit relies on the conductance properties of the rails. The rail conductance is influenced by the quality of the rails themselves (e.g., damaged rail, broken rail), the quality of the bonds in jointed track, and electrical influences of disturbance currents (e.g. saturated track due to high traction currents). In the track circuit model, the quality of the rails is modeled by the value of the impedance Z_R. Problems occur when this impedance is too high; in that case, the path via the ballast Z_B becomes more attractive and the current level at the receiver decreases.

B.3.4. BALLAST CONDITION

The condition of the ballast determines the resistance that currents encounter when flowing from one rail to the other rail or to the ground. Because the effect of a decreasing ballast resistance is similar to that of a train shunt, it is important that the ballast resistance is sufficiently high and constant. Due to environmental disturbances (mainly weather) and aging, the ballast resistance will fluctuate over time. Some degree of fluctuation is acceptable, but when the ballast resistance becomes too low, the section will be reported as occupied, even if there is actually no train present.

B.3.5. CIRCUIT-RELATED FAULTS

Although track circuits have a high reliability, their components (e.g., relays, cables, and power supply) can break. In this thesis, circuit-related faults are not treated further and it is assumed that the circuit itself functions properly.

C

ENERGY AND MASS BALANCES

This appendix describes energy and mass balances applicable to heating, ventilation, and air conditioning (HVAC) systems. This background information is in support of the concepts proposed in Chapter 3.

C.1. MASS BALANCES

For each hot water circuit in the HVAC, the following mass balance applies:

$$w_{\text{sw}}^{\text{b}}(t) = w_{\text{sw}}^{\text{a}_1}(t) + \ldots + w_{\text{sw}}^{\text{a}_{n_\text{a}}}(t) + w_{\text{sw}}^{\text{r}_1}(t) + \ldots + w_{\text{sw}}^{\text{r}_{n_\text{r}}}(t) \tag{C.1}$$

with $w_{\text{sw}}^{\text{b}}(t)$ the mass flow through the boiler at time t, and $w_{\text{sw}}^{\text{a}_1}(t) + \ldots + w_{\text{sw}}^{\text{a}_{n_\text{a}}}(t)$ and $w_{\text{sw}}^{\text{r}_1}(t) + \ldots + w_{\text{sw}}^{\text{r}_{n_\text{r}}}(t)$ the mass flows through the connected AHUs and radiators respectively at time t.

C.2. ENERGY BALANCES

Energy balances can be defined for each component in the HVAC system where energy is exchanged, e.g. the boiler, the radiator, and the AHU.

In the boiler, chemical or electrical energy is transformed into thermal energy. The heat generated is used to warm up the water in the hot water circuit. So, the following energy balance holds:

$$E_{\text{chem}}^{\text{b}}(t-\Delta) - E_{\text{chem}}^{\text{b}}(t) = \int_{t-\Delta}^{t} \left(E_{\text{sw,thermal}}^{\text{b}}(\tau) - E_{\text{rw,thermal}}^{\text{b}}(\tau) + E_{\text{loss}}^{\text{b}}(\tau) \right) d\tau \tag{C.2}$$

with $E_{\text{chem}}^{\text{b}}$ the energy in the available fuel, $E_{\text{rw,thermal}}^{\text{b}}$ the thermal energy of the water returning from the hot water circuit, $E_{\text{sw,thermal}}^{\text{b}}$ the energy in the water after it is heated by the boiler, $E_{\text{loss}}^{\text{b}}$ all energy originating from the fuel that is not converted to thermal energy of the water, and Δ a time shift.

In the radiator, part of the thermal energy of the hot water is transferred to the neighboring air, which has a relatively low temperature. The degree of energy exchange depends on the difference between the temperature of the hot water flowing through the radiator and the temperature of the zone air. The following energy balance applies:

$$E^{\mathrm{r}}_{\mathrm{sw,thermal}}(t) - E^{\mathrm{r}}_{\mathrm{rw,thermal}}(t) = Q^{\mathrm{r}}(t) + E^{\mathrm{r}}_{\mathrm{loss}}(t) \tag{C.3}$$

with $E^{\mathrm{r}}_{\mathrm{sw,thermal}}$ and $E^{\mathrm{r}}_{\mathrm{rw,thermal}}$ the thermal energy of the radiator supply and return water respectively, Q^{r} the heat transferred to the zone, and $E^{\mathrm{r}}_{\mathrm{loss}}$ the energy extracted from the water that is not transferred to the zone.

The energy exchange in the AHU is similar to that in the radiator, i.e. thermal energy of the water flowing through the coils is used to increase the thermal energy of the passing air:

$$E^{\mathrm{a}}_{\mathrm{sw,thermal}}(t) - E^{\mathrm{a}}_{\mathrm{rw,thermal}}(t) = E^{\mathrm{a}}_{\mathrm{sa,thermal}}(t) - E^{\mathrm{a}}_{\mathrm{ma,thermal}}(t) + E^{\mathrm{a}}_{\mathrm{loss}}(t) \tag{C.4}$$

with $E^{\mathrm{a}}_{\mathrm{sw,thermal}}$ and $E^{\mathrm{a}}_{\mathrm{rw,thermal}}$ the thermal energy of the AHU return and supply water respectively, $E^{\mathrm{a}}_{\mathrm{sa,thermal}}$ and $E^{\mathrm{a}}_{\mathrm{ma,thermal}}$ the thermal energy of the supply air and the mixed-air respectively, and $E^{\mathrm{a}}_{\mathrm{loss}}$ energy losses. In addition to the energy balances for the HVAC system components, energy balances apply to the zone(s):

C

$$m_{\mathrm{z}} c_{\mathrm{z}} \dot{T}^{\mathrm{z}}_{\mathrm{a}}(t) = -Q^{\mathrm{z}}(t) + Q^{\mathrm{r}}(t) + Q^{\mathrm{a}}(t) + Q^{\eta}(t) + \sigma(t) \tag{C.5}$$

with $T^{\mathrm{z}}_{\mathrm{a}}$ the zone air temperature, $m_{\mathrm{z}} c_{\mathrm{z}}$ the thermal capacity of the zone, Q^{z} heat losses to the outside/other zones, Q^{r} the heat produced by the radiators, Q^{a} the heat produced by the AHUs, Q^{η} the heat produced by people inside the room, and σ modeling and process noise.

BIBLIOGRAPHY

Agrawal, R., Ieong, S., and Velu, R. Timing when to buy. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 709–718, New York, NY, USA, 2011.

Ahmad, R. and Kamaruddin, S. An overview of time-based and condition-based maintenance in industrial application. *Computers & Industrial Engineering*, 63(1):135–149, 2012.

Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

Amari, S. V., McLaughlin, L., and Pham, H. Cost-effective condition-based maintenance using Markov decision processes. In *Proceedings of the Symposium on Reliability and Maintainability*, pages 464–469, Newport Beach, CA, USA, 2006.

Bertsekas, D. P. *Dynamic Programming and Optimal Control*, volume 1 of *2*. Athena Scientific Belmont, MA, 1995.

Billinton, R. and Huang, D. Aleatory and epistemic uncertainty considerations in power system reliability evaluation. In *Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems*, pages 1–8, Rincon, Bonaire, 2008.

Bishop, C. M. *Neural networks for pattern recognition*. Oxford university press, 1995.

Boudali, H. and Dugan, J. B. A discrete-time Bayesian network reliability modeling and analysis framework. *Reliability Engineering & System Safety*, 87(3):337–349, 2005.

Bouvard, K., Artus, S., Bérenguer, C., and Cocquempot, V. Condition-based dynamic maintenance operations planning & grouping. Application to commercial heavy vehicles. *Reliability Engineering & System Safety*, 96(6):601–610, 2011.

Camci, F. System maintenance scheduling with prognostics information using genetic algorithm. *IEEE Transactions on Reliability*, 58(3):539–552, 2009.

Castanier, B., Grall, A., and Bérenguer, C. A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliability Engineering & System Safety*, 87(1):109–120, 2005.

Celaya, J., Saxena, A., and Goebel, K. Uncertainty representation and interpretation in model-based prognostics algorithms based on Kalman filter estimation. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, pages 1–10, Minneapolis, Minnesota, USA, 2012.

Chan, H. and Darwiche, A. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163(1):67–90, 2005.

Charniak, E. Bayesian networks without tears. *AI Magazine*, 12(4):50–63, 1991.

Chatzi, E. N. and Smyth, A. W. The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing. *Structural Control and Health Monitoring*, 16(1):99–123, 2009.

Cheeseman, P. In defense of probability. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pages 1002–1009, Los Angeles, California, USA, 1985.

Chen, J. and Patton, R. J. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer Publishing Company, 2012.

Chen, J., Roberts, C., and Weston, P. Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems. *Control Engineering Practice*, 16(5):585–596, 2008.

Chen, N., Ye, Z.-S., Xiang, Y., and Zhang, L. Condition-based maintenance using the inverse Gaussian degradation model. *European Journal of Operational Research*, 243 (1):190–199, 2015.

Cherfi, Z., Oukhellou, L., Côme, E., Denoeux, T., and Aknin, P. Partially supervised independent factor analysis using soft labels elicited from multiple experts: application to railway track circuit diagnosis. *Soft Computing*, 16(5):741–754, 2012.

Cobb, B. R. and Shenoy, P. P. A comparison of methods for transforming belief function models to probability models. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 255–266. Springer, 2003a.

Cobb, B. R. and Shenoy, P. P. A comparison of Bayesian and belief function reasoning. *Information Systems Frontiers*, 5(4):345–358, 2003b.

Cobb, B. R. and Shenoy, P. P. On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41(3):314–330, 2006.

Cooper, G. F. and Herskovits, E. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.

Crowder, M. and Lawless, J. On a scheme for predictive maintenance. *European Journal of Operational Research*, 176(3):1713–1722, 2007.

Daly, R., Shen, Q., and Aitken, S. Learning Bayesian networks: approaches and issues. *The Knowledge Engineering Review*, 26(02):99–157, 2011.

Darwiche, A. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.

De Bruin, T., Verbert, K., and Babuška, R. Railway track circuit fault diagnosis using recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2016.

Dekker, R., Wildeman, R. E., and van der Duyn Schouten, F. A. A review of multi-component maintenance models with economic dependence. *Mathematical Methods of Operations Research*, 45(3):411–435, 1997.

Del Moral, P. Non-linear filtering: interacting particle resolution. *Markov Processes and Related Fields*, 2(4):555–581, 1996.

Dempster, A. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, pages 325–339, 1967.

Destercke, S. and Dubois, D. Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. *Information Sciences*, 181(18):3925–3945, 2011.

Dexter, A. L. and Ngo, D. Fault diagnosis in air-conditioning systems: a multi-step fuzzy model-based approach. *HVAC&R Research*, 7(1):83–102, 2001.

Dubois, D. Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51(1):47–69, 2006.

Dubois, D. and Prade, H. *Possibility Theory*. Plenum Press, New-York, 1988.

Dubois, D. and Prade, H. Possibility theory, probability theory and multiple-valued logics: A clarification. *Annuals of Mathematics and Artificial Intelligence*, 32(1-4):35–66, 2001.

Dubois, D., Prade, H., and Smets, P. Representing partial ignorance. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 26(3):361–377, 1996.

Dubois, D., Prade, H., and Sandri, S. On possibility/probability transformations. In Lowen, R. and Roubens, M., editors, *Fuzzy Logic: State of the Art*, pages 103–112. Springer Netherlands, Dordrecht, 1993.

Dubois, D., Moral, S., and Prade, H. Belief change rules in ordinal and numerical uncertainty theories. In Dubois, D. and Prade, H., editors, *Belief Change*, volume 3 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 311–392. Springer Netherlands, 1998.

Elwany, A. H., Gebraeel, N. Z., and Maillart, L. M. Structured replacement policies for components with complex degradation processes and dedicated sensors. *Operations research*, 59(3):684–695, 2011.

Engel, S., Gilmartin, B., Bongort, K., and Hess, A. Prognostics, the real issues involved with predicting life remaining. In *Proceedings of the IEEE Aerospace Conference*, pages 457–469, Big Sky, MT, USA, 2000.

Etzioni, O., Tuchinda, R., Knoblock, C. A., and Yates, A. To buy or not to buy: mining airfare data to minimize ticket purchase price. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 119–128, New York, NY, USA, 2003.

Feinberg, E. A. and Shwartz, A. *Handbook of Markov Decision Processes: Methods and Applications*, volume 40. Springer Science & Business Media, 2012.

Fekih, A., Xu, H., and Chowdhury, F. N. Neural networks based system identification techniques for model based fault detection of nonlinear systems. *International Journal of Innovative Computing, Information and Control*, 3(5):1073–1085, 2007.

Ferson, S. and Ginzburg, L. Different methods are needed to propagate ignorance and variability. *Reliability Engineering & System Safety*, 54(2):133–144, 1996.

Fisher, I. *The Theory of Interest*. Kelley, 1965.

Fletcher, R. *Practical Methods of Optimization*. John Wiley & Sons, 2013.

Frank, P., Garcıa, E. A., and Köppen-Seliger, B. Modelling for fault detection and isolation versus modelling for control. *Mathematics and Computers in Simulation*, 53(4):259–271, 2000.

Furuya, A. and Madanat, S. Accounting for network effects in railway asset management. *Journal of Transportation Engineering*, 139(1):92–100, 2012.

Gebraeel, N. Z., Lawley, M. A., Li, R., and Ryan, J. K. Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, 37(6):543–557, 2005.

Goldstein, M. Subjective Bayesian analysis: principles and practice. *Bayesian Analysis*, 1 (3):403–420, 2006.

Grall, A., Bérenguer, C., and Dieulle, L. A condition-based maintenance policy for stochastically deteriorating systems. *Reliability Engineering & System Safety*, 76(2):167–180, 2002.

Graves, A. and Jaitly, N. Towards end-to-end speech recognition with recurrent neural networks. In *Proceedings of the 31st International Conference on Machine Learning*, pages 1764–1772, Beijing, China, 2014.

Groves, W. and Gini, M. On optimizing airline ticket purchase timing. *ACM Transanctions on Intelligent Systems Technology*, 7(1):3:1–3:28, 2015.

György, K., Kelemen, A., and Dávid, L. Unscented Kalman filters and particle filter methods for nonlinear state estimation. *Procedia Technology*, 12:65–74, 2014.

Haenni, R. Ignoring ignorance is ignorant. Technical report, Center for Junior Research Fellows, University of Konstanz, Konstanz, Germany, 2003.

Haenni, R. and Lehmann, N. Implementing belief function computations. *International Journal of Intelligent Systems*, 18(1):31–49, 2003.

Heckerman, D. A tutorial on learning with Bayesian networks. In Jordan, M., editor, *Learning in Graphical Models*, volume 89, pages 301–354. Springer Netherlands, 1998.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

Holub, O. and Macek, K. HVAC simulation model for advanced diagnostics. In *Proceedings of the IEEE 8th International Symposium on Intelligent Signal Processing*, pages 93–96, Funchal, Madeira, Portugal, 2013.

Howie, D. *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. Cambridge University Press, 2002.

Huynh, K. T., Barros, A., and Bérenguer, C. Multi-level decision-making for the predictive maintenance of k-out-of-n :f deteriorating systems. *IEEE Transactions on Reliability*, 64(1):94–117, 2015.

Hwang, I., Kim, S., Kim, Y., and Seah, C. E. A survey of fault detection, isolation, and reconfiguration methods. *IEEE Transactions on Control Systems Technology*, 18(3):636–653, 2010.

Isermann, R. Model-based fault-detection and diagnosis: status and applications. *Annual Reviews in Control*, 29(1):71–85, 2005.

Isermann, R. *Fault-Diagnosis Applications: Model-Based Condition Monitoring: Actuators, Drives, Machinery, Plants, Sensors, and Fault-Tolerant Systems*. Springer Science & Business Media, 2011.

Jardine, A., Lin, D., and Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

Jaynes, E. T. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957a.

Jaynes, E. T. Information theory and statistical mechanics. ii. *Physical review*, 108(2):171, 1957b.

Kaipio, J. and Somersalo, E. *Statistical and Computational Inverse Problems*. Springer Science & Business Media, 2006.

Kalman, R. E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

Katipamula, S. and Brambley, M. R. Review article: methods for fault detection, diagnostics, and prognostics for building systems—a review, part i. *HVAC&R Research*, 11(1): 3–25, 2005.

Keizer, M. C. O., Teunter, R. H., and Veldman, J. Clustering condition-based maintenance for systems with redundancy and economic dependencies. *European Journal of Operational Research*, 251(2):531–540, 2016.

Kiureghian, A. D. and Ditlevsen, O. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2):105–112, 2009.

Klir, G. and Yuan, B. *Fuzzy Sets and Fuzzy Logic*, volume 4. Prentice Hall New Jersey, 1995.

Knowles, M., Baglee, D., and Wermter, S. Reinforcement learning for scheduling of maintenance. In *Research and Development in Intelligent Systems XXVII*, pages 409–422. Springer, 2011.

Kohlas, J. and Monney, P. *A Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence*. Springer-Verlag, 1995.

Kukal, J., Macek, K., Rojicek, J., and Trojanová, J. From symptoms to faults: Temporal reasoning methods. In *Proceedings of the International Conference on Adaptive and Intelligent Systems*, pages 155–159, Klagenfurt, Austria, 2009.

Lam, J. Y. J. and Banjevic, D. A myopic policy for optimal inspection scheduling for condition based maintenance. *Reliability Engineering & System Safety*, 144:1 – 11, 2015.

Laplace, P. *A Philosophical Essay on Probabilities*. New York: Dover Publications Inc., 1814.

Le, T. T., Chatelain, F., and Berenguer, C. Hidden Markov models for diagnostics and prognostics of systems under multiple deterioration modes. In *Proceedings of the European Safety and Reliability Conference*, pages 1197–1204, 2014.

Le, T. T., Chatelain, F., and Bérenguer, C. Multi-branch hidden markov models for remaining useful life estimation of systems under multiple deterioration modes. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2016.

Le, T., Berenguer, C., and Chatelain, F. Prognosis based on multi-branch hidden semi-markov models: A case study. In *Proceedings of the 9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, volume 48 of *21*, pages 91–96, Paris, France, 2015.

Lee, W.-Y., House, J. M., and Kyong, N.-H. Subsystem level fault diagnosis of a building's air-handling unit using general regression neural networks. *Applied Energy*, 77(2):153–170, 2004.

Li, H. and Braun, J. E. Decoupling features and virtual sensors for diagnosis of faults in vapor compression air conditioners. *International Journal of Refrigeration*, 30(3): 546–564, 2007.

Li, Z., Molodova, M., and Dollevoet, R. Detectability of isolated short wave rail surface defects by way of axle box acceleration. In *Proceedings of the 21st International Symposium on Dynamics of the Vehicles on Roads and Tracks*, Stockholm, Sweden, 2009.

Liang, J. and Du, R. Model-based fault detection and diagnosis of HVAC systems using support vector machine method. *International Journal of Refrigeration*, 30(6):1104–1114, 2007.

Lin-Hai, Z., Jian-Ping, W., and Yi-Kui, R. Fault diagnosis for track circuit using AOK-TFRs and AGA. *Control Engineering Practice*, 20(12):1270–1280, 2012.

Lindley, D. V. *Making Decisions*. Wiley New York, 1985.

Lindley, D. The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Statistical Sciences*, 2(1):17–24, 1987.

Längkvist, M., Karlsson, L., and Loutfi, A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42:11–24, 2014.

Lo, C., Chan, P., Wong, Y., Rad, A., and Cheung, K. Fuzzy-genetic algorithm for automatic fault detection in HVAC systems. *Applied Soft Computing*, 7(2):554–560, 2007.

Loewenstein, G., Read, D., and Baumeister, R. F. *Time and Decision: Economic and Psychological Perspectives of Intertemporal Choice*. Russell Sage Foundation, 2003.

Lu, S., Lu, H., and Kolarik, W. J. Multivariate performance reliability prediction in real-time. *Reliability Engineering & System Safety*, 72(1):39–45, 2001.

Mann, L., Saxena, A., and Knapp, G. Statistical-based or condition-based preventive maintenance. *Journal of Quality in Maintenance Engineering*, 1(1):46–59, 1995.

Medury, A. and Madanat, S. System-level optimization of maintenance and replacement decisions for road networks. In *11th International Conference on Applications of Statistics and Probability in Civil Engineering*, pages 235–242, Zurich, Switzerland, 2011.

Meyer-Baese, A. and Schmid, V. J. *Pattern Recognition and Signal Analysis in Medical Imaging*. Elsevier, 2014.

Molodova, M., Li, Z., and Dollevoet, R. Axle box acceleration: Measurement and simulation for detection of short track defects. *Wear*, 271(1–2):349–356, 2011.

Mrad, A. B., Delcroix, V., Piechowiak, S., and Leicester, P. From information to evidence in a Bayesian network. In *Probabilistic Graphical Models*, pages 33–48. Springer, 2014.

Mrad, A. B., Delcroix, V., Piechowiak, S., Leicester, P., and Abid, M. An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence*, 43(4):802–824, 2015.

Mulumba, T., Afshari, A., Yan, K., Shen, W., and Norford, L. K. Robust model-based fault diagnosis for air handling units. *Energy and Buildings*, 86:698–707, 2015.

Namburu, S. M., Azam, M. S., Luo, J., Choi, K., and Pattipati, K. R. Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers. *IEEE Transactions on Automation Science and Engineering,*, 4(3):469–473, 2007.

Nan, C., Khan, F., and Iqbal, M. T. Real-time fault diagnosis using knowledge-based expert system. *Process Safety and Environmental Protection*, 86(1):55–71, 2008.

Narasimhan, S., Roychoudhury, I., Balaban, E., and Saxena, A. Combining model-based and feature-driven diagnosis approaches-a case study on electromechanical actuators. In *Proceedings of the 21st International Workshop on the Principles of Diagnosis*, Portland, OR, 2010.

Nguyen, K.-A., Do, P., and Grall, A. Multi-level predictive maintenance for multi-component systems. *Reliability Engineering & System Safety*, 144:83–94, 2015.

O'Hagan, A. Dicing with the unknown. *Significance*, 1(3):132–133, 2004.

Ompusunggu, A., Papy, J.-M., and Vandenplas, S. Kalman filtering based prognostics for automatic transmission clutches. *IEEE/ASME Transactions on Mechatronics*, 21(1): 419–430, 2016.

Oosterom, M. and Babuška, R. Virtual sensor for fault detection and isolation in flight control systems-fuzzy modeling approach. In *Proceedings of the 39th IEEE Conference on Decision and Control*, volume 3, pages 2645–2650, Sydney, NSW, 2000.

Oukhellou, L., Debiolles, A., Denoeux, T., and Aknin, P. Fault diagnosis in railway track circuits using Dempster-Shafer classifier fusion. *Engineering Applications of Artificial Intelligence*, 23(1):117–128, 2010.

Pandey, M., Yuan, X.-X., and Van Noortwijk, J. The influence of temporal uncertainty of deterioration on life-cycle management of structures. *Structure and Infrastructure Engineering*, 5(2):145–156, 2009.

Papoulis, A. *Signal Analysis*. McGraw-Hill New York, 1978.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988.

Pearl, J. and Russel, S. Bayesian networks. In Arbib, M., editor, *Handbook of Brain Theory and Neural Networks*. MIT press, 2001.

Pearl, J. *Causality; Models, Reasoning, and Inference*. Cambridge University Press, 2000.

Peng, Y. and Dong, M. A prognosis method using age-dependent hidden semi-markov model for equipment health prediction. *Mechanical Systems and Signal Processing*, 25 (1):237–252, 2011.

Peng, Y., Dong, M., and Zuo, M. J. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, 2010.

Pérez-Lombard, L., Ortiz, J., and Pout, C. A review on buildings energy consumption information. *Energy and Buildings*, 40(3):394–398, 2008.

Piette, M. A., Kinney, S. K., and Haves, P. Analysis of an information monitoring and diagnostic system to improve building operations. *Energy and Buildings*, 33(8):783–791, 2001.

Powell, W. B. *Introduction to Markov Decision Processes*, pages 57–109. John Wiley & Sons, Inc., 2011.

Reuben, L. C. K. and Mba, D. Diagnostics and prognostics using switching Kalman filters. *Structural Health Monitoring*, 13(3):296–306, 2014.

Rigatos, G. G. Extended Kalman and particle filtering for sensor fusion in motion control of mobile robots. *Mathematics and Computers in Simulation*, 81(3):590–607, 2010.

Robelin, C.-A. and Madanat, S. M. Reliability-based system-level optimization of bridge maintenance and replacement decisions. *Transportation Science*, 42(4):508–513, 2008.

Sandidzadeh, M. and Dehghani, M. Intelligent condition monitoring of railway signaling in train detection subsystems. *Journal of Intelligent and Fuzzy Systems*, 24(4):859–869, 2013.

Sathaye, N. and Madanat, S. A bottom-up solution for the multi-facility optimal pavement resurfacing problem. *Transportation Research Part B: Methodological*, 45(7): 1004–1017, 2011.

Savage, L. *The Foundations of Statistics*. New York: John Wiley & Sons, 1954.

Schein, J. and Bushby, S. T. A hierarchical rule-based fault detection and diagnostic method for HVAC systems. *HVAC&R Research*, 12(1):111–125, 2006.

Schwabacher, M. and Goebel, K. A survey of artificial intelligence for prognostics. In *AAAI Fall Symposium*, pages 107–114, Arlington, Virginia, USA, 2007.

Shafer, G. *A Mathematical Theory of Evidence*, volume 1. Princeton University Press Princeton, 1976.

Shafer, G. Belief functions. In Shafer, G. and Pearl, J., editors, *Readings in Uncertain Reasoning*. Morgan Kaufman, 1990.

Shenoy, P. Valuation-based systems for Bayesian decision analysis. *Operations Research*, 40(3):463–484, 1992a.

Shenoy, P. P. A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, 3(5):383–411, 1989.

Shenoy, P. P. Valuation-based systems: A framework for managing uncertainty in expert systems. In Zadeh, L. A. and Kacprzyk, J., editors, *Fuzzy Logic for the Management of Uncertainty*, pages 83–104. John Wiley & Sons, Inc., New York, NY, USA, 1992b.

Si, X.-S., Wang, W., Hu, C.-H., and Zhou, D.-H. Remaining useful life estimation: a review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1):1–14, 2011.

Si, X.-S., Wang, W., Hu, C.-H., Chen, M.-Y., and Zhou, D.-H. A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation. *Mechanical Systems and Signal Processing*, 35(1):219–237, 2013.

Si, X.-S., Wang, W., Hu, C.-H., and Zhou, D.-H. Estimating remaining useful life with three-source variability in degradation modeling. *IEEE Transactions on Reliability*, 63 (1):167–190, 2014.

Sikorska, J., Hodkiewicz, M., and Ma, L. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 25(5): 1803–1836, 2011.

Smets, P. *Un Modéle Mathématico-Statistique Stimulant le Processus du Diagnostic Médical.* PhD thesis, Université de Bruxelles, 1978.

Smets, P. The transferable belief model and other interpretations of Dempster-Shafer's model. In *Proceedings of the 6th Annual Conference on Uncertainty in Artificial Intelligence*, pages 375–383, Amsterdam, the Netherlands, 1990.

Smets, P. Resolving misunderstandings about belief functions. *International Journal of Approximate Reasoning*, 6(3):321–344, 1992.

Smets, P. What is Dempster-Shafer's model? In Yager, R., Kacprzyk, J., and Fedrizzi, M., editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 5–34. John Wiley & Sons, Inc., 1994.

Smets, P. Decision making in a context where uncertainty is represented by belief functions. In Srivastava, R. and Mock, T., editors, *Belief Functions in Bussiness Decisions*. Physica-Verlag, 2002.

Smets, P. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005.

Smets, P. Application of the transferable belief model to diagnostic problems. *International Journal of Intelligent Systems*, 13(2-3):127–157, 1998.

Smets, P. and Kennes, R. The transferable belief model. *Artificial Intelligence*, 66(2):191–234, 1994.

Spoors, R. *General Information on Track Circuits.* Railroad PLC, London, UK, 1998.

Sun, J., Zuo, H., Wang, W., and Pecht, M. G. Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. *Mechanical Systems and Signal Processing*, 28:585–596, 2012.

Sun, S. and Zhao, H. Fault diagnosis in railway track circuits using support vector machines. In *Proceedings of the 12th International Conference on Machine Learning and Applications*, volume 2, pages 345–350, Miami, FL, 2013.

Tang, D., Yu, J., Chen, X., and Makis, V. An optimal condition-based maintenance policy for a degrading system subject to the competing risks of soft and hard failure. *Computers & Industrial Engineering*, 83:100 – 110, 2015.

Tian, Z., Jin, T., Wu, B., and Ding, F. Condition based maintenance optimization for wind power generation systems under continuous monitoring. *Renewable Energy*, 36(5): 1502–1509, 2011.

Tinga, T. Application of physical failure models to enable usage and load based maintenance. *Reliability Engineering & System Safety*, 95(10):1061–1075, 2010.

Tobon-Mejia, D., Medjaher, K., Zerhouni, N., and Tripot, G. A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *IEEE Transactions on Reliability*, 61(2):491–503, 2012.

Van Horenbeek, A. and Pintelon, L. A dynamic predictive maintenance policy for complex multi-component systems. *Reliability Engineering & System Safety*, 120:39–50, 2013.

Van Noortwijk, J. A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1):2–21, 2009. Maintenance Modeling and Application.

Van Otterlo, M. and Wiering, M. Reinforcement learning and Markov decision processes. In *Reinforcement Learning*, pages 3–42. Springer, 2012.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., and Kavuri, S. N. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3):293–311, 2003.

Verbeek, W. Condition monitoring for track circuits: A multiple-model approach. Master's thesis, Delft University of Technology, the Netherlands, 2015.

Verbert, K., De Schutter, B., and Babuška, R. Exploiting spatial and temporal dependencies to enhance fault diagnosis: Application to railway track circuits. In *Proceedings of the 2015 European Control Conference*, pages 3052–3057, Linz, Austria, 2015a.

Verbert, K., De Schutter, B., and Babuška, R. Reasoning under uncertainty for knowledge-based fault diagnosis: A comparative study. In *Proceedings of the 9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, pages 422–427, Paris, France, 2015b.

Verbert, K., De Schutter, B., and Babuška, R. Fault diagnosis using spatial and temporal information with application to railway track circuits. *Engineering Applications of Artificial Intelligence*, 56:200–211, 2016.

Vu, H. C., Do, P., Barros, A., and Bérenguer, C. Maintenance planning and dynamic grouping for multi-component systems with positive and negative economic dependencies. *IMA Journal of Management Mathematics*, 26(2):145–170, 2015.

Wang, L., Chu, J., and Mao, W. A condition-based order-replacement policy for a single-unit system. *Applied Mathematical Modelling*, 32(11):2274–2289, 2008.

Wang, L. Modeling and simulation of HVAC faulty operations and performance degradation due to maintenance issues. In *Proceedings of the Asia Conference of International Building Performance Simulation Association*, Shanghai, China, 2014.

Wang, S. and Xiao, F. AHU sensor fault diagnosis using principal component analysis method. *Energy and Buildings*, 36(2):147–160, 2004.

Wang, W. A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance. *International Journal of Production Research*, 38(6): 1425–1436, 2000.

Wang, W. and Wang, H. Preventive replacement for systems with condition monitoring and additional manual inspections. *European Journal of Operational Research*, 247(2): 459–471, 2015.

Wang, W., Carr, M., Xu, W., and Kobbacy, K. A model for residual life prediction based on brownian motion with an adaptive drift. *Microelectronics Reliability*, 51(2):285–293, 2011.

Wang, X. Wiener processes with random effects for degradation data. *Journal of Multivariate Analysis*, 101(2):340–351, 2010.

Whitmore, G. and Schenkelberg, F. Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Analysis*, 3(1):27–45, 1997.

Wiegerinck, W., Kappen, H., and Burgers, W. Bayesian networks for expert systems: Theory and practical applications. In Babuška, R. and Groen, F., editors, *Interactive Collaborative Information Systems*, volume 281 of *Studies in Computational Intelligence*, pages 547–578. Springer, 2010.

Wu, B., Tian, Z., and Chen, M. Condition-based maintenance optimization using neural network-based health condition prediction. *Quality and Reliability Engineering International*, 29(8):1151–1163, 2013.

Xiao, F., Zhao, Y., Wen, J., and Wang, S. Bayesian network based {FDD} strategy for variable air volume terminals. *Automation in Construction*, 41:106 – 118, 2014.

Xu, Z., Ji, Y., and Zhou, D. Real-time reliability prediction for a dynamic system based on the hidden degradation process identification. *IEEE Transactions on Reliability*, 57(2): 230–242, 2008.

Yager, R. On the Dempster-Shafer framework and new combination rules. *Information Science*, 41(2):93–137, 1987.

Yaghlane, B. B. and Mellouli, K. Inference in directed evidential networks based on the transferable belief model. *International Journal of Approximate Reasoning*, 48(2):399–418, 2008.

Yaghlane, B. B., Smets, P., and Mellouli, K. Directed evidential networks with conditional belief functions. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 291–305. Springer, 2003.

Yam, R., Tse, P., Li, L., and Tu, P. Intelligent predictive decision support system for condition-based maintenance. *The International Journal of Advanced Manufacturing Technology*, 17(5):383–391, 2001.

Yeo, H., Yoon, Y., and Madanat, S. Algorithms for bottom-up maintenance optimisation for heterogeneous infrastructure systems. *Structure and Infrastructure Engineering*, 9 (4):317–328, 2013.

Yin, S., Ding, S. X., Haghani, A., Hao, H., and Zhang, P. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control*, 22(9):1567–1581, 2012.

Zadeh, L. Review of a mathematical theory of evidence. *AI Magazine*, 5(3):81–83, 1984.

Zadeh, L. A. Fuzzy logic and approximate reasoning. *Synthese*, 30(3-4):407–428, 1975.

Zadeh, L. A. Generalized theory of uncertainty (GTU)-principal concepts and ideas. *Computational Statistics & Data Analysis*, 51(1):15–46, 2006.

Zadeh, L. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

Zadeh, L. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100, Supplement 1:9–34, 1999.

Zadeh, L. Is there a need for fuzzy logic? *Information Sciences*, 178(13):2751–2779, 2008.

Zhao, Y., Xiao, F., and Wang, S. An intelligent chiller fault detection and diagnosis methodology using bayesian belief network. *Energy and Buildings*, 57:278 – 288, 2013.

Zheng, J. F., Si, X. S., Hu, C. H., Zhang, Z. X., and Jiang, W. A nonlinear prognostic model for degrading systems with three-source variability. *IEEE Transactions on Reliability*, pages 1–15, 2016.

Zhou, Y., Sun, Y., Mathew, J., Wolff, R., and Ma, L. Latent degradation indicators estimation and prediction: A monte carlo approach. *Mechanical Systems and Signal Processing*, 25(1):222–236, 2011.

Zhu, Q. *Maintenance optimization for multi-component systems under condition monitoring*. PhD thesis, Eindhoven University of Technology, 2015.

Zogg, D., Shafai, E., and Geering, H. Fault diagnosis for heat pumps with parameter identification and clustering. *Control Engineering Practice*, 14(12):1435–1444, 2006.

GLOSSARY

LIST OF SYMBOLS

GENERAL

τ	time
μ	mean
σ	standard deviation
$N(\mu,\sigma)$	normal distribution with mean μ and standard deviation σ
$\mathbb{E}$	expectation
$\Theta_{\mathcal{V}}$	the set of values variable $\mathcal{V}$ can take
$P(\cdot)$	probability mass function
$p(\cdot)$	probability density function
$m(\cdot)$	Dempster-Shafer mass function
$u(\cdot)$	utility function
(G,D)	Bayesian network, where G is the structure and D the set of local probability functions;
$B(\cdot)$	standard Brownian motion

FAULT DIAGNOSIS

h	healthy mode
$f_{i,1},\ldots,f_{i,\ell_i}$	fault modes of system component i
$F_{i,j}$	binary fault variable indicating whether fault $f_{i,j}$ is present
$H_i(\tau)$	health state of component i at time τ (single fault scenario), $H_i(\tau) \in \Theta_H = \{h, f_{i,1},\ldots,f_{i,\ell_i}\}$
$\mathbf{H}_i(\tau)$	health state of component i at time τ (multiple fault scenario), $\mathbf{H}_i(\tau) = \{F_{i,1},\ldots,F_{i,\ell_i}\}$, $F_{i,j} \in \{0,1\}$
$X_{\mathrm{o},i}$	operating state of component i
M_i	monitoring signal vector of component i
$M_{\mathrm{env},i}$	effect of environmental disturbances on M_i
M'_i	M_i corrected for environmental disturbances
$C_1,\ldots,C_z$	diagnostic features
W_{C_k}	set of possible faults given C_k
$\mathcal{S}_1,\ldots,\mathcal{S}_z$	symptoms
K_i	intra-component dependencies of component i
T_i	temporal degradation behavior of component i
S_i	spatial dependencies of component i

G_i	spatio-temporal dependencies of component i
o	object
$\mathcal{P}_o$	path of object o
$\mathcal{N}_i$	local neighborhood of component i
$\mathcal{K}_i$	subset of $\mathcal{N}_i$ containing the healthy components
δ_{w}	window length

PROGNOSIS

$d_1, \ldots, d_{\mathrm{r}}$	degradation modes
$c_1, \ldots, c_p$	failure modes
$\mathbf{X}$	(multivariate) degradation process
$\mathbf{Y}$	noise-disturbed measurement process of degradation process $\mathbf{X}$
$m_1(\cdot), \ldots, m_\ell(\cdot)$	parametrized models
θ_j	vector of stochastic parameters associated with degradation model d_j
ϕ_j	vector of deterministic parameters associated with degradation mode d_j
λ_l	failure threshold for failure mode c_l
τ_{eol}	failure time
$\mathbf{S}_k^j$	degradation state at τ_k according to degradation model j
$g_l(\cdot)$	function of $\mathbf{X}$ defining failure in mode c_l
$P_{\mathrm{func},l}$	system reliability with respect to failure mode c_l
P_{func}	overall system reliability
$P_{\mathrm{func},l}^j$	prediction of the system reliability with respect to failure mode c_l conditional to degradation mode d_j
$\mathcal{F}_l$	binary variable indicating whether the system fails in failure mode c_l

MAINTENANCE OPTIMIZATION

A	discrete set of possible maintenance actions
a	maintenance action, $a \in A$
T	discrete set of possible maintenance times
t	maintenance time, $t \in T$
z_i	time needed to get the personnel and material at the maintenance location and to maintain component i
q_i	maximum possible time gap between any two consecutive out-off-service periods of component i
$c_{\mathrm{m}}(a)$	direct maintenance costs of action a
$C_{\mathrm{m}}(\cdot)$	lifetime-averaged direct costs of maintenance as function of a and t
$c_i(a,t)$	indirect costs of maintenance activity a at time t
$C_i(\cdot)$	lifetime-averaged indirect costs of maintenance as function of a and t

C_{c_l}	costs of failure in mode c_l
$C_{f_j}(a)$	penalty costs of (wrong) maintenance action a in case of fault f_j
$C_r(\cdot)$	costs of risk as function of a and t
$C(\cdot)$	total maintenance costs, $C(a,t) = C_m(a,t) + C_i(a,t) + C_r(a,t)$
S_m	set of states (Markov decision process)
s_m	state, $s_m \in S_m$
A_m	set of actions (Markov decision process)
a_m	action $a_m \in A_m$
P	transition function
R	reward function
γ	discounting factor
π	policy
$\mathcal{D}$	binary variable indicating whether maintenance is planned
$\mathcal{F}$	binary variable indicating whether the system fails
Δ	time interval between two consecutive monitoring instants
u_{max}	maximum utility for planning
δ_d	discounting rate
X	system-level maintenance strategy
$c_{m,1}(a)$	part of direct maintenance costs $c_m(a)$ that does not depend on economies of scale
$c_{m,2}(a)$	part of the direct maintenance costs $c_m(a)$ that can be shared between components simultaneously undergoing maintenance action a
$c_{m,3}$	part of the direct maintenance costs $c_m(a)$ that can be shared between all simultaneously maintained components
$C_{EOS}(X)$	cost reduction thanks to economies of scale under system-level strategy X
$C_{DT}(X)$	reduction in cost of downtime of system-level strategy X
$C_{LF}(X)$	penalty costs due to loss of functionality corresponding to system-level strategy X
C_{SL}	system-level optimization criterion

RAILWAY CASE

Z_R	rail impedance
Z_B	ballast admittance
Z_S	impedance of a train shunt
Z_N	impedance of neighboring section
s	train shunt
V_{rail}	voltage applied between the two rails at the transmitter
$I_{c,i}$	current measured at receiver of section i
α_1, α_2	threshold values to define system health
γ_1, γ_2	settings of the train detection system

HVAC CASE

E^{a}_{loss}	energy loss at the air handling unit [J]
$E^{a}_{ma,thermal}$	thermal energy of the mixed air [J]
$E^{a}_{rw,thermal}$	thermal energy of the AHU return water [J]
$E^{a}_{sa,thermal}$	thermal energy of the supply air [J]
$E^{a}_{sw,thermal}$	thermal energy of the AHU supply water [J]
E^{b}_{chem}	chemical energy consumed by the boiler [J]
E^{b}_{loss}	energy loss at the boiler [J]
$E^{b}_{rw,thermal}$	thermal energy of the boiler return water [J]
$E^{b}_{sw,thermal}$	thermal energy of the boiler supply water [J]
E^{r}_{loss}	energy loss at the radiator [J]
$E^{r}_{rw,thermal}$	thermal energy of the radiator return water [J]
$E^{r}_{sw,thermal}$	thermal energy of the radiator supply water [J]
F^{a}	AHU valve fault
F^{b}	boiler fault
F^{r}	radiator fault
X^{a}	position of the AHU valve
X^{r}	position of the radiator valve
T^{a}_{ma}	mixed air temperature [°C]
T^{a}_{rw}	temperature of the water returning from the AHU [°C]
T^{a}_{sa}	supply air temperature [°C]
$T^{a}_{sa,set}$	supply air temperature setpoint [°C]
T^{b}_{sw}	temperature of the water supplied by the boiler [°C]
$T^{b}_{sw,set}$	supply water temperature setpoint [°C]
T^{b}_{rw}	temperature of the water returning to the boiler [°C]
T^{o}_{a}	outside air temperature [°C]
T^{r}_{rw}	temperature of the water returning from the radiator [°C]
T^{r}_{sw}	temperature of the water supplied to the radiator [°C]
T^{z}_{a}	zone air temperature [°C]
$T^{z}_{a,set}$	zone air temperature setpoint [°C]
w^{a}_{sa}	air flow through the AHU [kg/s]
w^{a}_{sw}	mass flow through the air handling unit [kg/s]
w^{b}_{sw}	mass flow through the boiler [kg/s]
w^{r}_{sw}	mass flow through radiator [kg/s]
Q^{a}	radiation heat [J]
U^{a}	control signal for the AHU valve
U^{r}	control signal for the radiator valve

LIST OF ABBREVIATIONS

D-S	Dempster-Shafer
TBM	Transferable Belief Model
bba	basic belief assignment
AC	Alternating Current
DC	Direct Current
FN	False Negative
FP	False Positive
LSTM	Long-Short Term Memory
HVAC	Heating, Ventilation, and Air Conditioning

CURRICULUM VITAE

Kim Verbert was born on August 15th, 1987 in Roermond, The Netherlands.

She attended pre-university school from 1999 to 2005 at Stedelijk Lyceum in Roermond. In 2005 she started her studies in Human Kinetic Technology at the Hague University of Applied Science and obtained the Bachelor of Engineering degree (cum laude) in 2009. She continued her studies in Systems and Control at Delft University of Technology, where she obtained the Master of Science degree (cum laude) in 2012. Her master thesis was entitled "Adaptive friction compensation for motion control".

In 2012 she started her Ph.D. research on fault diagnosis and maintenance optimization at the Delft Center for Systems and Control under the supervision of prof. dr. ir. Bart De Schutter and prof. dr. Robert Babuška. In 2013, she obtained the DISC certificate from the Dutch Institute of Systems and Control. In the beginning of 2015, she spent two months at the Honeywell Prague laboratory, where she performed research on fault diagnosis methods for climate control systems.

Throughout her Ph.D. research she supervised several bachelor and master students, assisted in the B.Sc. course "Stochastic signal analysis" and the M.Sc. course "Control System Design", and participated in various national and international conferences.

List of publications

Journal papers

7. K. Verbert, B. De Schutter, R. Babuška, A multiple-model approach to system reliability prediction, submitted to *Reliability Engineering & System Safety*, September 2016.

6. K. Verbert, R. Babuška, B. De Schutter, Comparing Bayesian and Dempster-Shafer reasoning for knowledge-based fault diagnosis, submitted to *Engineering Applications of Artificial Intelligence*, May 2016.

5. K. Verbert, R. Babuška, B. De Schutter, Combining knowledge and historical data for system-level fault diagnosis of HVAC systems, submitted to *Engineering Applications of Artificial Intelligence*, April 2016.

4. K. Verbert, B. De Schutter, R. Babuška, Timely condition-based maintenance planning for multi-component systems, submitted to *Reliability Engineering & System Safety*, January 2016.

3. K. Verbert, B. De Schutter, R. Babuška, Fault diagnosis using temporal and spatial information with application to railway track circuits, *Engineering Applications of Artificial Intelligence*, 56: 200-211, 2016.

2. T. de Bruin, K. Verbert, R. Babuška, Railway track circuit fault diagnosis using recurrent neural networks, *IEEE Transactions on Neural Networks and Learning Systems*, 2016.

1. K. Verbert, R. Tóth, R. Babuška, Adaptive friction compensation: A globally stable approach, *IEEE/ASML Transactions on Mechatronics* 21(1): 351-363, 2016.

Conference papers

2. K. Verbert, B. De Schutter, R. Babuška, Reasoning under uncertainty for knowledge-based fault diagnosis: A comparative study, *Proceedings of the 9th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Paris, France, 422-427, 2015.

1. K. Verbert, B. De Schutter, R. Babuška, Exploiting spatial and temporal dependencies to enhance fault diagnosis: Application to railway track circuits, *Proceedings of the 2015 European Control Conference*, Linz, Austria, 3052-3057, 2015.