Dissertation

Ein verbesserter Optimierungsansatz für Pavement Management Systeme

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften
eingereicht an der Technischen Universität Wien, Fakultät für Bauingenieurwesen

---

Doctoral Thesis

Advanced Optimization Approach for Pavement Management Systems

submitted in satisfaction of the requirements for the degree of
Doctor of Science in Civil Engineering
of the Vienna University of Technology, Faculty of Civil Engineering

by

Dipl.-Ing. Valentin Donev
Student ID: 01129517
Laudongasse 28/14, 1080 Vienna, Austria

Examiner:

Univ.Prof. Dipl.-Ing. Dr.techn. Ronald Blab
Institute of Transportation, Vienna University of Technology
Gusshausstraße 28/230/3, 1040 Vienna, Austria

Univ.-Prof. Dr.-Ing. habil. Markus Oeser
Institute of Highway Engineering, RWTH Aachen University
Mies-van-der-Rohe-Straße 1, 52074 Aachen, Germany

Univ.-Prof. Dr.-Ing. J. Stefan Bald
Institute of Road and Pavement Engineering, Technical University of Darmstadt
Otto-Berndt-Straße 2, 64287 Darmstadt, Germany

Vienna, May 2020

Acknowledgments

My interest for the topic of this thesis was sparked during the writing of the master's thesis. The research work and the accompanying peer review process took four and a half years, during which I started as an external PhD student and later became an assistant at the Institute of Transportation of TU Wien. At this point, I would like to acknowledge the people who contributed to the scientific results and supported me throughout this period.

First of all, I would like to express my sincere gratitude to my supervisor Prof. Ronald Blab, the dean of the Civil Engineering faculty, for his valuable feedback, ongoing support and encouragement. I appreciate the chance to expand my scientific interests beyond pavement and asset management as a part of his research team.

I would also like to thank Prof. Markus Oeser and Prof. Stefan Bald for agreeing to review my thesis and to join the committee for my oral examination.

This thesis would not have been possible without the scientific mentoring and support of Priv.-Doz. Markus Hoffmann. I profited greatly from the opportunity to build on his innovative work in the field, yet I was always given the freedom to explore alternative methodologies and develop my own ideas. He dedicated enormous amount of time and effort to discuss not only our joint research but also my professional and personal development, something that I appreciate as a rare privilege. By including me in his research projects, he secured the necessary financial basis during my doctoral studies, and I was also able to gain insights into the implementation of applied research into the practice.

Furthermore, I would like to acknowledge the professors and lecturers from Bulgaria who motivated me at the beginning of my engineering education and helped me build the foundation needed for my research work. I am grateful that I had the chance to know and learn from Prof. Georg Stefanoff who was a model for me in many aspects.

My deepest gratitude goes to my family which, though from a distance, always showed me their unconditional support and loving understanding.

Vienna, May 2020 Valentin Ivanov Donev

Abstract

Pavement management systems (PMSs) are assisting road agencies in network-level budgeting and project-level decision making regarding the rational planning of maintenance and rehabilitation (M&R) treatments. Faced with a large amount of data provided by automated condition surveys, common PMSs reduce the complexity of the decision problem by data aggregation towards condition-homogeneous sections and overall performance indices. This thesis consists of five publications, addressing the consequences of data aggregation for M&R optimization as well as possible solutions to the identified problems based on a new approach.

The first publication is devoted to an extensive analysis of current approaches for condition survey, assessment and prediction in Germany, Austria and Switzerland. Subjective condition thresholds limiting service life and biased deterministic performance models are identified as serious limitations with major impact on prediction and optimization results. Further examination of the drawbacks of common performance models is provided in the second publication, which focuses on the deviations resulting from ignoring data censoring and correlated competing distress types in the estimation of service life. A comparison of approaches based on a realistic simulation study identifies two methods allowing for unbiased prediction, namely a developed section-specific condition model combined with survival analysis, and a copula approach.

The third publication introduces a unique work-zone optimization approach which eliminates the need for data aggregation, thus avoiding the limitations of homogeneous sections. Here, treatment type, timing and work zones are the results of the optimization of the total discounted life cycle costs. Moreover, the optimization approach is presented as a part of a holistic end-to-end framework together with the developed costs, performance and M&R duration models. Key features of this new approach are the specific impact of any given treatment on multiple correlated distress types and the consideration of scale economies. Using high-resolution condition data instead of aggregated data greatly improves the accuracy of the models, but also leads to a very complex optimization problem. This necessitated the development of an innovative problem-tailored solution algorithm combining genetic algorithms and other heuristics. Furthermore, the applicability of the entire methodology and the effectiveness of this new algorithm are demonstrated using a parametric case study. Extensive results of optimization with hard/soft constraints as well as with and without user costs are presented and discussed in detail.

The benefits of this advanced optimization approach are emphasized by the fourth publication, which examines the consequences of aggregating short survey sections to long homogeneous sections. It is shown that the formation of homogeneous sections leads to erroneous service life and failure cause predictions. Moreover, as the aggregation is done prior to condition prediction, the number of possible work zone solutions in the optimization is greatly reduced, limiting the potential for utilization of economies-of-scale benefits. Finally, the fifth publication analyzes the consequences of common benefit-maximization strategies in network-level M&R optimization compared to cost-minimization strategies. The results suggest that the maximization of benefits based on aggregated condition indices leads to substantially higher agency costs and favors the selection of expensive M&R treatments with earlier timing, irrespective of actual failure causes and service lives. In summary, the developed advanced optimization approach allows for a substantial improvement of current PMSs towards more efficient use of public funds while minimizing the impacts of M&R treatments on road users and environment.

Kurzfassung

Pavement Management Systeme (PMS) unterstützen Straßenverwaltungen in der Bestandsverwaltung, der Budgetierung und der Planung von betrieblichen und baulichen Erhaltungsmaßnahmen. Angesichts großer Datenmengen aus einer immer genaueren Zustandserfassung reduzieren übliche PMS die Komplexität des Entscheidungsproblems durch starke Vereinfachungen. Wesentliche Ansätze dazu sind homogene Abschnitte und die Aggregation erfasster Zustandsmerkmalen zu einem Gesamtzustand für die Optimierung. Die Dissertation besteht aus fünf Publikationen, welche die Konsequenzen dieser Vereinfachungen sowie mögliche Lösungen dieser Problematik auf Basis eines neuen Ansatzes zeigen.

Die erste Publikation konzentriert sich auf eine umfassende Betrachtung der derzeit verwendeten Ansätze in der Zustandserfassung, Bewertung und Prognose der DACH-Länder. Die Analyse zeigt, welche Auswirkung die verwendeten subjektiven Zustandsgrenzen und fehlerhaften deterministischen Zustandsmodelle auf Prognose und Optimierung haben. Die zweite Publikation widmet sich den systematischen Fehlern, die sich aus der Vernachlässigung der Datenzensur und Korrelation zwischen den relevanten Schadensmerkmalen ergeben. Basierend auf einer Fallstudie wird gezeigt, wie sich diese Fehler in Zustandsmodellen und prognostizierter Lebensdauer durch eine Überlebensanalyse sowie einen Copula-Ansatz vermeiden lassen.

Die dritte Publikation stellt einen neuartigen Ansatz zur Optimierung der Bauloslänge auf Projektebene vor, der die Notwendigkeit der Datenaggregation und homogener Abschnitte vermeidet. Maßnahmenwahl, Eingriffszeitpunkt und Bauloslänge sind hier das Ergebnis der Optimierung aus der Minimierung der Lebenszykluskosten. Der neuartige Ansatz erlaubt demgemäß eine durchgängige Optimierung von der Zustandserfassung bis zur Budgetierung auf Basis von Skalenerträgen. Die dafür erforderliche detaillierte Modellierung von Zustandsmerkmalen und Maßnahmenwirkung schafft ein wesentlich komplexeres Entscheidungsproblem, dass jedoch durch einen neuartigen, heuristisch-genetischen Lösungsalgorithmus auflösbar ist. Die umfassende Analyse des Ansatzes erfolgt über eine kalibrierte Parameterstudie aus Sicht von Betreiber bzw. Betreiber und Nutzer für unterschiedlichste Randbedingungen und Zinssätze.

Die Vorteile des neuartigen Ansatzes gegenüber herkömmlichen Methoden werden anhand einer Gegenüberstellung der Konsequenzen aus der Aggregation der kurzen Erfassungsabschnitte zu langen homogenen Abschnitten und Gesamtzuständen belegt. Es kann gezeigt werden, dass die Bildung homogener Abschnitte zu systematischen Fehlern in der Prognose von Lebensdauer und Schadensursache ungeachtet von der verwendeten Methodik führt. Da die Aggregation üblicherweise vor der Zustandsprognose erfolgt, kommt es zudem zu einer starken Einschränkung des Lösungsraumes in der Optimierung. Dementsprechend können die sich ergebende Maßnahmenwahl und Bauloslänge in bisher üblichen PMS-Ansätzen allein schon deshalb nicht optimal sein. In der fünften Publikation werden abschließend die Konsequenzen der Nutzenmaximierung auf Basis der Kostenwirksamkeit den Ergebnissen des neuen durchgängigen Lebenszykluskostenansatzes auf Netzebene gegenübergestellt. Die Resultate auf Basis einer Parameterstudie belegen, dass die im PMS übliche Nutzenmaximierung für den Betreiber zu substanziell höheren Kosten im Vergleich zur Minimierung der Lebenszykluskosten unter vergleichbaren Bedingungen führt. Zusammenfassend ermöglicht der neuartige Ansatz eine substanzielle Verbesserung gegenüber herkömmlichen Ansätzen im PMS. Dementsprechend kann eine Umsetzung dieses verbesserten Optimierungsansatzes zu einem wesentlich effizienteren Einsatz der öffentlichen Mittel aus Sicht von Betreiber, Nutzern und Umwelt führen.

Contents

CHAPTER 4 Optimization of pavement maintenance and rehabilitation activities, timing and work zones for short survey sections and multiple distress types 75

CHAPTER 1

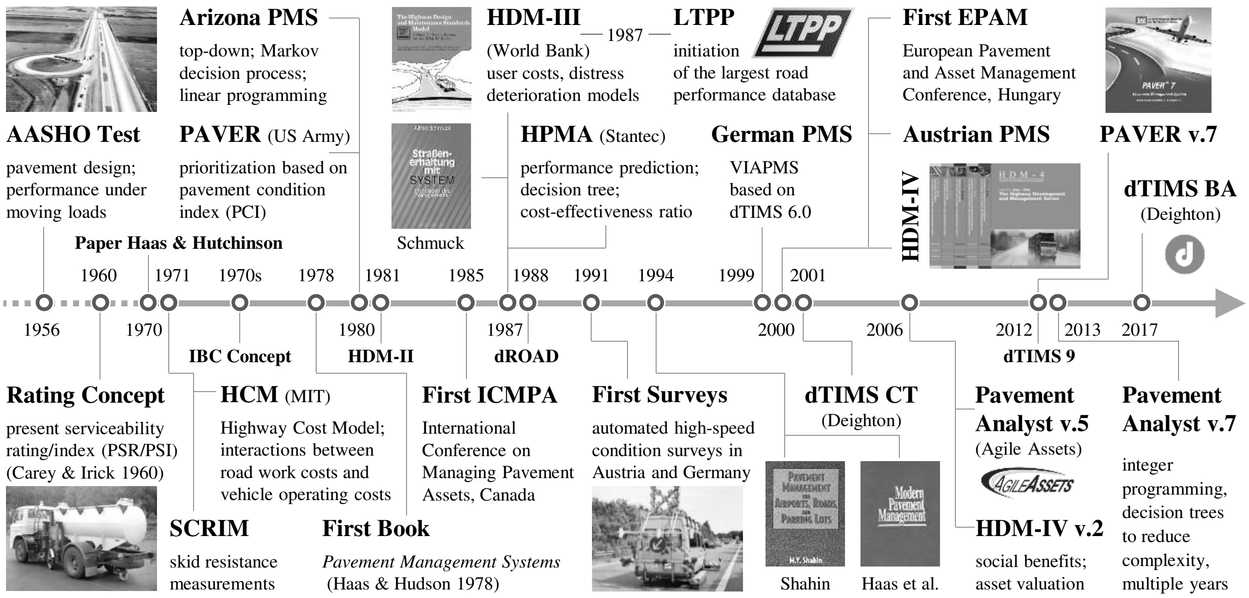
Introduction and summary

1.1 BACKGROUND, OBJECTIVES AND METHODOLOGY

1.1.1 Background

For the most part, road networks in developed countries have already been constructed, shifting the main focus of public agencies from construction to expansion and maintenance of the existing road infrastructure. While the construction of new pavements is a one-time expenditure, the maintenance of existing road networks of thousands of kilometers requires large investments each year, with the goal of extending service life and allowing safe use with minimum traffic interruptions and travel time delays. In light of the huge public attention and the scarcity of public funds, there is a demand for a systematic transparent and efficient decision-making approach based on objective criteria. To satisfy this demand, pavement management systems (PMSs) were developed. The decisions in a PMS are based on condition data from periodic surveys, mathematical models predicting future pavement conditions as well as the effects and costs of different maintenance and rehabilitation (M&R) treatments. The main network-level objective is either to minimize the total agency (and user) costs, to maximize the benefits (performance), to maximize reliability (bridges), or to minimize environmental impacts (emissions) given annual budget constraints. In more recent years, there have been increasing efforts to develop multi-objective optimization approaches (e.g. Kim and Frangopol 2017, Santos et al. 2018). Nevertheless, the achievement of the defined objective depends on the actual implementation of the optimal M&R program in practice. Thus, a PMS must also provide plausible location-specific treatment recommendations in terms of timing, type, work-zone length and layout that could be accepted by road engineers and applied with limited need for manual adaptation.

A brief overview of the development of pavement management concepts and software over the years is illustrated in Figure 1.1. The American Association of State Highway Officials (AASHO) Road Test from 1956 to 1960 was one of the largest full-scale road experiments ever conducted that led to significant advancements in pavement structural design and performance prediction models. Following the development of condition assessment concepts (e.g. Carey and Irick 1960), Haas and Hutchinson (1970) provided a framework and a comprehensive description of the main elements of a PMS with implementation recommendations. Parallel to that, methods and equipment for automated measurement of road surface characteristics at traffic speed were developed (e.g. Hosking and Woodford 1976, Benson et al. 1988). In the late 1970s and early



AASHO…	American Association of State Highway Officials	dTIMS…	Deighton Total Infrastructure Management System
SCRIM…	Sideway-force Coefficient Routine Investigation Machine	LTPP…	Long-Term Pavement Performance Program
HDM…	Highway Development and Management Model	HPMA…	Highway Pavement Management Application
HCM…	Highway Cost Model	IBC…	Incremental Benefit Cost

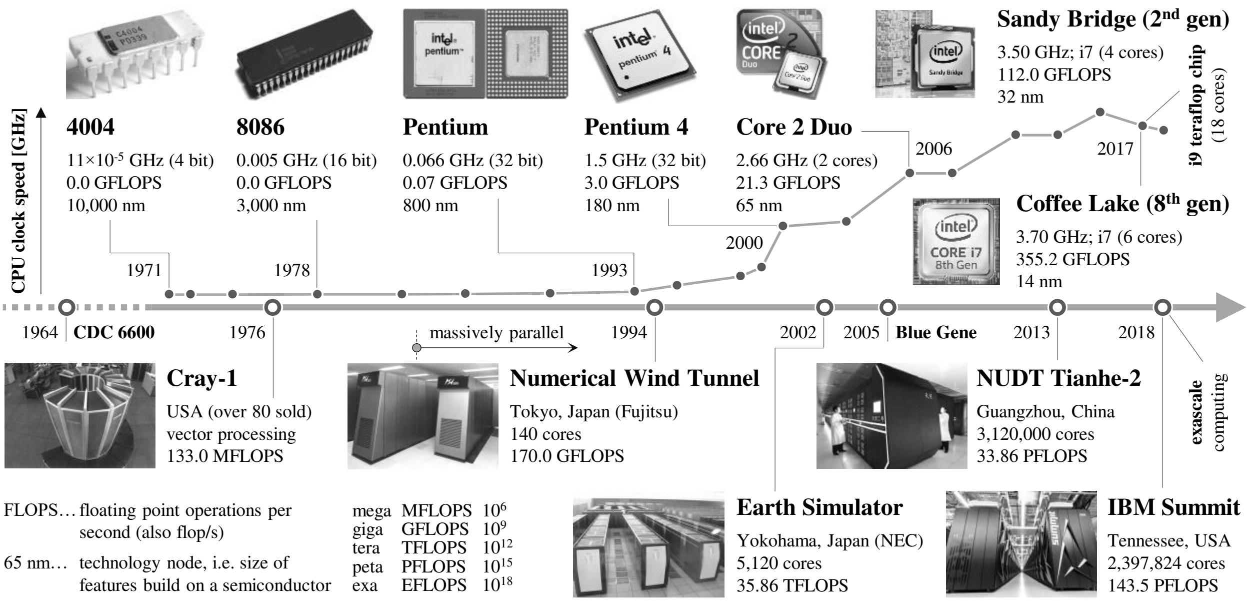
Note: The timeline is based on a subjective selection with no claims to completeness or accuracy. Logos, product and company names are registered trademarks of their respective holders, and are used for illustrative purposes only. The author has no affiliation with any company or involvement in any of the software products.

FIGURE 1.1 Timeline of selected concepts, approaches and software as well as international conferences, paper and book publications with significance for pavement management applications (1956-2017).

1980s, the first PMS implementations became reality, with notable examples like PAVER, Arizona PMS and Washington PMS (Shahin and Kohn 1981, Golabi et al. 1982, Nelson and LeClerc 1982). The Canadian software Deighton's Total Infrastructure Management System (dTIMS) has been in use by state agencies in the USA since 1990. The software was also used as a basis for the PMS implementations in Germany and Austria in 1999/2000 (Knepper 2001, Weninger-Vycudil 2001). Another prominent example of a PMS is the Highway Design and Maintenance Standards Model (HDM), which was funded by the World Bank and developed in multiple stages based on the results of several international field studies (Moavenzadeh et al. 1971, Watanatada et al. 1987, Kerali et al. 2006). Furthermore, the first books on PMSs were authored by Haas and Hudson (1978) in the USA and Schmuck (1987) in Germany. Initiated in 1987, the Long-Term Pavement Performance (LTPP) program has been providing researches with comprehensive data for the calibration and validation of countless performance models (FHWA 2015). Today, according to Haas et al. (2015), the leading vendors of PMS software worldwide are AgileAssets (Pavement Analyst) and Deighton (dTIMS).

It is important to note that this overview focuses mostly on common PMS software solutions implemented and used by road agencies. The contributions of individual researches to specific fields of pavement management like distress modeling or M&R optimization are not included for clarity reasons. However, a discussion of some of these models from the research literature is provided in Section 1.2 and Chapters 2 to 6.

PMSs are computer-based systems, allowing for storage, processing, analysis and reporting of pavement management data. Naturally, computer capabilities have had a major influence on the conception and development of PMSs. Thus, the evolution of pavement management concepts and software may also be analyzed in the context of the evolution of computer



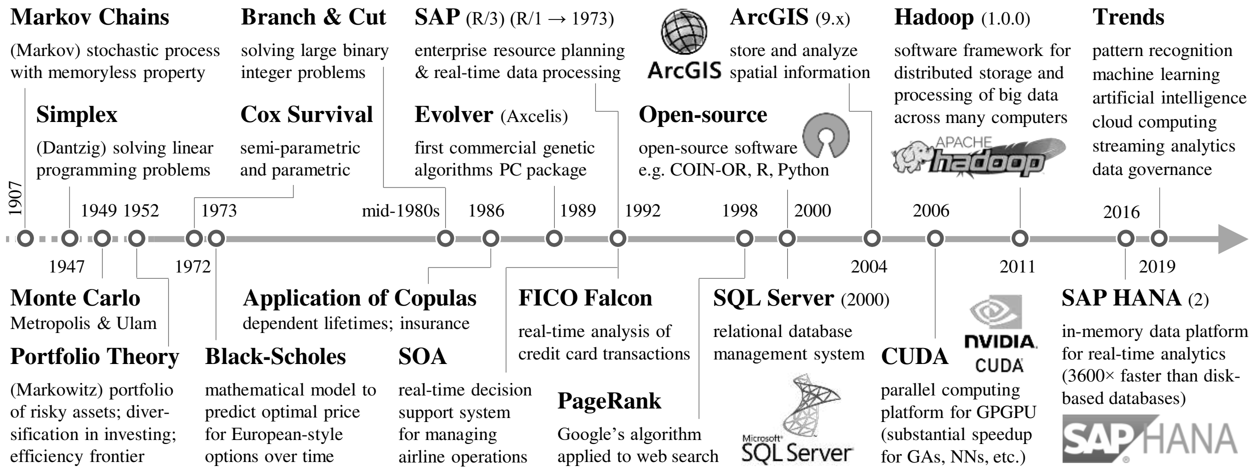
Note: The provided technical parameters may not be current, complete, accurate and/or comparable. Logos, product and company names are registered trademarks of their respective holders, and are used for illustrative purposes only. The author has no affiliation with any manufacturer or involvement in any of the products.

FIGURE 1.2 Timeline of the evolution of computer performance: selected commercially-available desktop microprocessors (e.g. Intel) vs. some of the world's most powerful supercomputers (1964-2018).

processing power, which is illustrated in Figure 1.2 (Intel 2012, Stallings 2016, Intel 2018, Strohmaier et al. 1993-2018). The figure shows that the clock speed of processors increases exponentially until 2006. After 2006, further increase in computing power has been achieved through multicore technology and parallelism. A crude estimation shows that the performance, measured in floating-point operations per second, has increased roughly 5000 times from 1993 (Pentium) to 2017 (Coffee Lake). Similar trends but on a higher order of magnitude can be identified by examining the evolution of supercomputers, where the number of processors has increased from hundreds to millions (massively parallel computers).

In addition to computing power, computer data storage has also gone through a remarkable development, referring both to primary (e.g. random-access memory) and secondary storage (e.g. hard drive) devices. The evolution of hard disk drives (HDD) is characterized by ever-increasing capacity and data transfer rates as well as ever-decreasing access time, physical dimensions and price per gigabyte storage. Nowadays, cloud computing offers practically unlimited storage capabilities with additional benefits of easier maintenance, sharing of information and access. Nevertheless, a typical pavement condition survey for a large road network will produce a few million data records, which will occupy only several hundred megabytes of disk storage.

The advancements in computer hardware, performance speed and storage capabilities go hand in hand with the progress in software, mathematical methods and algorithms. Figure 1.3 illustrates some of the achievements with relevance to management science and methods discussed in this thesis (Metropolis and Ulam 1949, Markowitz 1952, Cox 1972, Black and Scholes 1973, Carrière and Chan 1986, Rakshit et al. 1996, Page et al. 1998). Notably, the development of database management systems, artificial intelligence and parallel computing has opened a new set of opportunities for data analysis and optimization in PMSs. A major advancement in scientific computing was the development of general-purpose computing on graphics processing units (GPGPU). The highly parallel structure of GPUs makes them suitable for vector and matrix algebra computations and applications in image processing and data



GIS…	Geographic Information System
SQL…	Structured Query Language
GPGPU…	General-Purpose Computing on Graphics Processing Units
GA…	Genetic Algorithms (NN… Neural Networks)

Note: The timeline is based on a subjective selection of events with no claims to completeness or accuracy. Logos, product and company names are registered trademarks of their respective holders, and are used for illustrative purposes only. The author has no affiliation with any manufacturer or involvement in any of the products.

FIGURE 1.3 Timeline of selected mathematical models with real-world applications, decision support and database systems, big-data and software with significance for management science (1907-2016).

mining. Consequently, GPUs together with multicore processors have been used to dramatically reduce running time for neural networks and genetic algorithms, and also stimulating the development of new classes of algorithms like parallel metaheuristics (Alba et al. 2013, Tsutsui and Collet 2013). Furthermore, in-memory database management systems (SAP HANA) allow for a significantly faster data access and processing in comparison to traditional disk-based systems. Nevertheless, time is less critical factor in PMSs in contrast to, for example, winter road maintenance systems where a dynamic real-time optimization is required. As the optimization analysis in PMSs is typically conducted from time to time or when new survey data becomes available, computational time should not be the decisive factor in selecting algorithms.

In summary, computer computing power, storage, data processing and optimization algorithms have undergone a remarkable evolution in the last 30 years. Unfortunately, the same cannot be said about common PMSs where the software interface and visualization of data has improved but much of the simplifications made due to limited computer capabilities are still present. The two main simplification and drawbacks are the aggregation of condition survey data to long homogeneous sections and the optimization of M&R activities based on the area between the do-nothing and post-treatment performance models of an aggregated condition index (see Section 1.2.2 for details). Based on a comparison to new developed approaches, this work provides a quantitative analysis of the consequences from these simplifications, which were first systematically described together with other drawbacks of common PMSs by Hoffmann (2006).

Today, computer technology is no longer a limitation for the storage and processing of condition survey data. The challenging part remains solving the M&R optimization problem which if formulated without oversimplifications belongs to a difficult-to-solve class of problems. Nevertheless, this thesis presents the methodological foundations and exemplary application of new M&R optimization algorithm, capable of solving such problems. The developed approach completely eliminates the need for homogeneous sections and aggregated condition indices in PMSs. Moreover, it is possible to extend the approach to a full-stochastic cross-asset optimization accounting for risk that can be applied to any aging infrastructure system (Hoffmann 2018).

1.1.2 Objectives and methodology

The main objective of this thesis has been the development, programming and exemplary application of an approach for work-zone optimization based on short survey sections and holistic life cycle modeling, considering censored pavement data, multiple correlated distress types and multiple M&R treatment alternatives. This main objective has been divided into several tasks including an additional task, aiming to justify the necessity of the new approach by means of comparison to common state-of-the-art methods:

- Quantitative evaluation and systematic comparison of commonly used PMS approaches for generation of homogeneous sections, condition prediction and M&R optimization
- Systematic investigation of the consequences of data censoring and correlated competing risks for the estimation of service life, and identification of the most appropriate methods
- Development of data-generation procedure as well as calibration of cost and performance models necessary for the application of the new work-zone optimization methodology
- Development of an advanced work-zone optimization approach based on scale economies in a spreadsheet environment and design of solution algorithm for the formulated problem

These objectives were addressed based on a holistic end-to-end approach, with specific aspects being published in five peer-reviewed journal publications outlined in the following section (see Section 1.2.1).

This thesis uses synthetic computer-generated data for the demonstration of the applicability of the developed new approaches and as a basis for a ceteris paribus comparison and evaluation of common PMS methods. For this purpose, a general procedure for simulation of multiple distress types and short survey sections incorporating spatial correlation was developed. Nevertheless, available empirical data from different sources was used to calibrate all relevant input parameters, thus attaining a realistic case study for the comparison.

In general, simulation studies are widely recognized as a scientific tool for the development of new theories and a deeper understanding of complex systems. Simulation studies are used quite often in the literature on M&R treatment optimization, especially when demonstrating the applicability of new approaches (e.g. Fwa et al. 1996, Medury and Madanat 2011, Sathaye and Madanat 2012, Yeo et al. 2013). One of the reasons for this is that new approaches have to be tested for a variety of settings, and possible limitations have to be identified prior to any real-world data application. Sensitivity analysis and what-if scenarios help identify key factors, trade-offs and improvement potential of the overall system (e.g. design optimization, inspection intervals). Furthermore, a complete knowledge of service life enables model comparison and objective quantification of the deviations resulting from a specific model. In addition, a goodness-of-fit measure (e.g. R^2) alone does not prove the validity of a model and may be misleading due to representativeness issues and biases associated with empirical data (e.g. censoring). Moreover, traditional empirical research is limited to the existing data and cannot assess the benefits of new technologies. In any case, the combination of empirical data and simulation methods can provide a powerful tool for evaluation and comparison of optimization approaches.

In this thesis, all investigated and developed methods were manually implemented in Microsoft Excel and Visual Basic for Applications. One open-source and one commercial add-in for Excel have been used as solver engines for solving optimization problems. All graphical representations constitute original figures which were created manually and subsequently edited.

1.2 SUMMARY AND CONTRIBUTIONS

1.2.1 Overview of the thesis

Chapter 1 of this thesis provides an overview of the evolution of PMSs over the years in parallel with the advancements in computing power, mathematical algorithms, database management and software. The chapter highlights the research objectives as well as the connection between the different publications, of which this cumulative thesis consists. The most significant findings of the thesis are summarized in the following sections along two lines: analysis of the drawbacks of common PMSs (Section 1.2.2) and proposed innovative solution approaches (Section 1.2.3). The chapter ends by giving a summary of the author's contributions (Section 1.2.4) with conclusions and outlook for future work (Section 1.3). Figure 1.4 illustrates the main research issues treated in the thesis, linking them to the corresponding PMS module. Some key research topics for future work are identified as well. Notably, most of the work is concerned with the two main PMS elements: condition prediction and M&R optimization.

Chapter 2 (**Publication 1**) can be seen as an introductory chapter. The publication is mostly based on Donev (2014) but with some important extensions. The current approaches for condition survey, assessment and prediction in Germany, Austria and Switzerland are briefly discussed and compared. The results show that the survey and rating procedures differ in the number and type of surveyed distresses, measuring equipment, rating scales and definitions of aggregated condition indices. The drawbacks of aggregating individual distresses are highlighted.

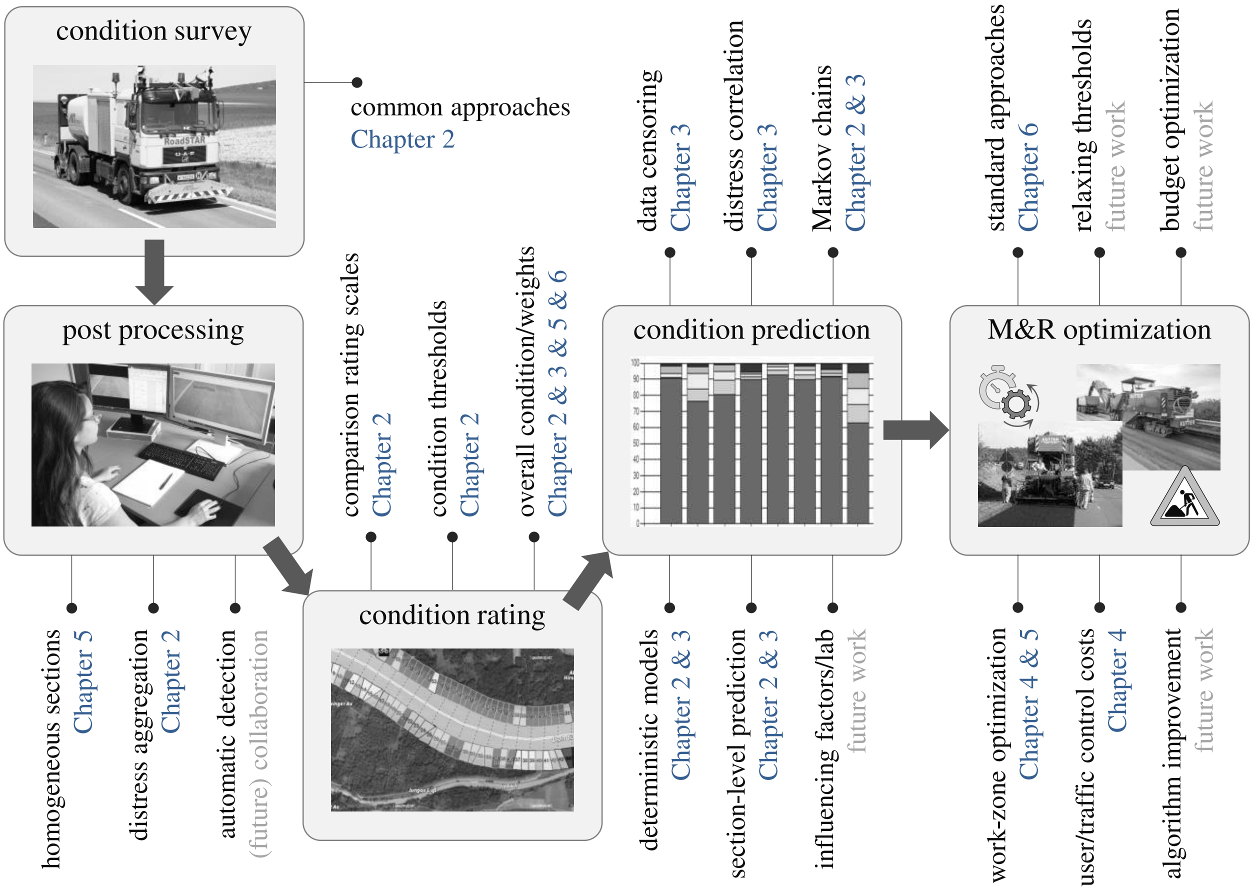


FIGURE 1.4 General overview of the basic modules in a typical PMS with key research issues and topics covered in the individual chapters of the dissertation together with some possible future objectives.

Furthermore, performance prediction models derived using multiple regression are discussed, with emphasis on multicollinearity issues. Within the chapter, two stochastic condition models are analyzed: homogeneous Markov chains and the Hoffmann process (Hoffmann 2018). An important contribution is the derivation of a stochastic performance model at the section level. Common section-level prediction approaches using shifting/scaling of a master function are compared to the developed stochastic model based on empirical time-series data for 100 sections.

Chapter 3 (**Publication 2**) focuses on the issue of data censoring in the context of pavement condition survey data with a discussion of common examples in practice. Computer-generated data is used to quantify the effect of neglecting censoring in the estimation of empirical performance models. The analysis is conducted for multiple distress types (competing risks), and by examining different possible distress correlations. Various models are applied to the artificially censored data including nonparametric and parametric survival analysis models, Markov chains and non-linear regression. While survival analysis models account for censoring, all models lead to biased distress-specific distributions in the presence of dependent competing risks. Two solutions to this problem are proposed, namely to combine the developed section-specific model from Chapter 2 with parametric survival analysis or to use copulas.

Chapter 4 (**Publication 3**) introduces a unique work-zone optimization approach which eliminates the need for aggregation, avoiding the limitations of homogeneous sections described in Chapter 5. Optimal treatment timing, type and work zones are the results of optimization which minimizes the total discounted life cycle costs. Key features of the approach are the consideration of multiple correlated distress types and economies-of-scale cost functions. The method requires unbiased service life estimates which can be obtained by taking into account the findings of Chapter 3. Nevertheless, employing scale economies and using detailed condition data instead of aggregated data leads to a much more complex optimization problem. The developed problem-tailored innovative algorithm combines different heuristics techniques, including genetic algorithms to identify close to optimal work-zone solutions. Moreover, the work-zone optimization approach is presented as a part of a holistic framework for a new PMS. For this purpose, all necessary cost (agency, user and temporary traffic control), performance and M&R duration models were developed. Extensive results of optimization with and without user costs and soft constraints for different discount rates are presented and discussed in detail.

Chapter 5 (**Publication 4**) and Chapter 6 employ the same case-study data as Chapter 4 to illustrate the limitations of state-of-art approaches for data aggregation and M&R programming. In Chapter 5, the consequences of aggregating short survey sections to long homogeneous sections are analyzed. It is shown that the formation of homogeneous sections leads to erroneous service life and failure cause predictions with threshold violations. Moreover, the sections shift with each survey, and the results are highly sensitive to the used algorithm. This identifies the need for alternative solutions, with one possibility being the approach from Chapter 4.

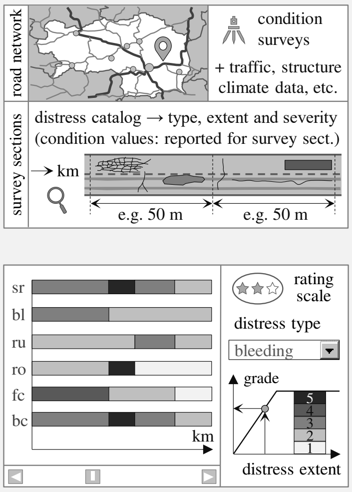
Chapter 6 (**Publication 5**) analyzes the consequences of common benefit-maximization strategies in network-level M&R optimization. In the given context, benefit is commonly defined as the area between the performance curves with and without treatments of an aggregated condition index. Strategies that maximize benefits are compared to strategies that minimize discounted life cycle costs for 1000 road sections and different budget scenarios. The results show that the maximization of benefits based on an aggregated condition index leads to substantially higher agency costs and favors the selection of expensive M&R treatments with earlier timing, irrespective of actual failure causes and service lives.

1.2.2 Limitations of common approaches in pavement management

Systematic analysis of methodological approaches in pavement management requires thorough knowledge of the individual PMS components and processes and their interaction. The impact of a specific method or approach should be investigated up to the optimization and final results. The majority of research literature, however, is dedicated to the development of new methods instead of systematic analysis and comparison of existing ones. Moreover, condition assessment, condition prediction and optimization have been the focus of separate groups of researchers. The commercial market is also split between providers of data collection equipment and PMS software vendors (Perera et al. 2008). Thus, the former are usually not acquainted with the advantages and disadvantages of the different prediction and optimization methods or how the collected data can be put to best use. There is also a discrepancy between the quality and density of data provided by automated condition surveys and the data actually used in PMS. A large part of this thesis is dedicated to a critical analysis of common PMS approaches. This section provides only a summary of the key conclusions (for details consult the individual chapters).

PMSs in German-speaking countries (DACH) rely on data from periodic automated condition surveys every 3 to 6 years. The condition assessment procedure includes dividing the road network into sections, and for each section rating the collected distress data based on distress type, extent and severity. To summarize and compare the results, the individual distresses are aggregated to an overall condition index (OCI). Although the survey and rating approaches in different countries are very similar, the results are not directly comparable due to different sectioning algorithms, differences in the number and type of surveyed distresses, used rating scales, as well as formulas and weights for computation of the OCI. Figure 1.5 summarizes the limitations of common condition survey and assessment procedures. The collected raw condition data is subjected to multiple levels of aggregation prior to its use as an input in condition

Condition survey and assessment procedures (DACH)



In the thesis:
Section 2.1

⚠ country-specific survey intervals, length of survey sections, distress catalogues, measuring equipment/principles, etc.

⚠ as well as different rating scales, aggregation rules & weights for overall index → subjectivity/low comparability of results

⚠ aggregating distress types with different causes (e.g. traffic/temperature) and progression to a distress group as well as

⚠ converting linear cracks to area & severity-based weighting affects prediction accuracy and effective treatment selection

⚠ arbitrary condition thresholds lead to unnecessary loss of service life (solution: trigger values as decision variables)

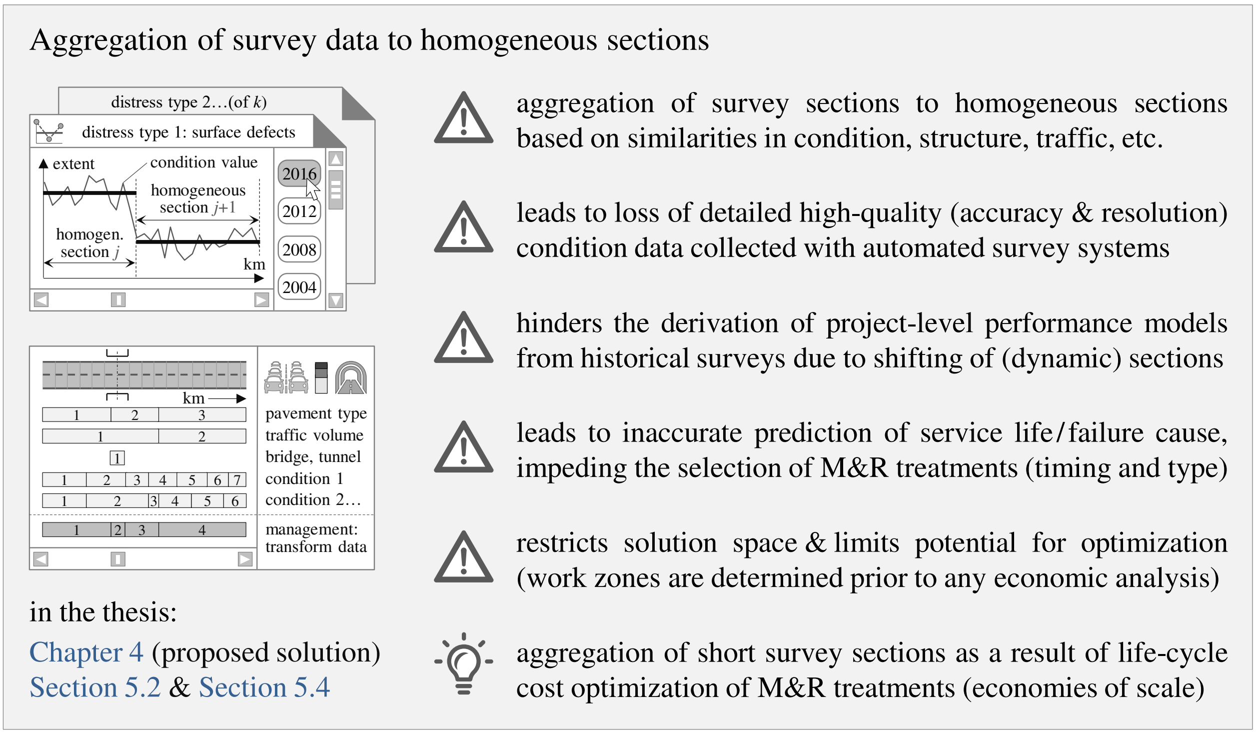
💡 condition survey covering distress extent and severity for each distress type based on short survey sections ($\leq$ 50 m)

FIGURE 1.5 Main limitations of common approaches for pavement condition survey and assessment.

prediction. For example, if a pavement section exhibits a distress type with multiple severity classes (e.g. low and moderate cracking), the total extent of this distress is computed using severity-based weighting. Another level of aggregation is when different distress types are combined into a distress group. Examples include aggregation of surface distresses (bleeding, raveling, delamination, etc.) and aggregation of alligator cracking and linear cracks (converted into cracking area), as stipulated in HDM and the Austrian PMS (Paterson 1987, Weninger-Vycudil 2001). However, the combination of distress types with different causes, initiation mechanisms (e.g. top-down vs. bottom-up cracks) and progression obviously leads to inaccurate performance prediction models and hinders the assignment of optimal treatments.

At the next stage of aggregation, different distress types are combined into an OCI, representing the condition of a pavement section with a single score. The overall condition can be used to provide an overview of the network condition for high-level decision makers and to identify the worst-performing road sections. However, if the OCI is used in condition prediction or as an objective in M&R optimization (e.g. maximizing benefits), this leads to serious drawbacks, which will be discussed later in this section. In the following, the spatial aggregation of data based on the generation of homogeneous sections will be reviewed.

The results of a condition survey are usually provided for short survey sections (e.g. 10-50 m). In the preparation of the data for PMS, these sections are aggregated to longer homogeneous sections to reduce the amount of data to be analyzed and to obtain practical lengths for M&R projects. The sections are combined based on similarities in condition (only last survey), pavement structure and traffic volume using an algorithm like cumulative difference approach or Bayesian segmentation algorithm (AASHTO 1993, Thomas 2004). The consequences of this aggregation for condition prediction and M&R optimization are hardly ever discussed in the literature. Figure 1.6 lists the main limitations of this approach, with evidence from a parametric case study provided in the thesis (see Chapter 5).



Aggregation of survey data to homogeneous sections

⚠ aggregation of survey sections to homogeneous sections based on similarities in condition, structure, traffic, etc.

⚠ leads to loss of detailed high-quality (accuracy & resolution) condition data collected with automated survey systems

⚠ hinders the derivation of project-level performance models from historical surveys due to shifting of (dynamic) sections

⚠ leads to inaccurate prediction of service life/failure cause, impeding the selection of M&R treatments (timing and type)

⚠ restricts solution space & limits potential for optimization (work zones are determined prior to any economic analysis)

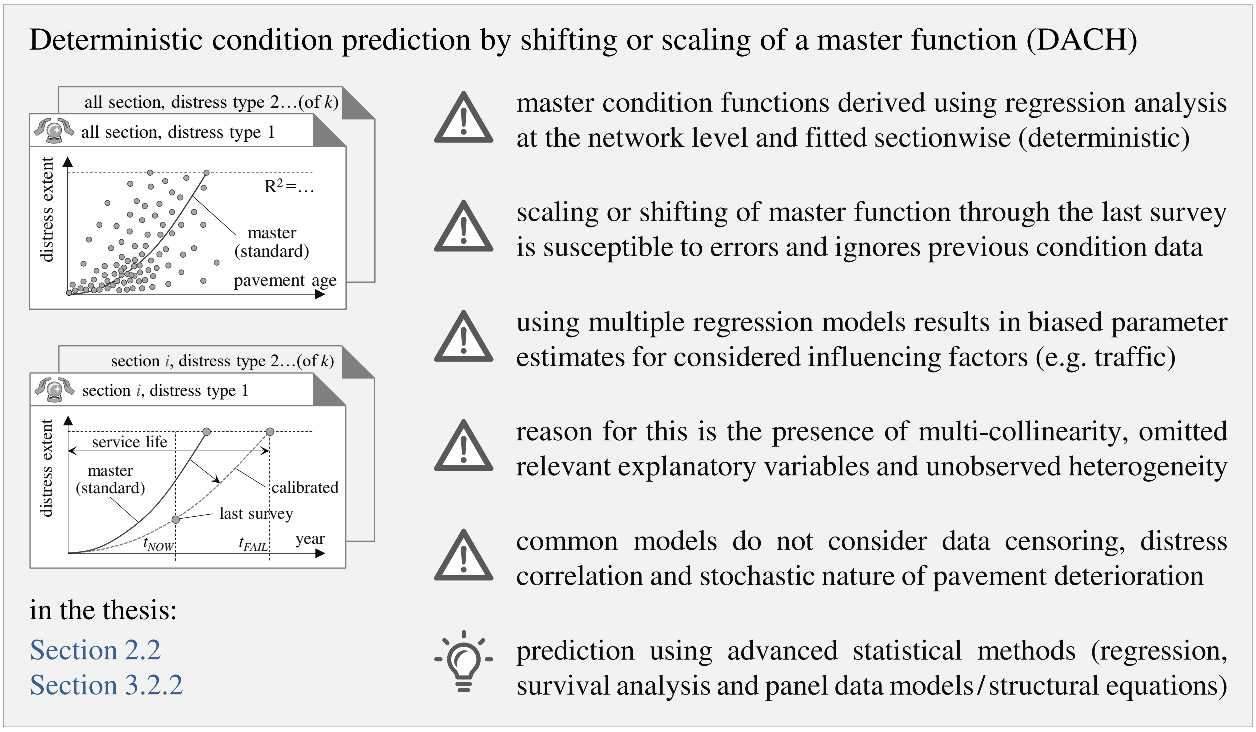
💡 aggregation of short survey sections as a result of life-cycle cost optimization of M&R treatments (economies of scale)

in the thesis:
Chapter 4 (proposed solution)
Section 5.2 & Section 5.4

FIGURE 1.6 Main consequences from the aggregation of short survey sections to long homogeneous sections prior to condition prediction and M&R optimization with reference to a proposed solution.

The formation of homogeneous sections has been an integral part of PMSs since their first introduction in the late 1970s and early 1980s. At that time, it was a necessary simplification due to limited computer capabilities but also due to a lack of methodology on how to combine short sections (with assigned treatments) into work zones. In the following years, much effort was focused on development and improvement of existing road-sectioning algorithms. However, not only each algorithm leads to different sectioning but also each survey produces different homogeneous sections. The shifting of dynamic homogeneous sections impairs the consistency of collected time-series condition data for a given road section. Thus, only the last survey can be used in performance prediction at the section level. Although fixed management sections do allow the use of historic condition data, they also cause averaging errors and loss of information (even more so than dynamic sections). Averaging errors in current condition manifest themselves in inaccurate condition predictions and potential future threshold violations for each individual distress type. Blurred predictions in terms of service life and failure cause hinder the selection of optimal timing and type for M&R treatments. Moreover, generating homogenous sections prior to condition prediction limits the number of possible work-zone solutions and the potential benefits of economies-of-scale cost optimization. An alternative approach that does not rely on homogeneous sections and data aggregation is presented in Chapter 4.

Apart from the averaging errors resulting from the input data (homogeneous sections), performance prediction models in common PMSs suffer from further methodological drawbacks. In German-speaking countries, empirical deterministic condition (master) functions have been derived for each distress type, using regression analysis at the network level. For a section-level prediction, these functions are shifted or scaled through the last measured condition value. A summary of the deviations resulting from the application of this approach is provided in Figure 1.7. First, the developed models ignored data censoring which can lead to biased model parameters and service life estimates. An example of ignoring (right) censoring is excluding road



Deterministic condition prediction by shifting or scaling of a master function (DACH)

⚠ master condition functions derived using regression analysis at the network level and fitted sectionwise (deterministic)

⚠ scaling or shifting of master function through the last survey is susceptible to errors and ignores previous condition data

⚠ using multiple regression models results in biased parameter estimates for considered influencing factors (e.g. traffic)

⚠ reason for this is the presence of multi-collinearity, omitted relevant explanatory variables and unobserved heterogeneity

⚠ common models do not consider data censoring, distress correlation and stochastic nature of pavement deterioration

in the thesis:
Section 2.2
Section 3.2.2

prediction using advanced statistical methods (regression, survival analysis and panel data models / structural equations)

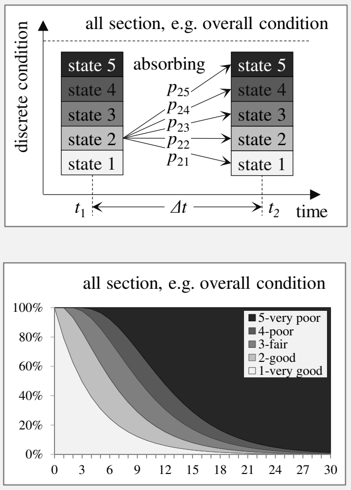
FIGURE 1.7 Main limitations of deterministic approaches for condition prediction based on scaling or shifting of a master function through the last survey applied in PMS in German-speaking countries.

sections with no signs of distress from the analysis. Second, common models suffer from several types of bias resulting from omitting relevant explanatory variables, unobserved heterogeneity and multicollinearity (high correlation between age and accumulated traffic). Third, over-simplified linear model forms do not describe realistically the actual distress progression. Finally, the shifting of the master function may give a plausible prediction for the first one or two years but it is not suitable for medium- and long-term predictions. Many of the described limitations can be avoided by using more-advanced statistical techniques like section-level regression (Chapter 2), survival analysis (Chapter 3) and panel data models (Madanat et al. 2010).

Due to the widely recognized drawbacks of deterministic prediction models, many researchers directed efforts towards the implementation of stochastic Markov chain models in pavement and bridge management (e.g. Butt et al. 1987, Jiang et al. 1988, Madanat and Ben-Akiva 1994, Li et al. 1996, Thompson et al. 1998, Mishalani and Madanat 2002, Abaza et al. 2004, Tsuda et al. 2006, Lethanh and Adey 2013). Markov chain models are popular in the literature because the concept of transition probabilities is easily understood and the models can be calibrated with fewer requirements on available condition data and influencing factors. In addition, the Markov decision process provides a framework for M&R modeling and the resulting optimization problem can be solved using standard techniques like dynamic programming. Nevertheless, Markov chain models exhibit several serious methodological drawbacks, as shown in Figure 1.8. Although a few researchers have to some extent accounted for performance history, age-dependent transition probabilities, censoring and influencing factors, one of the main drawbacks remains the inadequacy of a discrete-state model for prediction of the continuous pavement condition data, with the number of states having a major impact on the results.

State-of-the-art PMSs usually rely on single-objective M&R optimization, with the two most common objectives being minimization of agency and/or user costs and maximization of benefits. In the latter case, benefit (sometimes referred to as effectiveness) is defined as the area

Condition prediction at the network level using homogeneous Markov chains (stochastic)



in the thesis:
Section 2.2.3
Section 3.2.3

⚠ future road deterioration depends only on current condition (memoryless) ignoring performance and maintenance history

⚠ constant transition probabilities (homogeneous) ignore the cumulative deterioration due to ageing, traffic loading etc.

⚠ distortion of continuous condition data → the number of states has major impact on transition matrix and service life

⚠ no section-specific prediction (equal for all sections in the same state) → limited usability for PMS at the project level

⚠ similar average deterioration curves for all distress types due to exponential distribution and absorbing state

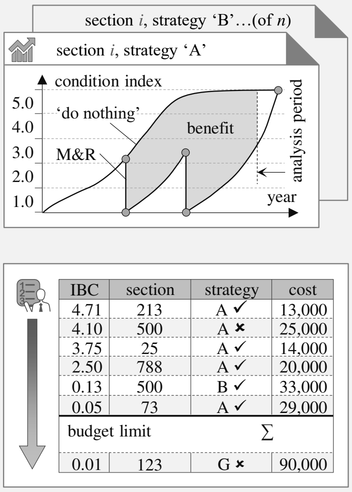
⚠ data censoring and distress correlation are neglected; standard formulations do not account for explanatory variables

FIGURE 1.8 Main limitations of using homogeneous discrete-time (discrete-state) Markov chain models in pavement condition prediction at the network level.

between the do-nothing and post-treatment performance curves of an OCI. The benefit-maximization approach has been the core element of several widely used pavement and bridge management systems for more than 30 years (Li et al. 1998, Thompson et al. 1998, Weninger-Vycudil 2001, Zavitski et al. 2008, AASHTO 2012). The concept of maximizing benefits is appealing and easily understood by both practitioners and decision makers. The "benefits" achieved in this manner, however, are not related to actual monetary or non-monetary benefits, resulting from the application of a specific M&R strategy. Another reason for the popularity of the approach is that the optimization can be conducted using the heuristic incremental benefit-cost (IBC) technique. IBC can be applied to a large road network, requiring significantly less computational effort (running time) in comparison to other available algorithms (Patidar et al. 2011). Using the efficiency frontier for screening of strategies at the section level, significantly reduces the number of M&R strategies considered at the network-level. This was an essential factor at the time when the above-referenced PMSs were first developed (see Section 1.1.1).

A comparison to strategies minimizing the total discounted life cycle costs (state-of-the-art in the research literature) clearly shows that the maximization of benefits is not an appropriate optimization objective in pavement management. The key findings from such comparison are listed in Figure 1.9. The case study in Chapter 6 shows that the maximization of benefits based on an OCI leads to a strong preference of major rehabilitation and replacement treatments with systematically earlier timing, irrespective of actual failure causes and service-life predictions. The reason for this is that extensive treatments yield a larger area between the curves but they also cost the most. Thus, for more generous annual budgets, following benefit-maximization strategies leads to substantially higher agency costs in comparison to strategies resulting from the minimization of life cycle costs. Furthermore, the results at the network level suggest that the annual agency costs do not exhibit a high sensitivity regarding the weights for the individual distresses in OCI, but at the road section level, the optimal M&R strategy might change.

M&R programming using benefit maximization based on overall condition index



IBC	section	strategy	cost
4.71	213	A ✓	13,000
4.10	500	A ✗	25,000
3.75	25	A ✓	14,000
2.50	788	A ✓	20,000
0.13	500	B ✓	33,000
0.05	73	A ✓	29,000
budget limit		Σ	
0.01	123	G ✗	90,000

⚠ network-level M&R programs maximizing benefits (area between the curves) with agency costs only as a constraint

⚠ give preference to expensive rehab/reconstruction treatments (larger benefit area), irrespective of actual failure causes

⚠ lead to treatments with earlier timing (larger benefit area, more frequent repairs), irrespective of actual service life

⚠ result in high annual agency costs (less binding budget) and inefficient use of funds, i.e. M&R backlog (tight budget)

⚠ the optimal M&R strategy for each project depends on (subjective) weights of a fictive overall condition index

💡 network-level M&R programs minimizing agency (and user) life cycle costs for multiple distress types and treatments

in the thesis:
Section 6.2
Section 6.4

FIGURE 1.9 Main consequences from the use of benefit-maximization approaches based on an overall (composite) condition index in the optimization of M&R programs at the network level.

1.2.3 Proposed solutions to common drawbacks

In addition to the systematic analysis and criticism of common approaches in condition assessment, performance prediction and M&R optimization, this thesis makes several innovative conceptual and methodological contributions to the-state-of-the-art in pavement management. The novel part includes the development of a comprehensive simulation approach, advancements in condition prediction and a holistic end-to-end framework for M&R optimization, which makes the homogeneous-sections approach obsolete.

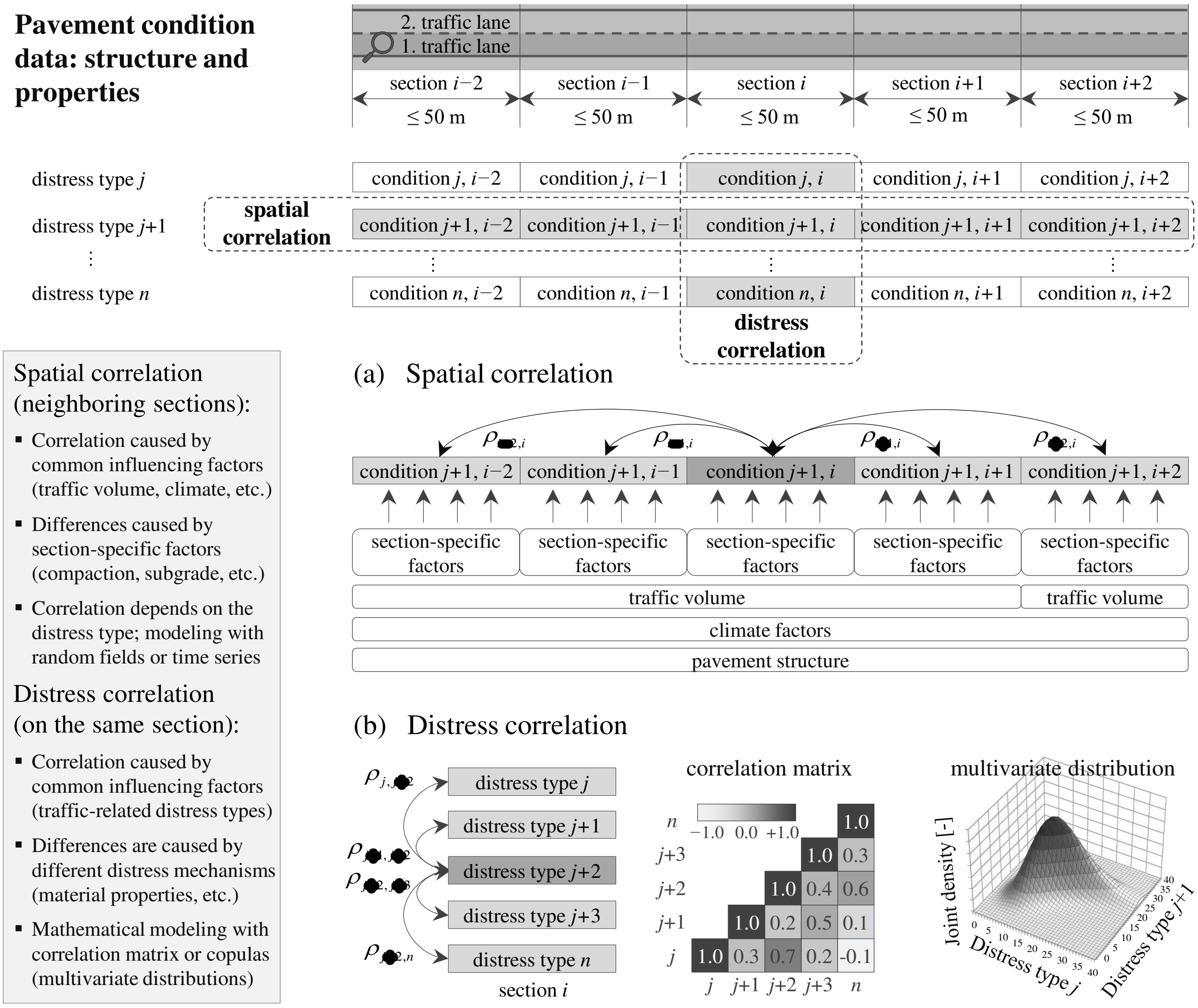
The importance and application area of simulation approaches in pavement management is stressed in Section 1.1.2. This work provides insides into the correlation structure of distress-specific service lives, which can be used not only in simulation studies but also as additional information in condition prediction, especially in the presence of data censoring (see below). The general correlation structure of pavement condition data is given in Figure 1.10. The developed simulation approach simultaneously considers both spatial correlation and distress correlation.

The first type of correlation, the spatial (serial) correlation between neighboring road sections, is illustrated schematically in Figure 1.10(a). This longitudinal spatial correlation considers that the condition on two neighboring sections is more likely to be similar than the condition on two sections that are in larger distance from each other. This work places the emphasis on the correlation between the observed (predicted) service lives instead of the correlation between measured current conditions. Both can be different, for example, if a distress initiation phase is present. This is something that the homogeneous-sections approach overlooks, as it assumes that sections which have similar current condition will have similar condition development (regardless of the age). Furthermore, the degree of spatial correlation significantly affects the length of work zones in the optimization of M&R activities, as discussed later in this section.

In this thesis, the longitudinal spatial correlation is modeled separately for each distress type, using a second-order autoregressive (AR) process. In contrast to time-series models, the observations are spaced at equal length intervals (and not time intervals). More advanced random-fields methods offer a continuous representation of the variability in multiple directions. Possible applications of random fields in the literature include estimation of missing condition values (Lethanh et al. 2016), as well as consideration of spatial variability of material properties and structural responses at the corresponding scales in stochastic finite element and multi-scale models (Lea 2010, Huber 2013, Lea and Harvey 2015, Zhou et al. 2019).

The analysis of empirical data from automated surveys in Austria has shown that the degree of spatial correlation depends on the distress type and the length of survey sections. For example, skid-resistance values correlate highly only with immediate neighbors, while for other distresses like alligator cracking the correlation is significant over a longer distance. The number of significant neighboring condition values (i.e. the order of the AR model) is determined based on the partial autocorrelogram. For the majority of road sections, the correlation is significant over a length of one to four sections (50 m), with two sections being the most common.

The second type of correlation considered in this thesis is the correlation between different distress types on the same road section, as depicted in Figure 1.10(b). Empirical studies suggest that distress types are either uncorrelated or positively correlated (Hajek and Haas 1987, Molzer et al. 2000). From the positive correlation follows that the presence of one distress type on a given section makes the existence of other distresses more likely. Similar to the spatial correlation, the distress correlation can be accounted for in condition prediction. State-of-the-art



Pavement condition data: structure and properties

Spatial correlation
(neighboring sections):

- Correlation caused by common influencing factors (traffic volume, climate, etc.)
- Differences caused by section-specific factors (compaction, subgrade, etc.)
- Correlation depends on the distress type; modeling with random fields or time series

Distress correlation
(on the same section):

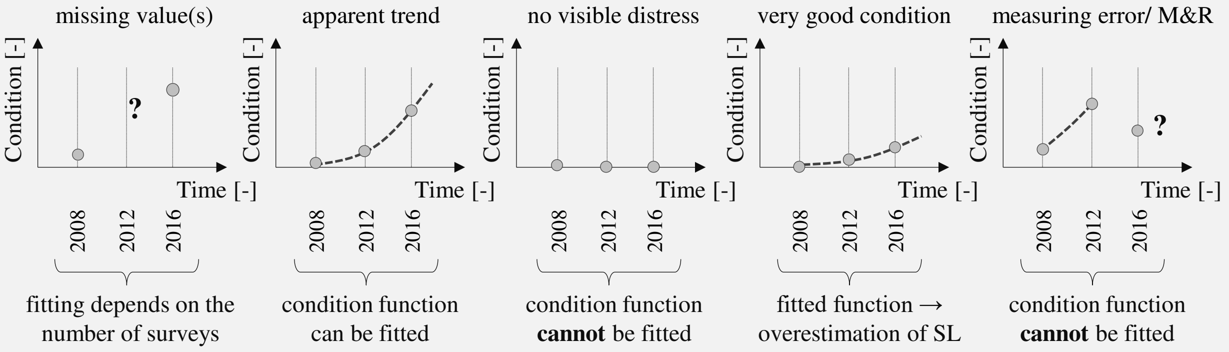
- Correlation caused by common influencing factors (traffic-related distress types)
- Differences are caused by different distress mechanisms (material properties, etc.)
- Mathematical modeling with correlation matrix or copulas (multivariate distributions)

FIGURE 1.10 Correlation structure of pavement condition data: correlation between neighboring road sections for a given distress type (a) and correlation between different distress types at section level (b).

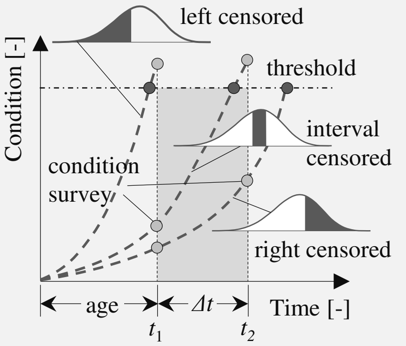
approaches, however, model the progression of the individual distress types as independent processes. Furthermore, distress correlation causes a special type of nonrandom censoring mechanism that leads to biased service-life estimates (see next page). If the marginal service-life distributions for the individual distress types and the correlation structure are known, then the multivariate (joint) distribution can be obtained using the copula approach, as demonstrated in Chapter 3. Usually, only the joint (overall) service life distribution (failure due to any cause) can be estimated using empirical data. If the copula is known, the marginal distributions can be estimated as well, by solving two simultaneous integral equations (see Zheng and Klein 1995). The limitations of the copula approach are related to the estimation of the copula parameter and the increasing complexity when more than two distress types are considered.

The already-mentioned problem of data censoring is a concern when estimating service life based on condition data collected from in-service pavements. Service life is censored when the beginning and/or end of the distress-specific service life are not exactly known. Aside from competing risks, censoring results from the fact that while deterioration is a continuously ongoing process, the condition surveys are conducted at discrete points in time. Moreover, the initial stages of the deterioration process for some distress types (e.g. bottom up-cracking) cannot be observed because they occur below the surface. More specifically, common cases of censor-

(a) Common cases of censoring and data deficiencies



(b) Survival analysis methods at the network level (Chapter 3)

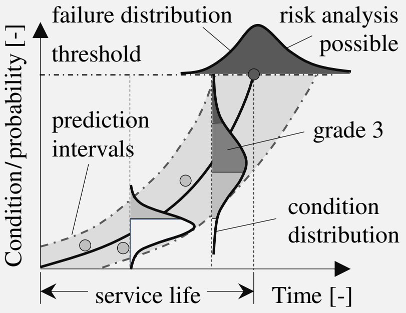


⚠ pavement condition data is always censored (due to periodic surveys) → neglecting censoring leads to biased predictions

⚠ survival analysis allows for a robust unbiased estimation of service life distribution for each distress at the network level

⚠ survival analysis accounts for influencing factors and distress correlation (disadvantage: no continuous deterioration model)

(c) Section-specific prediction based on regression analysis (Chapter 2.2.4 & Chapter 3.3.4)



⚠ regression at the section level considers all available condition surveys in contrast to scaling/ shifting of a master function

⚠ derived failure distribution allows for consideration of risk; additional surveys reduce the uncertainty (value of surveys)

⚠ a combination of survival analysis and section-specific condition model can be beneficial in pavement management

FIGURE 1.11 Common cases of data censoring in practice (a) together and benefits of combining the strengths of survival-analysis models (b) and regression-based curve fitting (c) for condition prediction.

ing which are relevant for pavement management are shown in Figure 1.11(a). For example, sections exhibiting no distresses at the time of the last survey are right censored. Such sections are typically excluded from network-level regression analysis, leading to biased mean service life estimate (Molzer et al. 2000). Missing values and measurement errors (if identified as such) indicate poor data quality and also cause censoring of service life. The case of applying M&R due to another failure cause (i.e. competing risk) was already discussed. Finally, left censoring occurs when a road section has failed prior to the first survey.

Data censoring is a well-known problem in many fields like engineering, biostatistics, sociology, econometrics, etc. Survival analysis (SA) provides the mathematical means to account for censored observations when estimating duration like distress-initiation time or service life (see Figure 1.11(b)). In general, SA methods can be classified into non-parametric (e.g. Kaplan-

Meier), semi-parametric (e.g. Cox regression) and parametric (e.g. Weibull) models (Kleinbaum and Klein 2012). In pavement management literature, SA techniques are considered primarily in the development of crack initiation models. In German literature, the topic of censoring is usually ignored. Furthermore, classical SA assumes random censoring. However, if the censoring mechanism is nonrandom (e.g. due to correlated competing risks), SA models may lead to biased parameter estimates. Another limitation of SA is that it does not incorporate a condition model, describing the states prior to failure. Although multi-state SA models are available, they are mostly based on Markov chains (Hougaard 2000). Nevertheless, SA is capable of capturing performance history as well as unobserved heterogeneity between sections using frailty models (Kleinbaum and Klein 2012). Furthermore, depending on the availability and explanatory power of covariates, section-specific survival curves can be estimated for each individual section. Thus, SA provides an estimate of service life and its distribution at both the network and the section level. The individual distributions can be further narrowed down given that the section has survived until a particular time, allowing for the estimation of remaining life and its distribution.

Pavement condition data has been collected on a regular basis in many countries for decades. If time-series data at the section level is available, a condition function can be fitted using nonlinear regression. In contrast to scaling and shifting of a master function (see Section 1.2.2), regression considers all available observations, is less susceptible to measurement errors and allows for the derivation of confidence and prediction bounds. Moreover, if a power function with constant for all sections power parameter is employed, this reduces the degrees of freedom by one, resulting in more stable prediction. The constant power parameter is based on the assumption that the condition development for each distress type follows a general characteristic form (e.g. degressive for rutting), whereby the deterioration can be accelerated or decelerated depending on section-specific factors (Archilla and Madanat 2000, Hoffmann 2006). Furthermore, this assumption transforms the power function into an intrinsically linear model that can be estimated using simple linear regression (Backhaus et al. 2015). Chapter 2 presents an extension of the common regression model, enabling the analytical derivation of condition and failure distributions at the section level under consideration of the full performance history. The condition distribution is based on the unstandardized student-t density but even if the distribution of the error terms is non-normal, a non-parametric empirical distribution can be estimated using bootstrapping (McCullough 1996). The stochastic model allows for full-stochastic reliability-based life cycle cost optimization of bridges or other critical assets at the single-facility level (see Hoffmann 2018). Chapter 3 shows how the section-specific regression model can be used together with survival analysis to narrow down the prediction intervals for censored sections.

Obtaining an accurate service-life estimate is of crucial importance in every PMS. Even the best optimization algorithm will be of limited use if it uses inaccurate condition predictions as basis for M&R decision planning. As mentioned, the aggregation to homogeneous sections is a huge source of inaccuracy and errors in condition prediction (Section 1.2.2). Still, the aggregation process is seen as indisputable necessity in all PMS applications, while the problem is generally neglected in the research literature. This thesis demonstrates the applicability of a fundamentally different methodology, redefining how condition information is analyzed in a PMS. Instead of losing a great amount of detailed data at the beginning of the analysis prior to any decision making, the new approach utilizes all available information up to the M&R optimization. Thus, the aggregation of short survey sections to work zones takes place at the M&R optimization level on the basis of economies-of-scale cost minimization. The difference between this new approach and the aggregation to homogeneous sections is depicted in Figure 1.12(a,b). The new approach allows for more accurate high-resolution condition predictions, by making use of the quality condition data provided by automated surveys. Incorporating this approach in a methodology where M&R treatments are prescribed based on any possible distress combination has the potential to significantly improve section-level M&R recommendations, as desired by road agencies (Wang and Pyle 2019). To achieve this goal is also essential to include a structural indicator in addition to the commonly surveyed surface characteristics (IRI, skid resistance, etc.). However, the incorporation, for example, of network-level deflection measurements within the homogeneous-sections approach is not meaningful and cannot provide additional benefits.

This thesis introduces a new method for work-zone optimization which eliminates the need for homogenous sections, and provides a holistic framework for a new PMS including, besides the approach, also all necessary cost and performance models. Economies-of-scale models for M&R costs, temporary traffic control costs (TTCC) and user costs were developed using bottom-up cost calculation and unit-cost estimates from contractor’s bid prices in Austria. Economies-of-scale savings result from the spreading of fixed costs over a larger work-zone area, leading to a decrease of total costs per square meter (degressive function). In contrast, common PMSs and the majority of research work employ average cost estimates (e.g. Weninger-Vycudil 2001). Disregarding economies-of-scale effects, however, may lead to an over- or underestimation of costs for any given road project. Alternatively, economies-of-scale cost functions can be derived statistically based on historical cost records for completed projects. The few empirical studies available in the literature show that scale economies are indeed present and their influence on treatment costs is substantial (Irfan et al. 2012, Hoffmann 2018, Qiao et al. 2019).

TTCC are rarely taken into account or only considered in the total average costs. A separate modeling of these costs enables optimization of the work-zone layout and the consideration of more realistic work zones, consisting of multiple M&R treatments. Furthermore, the research work on network-level optimization often neglects work-zone effects on users and models (if at all) only condition-related user costs as a function of a single condition indicator (IRI). Detailed models for user costs are used either for single projects or in work-zone scheduling optimization. The here developed economies-of-scale user cost model is based on parameters and statistics from the literature, standardized work-zone layouts, actual traffic volumes, demand-capacity analysis and an M&R activity duration model.

Another significant contribution of this work are the developed post-treatment performance models, simulating multiple correlated distress types. A simulation approach is relevant for modeling of the condition after the first and all consecutive M&R treatments in the life cycle. This follows from the fact that there is no survey data available for the future which can be used for empirical calibration of condition functions. In practice, treatment performance is often described using deterministic models with age as a primary explanatory variable. Due to this approach, all road sections with a specific treatment in a given year will have identical future condition progression, resulting in distorted condition and annual costs predictions. Furthermore, covariates like traffic, thickness, and freeze index are similar for a large number of contiguous survey sections, falling short in explaining the true spatial variability of exhibited distresses across sections. Alternatively, the simulation approach in this thesis considers that the distress and spatial correlation structure of the data will not change significantly after a treatment. Moreover, initial service life and post-treatment life are correlated as well, and the flexible approach allows for explanatory variables to be included in the distributions of the individual distresses.

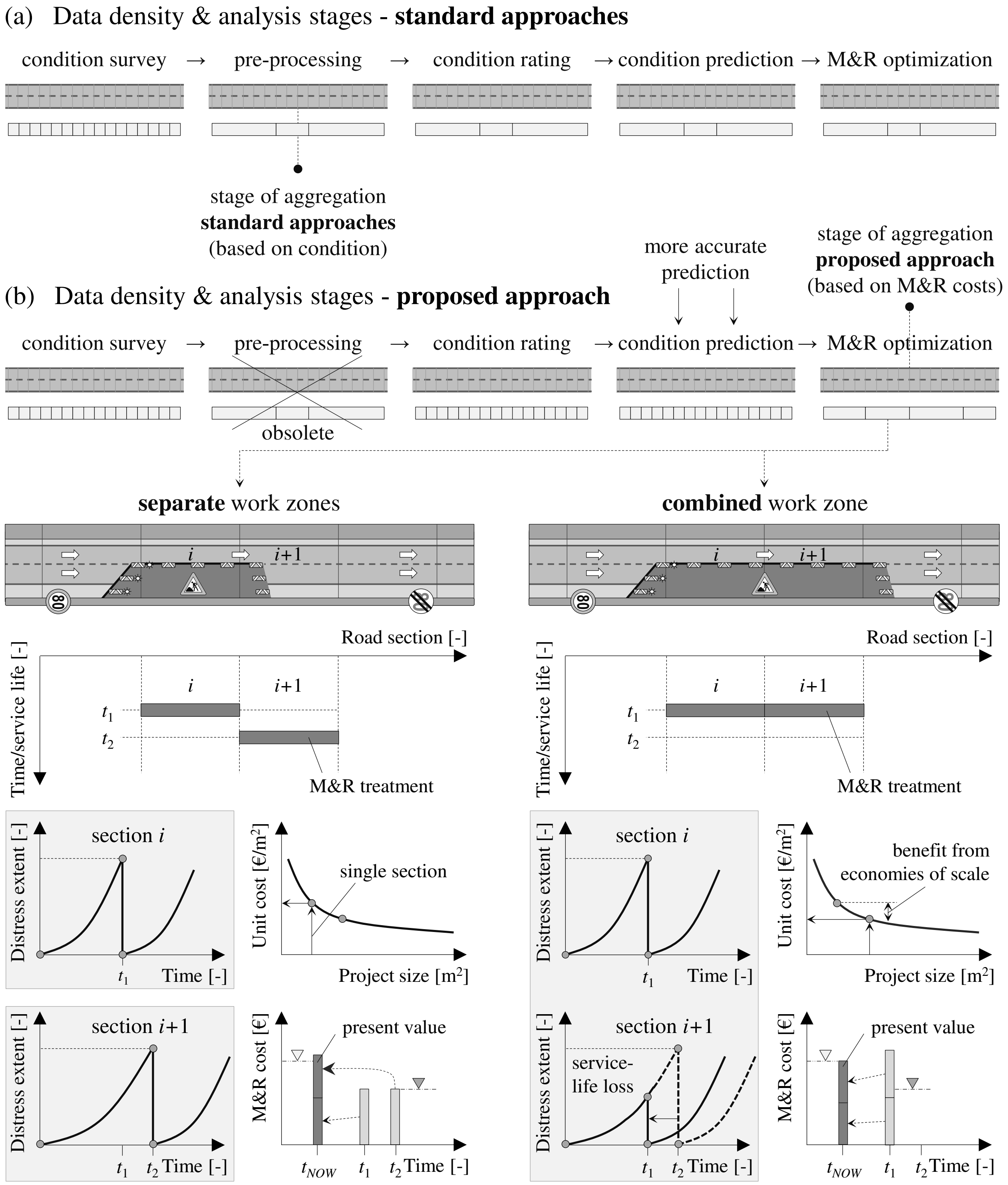


FIGURE 1.12 Aggregation of survey data in common PMSs prior to the condition assessment level (a) vs. proposed combination of short survey sections into work zones at the M&R optimization level (b).

The developed holistic M&R optimization approach simultaneously considers multiple distress types as well as the effect of each treatment on every possible distress combination. The complexity of the optimization problem is not affected by the number of considered distress types. In contrast, many researchers consider just one distress type or an aggregated condition index in M&R optimization. Some examples include international roughness index (IRI) (Sathaye and Madanat 2012, Lee and Madanat 2015), critical condition index (CCI) (Wu and Flintsch 2009, Santos et al. 2017), pavement condition index (PCI) (Li et al. 1998, Mathew and

Isaac 2014), pavement serviceability rating (PSR) (Yeo et al. 2013), pavement condition score (PCS) (Wang et al. 2003), and so on. Roughness is perhaps the most common condition indicator used in optimization due to the fact that there are many roughness-based performance and user costs models available in the literature. However, considering all observed distress types is indispensable for more targeted treatment assignment, allowing the selection of technically and economically most appropriate sequence of treatments for the particular distress combination.

The methodology for work-zone optimization based on scale economies and considering multiple distress types is illustrated using a simplified example in Figure 1.12(b). M&R treatments on two neighboring sections can be combined in a joint work zone, by moving forward the treatment on the section with the longer service life (*i*+1) to the timing of the other section (*i*), provided that the type of treatment is the same. Postponing the treatment on section *i* is not acceptable with regard to the defined condition thresholds. The two sections should be combined if the discounted cost savings due to scale economies are higher than the increase of present value resulting from shifted timing on section *i*+1. The initial M&R solutions are the result of single-section optimization (first stage) but the treatment type can still be changed in the work-zone optimization (second stage) if the condition and technical constraints are satisfied. Considering multiple treatments and a larger number of road sections leads to a complex optimization problem, namely a nonlinear, nonsmooth, nonconvex, general integer optimization problem.

Within this thesis, an innovate algorithm was developed specifically for the given optimization problem, using Visual Basic for Applications programming language. The algorithm called EMS (expand, merge and solve) consist of multiple steps and iterations, combining different heuristic techniques. The solving part is handled by a commercial add-in for Excel, employing steady-state genetic algorithms. The applicability of the algorithm is demonstrated using a realistic case study consisting of 1000 freeway survey sections. A visual inspection of the results showed excellent solution quality in almost all cases. The case study produces extensive results including work-zone solutions resulting from minimizing M&R agency costs, total agency costs (incl. TTCC), societal costs (agency and user costs) as well as a sensitivity analysis for variations in the discount rate. Furthermore, a hypothetical example shows that the initial service life which is influenced by pavement design has a major effect on the optimal M&R program.

An important issue discussed in this work is the dependence of service life on the definition of trigger and threshold values. A possible way to mitigate the influence of subjective thresholds is using soft constraints for some distress types like alligator cracking, allowing for a certain exceedance of thresholds depending on a penalty function. Nevertheless, hard constraints will still be needed for some distresses like rutting due to safety-related reasons. Soft constraints grant more flexibility in the optimization by expanding the solution space. The application of this approach to the case study shows that using soft constraints leads to greater work-zone lengths. Thus, the treatments on some sections are postponed, if this leads to longer work zones. The drawback of using soft constraints is that the estimation of penalty costs is not straightforward.

The work-zones lengths obtained in the case study are too short compared to typical lengths known from practice. The reason for this is that the approach does not yet account for some factors that may considerably increase work-zone lengths in a fully developed approach. Such factors include the consideration of aggregated work zones consisting of multiple different M&R activities, aggregation of two work zones with a small gap in between as well as the costs for other (in-house) activities with a high proportion of fixed costs (e.g. supervision and planning, safety audits, call-for-tender procedures, quality assurance and acceptance testing).

1.2.4 Contribution of the author

This thesis constitutes a cumulative work, consisting of an introductory chapter and five scientific articles that were published in international peer-reviewed journals. The individual contributions of the author to each publication are summarized as follows:

- **Publication 1** (Chapter 2): *Introduction of a new continuous time and state space stochastic process in condition prediction*

The author prepared the second half of the manuscript, conducted the literature analysis, prepared all figures, analyzed the multicollinearity issues and had a significant contribution to the derivation of the project-level regression model. The author conducted the computations for the examples, the simulation of correlated failure types and the case study based on LTPP data.

- **Publication 2** (Chapter 3): *Condition prediction and estimation of service life in the presence of data censoring and dependent competing risks*

The author prepared the manuscript, conducted the literature review, created all figures, conceptualized the simulation case study and generated correlated condition data, exhibiting different types of censoring. The author implemented the analyzed methods in Excel, applied them to the generated data, interpreted the results and proposed the copula approach.

- **Publication 3** (Chapter 4): *Optimization of pavement maintenance and rehabilitation activities, timing and work zones for short survey sections and multiple distress types*

The author prepared the manuscript, analyzed relevant literature and created all figures. The author implemented and substantially contributed to the methodological development of a holistic M&R modeling and work-zone optimization framework, building on the idea of Markus Hoffmann for bundling of short survey sections based on economies-of-scale costs. The author developed a simulation approach, simultaneously considering distress and spatial correlation, developed agency, temporary traffic control and user cost models as well as a treatment duration model. The author implemented an M&R performance and cost modeling of multiple distress types and multiple treatments types for a large number of road sections in Excel. One of the main contributions of the author lies in the design of the solution algorithm for the highly complex work-zone optimization problem using Visual Basic for Applications. Furthermore, the author conducted all computations, optimization runs, sensitivity analysis and evaluated the results.

- **Publication 4** (Chapter 5): *Aggregation of condition survey data in pavement management: shortcomings of a homogeneous sections approach and how to avoid them*

The author prepared the manuscript, created all figures, analyzed sectioning algorithms from the literature and implemented them in Excel. The author developed a framework for comparison, selected boundary conditions, applied the algorithms to generated data and analyzed the results.

- **Publication 5** (Chapter 6): *Benefit maximization based on aggregated condition indices: drawbacks for selection of pavement treatments*

The author prepared the manuscript, conducted the literature analysis, created all figures and conceptualized the framework for comparison by defining boundary conditions and scenarios. He implemented the analyzed methods in Excel, wrote the algorithms for the generation of strategies in Visual Basic, conducted all optimizations and formulated key drawbacks of common methods.

1.3 CONCLUSIONS AND OUTLOOK

1.3.1 Conclusions

The conclusions and contributions of this work can be summarized according to the individual publications. **Publication 1** (Chapter 2) shows that common condition survey and assessment approaches differ in the number and type of surveyed distress types, measurement equipment, survey-section lengths, road segmentation algorithms, inspection intervals, condition thresholds and definition of aggregated condition indices. A major limitation resulting in loss of service life is the use of subjective condition thresholds which are not based, for example, on a comprehensive analysis and safety considerations (e.g. braking distance, curve skidding). Also, the use of aggregated indices in M&R planning and optimization should be strongly discouraged, as entirely different distress types may lead to the same overall condition index, blurring underlying failure causes. Furthermore, common deterministic prediction models based on shifting and/or scaling of a master function are susceptible to measurement errors and ignore previous performance history. In addition, the estimated model parameters may suffer from bias resulting from multicollinearity effects, model misspecification, unobserved heterogeneity and ignoring data censoring. Another finding of the analysis is that homogeneous Markov chains are not appropriate for pavement condition modeling on account of distortions due to absorbing states, limited use at the section level, the dependence of the results on the number of discrete states, as well as unrealistic assumptions like the memoryless property and constant transition probabilities (see also Chapter 3). In contrast, the introduced continuous-time and continuous-state stochastic models both at the network level and the section level do not exhibit these limitations, allowing for the estimation of remaining service life, reliability, risk and prediction intervals.

The nature of pavement deterioration process and practical limits of condition survey procedures lead to censoring of empirical data. Although some researchers have used classical survival analysis to account for censoring, only the case of just one possible distress type was considered. In **Publication 2** (Chapter 3), different survival analysis models are applied for service life estimation in the presence of uncorrelated andf correlated competing distress types. The deviations of different methods are objectively quantified based on an artificially censored data generated using a known deterioration process. The results show that classical survival analysis models exhibit substantial bias in the presence of multiple positively or negatively correlated distress types. However, this bias can be reduced or completely avoided by using the complete performance history at the section level. Another promising approach is based on copulas, enabling a simultaneous modeling of joint and marginal (distress-specific) service life distributions.

Publication 3 (Chapter 4) presents a novel optimization approach which is unique in its capabilities to simultaneously account for multiple treatments, multiple distress types, stochastic treatment lives, scale economies, traffic control, work-zone layouts, capacity, road user and environmental impacts. Instead of forming long homogeneous sections, the proposed approach is based on short survey sections, allowing more accurate condition predictions for each distress type. Survey sections are then combined into longer work zones if the economies-of-scale benefits exceed the resulting loss of service life. In addition, work-zone related user costs are considered in order to select appropriate work-zone layouts. As a part of a holistic end-to-end framework, all necessary cost and performance models are developed and calibrated for flexible pavements on freeways in Austria. The formulated optimization problem exhibits exponential complexity with network length and is solved by a developed problem-specific algorithm.

The proposed methodology is successfully applied to a case study of a realistic road network consisting of 1000 pavement sections. According to the results, temporary traffic control costs are the primary driver for the formation of longer work zones. The use of soft constraints increases the flexibility of the approach, reducing the dependence of the results on condition thresholds and outliers. Furthermore, the conducted sensitivity analysis shows that discount rates have a significant impact on the length of work zones and the type of M&R treatment. Higher discount rates, for example, attach more weight to short-term decisions and favor less costly treatments, despite shorter service lives. The consideration of additional user costs due to work zones leads to larger work zones with a shorter duration. Finally, the designed structural life has a huge impact on the extent of subsequent treatments and life cycle costs.

The last two publications analyze the consequences of aggregation to homogeneous sections and benefit maximization strategies in pavement management. In particular, **Publication 4** (Chapter 5) examines the consequences of the aggregation of condition data to homogeneous sections. Three common methods for road segmentation are compared based on empirical case study. The results show that the prediction errors and smoothing of peak values arising from the aggregation are not negligible. Moreover, the produced homogeneous sections shift with each survey which hinders the use of performance history in condition prediction at the section level. Furthermore, the work zones resulting from homogeneous sections (condition based) are compared to the work zones resulting from the advanced optimization approach presented in Chapter 4 (M&R based). Under the same boundary conditions of no threshold violations, the new optimization approach outperforms homogenous sections with different level of aggregation (200 m, 600 m and 1000 m) in terms of loss of service life and total discounted costs.

In **Publication 5** (Chapter 6), strategies maximizing benefits and strategies minimizing costs are compared for different budget scenarios. According to the common practice, benefit is defined as the area between the do-nothing and post-treatment performance curves of an aggregated condition index. The results show that the maximization of benefits is not an appropriate objective in PMS, as it leads to almost exclusive preference of extensive (and expensive) M&R treatments with timing as early as possible, irrespective of actual service life and failure causes. Thus, in the provided example, benefits maximization leads to 62% higher costs in comparison to the cost-minimization solution for unlimited budget. For limited budgets this percentage varies between 4% and 37%. For road agencies with a tight budget, this approach therefore will lead to an inefficient use of funds, violations of condition thresholds and M&R backlogs. Thus, the objective of maximum benefits is misleading, as these benefits are not related to actual monetary or non-monetary benefits resulting from following a specific strategy.

The conclusions of this thesis contribute to both theory and practice of pavement management. In the context of aging infrastructure, increasing maintenance needs and limited public budgets, a better budget allocation and/or substantial savings can be achieved by minimizing agency (and user) costs instead of maximizing the area between curves. Applying the proposed performance models and the advanced optimization approach will improve the quality of condition prediction and the quality of project-level M&R treatment recommendations by making better use of the collected condition survey data. This will lead to considerable efficiency gains regarding availability of the road network as well as an improvement of the average network performance and/or substantial cost savings compared to current approaches.

1.3.2 Key research contributions

Based on the conclusions, the main research contributions of thesis can be summarized in the order of the publications as follows:

- Identification and demonstration of the main differences and shortcomings in common approaches for condition survey, assessment and prediction using empirical examples
- Derivation of failure and condition distributions for a stochastic condition model at the section level with comparison to scaling/shifting of a master function
- Multiple examples showing the many drawbacks of homogenous discrete-time Markov chain models for pavement condition prediction at the network level
- Demonstration of the limitations of regression analysis and classical survival analysis regarding the estimation of unbiased service life for correlated competing distress types
- Proposal of two approaches for competing-risks modeling: multivariate service life distributions (copula approach) and a combination of survival analysis and a condition model
- Development of general approach for simulation of realistic pavement condition data simultaneously exhibiting distress and spatial correlation for various applications
- Calibration of economies-of-scale cost models for M&R construction, temporary traffic control and road users based on literature and empirical data using bottom-up approach
- Elaboration and implementation of M&R treatment modeling methodology with specific impact of any given treatment on multiple distress types according to a decision tree
- Development of an advanced work-zone optimization approach eliminating the need for data aggregation and homogeneous sections based on economies-of-scale costs
- Design and implementation of an effective and efficient solution algorithm for the formulated (complex) work-zone optimization problem
- Systematic investigation and quantification of the impact of aggregation towards homogeneous sections on service life prediction and M&R costs based on parametric case study
- Extensive simulation results clearly showing that the maximization of benefits based on aggregated condition indices is not an appropriate optimization objective in PMSs

1.3.3 Future research

One of the challenges for future research will be the integration of even more detailed 3D road condition data based on laser scanning in the decision-making process. Regarding condition prediction, the developed empirical models should be combined with physical models considering mechanistic principles and laboratory testing. Combined methods will not only improve the quality of predictions but provide valuable knowledge about the impact of the various factors influencing deterioration. This will open the door for further optimization of design concepts and evaluation of new materials. Moreover, given the huge impact on the optimization results, a systematic analysis of the effect of the different M&R treatments on service life is necessary.

Research efforts will be also dedicated to the improvement and extension of the developed advanced optimization approach. The flexibility of the approach should be enhanced by including practical possibilities like the aggregation of work zones with a small gap in between and work zones consisting of different M&R treatments sharing traffic control. A topological model of the road network together with a traffic simulation model are required for a more realistic assessment of user costs with consideration of all traffic lanes, possible detours and redistribution of traffic. Furthermore, an improvement of the solution algorithm in terms of running time will be achieved by using parallel genetic algorithms and two-dimensional encoding of candidate solutions, which will be a pioneering feat in the field of pavement engineering. The next step is to extend the approach to a network-level cross-asset optimization considering other road assets like bridges and tunnels as well as different budget scenarios under risk of failure. At the network level, the reduction of greenhouse gas emissions should be incorporated in the optimization problem as well. As a final step towards practical implementation, the applicability of this stochastic cross-asset optimization approach has to be demonstrated using real-world data.

REFERENCES

AASHTO, 1993. *AASHTO guide for design of pavement structures*. III-19–III-24, and Appendix J. Washington, D.C.: American Association of State Highway Officials.

AASHTO, 2012. *Pavement management guide*. 2nd ed. Washington, DC: American Association of State Highway and Transportation Officials.

Abaza, K.A., Ashur, S.A. and Al-Khatib, I.A., 2004. Integrated pavement management system with a Markovian prediction model. *Journal of Transportation Engineering*, 130 (1), 24-33.

Alba, E., Luque, G. and Nesmachnow, S., 2013. Parallel metaheuristics: recent advances and new trends. *International Transactions in Operational Research*, 20 (1), 1-48.

Archilla, A.R. and Madanat, S., 2000. Development of a pavement rutting model from experimental data. *Journal of Transportation Engineering*, 126 (4), 291-299.

Backhaus, K., Erichson, B. and Weiber, R., 2015. *Fortgeschrittene multivariate Analysemethoden: Eine anwendungsorientierte Einführung* [Advanced multivariate analysis methods: An application-oriented introduction]. 3rd ed, Berlin: Springer (in German).

Black, F. and Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81 (3), 637-654.

Benson, K.R., Elkins, G.E., Uddin, W. and Hudson, W.R., 1988. Comparison of methods and equipment to conduct pavement distress surveys. *Transportation Research Record: Journal of the Transportation Research Board*, 1196, 40-50.

Butt, A.A., Shahin, M.Y., Feighan, K.J. and Carpenter, S.H., 1987. Pavement performance prediction model using the Markov process. *Transportation Research Record: Journal of the Transportation Research Board*, 1123, 12-19.

Carey, W.N., Jr. and Irick, P.E., 1960. The pavement serviceability-performance concept. *Highway Research Board Bulletin*, 250, 40-58.

Carrière, J.F. and Chan, L.K., 1986. The bounds of bivariate distributions that limit the value of last-survivor annuities. *Transactions of Society of Actuaries,* 1986, 38, 51-74.

Cox, D.R., 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34 (2), 187-220.

Crowder, H., Johnson, E.L. and Padberg, M., 1983. Solving large-scale zero-one linear programming problems. *Operations Research*, 31 (5), 803-834.

Donev, V., 2014. *Erhaltungsmanagement Straßenoberbau - Zustandsbewertung und Zustandsprognose* [Pavement management - condition assessment and condition prediction]. Wien: Mitteilungen der Professur für Straßen- und Flugbetriebsflächenbau (in German).

FHWA, 2015. *The long-term pavement performance program*. Report FHWA-HRT-15-049. Washington, DC: Federal Highway Administration.

Fwa, T.F., Chan, W.T. and Tan, C.Y., 1996. Genetic-algorithm programming of road maintenance and rehabilitation. *Journal of Transportation Engineering*, 122 (3), 246-253.

Golabi, K., Kulkarni, R.B. and Way, G.B., 1982. A statewide pavement management system. *Interfaces*, 12 (6), 5-21.

Haas, R. and Hudson, W., 1978. *Pavement management systems*. New York, NY: McGraw-Hill.

Haas, R., Hudson, W.R. and Falls, L.C., 2015. *Pavement asset management*. Salem, MA: Scrivener.

Haas, R.C.G. and Hutchinson, B.G., 1970. A management system for highway pavements. *Proceedings of the 5th Australian Road Research Board (ARRB) Conference*, Canberra, 124-160.

Hajek, J.J. and Haas, R.C.G., 1987. Factor analysis of pavement distresses for surface condition predictions. *Transportation Research Record: Journal of the Transportation Research*, 1117, 125-133.

Hoffmann, M., 2006. *Instandsetzung von Strassen* [Rehabilitation of roads]. Dissertation (PhD). Technische Universität Graz (in German).

Hoffmann, M., 2018. *Lebenszykluskosten der Straßeninfrastruktur: Optimierung von Investitionsstrategien und technischen Maßnahmen, Bau- und Betriebsweisen für Straßenanlagen in ihrem Lebenszyklus* [Life cycle costs of road infrastructures]. Habilitation thesis. Technische Universität Wien (in German).

Hosking, J.R. and Woodford, G.C., 1976. *Measurement of skid resistance part 1. Guide to the use of SCRIM*. TRRL Laboratory Report No. 737. Crowthorne: Transport and Road Research Laboratory.

Hougaard, P., 2000. *Analysis of multivariate survival data*. 1st ed, New York, NY: Springer.

Huber, M., 2013. *Soil variability and its consequences in geotechnical engineering*. Dissertation (PhD). University of Stuttgart.

Intel, 2012. *Intel Chips timeline* [online]. Intel Corporation. Available from https://www.intel.com/content/www/us/en/history/history-intel-chips-timeline-poster.html [Accessed 05 Jan 2020].

Intel, 2018. *Export compliance metrics for Intel microprocessors* [online]. Available from https://www.intel.com/content/www/us/en/support/articles/000005755/processors.html [Accessed 05 Jan 2020].

Irfan, M., Khurshid, M.B., Ahmed, A. and Labi, S., 2012. Scale and condition economies in asset preservation cost functions: case study involving flexible pavement treatments. *Journal of Transportation Engineering*, 138 (2), 218-228.

Jiang, Y., Saito, M. and Sinha, K.C., 1988. Bridge performance prediction model using the Markov chain. *Transportation Research Record: Journal of the Transportation Research Board*, 1180, 25–32.

Kerali, H.G.R., Odoki, J.B. and Stannard, E.E., 2006. *Overview of HDM-4*. The Highway Development and Management Series, Volume 1. Paris, France: World Road Association (PIARC).

Kim, S. and Frangopol, D.M., 2017. Efficient multi-objective optimization of probabilistic service life management. *Structure and Infrastructure Engineering*, 13 (1), 147-159.

Kleinbaum, D.G. and Klein, M., 2012. *Survival analysis: a self-learning text*. 3rd ed. New York, NY: Springer.

Knepper, S., 2001. Development of a pavement management system in Germany. *5th International Conference on Managing Pavements*, 11-14 August, Seattle, Washington, USA.

Lea, J.D., 2010. *The effect of spatial variability on the reliability of pavements*. Dissertation (PhD). University of California, Davis.

Lea, J.D. and Harvey, J.T., 2015. Using spatial statistics to characterise pavement properties. *International Journal of Pavement Engineering*, 16 (3), 239-255.

Lee, J. and Madanat, S., 2015. A joint bottom-up solution methodology for system-level pavement rehabilitation and reconstruction. *Transportation Research Part B*, 78, 106-122.

Lethanh, N. and Adey, B.T., 2013. Use of exponential hidden Markov models for modelling pavement deterioration. *International Journal of Pavement Engineering*, 14 (7), 645-654.

Lethanh, N., Richmond, C. and Adey, B.T., 2016. Investigation of the ability to estimate values of road section condition indicators based on their spatial correlation, *Journal of Infrastructure Systems*, 22 (3), 04016006.

Li, N., Xie, W.-C. and Haas, R., 1996. Reliability-based processing of Markov chains for modeling pavement network deterioration. *Transportation Research Record: Journal of the Transportation Research Board*, 1524, 203-213.

Li, N., Haas, R., and Huot, M., 1998. Integer programming of maintenance and rehabilitation treatments for pavement networks. *Transportation Research Record: Journal of the Transportation Research Board*, 1629, 242-248.

Madanat. S. and Ben-Akiva, M., 1994. Optimal inspection and repair policies for infrastructure facilities. *Transportation Science*, 28 (1), 55-62.

Madanat, S., Nakat, Z. and Jin, E.-J., 2010. Empirical modeling of pavement overlay crack progression with field data. *Journal of Infrastructure Systems*, 16(4), 292-298.

Markowitz, H., 1952. Portfolio Selection. *Journal of Finance*, 7 (1), 77–91.

Mathew, B.S. and Isaac, K.P., 2014. Optimisation of maintenance strategy for rural road network using genetic algorithm. *International Journal of Pavement Engineering*, 15 (4), 352-360.

Medury A. and Madanat S., 2011. System-level optimization of maintenance and replacement decisions for road networks. *In:* M.H. Faber, J. Köhler and K. Nishijima, eds. *Applications of Statistics and Probability in Civil Engineering*, 1-4 August 2011, Zürich. London: Taylor & Francis, 234-242.

Metropolis, N. and Ulam, S., 1949. The Monte Carlo method. *Journal of the American Statistical Association*, 44 (247), 335-341.

McCullough, B.D., 1996. Consistent forecast intervals when the forecast-period exogenous variables are stochastic. *Journal of Forecasting*, 15 (4), 293-304.

Mishalani, R.G. and Madanat, S.M., 2002. Computation of infrastructure transition probabilities using stochastic duration model. *Journal of Infrastructure Systems*, 8 (4), 139-148.

Moavenzadeh, F., Stafford J.H., Suhbrier J.H. and Alexander J., 1971. *Highway design study phase I: the model*. IBRD Economics Department Working Paper No 96. Washington DC: International Bank for Reconstruction and Development.

Molzer, C., Felsenstein, K., Viertl, R., Litzka, J. and Vycudil, A., 2000. *Statistische Methoden zur Auswertung von Straßenzustandsdaten* [Statistics for pavement performance modeling]. Wien: BMVIT, Straßenforschung Heft 499 (in German).

Nelson, T.L. and LeClerc, R.V., 1982. *Development and implementation of Washington State's pavement management system.* Report No. WA-RD 50.1. Olympia, WA: Washington State Department of Transportation.

Page, L., Brin, S., Motwani, R. and Winograd, T., 1998. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford, CA: Stanford University.

Paterson, W.D.O., 1987. *Road deterioration and maintenance effects: models for planning and management*. The Highway Design and Maintenance Standard Series. Baltimore, MD: Johns Hopkins University Press.

Patidar, V., Labi, S., Morin, T., Thompson, P.D. and Sinha, K.C., 2011. Evaluating methods and algorithms for multicriteria bridge management at the network level. *Transportation Research Record: Journal of the Transportation Research Board*, 2220, 38-47.

Perera, R., Pulipaka, A.S. and Kohn, S.D., 2008. *Pavement management catalog*. Washington, DC: Federal Highway Administration.

Qiao, Y., Labi, S., Fricker, J. and Sinha, K.C., 2019. Costs and effectiveness of standard treatments applied to flexible and rigid pavements: case study in Indiana, USA. *Infrastructure Asset Management*, 6 (1), 15-29.

Rakshit, A., Krishnamurthy, N. and Yu, G., 1996. System Operations Advisor: A real-time decision support system for managing airline operations at United Airlines. *Interfaces*, 26 (2), 50-58.

Santos, J., Ferreira, A. and Flintsch, G., 2017. An adaptive hybrid genetic algorithm for pavement management. *International Journal of Pavement Engineering.* doi: 10.1080/10298436.2017.1293260.

Santos, J., Ferreira, A., Flintsch, G. and Cerezo, V., 2018. A multi-objective optimisation approach for sustainable pavement management. *Structure and Infrastructure Engineering*, 14 (7), 854-868.

Sathaye, N. and Madanat, S., 2012. A bottom-up optimal pavement resurfacing solution approach for large-scale networks. *Transportation Research Part B*, 46 (4), 520-528.

Schmuck, A., 1987. *Straßenerhaltung mit System: Grundlagen des Managements* [Pavement management with a system: fundamentals of management]. Bonn: Kirschbaum.

Shahin, M.Y. and Kohn S.D., 1981. *Pavement maintenance management for roads and parking lots*. Technical Report No. M-294. U. S. Army Construction Engineering Research Laboratory.

Stallings, W., 2016. *Computer organization and architecture: designing for performance*. 10th ed. Hoboken, NJ: Pearson.

Strohmaier, E., Dongarra, J., Simon, H., Meuer, M. and Meuer, H., 1993-2018. *TOP500 Supercomputer Sites* [online]. Available from https://www.top500.org/statistics/sublist/ [Accessed 05 Jan 2020].

Thomas, F., 2004. Generating homogeneous road sections based on surface measurements: available methods. *2nd European Pavement and Asset Management Conference*, Berlin, Germany.

Thompson, P.D., Small, E.P., Johnson, M. and Marshall, A.R., 1998. The Pontis bridge management system. *Structural Engineering International*, 8 (4), 303-308.

Tsuda, Y., Kaito, K., Aoki, K. and Kobayashi, K., 2006. Estimating Markovian transition probabilities for bridge deterioration forecasting. *Journal of Structural Engineering and Earthquake Engineering*, 23, 241-256.

Tsutsui, S. and Collet, P., eds., 2013. *Massively Parallel Evolutionary Computation on GPGPUs*. Berlin: Springer.

Wang, F., Zhang, Z. and Machemehl, R.B., 2003. Decision-making problem for managing pavement maintenance and rehabilitation projects. *Transportation Research Record: Journal of the Transportation Research Board*, 1853, 21-27.

Wang, Z. and Pyle, T., 2019. Implementing a pavement management system: The Caltrans experience. *International Journal of Transportation Science and Technology*, 8 (3), 251-262.

Watanatada, T., Harral, C.G., Paterson, W.D.O., Dhareshwar, A.M., Bhandari, A. and Tsunokawa, K., 1987. *The highway design and maintenance standards model volume 1: Description of the HDM-III model*. The Highway Design and Maintenance Standard Series. Baltimore, MD: Johns Hopkins University Press.

Weninger-Vycudil, A., 2001. *Entwicklung von Systemelementen für ein österreichisches Pavement Management System* [Development of system elements for the Austrian pavement management system]. Dissertation (PhD). Technische Universität Wien (in German).

Wu, Z. and Flintsch, G.W., 2009. Pavement preservation optimization considering multiple objectives and budget variability. *Journal of Transportation Engineering*, 135 (5), 305-315.

Yeo, H., Yoon, Y. and Madanat, S., 2013. Algorithms for bottom-up maintenance optimisation for heterogeneous infrastructure systems. *Structure and Infrastructure Engineering*, 9 (4), 317-328.

Zavitski, J.L., Rose, R.T., and Kuhl, G., 2008. Integrating pavement management into a comprehensive strategic asset management system for the state of Utah Department of Transportation. *7th International Conference on Managing Pavement Assets*, 23-28 June, Calgary, Alberta, Canada.

Zhou, H., Li, J. and Spencer, B.F., Jr., 2019. Multiscale random fields-based damage modeling and analysis of concrete structures. *Journal of Engineering Mechanics*, 145 (7), 04019045.

Zheng, M. and Klein, J.P., 1995. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82 (1), 127-138.

CHAPTER 2

Introduction of new continuous time and state space stochastic process in condition prediction

by Markus Hoffmann & Valentin Donev

in *International Journal of Pavement Engineering*
Published online: 25 Mar 2016
Published in print: Volume 19, Issue 4, 2018
DOI: 10.1080/10298436.2016.1162304

ABSTRACT

Periodic condition assessments of pavements together with condition predictions are the basis for investment decisions in every pavement management system (PMS). Typical approaches include surveys of distress types every 3–6 years with condition rating and calculation of aggregated indices for road safety and/or structural health. Furthermore, advanced PMS prediction models allow a comparison of maintenance alternatives and an optimization of investment strategies. This paper presents an overview of current survey and rating approaches in Germany, Switzerland and Austria, together with an impact analysis of different methods, utilized deterministic performance functions and condition threshold (trigger) values for all major distress types. The core of this paper is a comparison of common deterministic condition prediction models with discrete stochastic approaches (Markov chains), prediction models based on advanced regression techniques, and an innovative stochastic continuous-time and continuous-state-space process (Hoffmann process). All prediction models are applied to real-world data from condition surveys in Austria and the long-term pavement performance database (USA) for single sections and at the network level. The paper provides evidence why deterministic prediction approaches are leading to substantial bias in condition distribution and remaining service life as they do not account for the stochastic nature of pavements. Classic Markov-chain approaches do not account for censoring of survey data and neglect changes in transition probabilities with increasing age. Applying common bivariate and multiple regression techniques may also lead to certain bias due to collinearity effects and specification bias. The paper provides mathematical evidence on ways to avoid these shortcomings based on the presented innovative stochastic process, leading to a higher reliability in condition assessment and an improved accuracy of condition predictions. The aspects of censoring, distress specific assignment and optimization of treatments with this new Hoffmann process will be covered in forthcoming papers.

KEYWORDS: pavement management, distress survey, prediction models, condition assessment procedures, service life.

2.1 PAVEMENT CONDITION ASSESSMENT

2.1.1 Condition surveying and evaluation

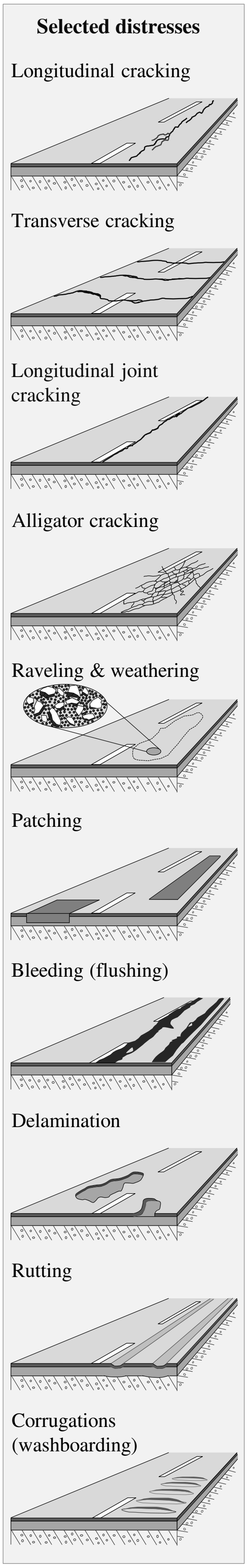
The combined effects of traffic and weather conditions lead to a deterioration of pavement condition over time, manifesting in various forms of distresses. The choice of material type, layer structure and thickness, as well as construction quality together with other local observed and unobserved factors have a major impact on deterioration rates and distress types. For any systematic manual or automated pavement condition survey all relevant distress types must be defined and categorized. Typical categories depend on the underlying deterioration process, impact on structural service life or road user safety and comfort, as well as related maintenance and rehabilitation (M&R) actions. Major distress categories for flexible pavements include cracking, surface defects, rutting and roughness as well as skid resistance and bearing capacity. Rigid pavements exhibit additional types of distress like corner breaks, shoulder drop-off, pop outs, pumping, cracks and blow-ups at joints, but are less susceptible to rutting and alligator cracking (Mallick and El-Korchi 2013, Haas et al. 2015).

Surveys of pavement condition are the basis for any investment decision and are related to remaining service life, asset value and road user costs. In German-speaking countries periodic routine condition surveys are conducted every 3–6 years on the entire road network. More detailed surveys and material testing are considered only for specific sections prior to possible M&R treatments and/or after implementation of selected treatments. Periodic condition surveys are used for calculating budget scenarios, development of treatment strategies and setting priorities. Detailed surveys at the section level are the basis for specific planning of treatments, comparison of maintenance alternatives and quality assessment after implementation.

Switzerland has the highest number of surveyed distress types with visual assessment of severity and extent, while measurements are only performed in detailed surveys or research projects. In Germany automated distress surveys are mainly based on measured failure extent and a high level of aggregation with more generalized definitions of failure types. In Austria distress type, extent and severity are determined using both visual survey (video) and measuring equipment (rutting, skid resistance, etc.). Figure 2.1 provides an overview of routine pavement survey approaches in Austria, Germany and Switzerland with recorded distress types (SN 640 925b 2003, SN 640 925 b Anhang 2003, ZTV ZEB-StB 2006, Weninger-Vycudil et al. 2009).

Although layer structure, design procedures and distress types are quite similar, the analyzed periodic survey approaches differ in defined units for severity and extent, data collection techniques, as well as sectioning and data aggregation. These differences are mainly due to historical development, expert opinions, survey approaches and implemented pavement management system (PMS), rather than a systematic assessment of needs for optimal decisions. This paper will provide evidence on how these differences affect calculated performance functions and service lives based on standard models in PMS compared to a new continuous-time and state-space stochastic process on a road section and at the network level.

Figure 2.2 provides an example of condition survey and evaluation approaches for the general distress type "cracking" in Austria, Germany and Switzerland. In Austria linear (longitudinal and transverse) cracks are converted into a cracking area by multiplying the crack length with a weighting factor for the severity level and with an influence width of 0.5 m. The resulting area is then added to the area of alligator cracking and divided by the area of the surveyed section.



Austria (AT)

A1. CRACKING (ZG_{RI})

Longitudinal cracking, transverse cracking and irregular cracking (Extent: m*; Severity levels: S1, S2, S3)

Open joints, longitudinal joint cracks (Extent: m*; Severity levels: S1, S2, S3)

Alligator cracking (Extent: m²; Severity levels: S2, S3)

A2. SURFACE DEFECTS (ZG_{OS})

Raveling and weathering (Extent: m²; Severity levels: S1)

Patching (Extent: m²; Severity levels: S1)

Bleeding (Extent: m²; Severity levels: S1)

Delamination (Extent: m²; Severity levels: S2)

Potholes (Extent: m²; Severity levels: S2)

A3. LONGITUDINAL EVENNESS (ZG_{LE})

International Roughness Index IRI [m/km] for 50 m sections (RoadSTAR - measuring vehicle)

A4. RUTTING (ZG_{SR})

Maximum of the rut depth [mm] in the left and the right wheel path under a simulated 2 m straight edge; averaged for 50 m sections (RoadSTAR - measuring vehicle)

A5. SKID RESISTANCE (ZG_{GR})

Longitudinal friction coefficient μ [-] of a wetted pavement surface in the right wheel path; test wheel with constant slip ratio; test speed 60 km/h; speed correction; averaged for 50 m sections (RoadSTAR)

A6. LOAD-BEARING CAPACITY (ZG_{Tragf})

Measurements at the project level; Estimation of remaining service life (RSL) [%]; non-destructive with e.g. falling weight deflectometer (FWD)

Comments:

*Linear cracks are multiplied by 0.5 m influence width and converted into cracking area

Estimation condition value cracking:

$$CV_{RI} = \left(\frac{0.5 \times \sum_i (AML_{RI,i} \times G_{RI,i})}{A_B} + \frac{\sum_i (AMA_{RI,i} \times G_{RI,i})}{A_B} \right) \times 100[\%]$$

Estimation condition value surface defects:

$$CV_{OS} = \frac{\sum_i (AMA_{OS,i} \times G_{OS,i})}{A_B} \times 100[\%]$$

G_i… Weighting factor depending on the severity level

A_B … Area of pavement section [m²]

Germany (DE)

G1. SUBSTANCE-RELATED SURFACE CHARACTERISTICS

Alligator cracking, linear cracks RISS (Extent: area affected in %)

Patching AFLI (thin repairs to cracking and surface defects) (Extent: area affected in %)

Patching EFLI (repairs to utility cuts) (Extent: area affected in %)

Raveling AUS* (Extent: area affected in %)

Open joints ONA* (Extent: total length in m)

Bleeding BIN* (Extent: area affected in %)

G2. GENERAL UNEVENNESS

Power spectral density AUN [cm³] for characterization of general unevenness for 100 m sections (EFA - measuring vehicle)

Effective longitudinal evenness index LWI [-]† for characterization of the impact of single obstacles and periodic unevenness on vehicle, driver and cargo for 100 m sections (EFA - measuring vehicle)

Weighted longitudinal profile BLP [mm]‡ (new method)

G3. RUTTING

Maximum of the rut depth MSPT [mm] in the left and the right wheel path under a simulated 2 m straight edge; averaged for 100 m sections (EFA - measuring vehicle)

Maximum of the fictive water depth MSPH [mm] in the left and the right wheel path averaged for 100 m sections (EFA - measuring vehicle)

G4. SKID RESISTANCE

Sideway-force coefficient (SFC) [-] of a wetted pavement surface in the right wheel path with test wheel running under slip angle (20°); test speed 40/60/80 km/h; speed and temperature correction; calculated for 100 m sections (SKM - measurement system)

G5. LOAD-BEARING CAPACITY

For local investigations at the project level (FWD)

Comments:

*The pavement distresses raveling (AUS), open joints (ONA) and bleeding (BIN) are not rated or taken into account in the formation of the overall condition value

†The condition parameter effective longitudinal evenness index LWI is currently rated but not taken into account by the formation of the overall condition value

‡Recent approach for evaluation of the longitudinal road roughness (measurement and rating procedure under research)

Switzerland (CH)

S1. SURFACE SMOOTHNESS

Polished aggregate (Unit: m)

Bleeding (Unit: m or m²)

S2. SURFACE DISTRESSES

Wear (in the wheel pats) (Unit: m)

Weathering (loss of asphalt binder) (Unit: m²)

Raveling (loss of aggregate particles) (Unit: m²)

Delamination (Unit: m²)

Potholes (Unit: m²)

Open joints (Unit: m)

Transverse cracking (Unit: number)

"Wild" cracking (Unit: m²)

S3. SURFACE DEFORMATION

Rutting

Pavement uplift (along the sides of the rut) (Unit: m)

Corrugations (Unit: m²)

Shear deformation (Unit: m²)

S4. STRUCTURAL DISTRESSES

Settlement cracking (Unit: m²)

Settlement, depressions (Unit: m²)

Edge breaks (Unit: m²)

Frost heave (Unit: m²)

Longitudinal cracking (Unit: m)

Alligator cracking (Unit: m²)

Edge cracking (Unit: m)

S5. PATCHING

Patching (Unit: m²)

S6. LONGITUDINAL EVENNESS

Standard deviation s_W [‰] of angle values W for 250 m (angle measuring method) (Goniometer)

S7. RUTTING

Rut depth T [mm] under a simulated 4 m straight edge (wire method)

S8. SKID RESISTANCE

Friction coefficient μ [-] of a wetted pavement surface with locked wheel for 100 m sections (Skiddometer)

S9. LOAD-BEARING CAPACITY

Deflection measurements (Benkelman beam, Lacroix deflectograph, FWD)

Distress extent classes:

Class	Distress extent A	Pct area/ length
A0	No distress	0
A1	Rarely occurring	< 10%
A2	Occurring here and there	10…50%
A3	Very frequently occurring	> 50%

Distress severity classes:

Class	Severity levels S
S1	Low
S2	Moderate
S3	High

FIGURE 2.1 Summary and overview of distress types in asphalt concrete pavements as they are recorded in periodic routine surveys in Austria, Germany and Switzerland.

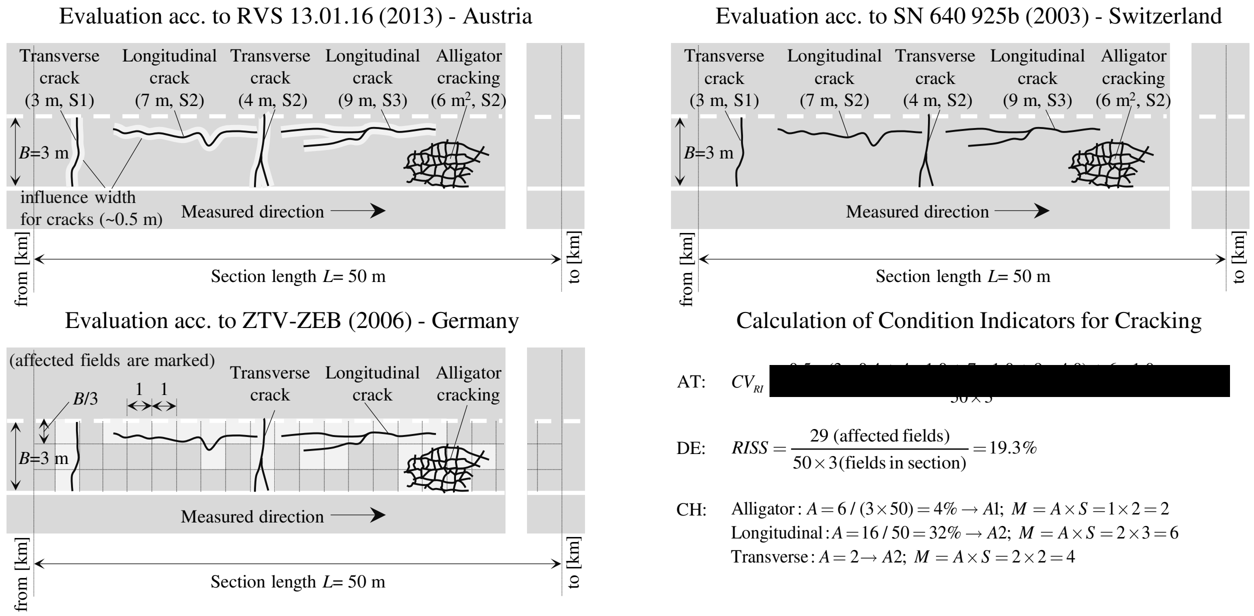
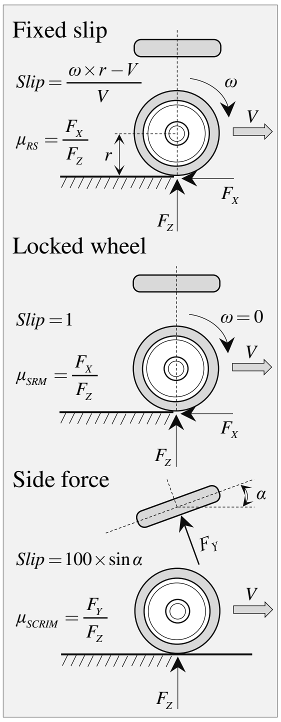


FIGURE 2.2 Evaluation of recorded cracking type, extent and severity according to guidelines and common PMS – approaches in Austria, Germany and Switzerland.

In Germany the area of cracking is determined by counting the fields with cracks on a defined grid in a road section without distinction between crack width and type of cracking. The aggregation approach in Germany and Austria may lead to the same cracking area despite different causes, types, extent and severity of cracking. This loss of information regarding different underlying distress causes, consequences and their progression over time is a major hindrance for the selection of appropriate treatments in M&R optimization. In Switzerland the represented failure severity is assessed on a qualitative basis for each type of cracking with no aggregation of distress types, while treatments are being selected manually.

Figure 2.3 provides an overview of the different principles and measurement equipment for wet-road skid resistance surveys in these countries, resulting in different measured values with different physical meaning (Do and Roe 2008, Donev 2014). Thus, the survey data cannot be directly compared. Obtained rating scales and relations regarding driving speed, design parameters and safety cannot be directly transferred from one country to another. In order to overcome these limitations, it has been a long-standing practice to conduct comparison measurements with different equipment or approaches on the same road sections (Fuchs 1996, Descornet 2006, Do and Roe 2008). Visual surveys with discrete rating scales are of subjective nature with the results depending on the survey approach and the number of condition classes. Therefore, regional or national survey methods are mainly used in PMS without conversion or comparison.

Traditionally, survey methods have been developed prior to related prediction models and are considered and improved separately. Rating scales and prediction models are chosen based on the type of provided survey data, as any change in survey method would make the historical data inconsistent and complicate the analysis of pavement performance. The differences in survey procedures, rating, weighting factors, aggregation rules and their impact on derived performance functions and resulting service lives are rarely considered in the literature and are not commonly known in practice (Kuhn 2012). Based on representative examples from German-speaking countries this paper provides a systematic overview of methodical aspects of these approaches and analyzes possible consequences in performance prediction.



Standard test conditions:

Single values: 0.15 m
Assessment length: 50 m
Test speed: 60 km/h
Test tire: PIARC, ribbed
Tire pressure: 2.0 bar
Static wheel load: 3 500 N
Water film thickness: 0.5 mm
Wheel path: right wheel track

Indicator: *Longitudinal friction coefficient* μ_{RS} [-]

Standard test conditions:

Measurement: discontinuous
Assessment length: 50 m
Test speed: 40, 60 or 80 km/h
Test tire: PIARC, ribbed
Tire pressure: 1.5 bar
Static wheel load: 3 500 N
Water film thickness: 0.5 mm
Wheel path: right/left track/center

Indicator: *Sliding friction coefficient* μ_{Skiddo} [-]

Standard test conditions:

Measurement: continuous
Assessment length: 100 m
Test speed: 40, 60 or 80 km/h
Test tire: SKM, smooth
Tire pressure: 3.5 bar
Static wheel load: 1 960 N
Water film thickness: 0.5 mm
Wheel path: right wheel track

Indicator: *Sideway-force coefficient* μ_{SCRIM} [-]

Fixed slip (18%)
e.g. RoadSTAR (Austria)

Locked wheel (100% slip)
e.g. Skiddometer BV 8 (Switzerland)

Side force (20° = ca. 34% slip)
e.g. SKM/SCRIM (Germany)

Legend

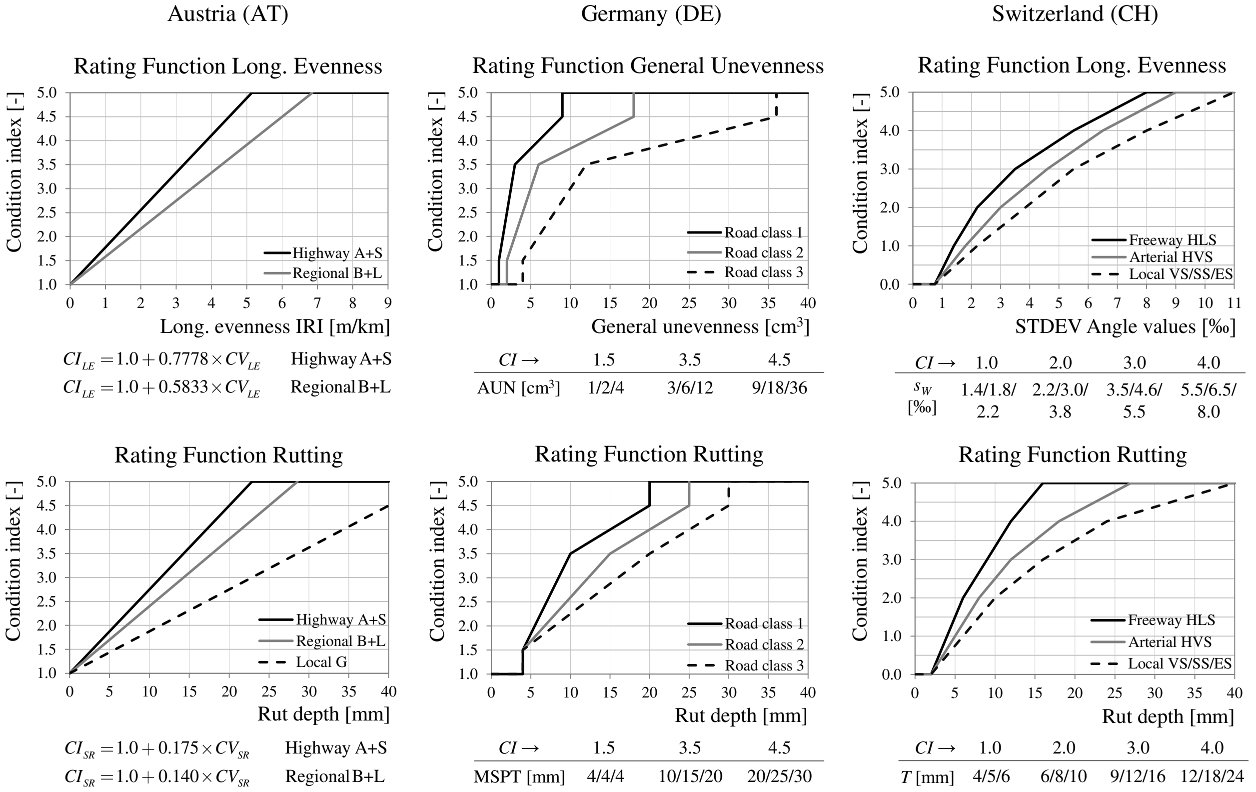
ω… Angular speed of the wheel [rad.s^{-1}]
F_Z… Wheel load [kN]
F_X… Braking force [kN]
V… Measuring speed [km/h]
α…Slip angle [°]

FIGURE 2.3 Standard test methods and devices used for network-wide measurements of skid resistance of pavements in Austria, Germany and Switzerland.

2.1.2 Condition assessment and rating scales

Measured condition values obtained from surveys do not provide information whether the pavement condition of a single section is relatively good or relatively bad or whether maintenance treatments should be applied. Typical rating approaches in condition assessment are mainly employing weighted scoring systems which enable a relative comparison of sections with different combinations of distress types based on their extent and severity. For a better understanding, these condition ratings are color coded allowing an overview of the pavement condition at the network level and an identification of sections with treatment priority by high-level decision makers. Due to their simplicity these rating systems are quite popular and play a major role in any PMS (e.g. condition triggered treatment). However, it can be shown that using such aggregated scores (e.g. based on common cost efficiency criteria) to compare treatment alternatives or in M&R optimization will always lead to subjective results.

The rating procedure consists of a transformation of measured conditions as a first step and a weighted aggregation towards partial and overall scores. In the transformation process the measured condition values with their associated units are converted into condition indices with no dimension by using distress-specific rating functions or scales. In most European countries these functions (scales) are based on the academic grading system, ranging, for example, from 1 (very good) to 5 (very bad) in Austria and Germany, respectively from 0 to 5 in Switzerland. Rating functions for different functional classes of roads are usually defined by changing their slope and shape. Characteristic points on these rating functions are a condition index of 4.5 (threshold value) and an index of 3.5 (warning value). Furthermore, condition index 1.5 usually corresponds to acceptance limits for quality control and assurance of new pavements. Exceeding the warning value requires intensive monitoring and possibly planning of maintenance treat-



$CI_{LE} = 1.0 + 0.7778 \times CV_{LE}$ Highway A+S

$CI_{LE} = 1.0 + 0.5833 \times CV_{LE}$ Regional B+L

$CI \rightarrow$	1.5	3.5	4.5
AUN [cm³]	1/2/4	3/6/12	9/18/36

$CI \rightarrow$	1.0	2.0	3.0	4.0
s_W [‰]	1.4/1.8/ 2.2	2.2/3.0/ 3.8	3.5/4.6/ 5.5	5.5/6.5/ 8.0

$CI_{SR} = 1.0 + 0.175 \times CV_{SR}$ Highway A+S

$CI_{SR} = 1.0 + 0.140 \times CV_{SR}$ Regional B+L

$CI \rightarrow$	1.5	3.5	4.5
MSPT [mm]	4/4/4	10/15/20	20/25/30

$CI \rightarrow$	1.0	2.0	3.0	4.0
T [mm]	4/5/6	6/8/10	9/12/16	12/18/24

FIGURE 2.4 Examples of condition rating functions for longitudinal evenness and rutting of pavements in Austria, Germany and Switzerland.

ments. Reaching the threshold value will normally trigger the application of treatments or traffic restrictions. Figure 2.4 provides an overview of currently used rating scales in German-speaking countries (SN 640 925b 2003, ZTV ZEB-StB 2006, Socina 2007, Weninger-Vycudil et al. 2009).

The derivation of rating scales is based on minimum standards, condition distributions, safety considerations and expert opinions. As an example, maximum values for rutting range from 12 mm (Switzerland) to 30 mm (Poland). The minimum values for skid resistance are commonly defined based on the 5th percentile of the empirical distribution of the network-wide measured skid resistance with a maximum of 5% of all sections below this limit. Since 5th percentiles depend on measurement method, material properties, applied treatments, paving technology and age distribution, the obtained values may change over time and are not transferable to other road networks. In general, threshold values in Austria are based on a harmonization of regional standards and existing limits in Germany. The width of the condition classes has been originally determined by cluster analysis. Threshold values in Germany are based on field experience and general acceptance (e.g. rutting, cracking). By contrast, the threshold and warning values for a fictive water depth are determined by simulation of hydrodynamic effects of ruts filled with water. Threshold values for general unevenness are mainly established with reference to generally accepted limits of human exposure to vibrations (e.g. Donev 2014).

The second step in the typical rating procedure is the combination of different distress types by applying weighting factors resulting in two partial indices, taking into account road user aspects (safety and comfort) and agency aspects (structural condition). Modifications of these approaches may also include the definition of critical values overriding the weighting process. These two indices are then combined into one overall condition index. In the Austrian and German PMS this overall condition index is both used for project prioritization as well as for optimization of M&R activities together with annual costs of their application. In Switzerland the maintenance treatments are manually assigned to distress types without use of any optimization algorithm (ZTV ZEB-StB 2006, Socina 2007, Weninger-Vycudil et al. 2009, Donev 2014).

2.1.3 Application on single section and network level

Apart from rutting, distress types cannot be compared directly in terms of survey method, condition assessment and rating scales. Therefore, rutting data on regional roads in Styria is used to demonstrate the impact of different rating scales at the road section and network level (Figure 2.5 and Figure 2.6). The results show that the requirements regarding rutting are far more rigorous in Switzerland as compared to Germany and Austria, leading to a higher amount of sections with perceived need for immediate treatment. However, if rutting is not considered being critical (e.g. Poland ≤ 30 mm), then there are almost no sections in very bad condition. From the right skewed distribution of rutting on the network level it can be concluded that

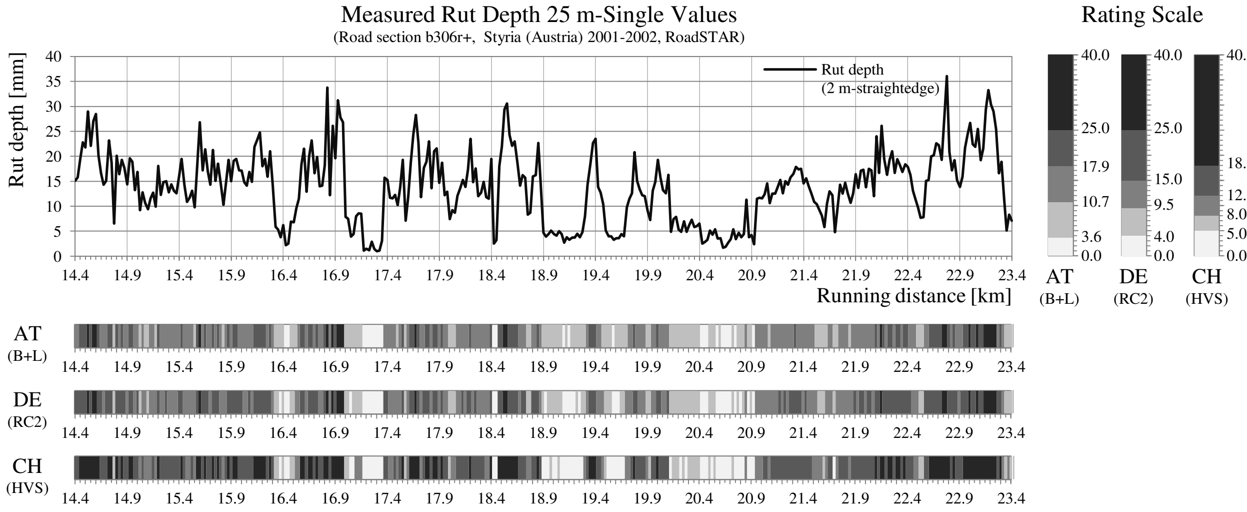


FIGURE 2.5 Comparison of results from different condition rating functions from Austria, Germany and Switzerland based on measurements of rut depth in Styria (single section).

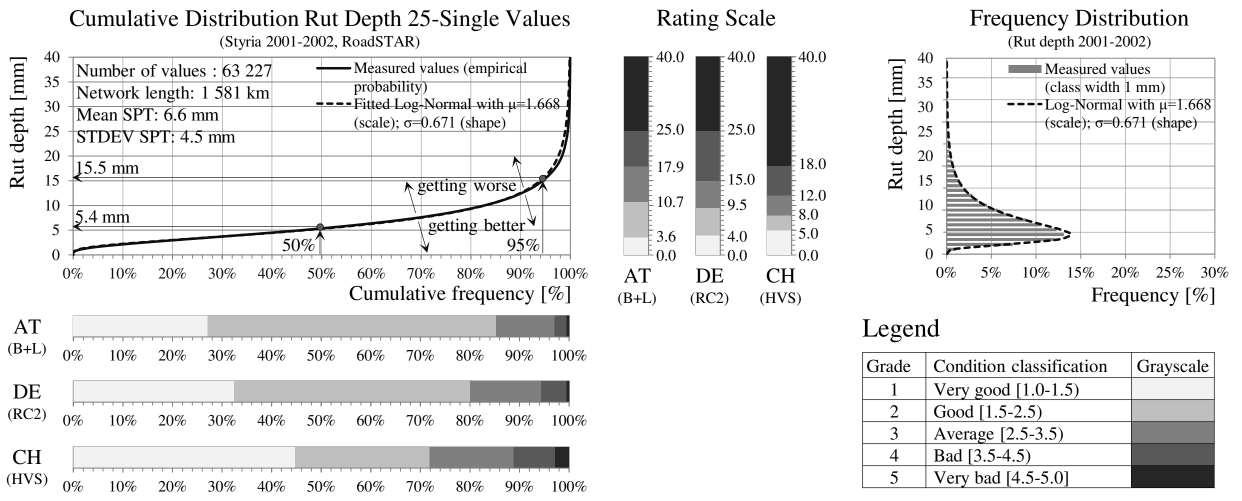


FIGURE 2.6 Comparison of results from different condition rating functions from Austria, Germany and Switzerland based on measurements of rut depth in Styria (network level).

rutting is rarely a critical issue with some initial rut depth after construction and decreasing degradation rate during service life (Donev 2014).

Furthermore, condition thresholds prior to defined failure criteria are used for release of guaranties defining the boundary of the construction risk between principal and agent. Rating in itself is a discretization and simplification of the condition distribution. Whether rating provides an accurate picture of this condition distribution depends on the number of condition classes and their distance. In summary, rating scale and condition distribution have a crucial impact on the resulting dominant failure types and corresponding treatments at the network level. Whether these ratings will lead to any practical consequences is dependent on the legal situation, available budgets and the chosen optimization approach.

2.1.4 Selected shortcomings of common survey and rating approaches

If the main goal of survey and rating approaches would be to provide relative condition information to decision makers or laymen, current approaches would suffice. However, if the target lies beyond color coding of roads, aiming at optimal decisions in an asset management or life cycle costing framework, there are several shortcomings in current approaches. As an example, the impact of predefined condition thresholds on service life and dominant failure types with subsequent consequences for treatment selection and budget are neglected in common PMSs (Figure 2.7). In addition, aggregation procedures and results of rated distress types towards partial or overall condition indices differ substantially, both in Austria and Germany as well as in other countries. Furthermore, it is possible to arrive at the same total condition index for sections with structural damages and sections with surface or smoothness deficits (Figure 2.8). Optimal treatments with structural rehabilitation activities in the first case and surface treatments on the latter lead to entirely different costs and service lives. Therefore, using condition indices instead of the original information on dominant distress causes will make an assignment of optimal treatments almost impossible.

For pavement performance modelling, this opens the possibility to either derive condition functions from original distress extent and severity or performance functions based on rated and aggregated condition indices. With distress types being related to a combination of stress and deterioration, the first approach will keep deviations from performance-based testing and field surveys at a minimum. However, if performance functions are fitted to already rated and aggregated condition indices (e.g. Germany), the impact of dominant distress and related failure type is blurred due to averaging effects resulting in a deformation of characteristic distress progression over time (Figure 2.9).

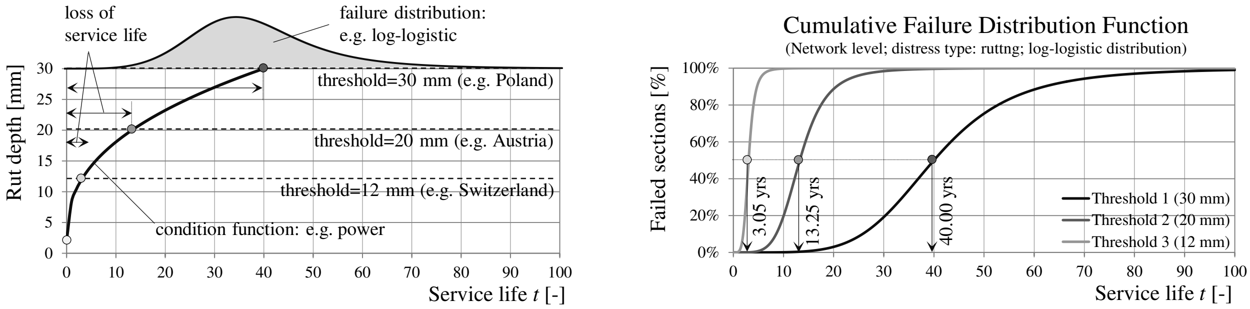
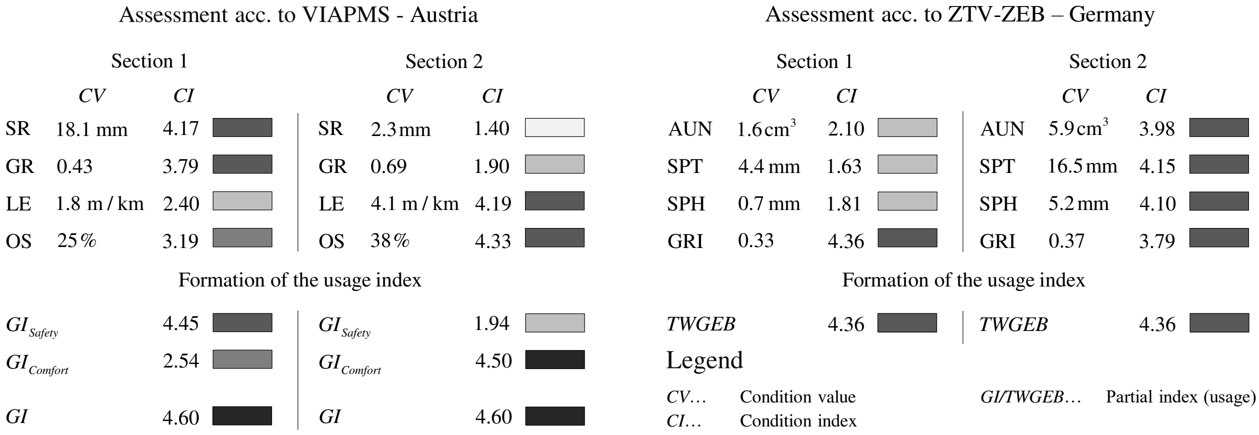


FIGURE 2.7 Service life being directly related to (arbitrary) condition threshold values with increasing thresholds leading to increased service life and variance of failure distributions.



Assessment acc. to VIAPMS - Austria

	Section 1 CV	CI			Section 2 CV	CI	
SR	18.1 mm	4.17	▓	SR	2.3 mm	1.40	□
GR	0.43	3.79	▓	GR	0.69	1.90	░
LE	1.8 m / km	2.40	░	LE	4.1 m / km	4.19	▓
OS	25 %	3.19	▒	OS	38 %	4.33	▓

Formation of the usage index

GI_{Safety}	4.45	▓	GI_{Safety}	1.94	░
$GI_{Comfort}$	2.54	▒	$GI_{Comfort}$	4.50	█
GI	4.60	█	GI	4.60	█

Assessment acc. to ZTV-ZEB – Germany

	Section 1 CV	CI			Section 2 CV	CI	
AUN	1.6 cm^3	2.10	░	AUN	5.9 cm^3	3.98	▓
SPT	4.4 mm	1.63	░	SPT	16.5 mm	4.15	▓
SPH	0.7 mm	1.81	░	SPH	5.2 mm	4.10	▓
GRI	0.33	4.36	▓	GRI	0.37	3.79	▓

Formation of the usage index

TWGEB	4.36	▓	*TWGEB*	4.36	▓

Legend

CV… Condition value
CI… Condition index
GI/TWGEB… Partial index (usage)

FIGURE 2.8 Identical partial and overall condition indices from entirely different distress types, extent and severity as result of typical rating processes.

If M&R treatments are triggered by the threshold value of a weighted overall index and not by the thresholds for the individual distress types, this will lead in most cases to an overestimation of service life and an underestimation of costs for non-redundant systems. From a deterministic point of view, these effects are mainly related to the weighted averaging of critical and non-critical service lives. However, it can be shown that for non-redundant systems the overall service life is on average shorter as the service life of the most critical distress (see Figure 2.17).

In reality, an optimal timing of treatments for distress types with no relation to traffic safety or comfort is a purely economic problem of the road authority or owner (Hoffmann 2006). In order to overcome these limits and avoid negative repercussions for the optimization, it is proposed to address different distress types with specific or combined treatments affecting one or several distress types. This can be accomplished if the distress extent is linked to the extent of treatments with severity as an indicator whether the treatments are technically efficient and necessary for safety reasons. Furthermore, calculating treatment costs based on costs functions and treatment extent allows to account for economies of scale and optimize work-zone lengths at the same time. Thus, the problem of averaging effects due to forming of homogenous sections (already at the condition assessment level) can be avoided as well. With the main emphasis on condition prediction and introduction of a new stochastic process in this paper, the task of handling multiple distress types and treatments in optimization will be covered in a future paper.

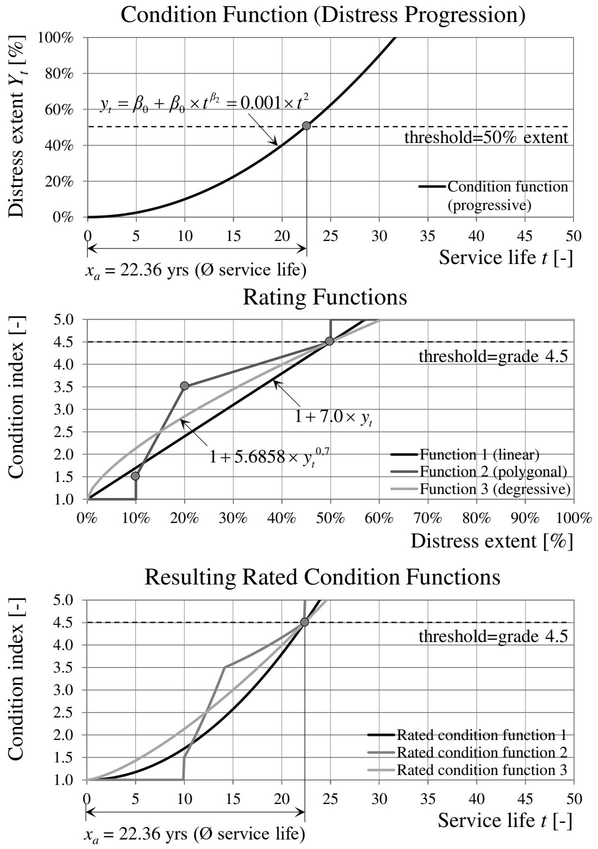


FIGURE 2.9 Deformation of characteristic distress progression function due to typical transformation in the rating process.

2.2 CONDITION PREDICTION MODELS

2.2.1 Condition prediction in existing systems

In general, condition prediction approaches are based on mechanistic principles or empirical data, and are divided into deterministic and stochastic models, depending on whether uncertainty is included or not (Nakat et al. 2008). In German-speaking countries only deterministic condition functions based on empirical data are used in PMSs (Figure 2.10). These averaged condition functions act as master curves and are calibrated for single sections to match data from observed conditions (see Section 2.2.5). At the network level the resulting condition distribution is the sum of all predictions on single sections. Probabilistic prediction models are mainly used in scientific studies, though they are state of art in other countries (Socina 2007).

In Austria the main PMS is based on the widely used Canadian software product VIAPMS (dTIMS) which has been in operation since 2000. Multiple regression models yielding deterministic predictions are used for cracking, rutting, surface defects, longitudinal evenness and skid resistance. These models have been estimated using empirical data from the first nationwide profile and skid resistance condition surveys from 1991 to 1996, the visual survey in 1995, the Austrian SHRP-sections (Strategic Highway Research Program) together with expert opinions (Molzer et al. 2000). The estimated model parameters differ according to the pavement type (considering applied maintenance treatments) and road class. Although other explanatory variables such as cumulative traffic loading (cumulative number of equivalent single axle loads, ESALs), freeze index (FI) and design index (DI) are included in the multiple regression analysis, the critical variable in all models is pavement age. Due to multicollinearity and other effects, even doubling the traffic loading has almost no consequences for the expected service life (Donev 2014). With the exception of cracking models (logarithmic function), all distress

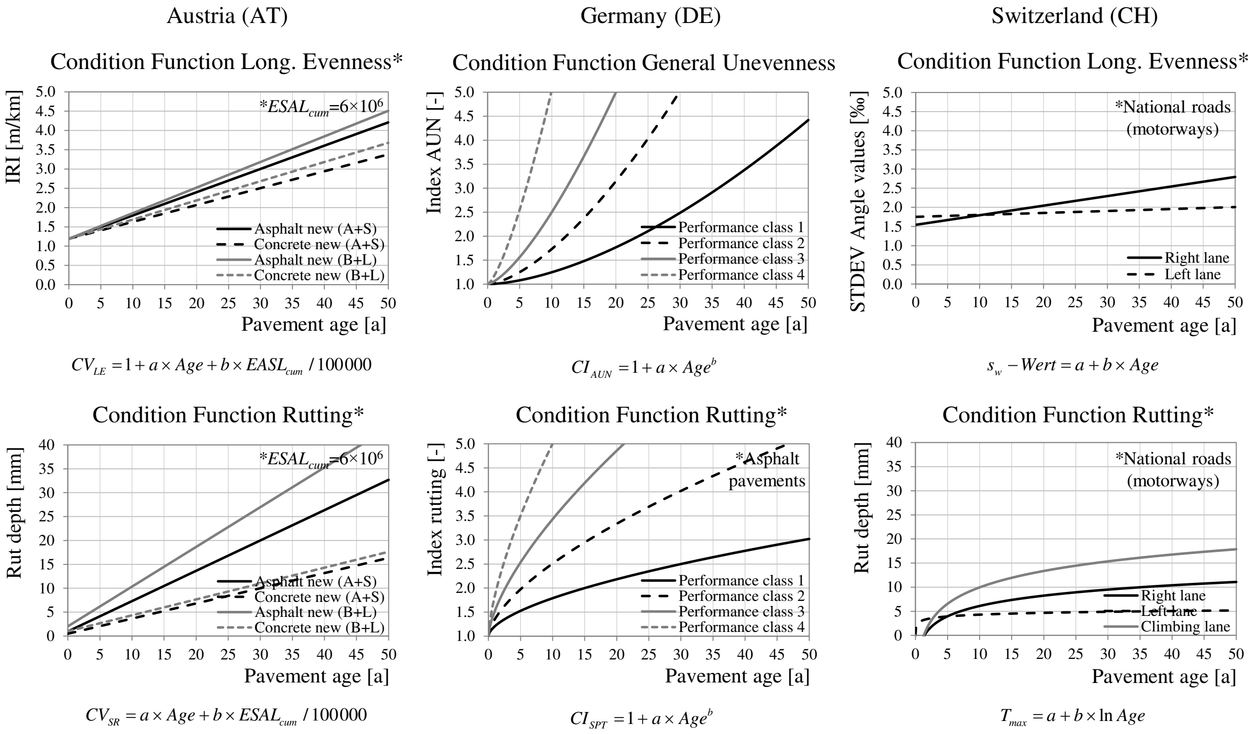


FIGURE 2.10 Commonly used deterministic condition functions in Austria, Germany and Switzerland for the same distress types (e.g. longitudinal evenness and rutting).

propagation functions are assumed to be linear. However, extensive evidence based on the analysis of empirical data from the long-term pavement performance (LTPP) program, mechanistic modelling, accelerated field and laboratory testing suggest otherwise.

In Germany the commonly used PMS is based on the Canadian software VIAPMS as well, and it was first implemented in 1999. Condition prediction functions have been derived from the first nationwide survey and expert opinions about the form of the distress progression functions. Even though condition surveys have been conducted every 4 years since 1992, these functions have not been updated in PMS yet. All distress models include only pavement age as an explanatory variable. Condition prediction in Germany is based on condition indices instead of condition values. This is a major limitation, which is well-known and discussed in the technical literature (Socina 2007, Donev 2014). To account for different rates of deterioration, instead of one master function four different standard performance functions are provided. Based on age and the latest known condition the function with the smallest vertical distance is used (Maerschalk and Socina 2008). Thus, almost similar calculated condition indices may result in the use of different performance functions and entirely different service lives (Donev 2014).

In Switzerland the PMS for motorways is still under development. The presented condition functions for all condition indices and road classes are the results from a research project on asphalt pavements (Scazziga 2008). The proposed deterministic models are derived using simple regression with pavement age as explanatory variable based on data from condition surveys, covering a few thousands of kilometers. The collected data is stratified by categories of additional predictors such as traffic loading, lane (for highways) and bearing capacity. Simple regression models are then derived separately for each combination of categories. As these models use old data and do not represent influencing factors like traffic, climate and structure in a mathematically correct manner, the estimated model parameters are expected to be biased.

Generally, practically applied prediction models and the German scientific literature disregard the presence of censored data almost entirely. Censoring appears, for example, if applied treatments during the service life are not documented. This is almost always the case, if empirical data in common PMSs is used, as age and treatment history are rarely well documented. Without the use of appropriate statistical techniques (e.g. survival analysis accounting for competing risks) the estimated regression parameters may be biased. The most likely effect will be an overestimation of service lives and an underestimation of costs as surviving sections are considered together with failed and replaced sections. These aspects will be accounted for using survival analysis and the new stochastic process, which will be covered in a forthcoming paper.

2.2.2 Bivariate and multiple regression models

Since it is not possible to consider all influencing factors in a regression model, some bias due to misspecification is always to be expected. As an example, Figure 2.11 considers a “true” model, consisting of two explanatory variables and a model with one variable being omitted. Since age and structural number (SNC) are correlated, the bivariate model violates the exogeneity assumption and the ordinary least squares (OLS) estimator is biased. The sign and size of this bias can be estimated accordingly (Wooldridge 2013, Donev 2014, Stocker 2014). The multiple regression model has a higher coefficient of determination, indicating a better fit in the sample. Contour plots are a convenient way to represent models with two explanatory variables, since they allow an assessment of the contributions of each variable to distress prediction.

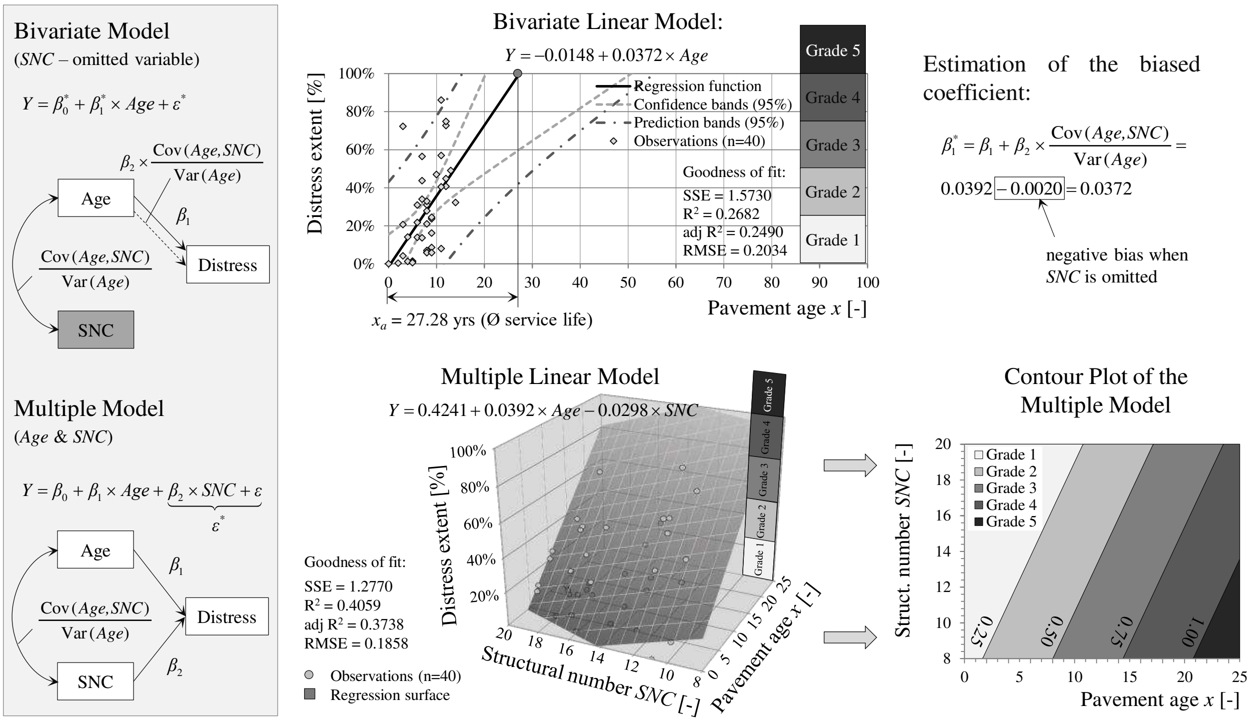
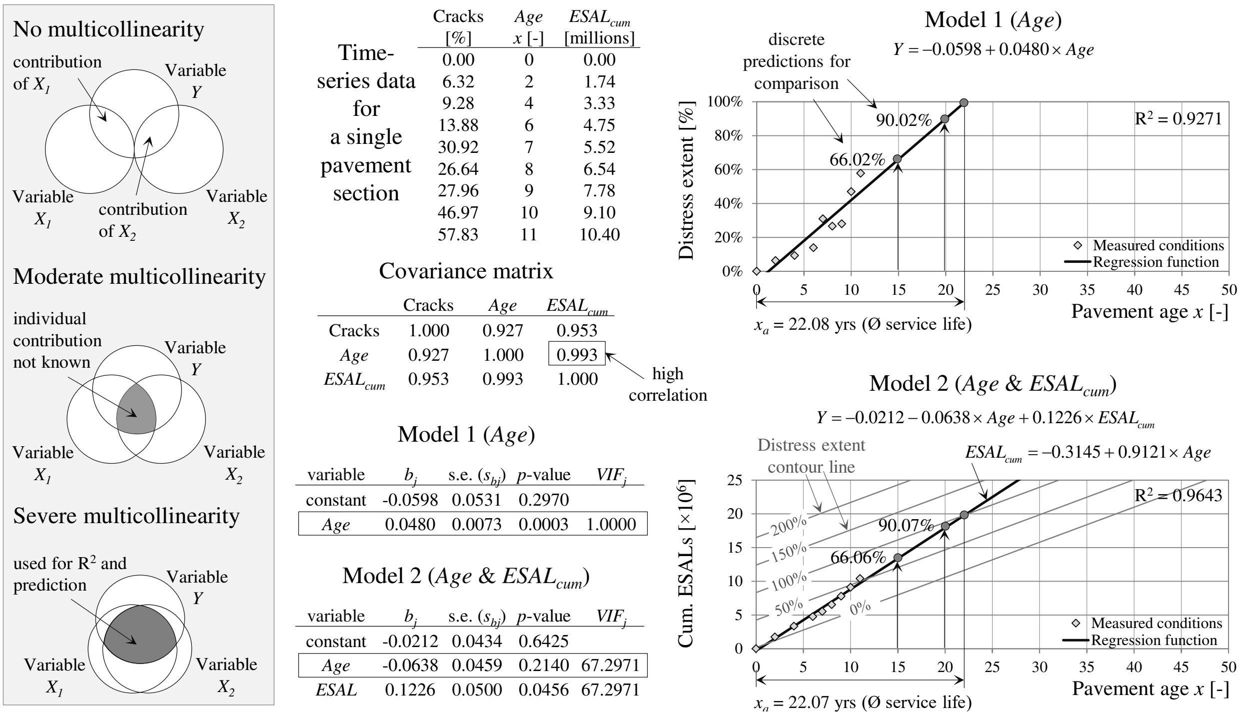


FIGURE 2.11 Simple/multiple regression and impact of an incomplete specification (omission of a relevant variable) on the parameter estimates and on the prediction (sub-network).



Time-series data for a single pavement section

Cracks [%]	*Age* *x* [-]	$ESAL_{cum}$ [millions]
0.00	0	0.00
6.32	2	1.74
9.28	4	3.33
13.88	6	4.75
30.92	7	5.52
26.64	8	6.54
27.96	9	7.78
46.97	10	9.10
57.83	11	10.40

Covariance matrix

	Cracks	*Age*	$ESAL_{cum}$
Cracks	1.000	0.927	0.953
Age	0.927	1.000	$\boxed{0.993}$
$ESAL_{cum}$	0.953	0.993	1.000

high correlation

Model 1 (*Age*)

variable	b_j	s.e. (s_{bj})	*p*-value	VIF_j
constant	-0.0598	0.0531	0.2970	
Age	0.0480	0.0073	0.0003	1.0000

Model 2 (*Age* & $ESAL_{cum}$)

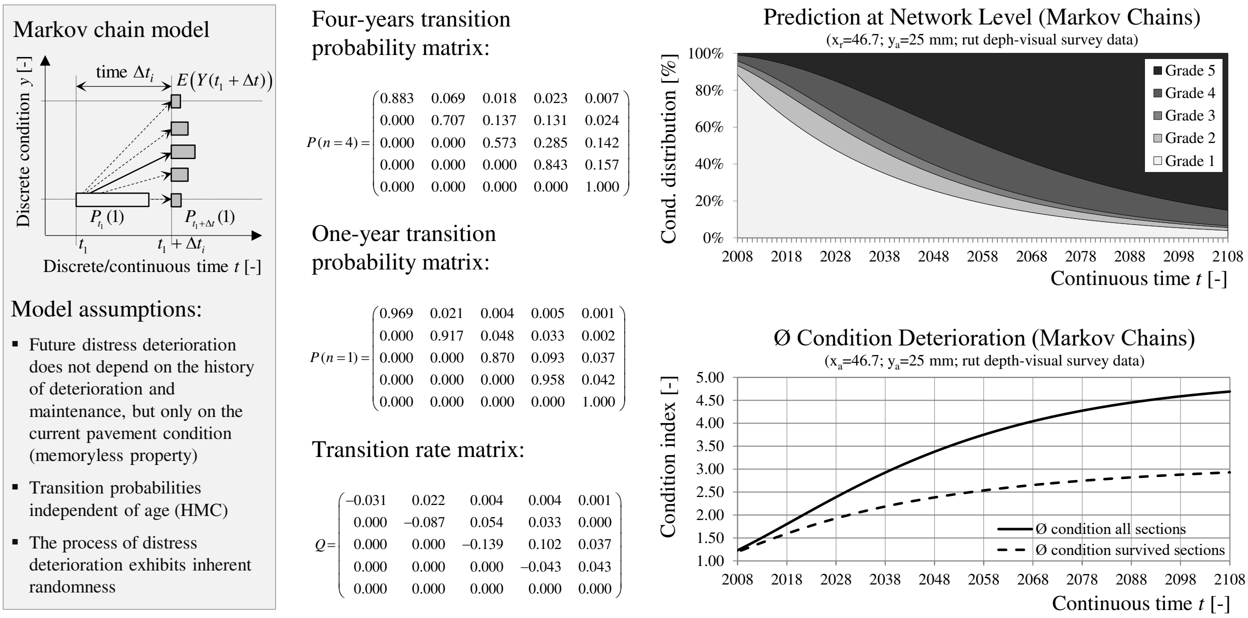
variable	b_j	s.e. (s_{bj})	*p*-value	VIF_j
constant	-0.0212	0.0434	0.6425	
Age	-0.0638	0.0459	0.2140	67.2971
ESAL	0.1226	0.0500	0.0456	67.2971

FIGURE 2.12 Simple/multiple regression and impact of severe multicollinearity on estimation, inference and prediction (single section). Graphical representation using Venn diagrams.

Multicollinearity refers to the existence of a high correlation among explanatory variables. Such high correlations, for example, between pavement age and cumulated variables (traffic, freeze, precipitation, etc.) during service life, are common at the project level. Possible consequences of severe multicollinearity are illustrated in Figure 2.12. For the given section, the explanatory variables age and cumulative ESALs are highly correlated (correlation coefficient of 0.993). In the model containing both variables, the regression coefficient for age changes drastically and is no longer physically plausible as the distress extent decreases with age. The standard error of this coefficient increases when compared to the model without ESALs. The null hypothesis cannot be rejected, and the coefficient is statistically insignificant. The variance inflation factor (VIF) in the model with both variables suggests that there might be some issues with multicollinearity (but not necessarily). For the goal of obtaining accurate condition predictions, multicollinearity in itself is not a problem as both models produce almost the same results regarding condition and average service life. However, multicollinearity limits conclusions concerning the influence of explanatory variables due to instability of the resulting parameters. Furthermore, the model cannot be applied to other data sets with different correlations.

2.2.3 Stochastic discrete state-space processes (e.g. Markov chains)

Prediction models based on Markov chains are stochastic models in discrete state space, not requiring assumptions about any functional form of distress progression, explanatory variables, failure and condition distribution. A homogeneous Markov chain is fully defined by the initial condition distribution vector and transition probability matrix ***P*** or generator matrix ***Q*** in discrete time or in continuous time, respectively (Stewart 2009). The resulting condition distribution provides information on the total share of sections in any condition class at any point in time (network level). Furthermore, it provides an “a priori” probability for reaching a certain condition for any single section (project level). For the derivation of transition probabilities and application of these model classes at least data from two successive surveys are needed.



Model assumptions:

- Future distress deterioration does not depend on the history of deterioration and maintenance, but only on the current pavement condition (memoryless property)
- Transition probabilities independent of age (HMC)
- The process of distress deterioration exhibits inherent randomness

Four-years transition probability matrix:

$$P(n=4)=\begin{pmatrix} 0.883 & 0.069 & 0.018 & 0.023 & 0.007 \\ 0.000 & 0.707 & 0.137 & 0.131 & 0.024 \\ 0.000 & 0.000 & 0.573 & 0.285 & 0.142 \\ 0.000 & 0.000 & 0.000 & 0.843 & 0.157 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{pmatrix}$$

One-year transition probability matrix:

$$P(n=1)=\begin{pmatrix} 0.969 & 0.021 & 0.004 & 0.005 & 0.001 \\ 0.000 & 0.917 & 0.048 & 0.033 & 0.002 \\ 0.000 & 0.000 & 0.870 & 0.093 & 0.037 \\ 0.000 & 0.000 & 0.000 & 0.958 & 0.042 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{pmatrix}$$

Transition rate matrix:

$$Q=\begin{pmatrix} -0.031 & 0.022 & 0.004 & 0.004 & 0.001 \\ 0.000 & -0.087 & 0.054 & 0.033 & 0.000 \\ 0.000 & 0.000 & -0.139 & 0.102 & 0.037 \\ 0.000 & 0.000 & 0.000 & -0.043 & 0.043 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \end{pmatrix}$$

FIGURE 2.13 Application of homogeneous Markov chain to rutting data of in-service pavements in Styria with resulting condition distribution, average progression and transition matrices.

Typically, these models are applied on electronic components with failure occurring independent of age due to random events like voltage fluctuations. However, in cases with deterioration being a result of cumulative loading and aging processes, the assumption of constant transition probabilities during service life is not valid (Guignier and Madanat 1999).

As an example, the results of a Markov chain model for rutting from two consecutive visual surveys in Styria is shown in Figure 2.13. The average deterioration function shows linear progression with an asymptotic behavior towards the absorbing failure condition if a grading system is used. This functional form is typical for Markov chains, irrespective of distress type. As a workaround, transition probabilities for different age groups can be estimated (similar to cohort survival models). However, these approaches rarely work due to the usually very limited number of assets in very bad condition or old age with almost no information on previous failures or treatments (censoring). Thus, it is nearly impossible to accurately model condition development, with expected remaining service life (e.g. rutting = 46.7 years) always being heavily overestimated and thus costs being underestimated. This holds especially true if the underlying deterioration process is non-linear by nature or dependent on other parameters that may vary over time (e.g. traffic volume). Therefore, such models provide very limited accuracy in condition prediction at the network level and fail to consider actual condition development on single sections that may deviate considerably from the average pavement performance.

2.2.4 Continuous-time and state-space processes (e.g. Hoffmann process)

In principle, there are endless possibilities for modelling stochastic processes. The presented stochastic continuous-time and state-space process shown in Figure 2.14 is only one of these possibilities but has several interesting properties. It is based on the assumptions of probabilistic development of failure extent and severity with distinct failure distribution and scalable condition progression, which is characteristic for each distress type. Both characteristic condition development and failure distribution may take on any form but should be based on survival analysis, multiple regression or better structural equations to avoid bias.

The "a priori" probability of arriving at a specific condition as well as the total amount of sections in a certain condition is calculated as integral of the probability density function of the failure distribution over the interval from $x_{l,i}$ to $x_{u,i}$ (scaled condition functions). Considering only surviving sections at any given time will yield the average unbiased remaining service of all surviving sections at a certain age, regardless of actual condition. Depending on derived failure distribution this remaining service life of surviving sections deviates substantially with increasing age from average service life of all sections. Based on the same approach it is also possible to calculate an average remaining life for sections in a specific condition state at any given time (Figure 2.15). Thus, the presented process is also capable of handling censored data (e.g. failed sections or applied M&R treatments). As censoring may lead to substantial bias and is commonly neglected, this issue will be addressed in detail in a forthcoming paper.

For condition prediction on single sections, master functions are fitted to condition data from surveys. These master functions must be calculated from data at the network level to avoid bias. However, in contrast to common deterministic approaches, using prediction intervals allows to calculate failure distributions at the road section level. Thus, any "a priori" applied failure distribution on the network level may be narrowed down considerably using section-specific information from condition surveys (Figure 2.16).

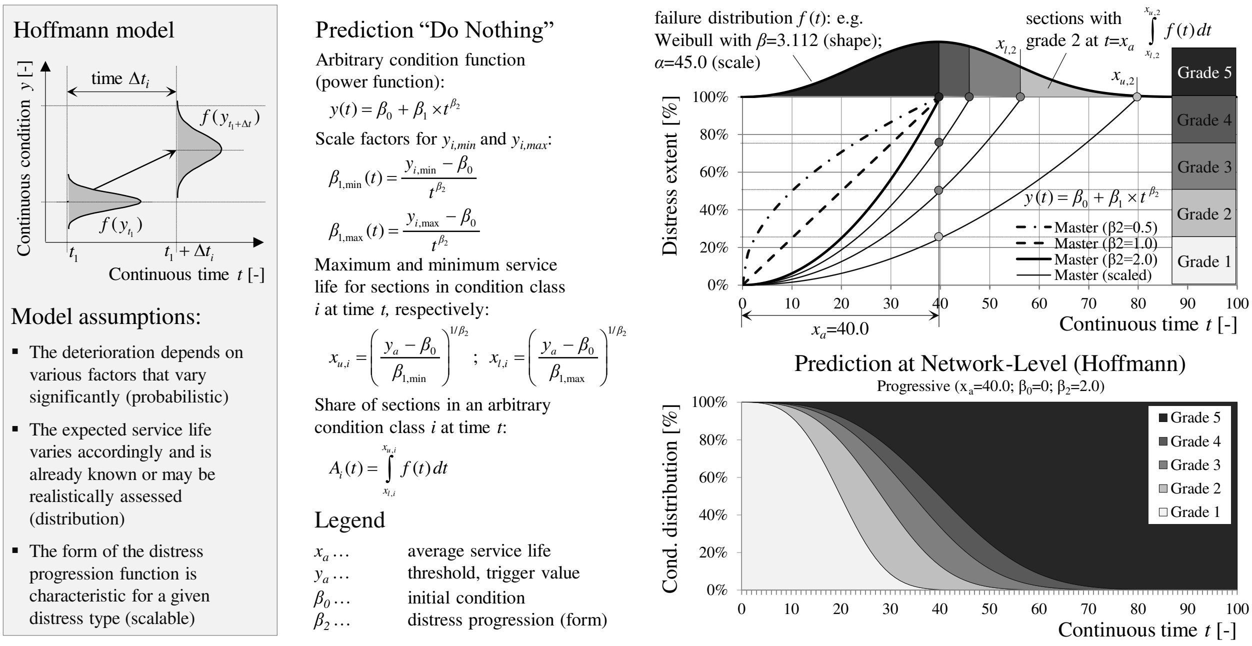


FIGURE 2.14 Application of a continuous time and state space process based on stochastic distributed characteristic condition development at the network level with related failure and condition distribution.

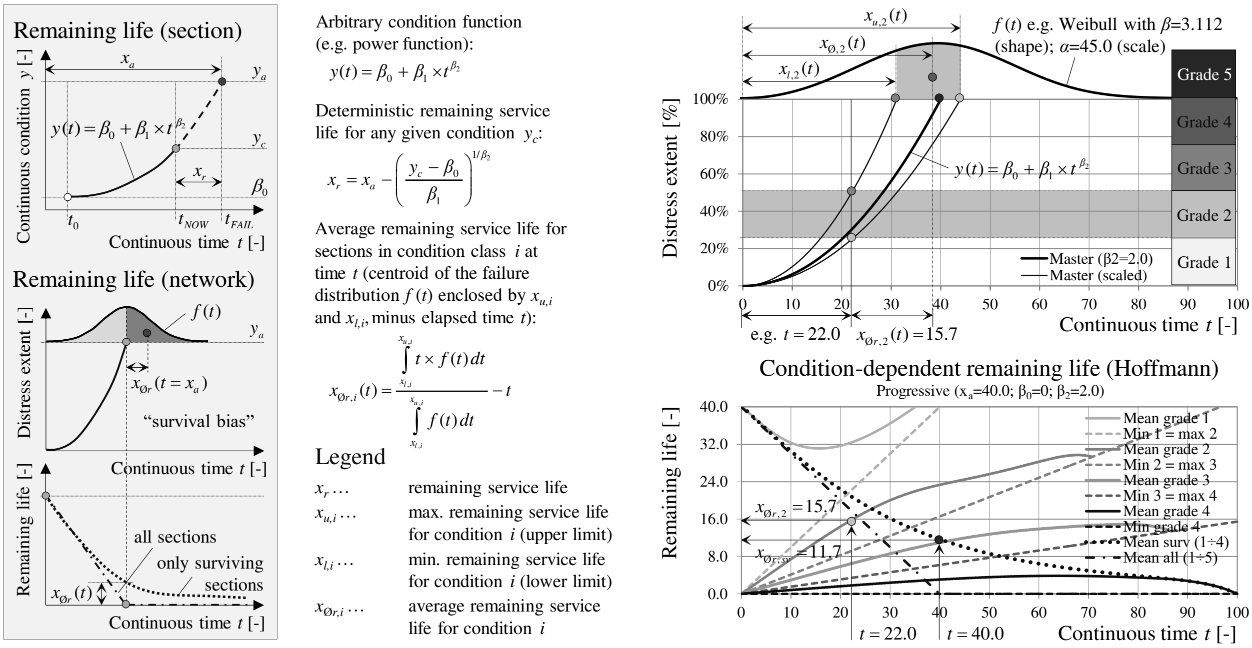
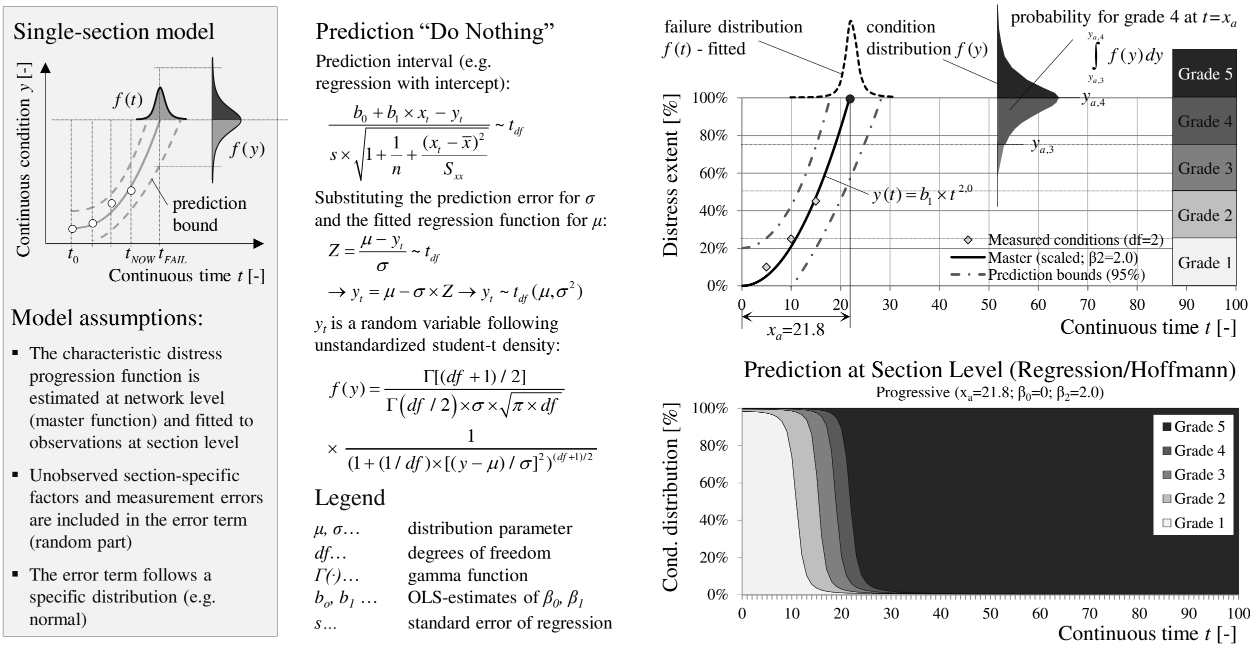


FIGURE 2.15 Calculation of remaining service lives of all surviving sections as well as the average remaining life of sections in a specific condition threshold at any given time based on the continuous time and state space process.

Due to the inherent stochastic nature of condition development of different uncorrelated distress types on one single section with condition thresholds, the average service life or time to treatments is on average shorter than an "a priori" critical failure type due to superposition of failure distributions (series systems). Based on the same principles it is also possible to calculate time to treatment for correlated distress types. Commonly used aggregated indices weighting individual distress types therefore result in an overestimation of service life and an underestimation of costs for non-redundant systems, with the opposite being true for redundant systems (Figure 2.17). The presented process can be extended to a full life cycle approach taking into account the stochastic nature of treatment impact, costs and failure-related risks (Hoffmann 2006, Hoffmann and Donev 2015).



Single-section model

Model assumptions:

- The characteristic distress progression function is estimated at network level (master function) and fitted to observations at section level
- Unobserved section-specific factors and measurement errors are included in the error term (random part)
- The error term follows a specific distribution (e.g. normal)

Prediction "Do Nothing"

Prediction interval (e.g. regression with intercept):

$$\frac{b_0 + b_1 \times x_t - y_t}{s \times \sqrt{1 + \frac{1}{n} + \frac{(x_t - \overline{x})^2}{S_{xx}}}} \sim t_{df}$$

Substituting the prediction error for σ and the fitted regression function for μ:

$$Z = \frac{\mu - y_t}{\sigma} \sim t_{df}$$

$$\rightarrow y_t = \mu - \sigma \times Z \rightarrow y_t \sim t_{df}(\mu, \sigma^2)$$

y_t is a random variable following unstandardized student-t density:

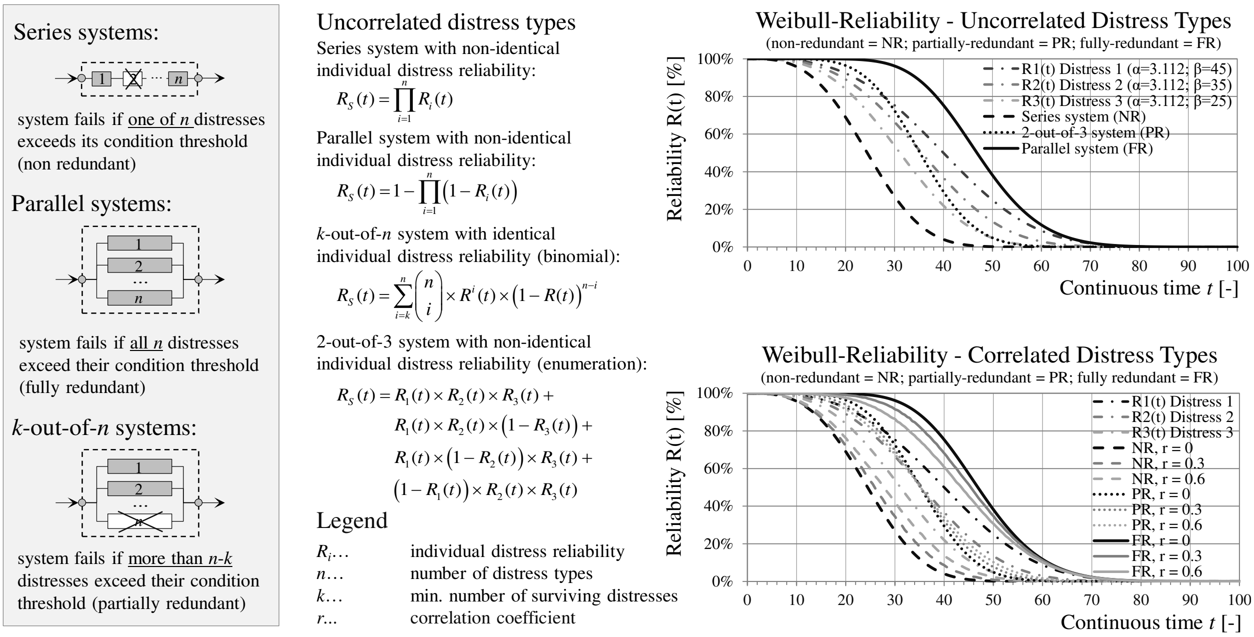
$$f(y) = \frac{\Gamma[(df+1)/2]}{\Gamma(df/2) \times \sigma \times \sqrt{\pi \times df}}$$

$$\times \frac{1}{(1 + (1/df) \times [(y-\mu)/\sigma]^2)^{(df+1)/2}}$$

Legend

μ, σ…	distribution parameter
df…	degrees of freedom
$\Gamma(\cdot)$…	gamma function
b_o, b_1 …	OLS-estimates of β_0, β_1
s…	standard error of regression

FIGURE 2.16 Application of the continuous time and state space process for a condition prediction at road section level with section-specific failure distribution derived using confidence intervals from fitting of the master function (regression analysis).



Series systems:

system fails if one of n distresses exceeds its condition threshold (non redundant)

Parallel systems:

system fails if all n distresses exceed their condition threshold (fully redundant)

k-out-of-*n* systems:

system fails if more than n-k distresses exceed their condition threshold (partially redundant)

Uncorrelated distress types

Series system with non-identical individual distress reliability:

$$R_S(t) = \prod_{i=1}^{n} R_i(t)$$

Parallel system with non-identical individual distress reliability:

$$R_S(t) = 1 - \prod_{i=1}^{n} \left(1 - R_i(t)\right)$$

k-out-of-*n* system with identical individual distress reliability (binomial):

$$R_S(t) = \sum_{i=k}^{n} \binom{n}{i} \times R^i(t) \times \left(1 - R(t)\right)^{n-i}$$

2-out-of-3 system with non-identical individual distress reliability (enumeration):

$$\begin{aligned} R_S(t) = {} & R_1(t) \times R_2(t) \times R_3(t) + \\ & R_1(t) \times R_2(t) \times \left(1 - R_3(t)\right) + \\ & R_1(t) \times \left(1 - R_2(t)\right) \times R_3(t) + \\ & \left(1 - R_1(t)\right) \times R_2(t) \times R_3(t) \end{aligned}$$

Legend

R_i…	individual distress reliability
n…	number of distress types
k…	min. number of surviving distresses
r...	correlation coefficient

FIGURE 2.17 Failure distributions for uncorrelated/correlated distress types and any given threshold in series systems, parallel systems and *k*-out-of-*n* systems based on continuous-time and state-space process.

2.2.5 Comparison of prediction models for a single section and at the network level

The following comparison of prediction models is based on LTPP data from consecutive condition observations of transverse cracking as dominant failure mode under low temperatures (Donev 2014). Distress extent for transverse cracking is defined as the sum of transverse cracks in meters, expressed as a percentage of the section length with no weighting for the severity of cracks. With the main emphasis on comparing prediction approaches, a simple power function ($\beta_0 + \beta_1 \times x^{\beta_2}$) with pavement age (time since the last major rehabilitation) as explanatory variable is selected as a cracking progression model. The procedure for calculation of parameters for this master function is based on the approach from Section 2.2.4 of this paper.

First, nonlinear regression is conducted individually for all sections (parameters β_1 and β_2 are free), yielding a distribution for the exponent β_2 (form parameter) with a mean value of 2.905. If the variance of this distribution is low, it is safe to assume that the mean condition progression is characteristic for this distress type. Repeating the regression with a restricted $\beta_2 = 2.905$ as a second step yields a failure distribution. The scaling factor $\beta_1 = 0.000361$ for the master function may then be calculated with the resulting average service life ($x_a = 15.31$ years). This new master function is adapted to all sections in a third step by standard techniques for single-section calibration – namely scaling, vertical shifting, horizontal shifting, and regression. The prediction at the network level is calculated as the sum of all predicted conditions at single sections for each year and compared to the results based on Section 2.2.4 (Figure 2.18).

Common scaling (stretching or compressing) of the master function is performed by calculating the model parameter β_1 based on a starting point (e.g. age zero, no damage) and the last observed condition. However, such approaches entirely neglect previous condition development, giving weight only to the last observation and ignoring measurement errors. Nonetheless, this calibration procedure has been used for cracking and rutting in Austria (VIAPMS) for more than a decade. The results of this approach on the network level in the provided example yield an unrealistically long service life for 10% of the LTPP sections being in relatively good health at the time of the last survey.

Commonly used horizontal or vertical shifting of deterministic master functions through the last observed condition state does not change model parameters and requires only one observed condition and pavement age. Moreover, predictions based on horizontal shifting require no knowledge of age, producing identical predicted condition states (also identical deterioration rates) for all sections with the same measured condition values. Vertical shifting is used for all distress types in the German PMS as a next step after selecting one of multiple standard master functions. Shifting is also used on highways and regional roads in Austria (VIAPMS) for the prediction of surface defects and longitudinal evenness (roughness). However, by shifting the master curve upwards or to the left, distress extent at time $t = 0$ yields large errors. If the function is instead shifted downwards or to the right, the typical continuous distress progression from $t = 0$ is interrupted, resulting in a distortion of age and condition distribution at the network level (see Figure 2.18). In summary, these commonly employed prediction models in PMSs may be simple to compute but yield large errors, especially if the prediction horizon extends beyond the next planned condition survey. This might not seem too bad with the deviations being corrected at the next survey but is a deciding factor in a full life cycle costs analysis for optimizing consecutive M&R treatments and addressing different failure types at the same time.

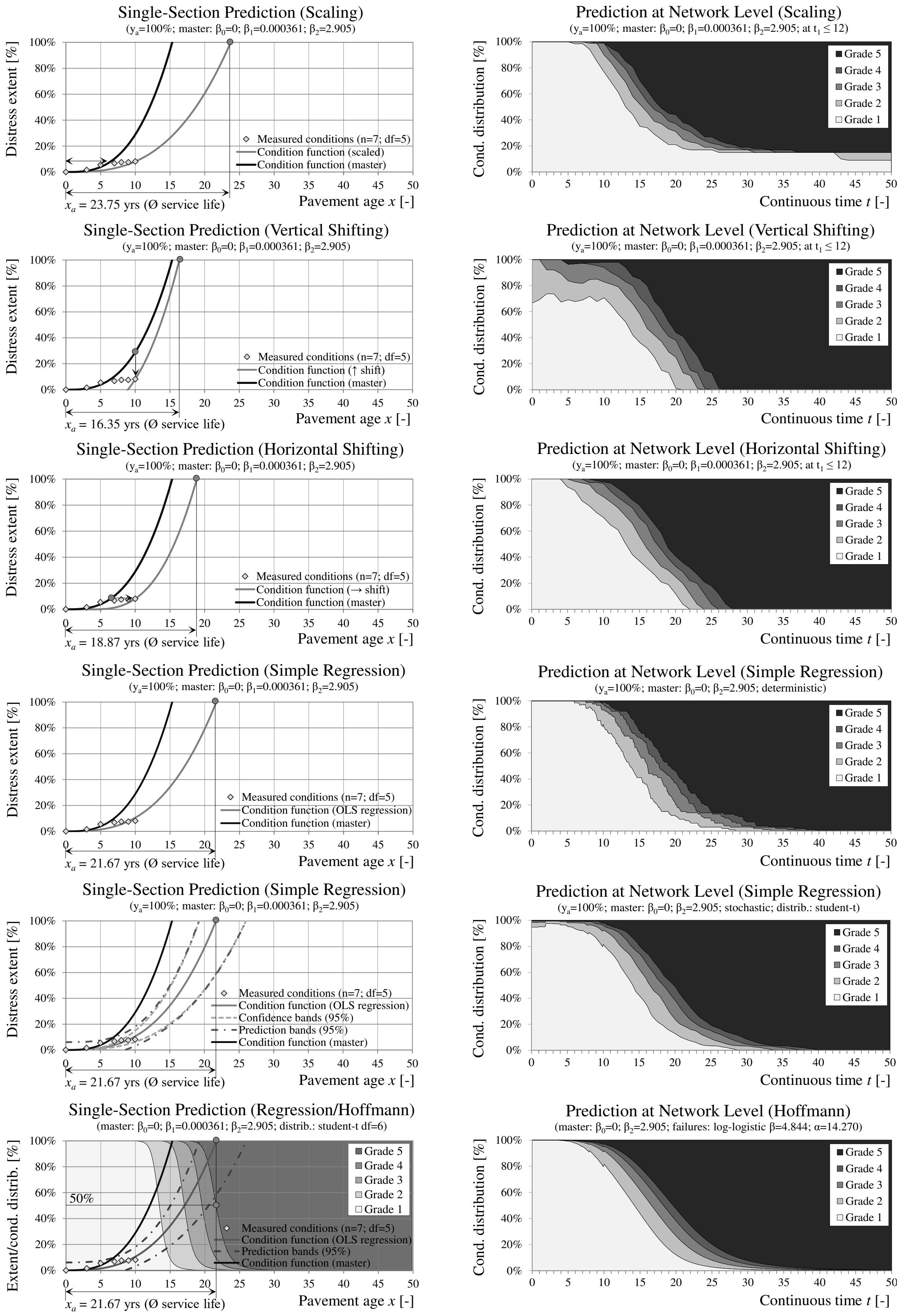


FIGURE 2.18 Comparison of condition prediction models for a single section and at network level based on consecutive observations of transverse cracking (LTPP data).

By contrast, any prediction based on regression (deterministic/stochastic) requires time-series data with a series of observations over time. The master function is fitted either by OLS or maximum-likelihood estimation (MLE) method yielding the scaling parameter β_1. The resulting model is intrinsically linear, allowing an analytical computation of prediction intervals, i.e. calculation of a probability for a certain condition threshold at any time (condition distribution) at the section level. With data from additional surveys, the master function may be automatically adjusted narrowing the “a priori” distribution of expected service life. Compared to regression models, the afore-mentioned deterministic prediction approaches at the section level yield large deviations of resulting service life, having a major impact on the selection of treatment timing. The condition distribution at the network level based on the sum of the predictions at the section level (with regression) will be accurate enough, due to the central limit theorem, if the number of sections is sufficiently large.

At the network level, the process developed by Hoffmann (based on master function and failure distribution) yields similar results to the prediction with regression (sum of single predictions) but has the advantage of describing the distribution of the entire network in closed mathematical form. Furthermore, this new model allows for estimation of condition-based remaining service life of surviving sections and unbiased average service life in the presence of censored data. In contrast to deterministic regression, failure distribution allows an attribution of failure risk and costs on road section and at network level alike. By including impact and costs of possible treatments in this entirely stochastic model, a prediction and life cycle costs (LCC) optimization of any simple or complex aging and repairable system becomes feasible. With the focus on comparing common prediction models with the capabilities of this new approach, these topics are not covered in this paper.

2.3 CONCLUSIONS

In German-speaking countries as well as in most European countries, M&R decisions in PMSs are based on condition or performance functions which are derived from data of periodic condition surveys. Rating or grading systems have been developed to compare the condition on sections with different distress types, severity and/or extent. Apart from “color coding” the road network for decision makers, certain condition threshold or resulting total condition grades are often used for an optimization of treatments. This paper provides an overview of commonly surveyed distress types, rating approaches and failure criteria together with an analysis of their drawbacks. However, the main emphasis of the paper is a comparison of deterministic prediction models in common PMSs with more advanced deterministic and stochastic approaches. With the application of these models to real-world data from condition surveys in Austria and the US (LTPP data), both on a road section and at the network level, the shortcomings of common deterministic approaches are analyzed in detail.

In summary, any approach in a PMS that neglects the presence of censored data, relies on aggregated condition indices, arbitrary condition thresholds and deterministic prediction will risk heavy bias and fail to achieve optimal results in selecting treatment type and timing from a life cycle costing perspective. The proposed new stochastic continuous-time and state-space process avoids most of these shortcomings. The paper addresses how this new process can be applied in stochastic predictions at the single section and network level both for uncorrelated and correlated distress types. Furthermore, the process can be extended to a full life cycle including M&R

treatments, allowing to address individual distress types based on extent and severity. Thus, several other methodical problems resulting from aggregated condition indices in optimization can be avoided allowing a unique entirely probabilistic prediction and LCC optimization of any complex ageing and repairable system.

REFERENCES

Descornet, G., 2006. *Harmonisation of European Routine and Research Measurement Equipment for Skid Resistance*. Brussels: FEHRL report.

Do, M.-T. and Roe, P., eds., 2008. *Report on the state-of-the-art of test methods. Deliverable D04, European Project TYROSAFE*. Brussels: Forum of European National Highway Research Laboratories.

Donev, V., 2014. *Erhaltungsmanagement Straßenoberbau - Zustandsbewertung und Zustandsprognose* [Pavement management – condition assessment and condition prediction]. Wien: Mitteilungen der Professur für Straßen- und Flugbetriebsflächenbau (in German).

Fuchs, M., 1996. *Internationaler PIARC Ringversuch Straßengriffigkeit* [International PIARC experiment to compare and harmonize skid resistance measurements]. Wien: Bundesministerium für Wirtschaftliche Angelegenheiten, Straßenforschung Heft 456 (in German).

Guignier, F. and Madanat, M., 1999. Optimization of infrastructure systems maintenance and improvement policies. *Journal of Infrastructure Systems*, 5(4), 124-134.

Haas, R., Hudson, W. and Falls, L., 2015. *Pavement asset management*. Salem, MA: Scrivener Publishing.

Hoffmann, M., 2006. *Instandsetzung von Strassen* [Rehabilitation of roads]. Dissertation (PhD). Technische Universität Graz (in German).

Hoffmann, M. and Donev, V., 2015. Probabilistic modeling of flexible pavement performance and LCC – based service lives with failure distributions. *In*: H. Furuta, D.M. Frangopol and M. Akiyama, eds. *Life-cycle of structural systems: design, assessment, maintenance and management*, 16-19 November 2014 Tokio. London: Taylor & Francis, 1981-1988.

Kuhn, K., 2012. Pavement network maintenance optimization considering multidimensional condition data. *Journal of Infrastructure Systems*. 18(4), 270-277.

Maerschalk, G. and Socina, M., 2008. *Weiterentwicklung der Bewertung des Pavement Management Systems (PMS) um ein Verfahren für die Umsetzung von Qualitätszielen*. [Extension of the evaluation functions of the pavement management system (PMS) by a system for the implementation of quality goals]. Bonn: Forschung Straßenbau und Straßenverkehrstechnik, Heft 1001 (in German).

Mallick, R. and El-Korchi, T., 2013. *Pavement engineering: principles and practice*. Boca Raton, FL: CRC Press.

Molzer, C., Felsenstein, K., Viertl, R., Litzka, J. Vycudil, A., 2000. *Statistische Methoden zur Auswertung von Straßenzustandsdaten*. [Statistics for pavement performance modeling]. Wien: BMVIT, Straßenforschung Heft 499 (in German).

Nakat, Z., Madanat, S., Farshidi, F. and Harvey, J., 2008. *Development of an empirical-mechanistic model of overlay crack progression using data from the Washington state PMS database*. Berkeley, CA: University of California Pavement Research Center, Research Report UCPRC-RR-2006-15.

Scazziga, I., 2008. *Massnahmenplanung im Erhaltungsmanagement von Fahrbahnen: Schadensprozesse und Zustandsverläufe* [Action planning within the management of road maintenance: pavement damage processes and performance curves]. Bern: Schriftenreihe 1304 (EP2), UVEK (in German).

SN 640 925b, 2003. *Erhaltungsmanagement der Fahrbahnen (EMF) - Zustandserhebung und Indexbewertung* [Maintenance management of roads (EMF) - condition survey and index rating]. Zürich: Schweizerischer Verband der Strassen- und Verkehrsfachleute (in German).

SN 640 925b Anhang, 2003. *Erhaltungsmanagement der Fahrbahnen (EMF) – Anleitung zur visuellen Zustandserhebung und Indexbewertung mit dem Schadenkatalog* [Maintenance management of roads (EMF) – guidance on the visual condition survey and index rating based on a distress catalog]. Zürich: Schweizerischer Verband der Strassen- und Verkehrsfachleute (in German).

Socina, M., 2007. *Erweiterung der Nutzung von Straßenzustandsdaten in der Systematischen Straßenerhaltung* [Extension of the use of pavement condition data in the systematic pavement maintenance]. Dissertation (PhD). Technische Universität Wien (in German).

Stewart, W.J., 2009. *Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling*. Princeton, NJ: Princeton University Press.

Stocker, H., 2014. *Einführung in die angewandte Ökonometrie* [An introduction to applied econometrics]. Manuskript zur Vorlesung „Methoden der Empirischen Wirtschaftsforschung“. Innsbruck: Universität Innsbruck (in German).

Weninger-Vycudil, A., Simanek, P., Haberl, J. and Rohringer, T., 2009. *Handbuch Pavement Management in Österreich 2009* [Manual pavement management in Austria 2009]. Wien: BMVIT, Straßenforschung Heft 584 (in German).

Wooldridge, J.M., 2013. *Introductory econometrics: a modern approach*. 5th ed. Mason, OH: South-Western, Cengage Learning.

ZTV ZEB-StB, 2006. *Zusätzliche Technische Vertragsbedingungen und Richtlinien zur Zustandserfassung und -bewertung von Straßen* [Additional technical terms of contract and guidelines for determining and assessing the condition of roads]. Köln: Forschungsgesellschaft für Straßen- und Verkehrswesen (in German).

CHAPTER 3

Condition prediction and estimation of service life in the presence of data censoring and dependent competing risks

by Valentin Donev & Markus Hoffmann

in *International Journal of Pavement Engineering*
Published online: 21 Feb 2017 (open access)
Published in print: Volume 20, Issue 3, 2019
DOI: 10.1080/10298436.2017.1293264

ABSTRACT

An accurate estimation of service life is of primary interest in pavement management systems (PMSs) limiting the time frame for maintenance and rehabilitation (M&R) treatments. Common condition prediction models are derived by regression analysis at the road network level based on empirical data from periodic condition surveys. If a particular section has not failed prior to the last survey or the condition has improved (e.g. due to treatment), it is considered as censored. If censoring is neglected the performance functions, service lives and estimated costs may show substantial bias. The authors who acknowledge this problem have used standard statistical (survival analysis) techniques accounting for censoring. However, any road section may fail due to different but dependent competing failure causes (risks), each leading to treatments. This constitutes a special type of censoring that cannot be addressed with traditional survival analysis (SA) methods relying on the assumption of independent censoring. As the number of failure causes usually exceeds one (e.g. fatigue, permanent deformation, thermal cracking), this case is quite common. Moreover, the time until a first failure depends on the sign and degree of correlation between present failure types being modelled by the overall survival function. This paper presents a critical review and comparison of common regression, Markov chain and survival analysis models with and without correlated competing risks based on computer-generated data. Using performance history and distress progression models at the section level in combination with survival analysis improves the accuracy of predictions in comparison. Furthermore, the paper proposes a simultaneous modeling of joint and marginal service life distributions based on copula functions as generalized solution accounting for dependence between competing risks. As the focus of this paper is on condition prediction with censored data, the distress-specific planning and optimization of treatments will be covered in forthcoming papers.

KEYWORDS: pavement management, survival analysis, data censoring, dependent competing risks, service life.

3.1 INTRODUCTION

3.1.1 Data censoring in pavement management

Performance models predict when, where and how (cause of failure) a pavement section will fail (deterministic) or may fail including uncertainties (probabilistic). In addition, performance models may be based on material behavior and pavement response (mechanistic) or survey data and statistical methods (empirical), or a combination of both (empirical-mechanistic). Currently, mainly pure empirical or empirical-mechanistic (E-M) models are applied for prediction in pavement management at the network level (Schram and Abdelrahman 2009, Kargah-Ostadi and Stoffels 2015). However, due to technical, economical and practical reasons, it is not possible to continuously observe the condition deterioration for all road sections during the entire service life. On the contrary, in most cases only data from two consecutive surveys is available. Moreover, survey methods change over time due to technological advancements and maintenance history is not always properly documented, especially for regional and local roads. Thus, the available data for an estimation of service life and development of performance models is censored. If this property of the empirical survey data is neglected, the estimated service life will be over- or underestimated depending on the type of censoring.

In this paper failure is defined as exceeding a given condition threshold triggering the application of maintenance and rehabilitation treatments. In general, condition thresholds differ among countries, agencies and road classes and have been set more or less arbitrarily (Hoffmann and Donev 2016). Figure 3.1 provides an overview of all common types of censoring in pavement management. If a road section has failed prior to a first survey, it is considered left censored, i.e. the true (latent) service life is less than the observed service life. Interval censoring

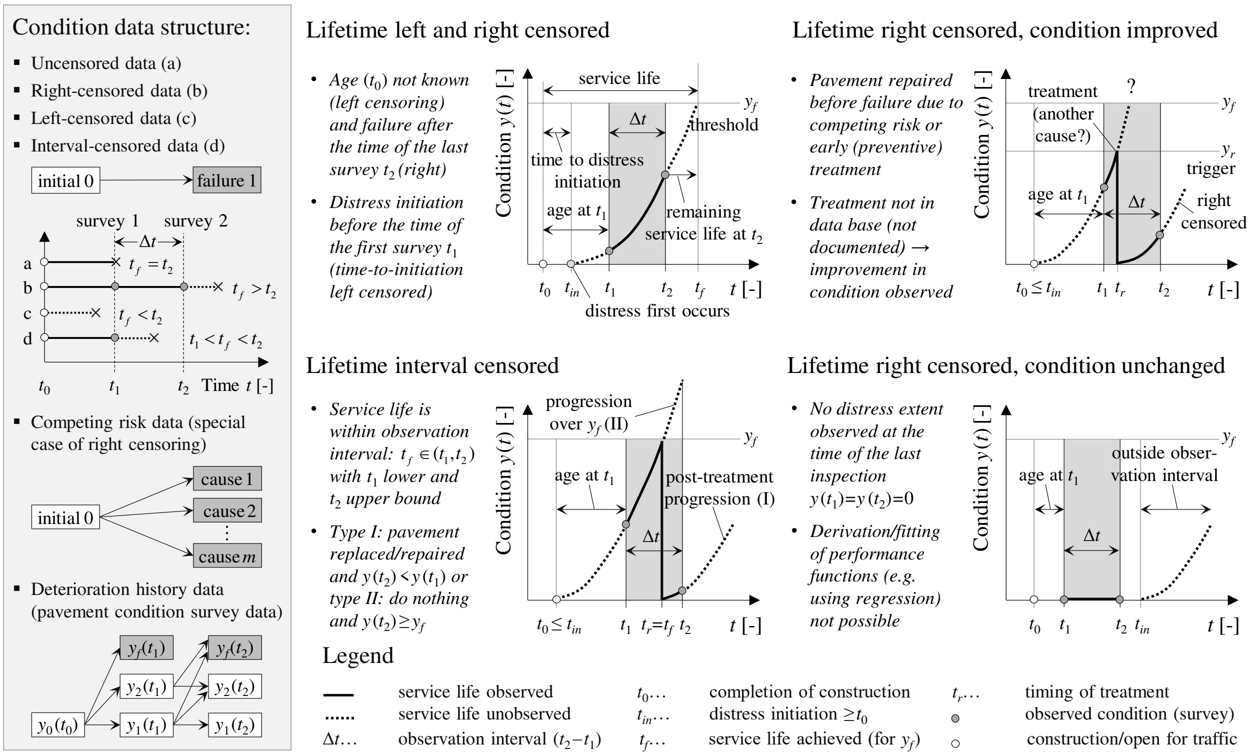


FIGURE 3.1 Common types of censoring of the lifetime (service life) of pavements in the context of continuous condition deterioration, application of treatments and periodic condition surveys.

occurs if the true lifetime is within a known interval of time (e.g. the survey interval Δt). If a pavement has not failed by the time of the last survey, it is considered right censored, i.e. the observed service life is less than the true service life.

In addition, there are a few special features of condition survey data that have to be considered as well. First, the pavement condition deteriorates continuously reaching different intermediate states before failure in contrast to typical failure-time data with only two discrete states (failure and non-failure). Thus, a model that incorporates this section-specific degradation history will always be superior to other models. Second, a flexible pavement may fail due to one or more failure causes/mechanisms (e.g. rutting, fatigue, low temperature) and a service life can be computed with respect to each failure type as well as an overall service life (time until failure of any cause). Thus, a pavement may be repaired before reaching the end of its service life due to a competing risk failure or due to an early preventive treatment. In this case the failure event cannot be observed in future condition surveys. Finally, for a network with high maintenance standards (highways) the majority of the sections will be in good condition exhibiting no signs of distress. For this case considering only failing sections in the analysis will lead to an underestimation of service life and an overestimation of investment needs. In summary, the provided examples of data censoring are quite common and have to be addressed with appropriate statistical methods.

The reminder of this paper is organized as follows. Sections 3.1.2 and 3.1.3 provide an overview of the common survival analysis methods and some applications of these methods in pavement condition data analysis. In Section 3.2.1, the reasons for using a simulation approach and key settings are explained. In Sections 3.2.2 to 3.3.3, the deviations resulting from the application of standard approaches (regression, Markov chains, SA) to the generated data are presented and analyzed. Sections 3.3.4 and 3.3.5 present innovative ways to overcome the described drawbacks of common models leading to an improved estimation of true service lives. The conclusions regarding comparative analysis and practical application are presented in Section 3.4.

3.1.2 Common survival analysis models

Survival analysis (biostatistics), failure-time analysis (engineering), event-history analysis (sociology) or duration analysis (econometrics) provide statistical techniques to estimate the expected time until the occurrence of an event. This event may be an initiation of a distress or failure of a pavement section exceeding a specified condition threshold for a given type of distress. The survival function of a section regarding such an event gives the probability of surviving until time t, while the hazard function gives the instantaneous potential for failure, given survival up to time t (Figure 3.2(a)). The relationship between survival function $S(t)$, probability density $f(t)$ and hazard function $h(t)$ is of fundamental importance in survival analysis (Kalbfleisch and Prentice 2002, Crowder 2012, Kleinbaum and Klein 2012).

The construction of the parametric likelihood as illustrated in Figure 3.2(b) is a necessary foundation for understanding the mathematics of survival analysis and correct interpretation of results. The likelihood function L is a function of the observed data, unknown parameters and the probability density of the assumed lifetime distribution. For right-censored observations with failure occurring after the time of observation t, the likelihood contribution is given by the survival function $S(t)$. Left-censored observations (already failed sections) are included in L

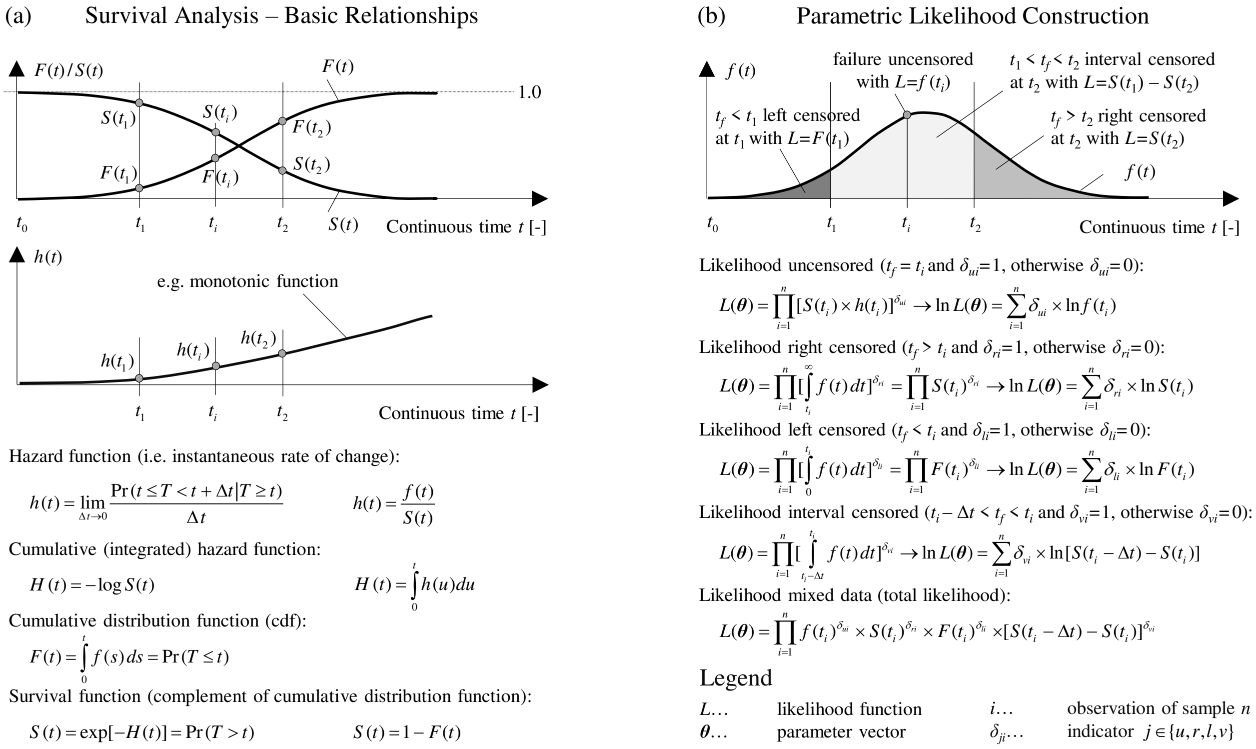


FIGURE 3.2 Relationship between probability density, hazard and survival function (a). Parametric likelihood estimation accommodating uncensored, left-, interval- and right-censored data (b).

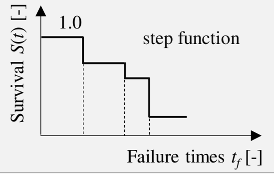
using the cumulative distribution function $F(t)$. For interval-censored data the likelihood is the difference between survival function estimated at the beginning and the end of the interval. The total likelihood is calculated as a product of n individual likelihood contributions assuming independence between different contributions, i.e. independence between the lifetimes or censoring times of the individual sections. The maximum likelihood estimation (MLE) method maximizes the total likelihood defined as the joint probability of obtaining the observed data with respect to the set of parameters.

Figure 3.3 provides an overview of common survival-analysis approaches including examples based on Kleinbaum and Klein (2012). The non-parametric Kaplan-Meier (KM) product-limit estimator is based on conditional probability and makes no assumptions about the form of the lifetime distribution. As with other survival-analysis models the validity of the estimation relies on the assumption that the censoring mechanism is statistically independent of the failure process (independent censoring assumption). As attractive as a nonparametric approach might seem, the Kaplan-Meier method cannot accommodate left- and interval-censored data. Out-of-sample prediction or estimation of survival probability beyond the last failure time is also not possible. As a further disadvantage, the model can only account for time-constant categorical explanatory variables (e.g. freeze/non-freeze zone) by stratifying the data and separate analysis.

Another popular method is the semi-parametric Cox proportional hazard (PH) model allowing for simultaneous control of covariates without specifying the baseline hazard function. Continuous time-independent variables and interaction terms (product of variables) may be used as explanatory variables. The proportional hazard assumption states that a covariate has a multiplicative effect on the hazard rate that is constant over time. In other words, the ratio of the hazard functions for two different sets of explanatory variables (hazard ratio, HR) does not depend on the time. A time-dependent predictor variable may be accommodated by the extended

Non-parametric models:

- Handle only right-censored data
- Do not allow simultaneously controlling for covariates



Kaplan-Meier estimate (Product limit formula):

$$S(t_f)=\prod_{i=1}^{f}\Pr\left(T>t_i \middle| T\geq t_i\right)$$

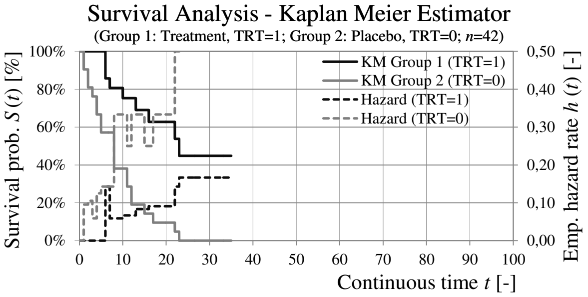
$$S(t_f)=S(t_{f-1})\times\Pr\left(T>t_f \middle| T\geq t_f\right)$$

95% Confidence intervals (CI) for KM curves:

$$\hat{S}(t)\pm 1.96\times\sqrt{V\hat{a}r[\hat{S}(t)]}$$

Greenwood’s variance formula:

$$V\hat{a}r[\hat{S}(t)]=\left(\hat{S}(t)\right)^2\times\sum_{t_f<t}\frac{m_f}{n_f\times(n_f-m_f)}$$



Semi-parametric models:

- Handle only right-censored data
- Estimated survival function may not go down to 0 (applies for KM)
- Baseline hazard not specified → computation difficult (requires approximation and iteration)
- Partial likelihood depends only on the ordering of survival times, but not on actual values
- *L* with ties: Breslow, Efron, exact

Cox hazard and survival function:

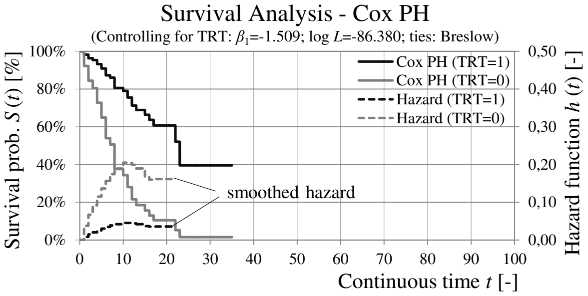
$$h(t,\boldsymbol{X})=h_0\times\exp\left(\sum_{j=1}^{k}\beta_i\times X_i\right)$$

$$S(t,\boldsymbol{X})=S_0(t)\times\exp\left(\sum_{j=1}^{k}\beta_i\times X_i\right)$$

$$S_0(t)=\exp\left[-\int_0^t h_0(u)du\right]=\exp[-H_0(t)]$$

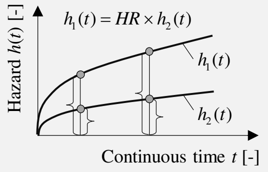
Partial Likelihood (no ties):

$$L_p(\beta)=\prod_{i=1}^{n}\left[\frac{\exp\left(\sum_{j=1}^{k}\beta_i\times X_i\right)}{\sum_{l\in R(t_i)}\exp\left(\sum_{j=1}^{k}\beta_i\times X_i\right)}\right]^{\delta_i}$$

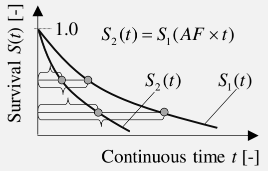


Parametric models:

- The true survival time follows a specific distribution
- PH: ratio of hazards for 2 sets of predictors is constant over time



- AFT: ratio of survival times is constant for fixed values of $S(t)$



- The hazard function of Weibull (Log-Logistic) model is monotonic (nonmonotonic)

Weibull PH model:

$$\lambda=\exp\left(\beta_0+\sum_{j=1}^{k}\beta_i\times X_i\right)$$

Weibull AFT model:

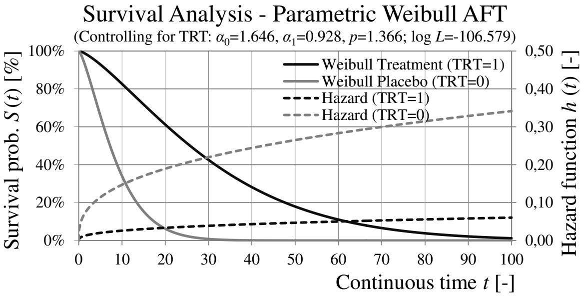
$$\frac{1}{\lambda^{1/p}}=\exp\left(\alpha_0+\sum_{j=1}^{k}\alpha_i\times X_i\right)$$

Weibull survival function:

$$S(t)=\exp[-(\lambda\times t)^p]$$

Weibull hazard function:

$$h(t)=\lambda\times p\times(\lambda\times t)^{p-1}$$



Log-Logistic AFT model:

$$\frac{1}{\lambda^{1/p}}=\exp\left(\alpha_0+\sum_{j=1}^{k}\alpha_i\times X_i\right)$$

Log-Logistic Proportional Odds (PO):

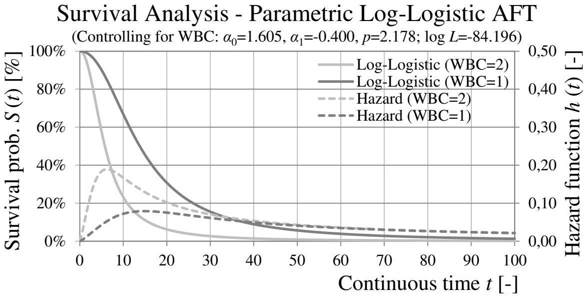
$$\lambda=\exp\left(\beta_0+\sum_{j=1}^{k}\beta_i\times X_i\right)$$

Log-Logistic survival function:

$$S(t)=\frac{1}{1+(\lambda\times t)^p}$$

Log-Logistic hazard function:

$$h(t)=\frac{\lambda\times p\times(\lambda\times t)^{p-1}}{1+(\lambda\times t)^p}$$



Legend

h_0…	baseline hazard function (not specified)	m_f…	number of failures at each time t	δ_i…	indicator variable δ_i=1 if t_i is uncensored and δ_i=0 otherwise	n…	total number of observations
β_i…	regression coefficients	n_f…	sections in the risk set at start of interval			HR…	hazard rate
X…	vector of k covariates			$R(t_i)$…	risk set at time t_i	AF…	acceleration factor

FIGURE 3.3 Overview of classic non-parametric, semi-parametric and parametric models in survival analysis together with their characteristics and limitations.

Cox model that no longer fulfils the PH assumption. Apart from the inclusion of covariates, the Cox model exhibits the same limitations as the KM method. If no covariates are used, as in this paper, the Cox model reduces to the baseline survival function $S_0(t)$, which is identical to the KM estimate. For this reason, semi-parametric models are not analyzed any further.

In parametric survival analysis the survival and the hazard are fully specified smooth functions being convenient for subsequent optimization and simulation. Non-parametric estimates, laboratory tests or theoretical assumptions may deliver insights about the selection of an appropriate model form with the exponential, Weibull and log-logistic distributions being very common. Parametric models account for left-, interval- and right-censoring providing even out-of-sample estimates beyond the last failure time. The survival time may be modelled as a function of explanatory variables based on underlying proportionality assumptions (e.g. proportional hazards, accelerated failure time or proportional odds). Parametric models may also incorporate a frailty component to account for unobserved heterogeneity among pavement sections (Kleinbaum and Klein 2012).

In PMS survival analysis may be used as an alternative or better in addition to regression analysis (or other continuous time and state space models). The process of crack initiation as well as the occurrence of some random distresses (e.g. settlement, edge breaks) can be described properly with survival and not regression analysis. However, duration models may be employed also for an estimation of the whole service life. Survival-analysis models require less survey data (one observation for a section) and less precision (status failed/not failed). Thus, with regard to the estimation of service life, SA is less susceptible to inconsistent survey values over time (non-monotonic progression), measurement errors and outliers compared to regression analysis. The obtained robust estimate of the mean service life may then be used for correction and calibration of performance functions developed with regression if the functional form is known. Providing only a survival probability (percentage of surviving/failed section at network level) and no information about the condition progression before failure may be seen as a disadvantage of survival analysis. However, combining both approaches or using even more advanced techniques may help to overcome these limitations.

3.1.3 Literature review

There is extensive literature on data censoring and application of survival models in pavement management starting with the cracking-initiation model in the highway development and management model (HDM) of the World Bank (Paterson 1987, Van Dam et al. 1997). The outcome variable in all proposed models was either time (age) until failure/distress initiation or the number of cumulative equivalent single-axle loads (ESALs). Most of the research work was primarily concerned with right censoring and incorporated a large number of explanatory variables selected by statistical tests. A number of researchers estimated the time until crack initiation using parametric (Shin and Madanat 2003, Loizos and Karlaftis 2005) or semi-parametric models (Nakat and Madanat 2008) based on data either from accelerated pavement tests (experimental data) or from condition surveys of in-service pavements (field data). Cracking-initiation models were also derived by joint estimation using both experimental and field data (Christofa and Madanat 2010, Reger et al. 2013). A fewer studies investigated the time to failure (service life) of pavements with failure defined as time from construction/overlay until exceeding either a terminal present serviceability index (PSI) (Prozzi and Madanat 2000) or a fatigue cracking threshold value (Wang et al. 2005). Dong and Huang (2015) evaluated the effect of preventive maintenance treatments separately for different types of distress but did not address overall service life. Yang and Kim (2015) proposed a non-parametric approach with cracking, rutting and roughness as competing risks, but did not attempt to derive marginal distributions nor did they account for correlation between considered types of distress. In summary, accounting for competing risks is the exemption and there are no known models accounting for multiple failure mechanisms based on marginal distributions.

Apart from very few exceptions, the problem of data censoring is generally neglected in German literature. The pavement management systems in Austria and Germany are based on the Canadian software VIAPMS (dTIMS) and employ deterministic prediction models derived with regression analysis without accounting for censoring excluding road sections exhibiting no distresses (Molzer et al. 2000). Rübensam et al. (2005) compared regression and survival analysis for the evaluation of long-term performance of pavements. Unfortunately, the analysis was limited only to Kaplan-Meier estimates and did not consider correlations between distress types. Nevertheless, the data showed extensive amounts of censoring within the defined failure criteria. Despite the principal agreement of all authors that censoring may lead to bias there is no paper with a systematic investigation of censoring impacts and analysis methods on resulting service lives of pavements available.

3.2 STATE-OF-THE-ART ESTIMATION OF SERVICE LIFE

3.2.1 Distress modelling and data-generation procedure

When analyzing real-world condition data, the true survival time as well as the underlying degradation process are always unknown. Goodness-of-fit measures (e.g. R^2, AIC) describe how well the model fits the observed data but do not account for unobserved data or the validity of model form and specification. Therefore, it is not possible to determine a priori which model will perform better under given circumstances. With the focus on model comparison and bias identification, this paper adopts an approach based on computer-generated data (Donev and Hoffmann 2016). The main advantage of this approach is that when applying different models to the artificially censored simulated data, the deviations from the “true” complete data are known and can be quantified. Once established, the better performing models can then be applied to real-world condition data allowing for practical conclusions about pavement deterioration and service life. Furthermore, a simulation approach allows for the comparison of models under different settings (e.g. different network age distributions, inspection intervals, number of observations) by means of sensitivity analysis. In contrast, the results of empirical-data applications (even with large sample sizes) often apply for the specific case and cannot not be generalized or transferred to other cases. The simulation as well as the subsequent analysis and all computations are conducted semi-manually using Microsoft Excel 2010.

This study concentrates on two types of distress, fatigue cracking and rutting, as dominant failure causes for flexible pavements at intermediate and high temperatures (Mallick and El-Korchi 2013, Norouzi et al. 2016). The approaches described here may be extended to three or more distress types. The individual steps in modelling and simulation are explained in Figure 3.4. The distress type cracking is modelled as a combination of two stochastic processes (i.e. initiation and progression) as commonly suggested in the literature (e.g. Paterson 1987, Nakat and Madanat 2008). The time-to-failure (service life) with regard to cracking t_{CR} is drawn from a Weibull distribution with assumed parameters for each of 1000 fictive sections using Monte Carlo simulation (MCS). The time to distress initiation t_{INI} is generated for the same 1000 sections using a different Weibull distribution (conveniently with the same shape parameter). The graphic on the bottom left (Figure 3.4(c)) shows the cumulative distribution functions (cdf)

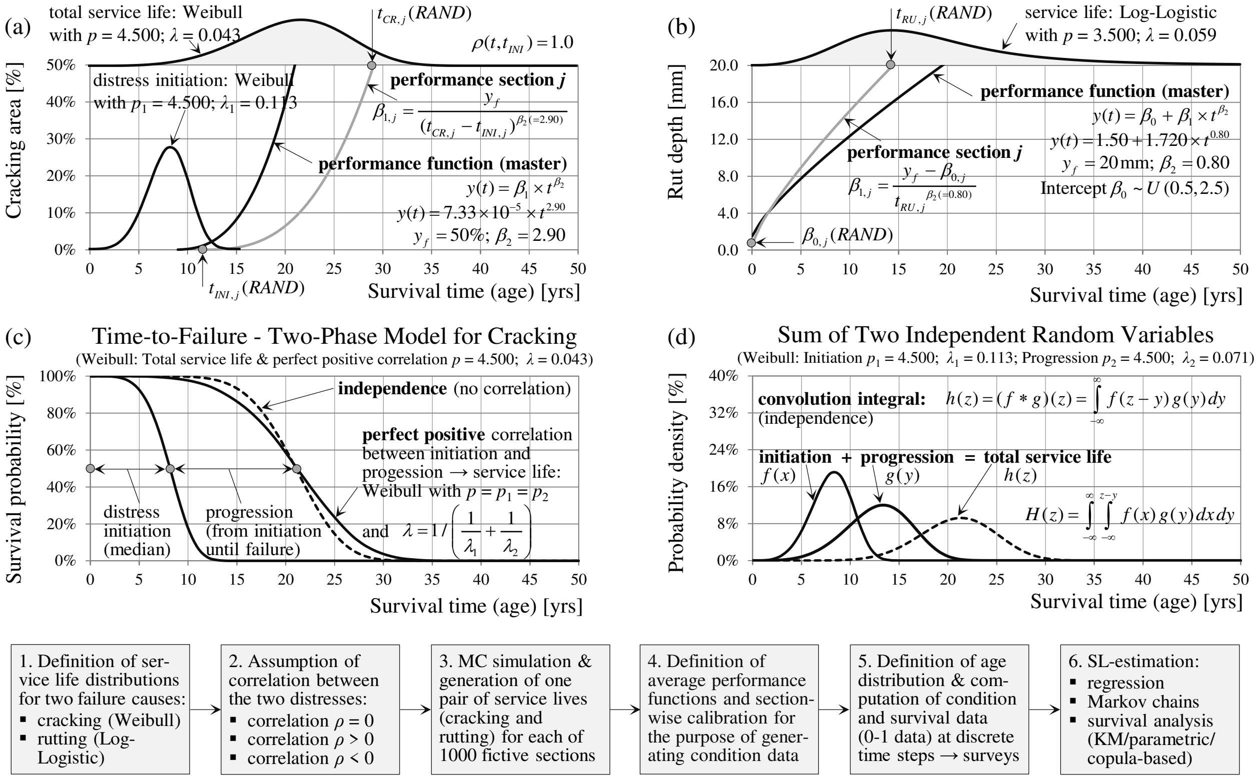


FIGURE 3.4 Data-generation procedure based on assumed service life (failure) distributions and performance functions with age as a dependent variable (a, b). Modeling of pavement cracking as a combination of two processes - initiation and progression (c, d).

for the time until distress initiation as well as the whole service life. The process of crack progression follows a third distribution being the difference between total service life and crack initiation. As shown in Figure 3.4(d), the total service life $h(z)$ is the sum of initiation $f(x)$ and progression $g(y)$ processes. Moreover, the initiation and progression distribution are usually correlated, since factors causing earlier initiation of cracks will lead to faster progression (e.g. traffic loads). If the two processes are assumed to be independent, the total service life can be computed using the convolution integral (Ross 2014). For simplicity, we assume perfect positive correlation between initiation and cracking. Thus, for a given section j, the same random number is used for the calculation of the time to distress initiation and the total service life (or the time for distress progression). As shown in Figure 3.4(c) the difference between independence (no correlation) and perfect correlation is not very large regarding the cdf.

Furthermore, the distress progression and the intermediate conditions until failure can be described with a power function – an approach based on the Hoffmann-process (Hoffmann and Donev 2015, Hoffmann and Donev 2016). The power parameter β_2 is assumed to be characteristic for each distress type and constant for all sections. In contrast, the scale parameter β_1 represents the different rate of distress progression among different sections. The parameter β_1 is estimated using β_2 (2.90), the condition threshold y_f (50%) and the generated section-specific total service life $t_{CR,j}$ together with initiation time $t_{INI,j}$ (Figure 3.4(a)). Thus, the extent of cracking can be computed for each section at any time. The same procedure is applied for the distress type rutting but without initiation process (Figure 3.4(b)). The service life is based on a log-logistic distribution (arbitrary) with the corresponding assumptions for β_2 (0.80) and y_f (20 mm). However, a different set of random numbers was used this time. One difference is that the

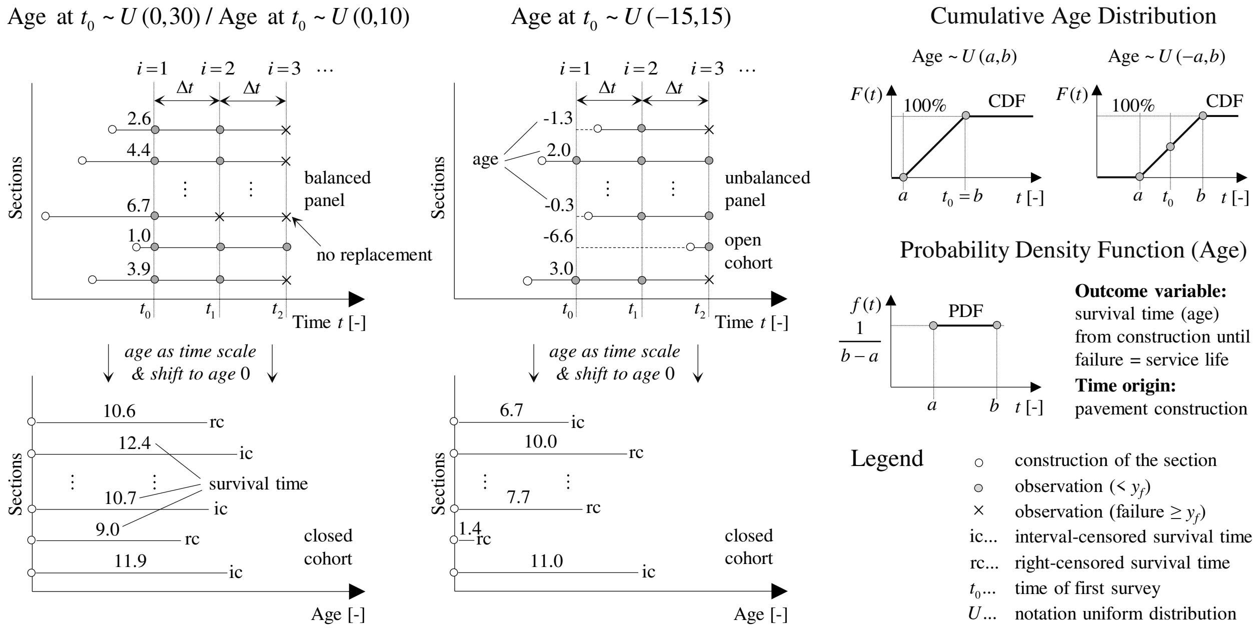


FIGURE 3.5 Generation of different age distributions (uniform). Definition of survival time as time from construction until failure with the corresponding transformation to age as the time scale.

power function for rutting also includes an intercept term (β_0), which is assumed to be uniformly distributed among sections. The assumed parameters of the service-life distributions and performance functions are selected to be in accordance with the existing literature. However, it is important to point out that the aim here is to generate data as a basis for a thorough comparison of methods and analysis of resulting bias. Thus, the paper provides readers with the knowledge and means to obtain less biased results from their real-world data but draws no conclusions on actual service lives of random pavement data as (out of scope). As a next step, age (i.e. the explanatory variable in the performance function) is generated for each section using a uniform distribution (Figure 3.5). Consequently, cracking area and rut depth are calculated for (seven) different discrete moments in time starting from t_0 representing consecutive condition surveys. The computations are carried out for three different survey intervals and three different age distributions. The age distribution with a negative minimum value represents an unbalanced panel, since each section is observed for a different number of times and only half of the sections are constructed at time t_0. However, the analysis is further conducted using age as the time scale, even though sections reach the same age at different calendar times.

Until now the question about the relationship between cracking and rutting failure times at section level has been neglected here. This study considers three different cases of dependence between the two types of distress illustrated by the scatter plots in Figure 3.6. The first case assumes independence between service life with regard to rutting and cracking (Figure 3.6(a)). It must be noted that independence implies zero correlation, but the converse does not necessarily hold true. The second case, as shown in Figure 3.6(b), is based on mechanistic principles suggesting negative dependence (Smith et al. 2007, Ozer et al. 2016). In practice stiffer asphalt binders are associated with higher cracking potential but better resistance to rutting. Less stiff asphalt binders are more prone to rutting but less susceptible to cracking. The dynamic shear rheometer (DSR) is used for testing the cracking and rutting potential of asphalt binder under laboratory conditions (Mallick and El-Korchi 2013). The third case of positive dependence has

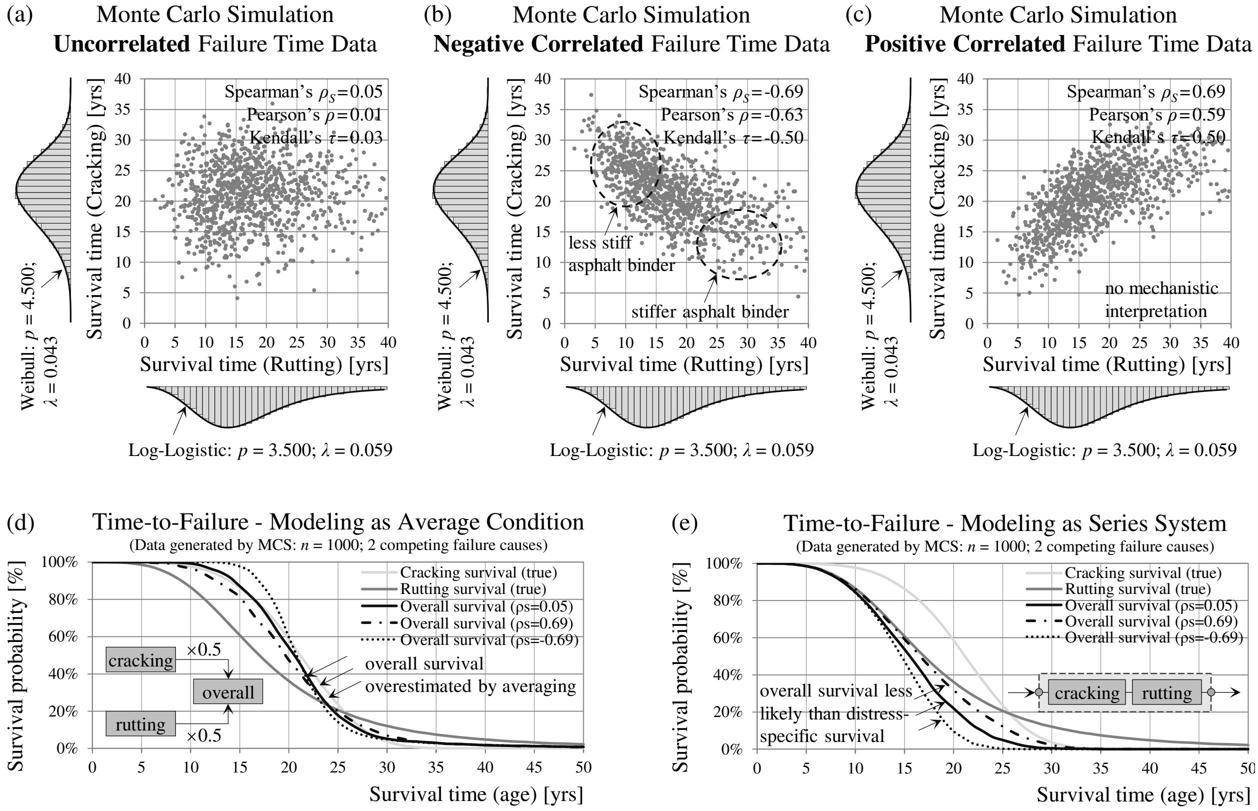


FIGURE 3.6 Scatter-plot of the generated rutting vs. cracking survival times given the same marginal (distress-specific) distributions, but different correlation. Modeling the overall (system) pavement survival as a series system or as a composite index aggregating single distresses (standard approach).

no meaning for these failure types from a mechanistic point of view, but it is considered for reasons of comparison (Figure 3.6(c)). However, most distresses are expected to be positively correlated, because they share common influencing factors. It is important to point out that the used marginal lifetime distributions for the distress types cracking (Weibull) and rutting (log-logistic) are the same in all three cases. The main difference relates to the ordering of the failure times for individual sections. Thus, in the case of negative correlation, long-lived sections in regard to cracking are more likely to be short-lived in regard to rutting and vice versa.

In a situation with competing failure causes the time until a first failure (i.e. the overall survival) might be of interest. Common PMSs use rating and aggregation procedures of subjective nature in order to obtain one total condition for the road section as a basis for the optimization of treatments. In these systems, as a first step, distress extent or severity are converted into a dimensionless index (school grade) using distress-specific rating scales. In a second step these ratings for different distresses are weighted and combined into one single overall condition index (Hoffmann and Donev 2016). In this example the indices for cracking and rutting are combined by taking the arithmetic average. Alternatively, the pavement as a whole might be modelled with the joint survival function, or what is known as a series system in reliability theory (Figure 3.6(e)). In that case the joint (system/overall) survival depends on the correlation between the two distresses and falls below the marginal survival probabilities in all three cases with negative correlation leading to the most unfavorable outcomes. As shown in Figure 3.6(d), the overall survival based on the arithmetic average leads to an overestimation of service life in comparison to the true series-system survival. This holds true even for a different combination of distress weights and marginal distributions. Thus, a systematic overestimation of service life and underestimation of M&R costs is to be expected in any PMS using such overall condition index in the objective function for optimization (Donev and Hoffmann 2016).

3.2.2 Nonlinear regression and performance prediction

Observations in the generated data sets are already left- and right-censored depending on the age distribution. All failed sections are interval-censored because of the survey intervals. The next step is to artificially censor sections due to competing risks. With competing risks only the earliest (first) failure can be observed and failures due to other causes are treated as censored. This holds true if reaching a defined failure threshold for cracking or rutting triggers an immediate application of a treatment restoring or improving the condition. In reality, treatments may be applied only after actually observing the failure (i.e. after periodic surveys) or based on an accurate prediction model at the section level. Furthermore, the pavement may fail due to two causes occurring simultaneously or one after the other (especially in the case of positive correlation). In this paper it is assumed that failures can occur due to only one cause at a time and section, but the problem is quite complex even with this minor simplification.

A standard approach in PMS consists of fitting a regression function to observed continuous data (Figure 3.7) using the ordinary least squares (OLS) method (Wooldridge 2013). In a regression model the dependent variable is not service life but condition $y(t)$ with censoring affecting the condition distribution. For example, in the case of rutting all observations ≥ 20 mm (failure threshold) are censored. In addition, on sections where cracking failure precedes rutting failure, all rutting observations after the cracking failure are censored (competing event). In the case of positive correlation (Figure 3.7(a,b)) short-lived sections have in average fewer uncensored observations until failure, resulting in enormous overestimation of the service life for rutting and cracking (+15 and +10 years, respectively). In the case of negative correlation (Figure 3.7(c,d)) long-lived sections with respect to rutting and cracking are censored as well, reducing the overall bias. Figure 3.7(e) shows that using all observations beyond the failure threshold (no censoring) in regression analysis leads to less biased results. Furthermore, prediction intervals, condition distribution and survival function can be estimated under the assumption of normally distributed error terms allowing for a comparison with the true survival function. In summary, standard regression analysis (OLS) does not account for data censoring resulting in erroneous performance functions and service-life estimates. Moreover, the process of crack initiation cannot be properly modelled with regression analysis leading to additional bias.

3.2.3 Homogeneous Markov chains

Markov chains are a class of stochastic processes describing the transitions of a system between a finite number of states, where the future evolution of the process depends only on the current state and not on the past history (memoryless property). The transition probabilities may be assumed constant (homogeneous) or changing over time (non-homogeneous). Transitions may occur at any time (time continuous) or only at specified time steps (time discrete). The survival time (service life) is defined as the time spent in all states prior to entering the absorbing state. An absorbing state is the failure state ($\geq y_f$), from which no further transitions to other states are possible (no treatments). Computational details for Markov chains can be found in the literature

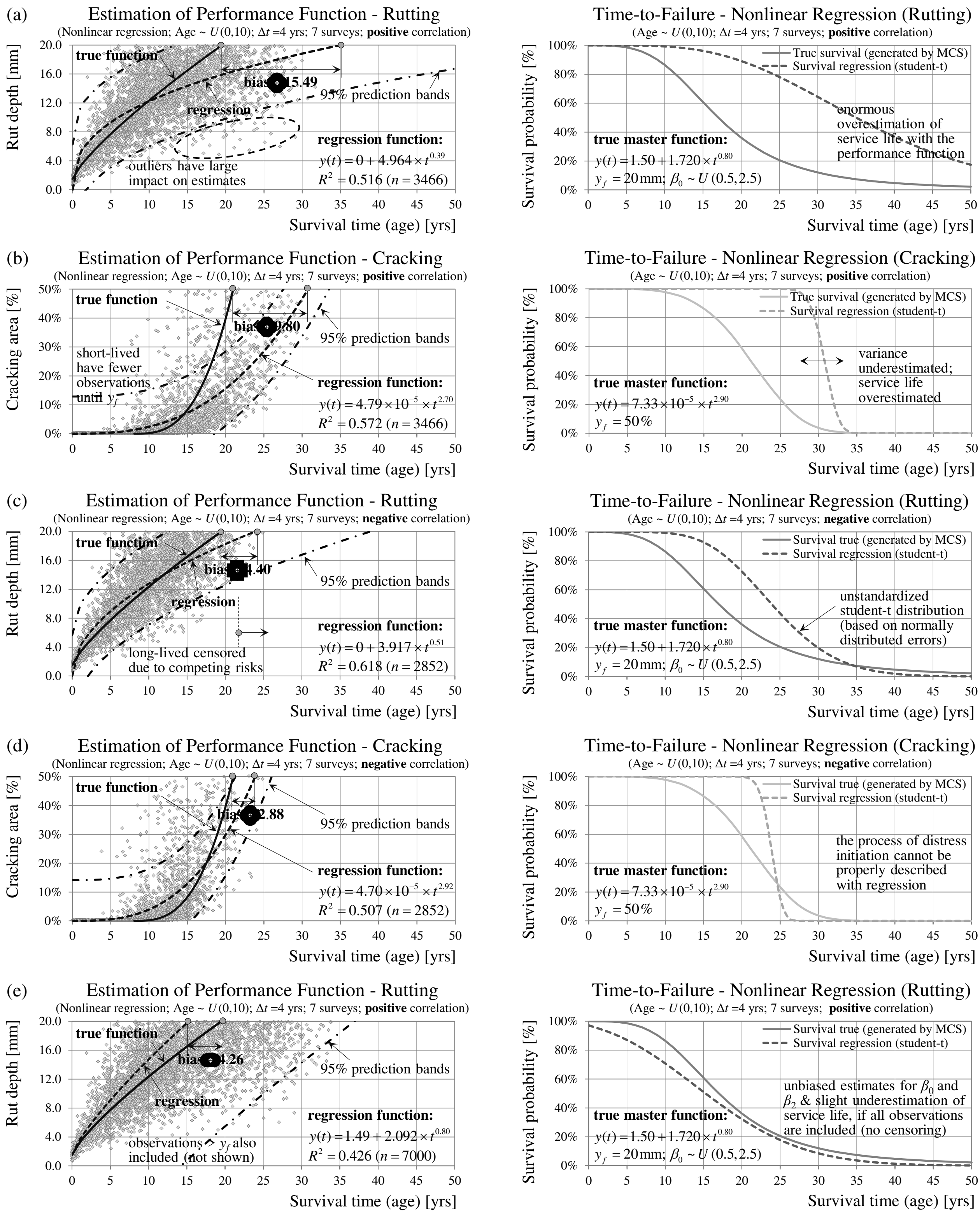


FIGURE 3.7 Estimation of performance and survival functions with nonlinear regression using the generated (censored) condition data and resulting bias with respect to true marginal distributions, mean lifetimes and performance functions.

and are not provided here (e.g. Crowder 2012, Ross 2014). Figure 3.8 illustrates a relationship between Markov chains and survival analysis based on a simple two-state (failure/non-failure) inhomogeneous Markov chain (Aalen et al. 2008). The survival probability is equal to the transition probability of remaining in a non-failure state $p_{00}(t)$. The transition probability for the

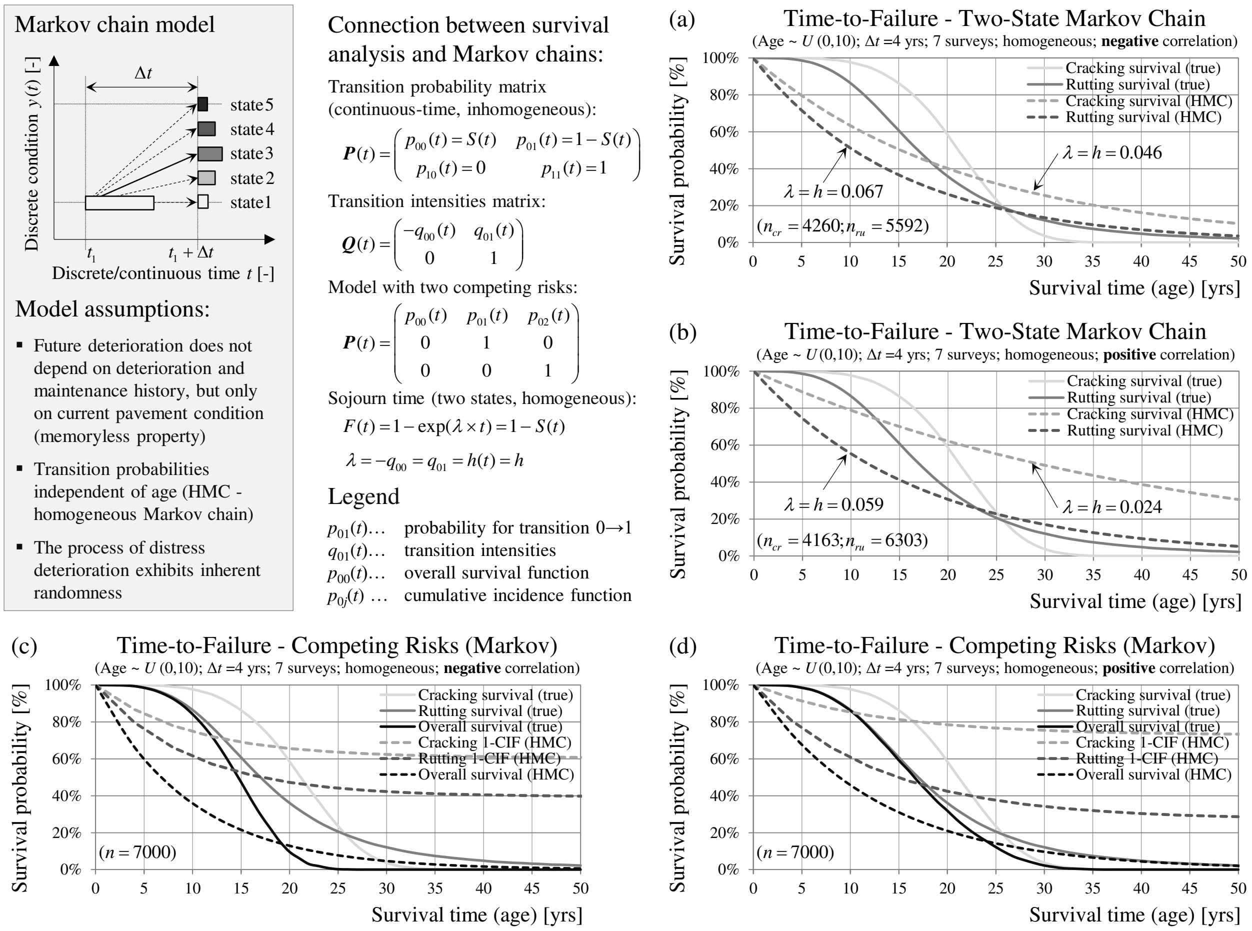


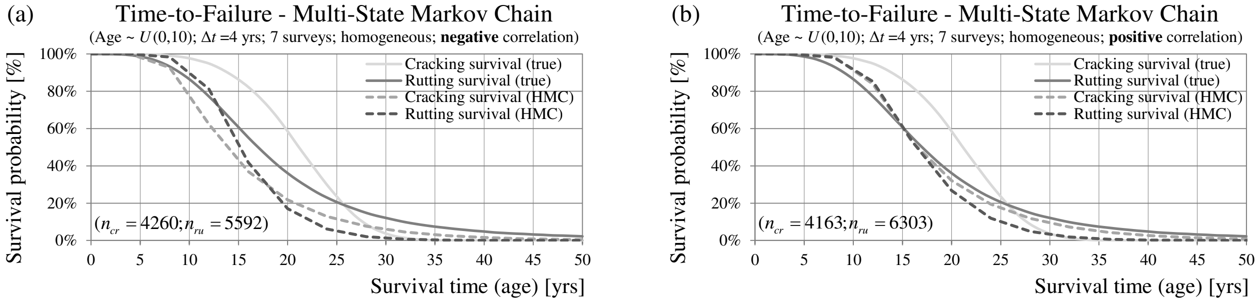
FIGURE 3.8 Relationship between survival analysis and Markov chains based on the example of a two-state (a, b) and a competing-risks (c, d) Markov chain.

complementary failure event is the cumulative distribution function $F(t)$. The transition intensity (i.e. the rate at which transitions occur) of changing state $q_{01}(t)$ correspond to the hazard function $h(t)$. Moreover, in the case of constant transition probabilities and continuous time, the time spent in a non-failure state (i.e. sojourn time) is exponentially distributed with parameter $-q_{00}$. Thus, the survival time in the two-state homogeneous Markov chain follows an exponential distribution with constant hazard function over time.

Figure 3.8(a,b) provides an application of such a two-state homogeneous Markov chain model to the generated distress data with resulting survival functions. The computation is based on the full set of 7 surveys (continuous condition) and 4 years inspection interval (transition step). For a section where failure has occurred due to one cause, further observations of all other distress types are censored (competing risk censoring). The transition probabilities are estimated by counting of sections. In the case of negative and positive correlation the deviations from the true model are enormous due to the inflexibility of the exponential distribution. In the case of negative correlation, the estimated hazards are higher.

The Markov chain model can be formulated as a competing-risks model with more than one absorbing state (Figure 3.8(c,d)). The probability of remaining in a non-failure state is defined as overall survival function. The transition probability of reaching one failure type is defined as cumulative incidence function for this type but has limited practical value (see Section 3.3.3). An improvement of the resulting estimates of service life may be achieved by using information about the condition states prior to failure. Figure 3.9(a,b) shows a multi-state model with five condition states based on a rating scale. At this point it has to be noted that Markov chain models do not account for interval censoring and dependencies between competing failure causes. Left-censored observations are included in the absorbing state, which therefore plays no role in the computation. Observations of right-censored sections are included in the estimation of the transition probabilities, but the model does not explicitly account for right-censoring either. The example also shows that the number of states has a significant impact on the predicted service life. However, any approach for state classification based on rating scale or else is more or less arbitrary. Despite these shortcomings further slight improvements of model accuracy can be achieved by calculation of separate transition matrices for different age groups quite similar to cohort models (Figure 3.9(c,d)).

The Markov chain approach is popular because it allows for modeling of M&R treatments and the use of standard optimization techniques (dynamic programming). An often made assumption is that within one transition the condition will remain in the current state or will change to next worse state (Butt et al. 1987, Abaza 2017). Although this assumption seems plausible, it is often made without justification in order to reduce the number of unknown probabilities and to simplify the computations. In general, limitations of Markov models include cumbersome estimation of transition probability matrices, small number of observed transitions to states with poorer condition and no section-specific prediction. Explanatory variables may be considered only by further segmentation of the observations in homogeneous groups at the cost of reducing the sample size (Madanat et al. 1995). Furthermore, the assumption of constant transition probabilities in the homogeneous case does not hold true for pavement deterioration due to cumulative traffic and ageing (Li et al. 1996, Hoffmann and Donev 2016, Abaza 2017).



$$P = \begin{pmatrix} 182 & 409 & 42 & 4 & \boxed{1} \\ 0 & 287 & 626 & 71 & \boxed{6} \\ 0 & 0 & 209 & 582 & \boxed{94} \\ 0 & 0 & 0 & 87 & \boxed{566} \\ 0 & 0 & 0 & 0 & \boxed{29} \end{pmatrix} \begin{matrix} \rightarrow \text{failed (rutting, positive correlation) and interval-censored} \\ \rightarrow \text{left-censored} \end{matrix}$$

$$P = \begin{pmatrix} 0.285 & 0.641 & 0.066 & 0.006 & 0.002 \\ 0.000 & 0.290 & 0.632 & 0.072 & 0.006 \\ 0.000 & 0.000 & 0.236 & 0.658 & 0.106 \\ 0.000 & 0.000 & 0.000 & 0.133 & 0.867 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{pmatrix}$$

State	Cracking [%]	Rutting [mm]
State 1	0-0.9	0-7.6
State 2	0.9-6.7	7.6-12.1
State 3	6.7-21.7	12.1-16.2
State 4	21.7-50.0	16.2-20.0
State 5	≥50.0	≥ 20.0

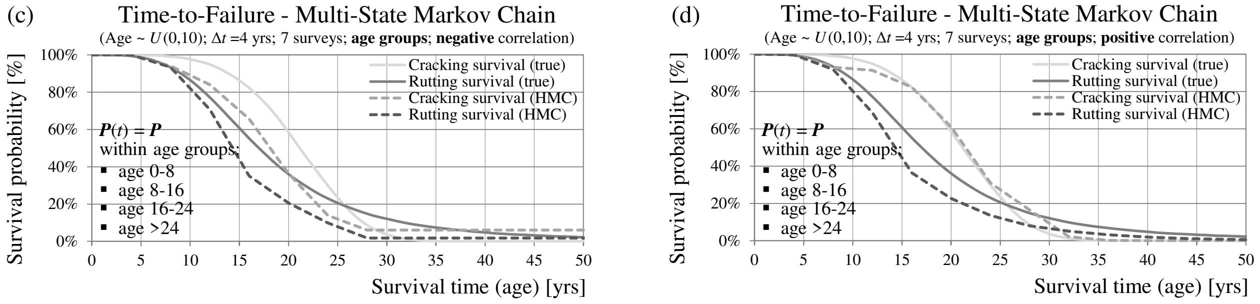


FIGURE 3.9 Survival probabilities based on a five-state Markov chain with transition probabilities being constant for the full analysis period (a, b) or within age groups (c, d).

3.3 SURVIVAL ANALYSIS FOR ESTIMATION OF SERVICE LIFE

3.3.1 Survival analysis in the absence of competing risks

In this section mathematical models accounting for censoring are applied to the generated data. At first it is assumed that failure may occur only due to cracking (i.e. no competing risks). The aim is to investigate the influence of survey interval, number of surveys, age distribution and censoring pattern on the estimation accuracy under ceteris paribus conditions with the non-parametric Kaplan-Meier method. Figure 3.10(a,b) shows that wider survey intervals and a higher number of surveys are desirable to obtain a complete estimate of the survival function, but also increase the share of interval-censored sections. Furthermore, Figure 3.10(c) shows that fitting a parametric Weibull model accounting for interval censoring results in unbiased estimate of the survival function. However, incorrect assumption regarding the distribution (e.g. log-logistic) for cracking lifetime leads to bias (in the lower tail) of the survival function. Using only failed sections and ignoring interval censoring as well results in an underestimation of service life. Figure 3.10(d) shows that different age distributions are associated with different amount of sections being left-, interval- and right-censored. The age distribution $U(0,30)$ generates a considerable amount of already failed sections (33%) leading to overestimation, since KM cannot accommodate left-censoring.

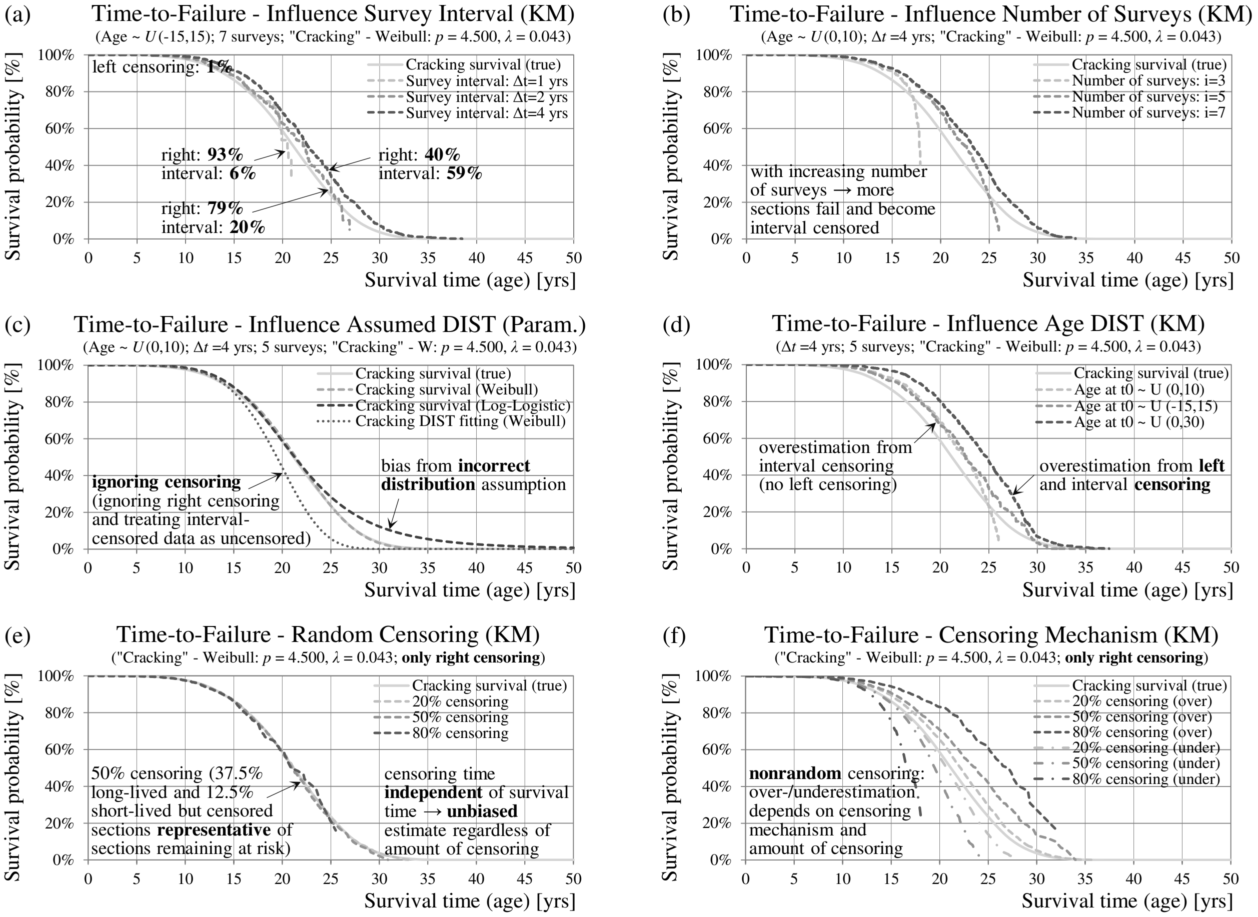
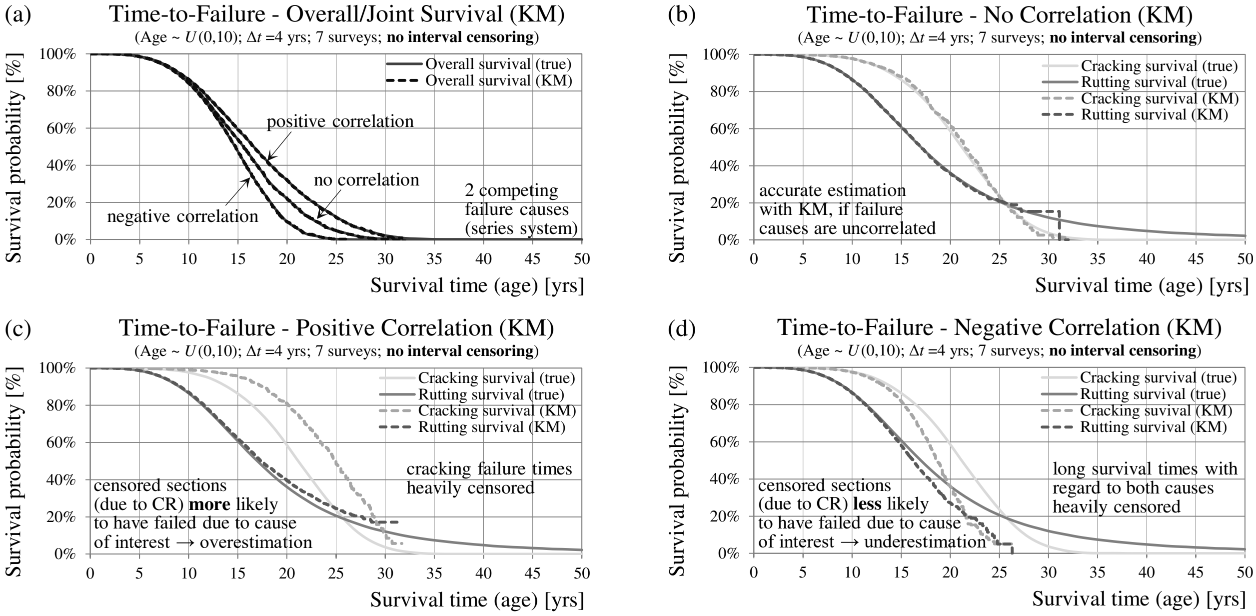


FIGURE 3.10 KM estimates of cracking survival time in the absence of competing risks - influence of survey interval length (a), number of consecutive surveys (b), assumed lifetime distribution (c), age distribution (d) and (in)dependence of the censoring mechanism (e, f).

The censoring process may be modelled by two random variables T and C representing the distribution of the failure times and the censoring times, respectively. Thus, for each section only the minimum of the two variables can be observed. In this case the data will only consist of uncensored and right-censored observations making KM an appropriate estimation method. Figure 3.10(e) shows the special case where the two random variables are statistically independent. Depending on parameters of both distributions different shares of sections will be censored. However, this does not affect the survival function estimate, which is unbiased even with 80% of the sections being censored. In contrast, if the censoring and failure mechanisms are correlated, the KM estimates will be biased, with direction and size of the bias depending on the sign of the correlation and amount of censoring (Figure 3.10(f)). This censoring scheme resembles censoring due to dependent competing risks and will be discussed in detail in the next section.

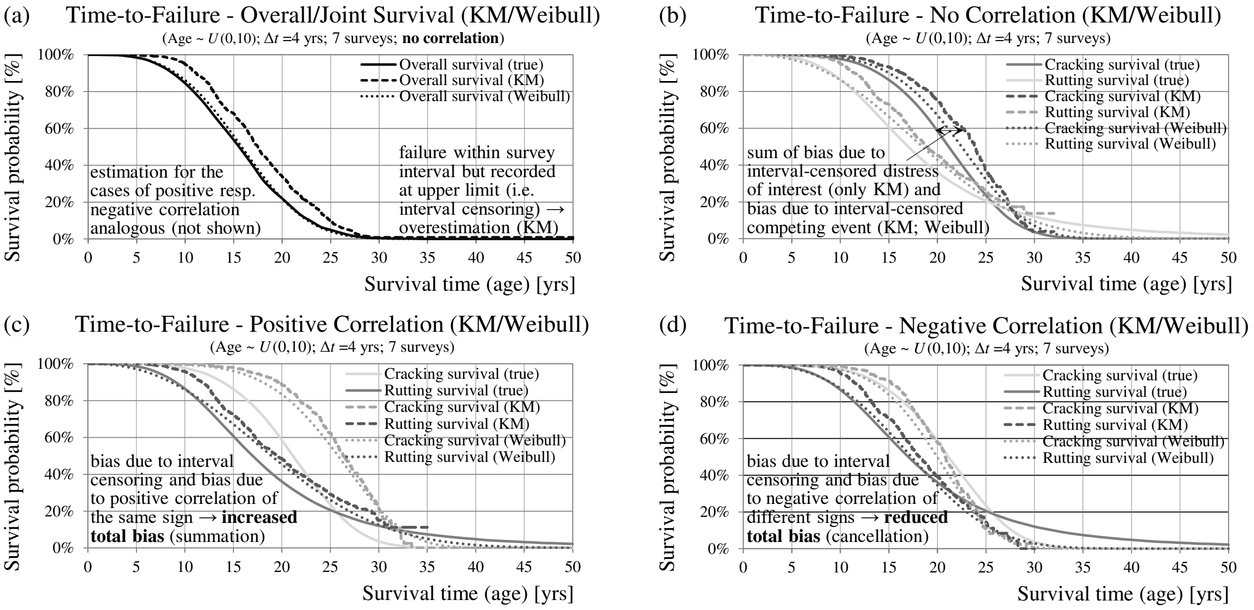
3.3.2 (Non-)parametric survival analysis in the presence of competing risks

In the next step the two failure types (cracking and rutting) are simultaneously analyzed in the cases of independence and positive as well as negative dependence (Figure 3.11). For the purpose of having only censoring due to competing risks, the generated data is manipulated one more time by replacing all interval-censored observations with the true uncensored lifetimes. Furthermore, with the use of a non-parametric approach any bias due to distribution assumptions will be excluded too. Thus, the overall survival function (failure from any cause) is estimated accurately in all three cases of dependence (Figure 3.11(a)). Figure 3.11(b) shows that the KM estimates of the marginal distributions for cracking and rutting are also unbiased, if the assumption of independent competing risks is met (Kleinbaum and Klein 2012). Independence means



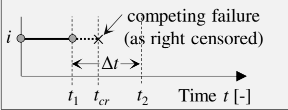
Competing causes of failure	**Classification based on median survival time**		**No correlation**				**Positive correlation**				**Negative correlation**			
			CR censored		Right censored		CR censored		Right censored		CR censored		Right censored	
Cracking	Short-lived sections	500	258	51%	0	0%	371	**74%**	0	0%	155	31%	0	0%
	Long-lived sections	500	382	**77%**	15	2%	325	**65%**	51	5%	463	**93%**	2	0%
Rutting	Short-lived sections	500	35	7%	0	0	33	7%	0	0%	6	1%	0	0%
	Long-lived sections	500	310	**62%**	15	2%	220	44%	51	5%	374	**75%**	2	0%

FIGURE 3.11 Unbiased KM estimates (no interval censoring) of joint and marginal (no correlation) survival probabilities (a, b) in contrast to service life over/underestimation in the case of positive/negative correlated competing risks (c, d).



Competing causes of failure	No correlation (Weibull)		Positive correlation (Weibull)		Negative correlation (Weibull)	
	p	λ	p	λ	p	λ
Overall/Joint	3.356	0.057	3.037	0.052	4.068	0.062
Cracking	4.669	0.041	5.256	0.036	4.774	0.046
Rutting	2.518	0.047	2.384	0.045	2.971	0.051

Bias due to interval censoring of the competing event



unbiased estimate: $L_i = S(t_{cr})$
biased estimate: $L_i = S(t_2)$

FIGURE 3.12 KM/Weibull estimates of the overall (a) and marginal survival probabilities in the case of uncorrelated (b), positive (c) and negative (d) correlated cracking and rutting.

that sections censored due to the competing event are no more or less likely to fail due to the investigated cause of interest. In the case of positive correlation censored sections are more likely to have failed than those not censored leading to an overestimation of the survival probabilities (Figure 3.11(c)). In the case of negative correlation short-lived sections with regard to cracking are also long-lived with regard to rutting, and vice versa. Thus, censored sections are less likely to have failed due to the cause of interest resulting in an underestimation of survival probabilities (Figure 3.11(d)). The size of the under-/overestimation in each case depends on the amount of censored sections and the leading cause of failure.

Next, the computations are repeated with the interval-censored data using parametric models. With KM the overall survival function is overestimated in average by 2 years or half of the inspection interval, since the failures are recorded at the upper limit of the interval (Figure 3.12(a)). The same overestimation is observed for marginal distributions in the case of no correlation (Figure 3.12(b)). Surprisingly, the estimates with a parametric model are not identical to the true marginal survival functions. The reason for the remaining bias is interval censoring of the competing event. Censored sections due to competing risks are treated as right-censored with the integral of the survival function as contribution to the likelihood. The lower limit of this integral is assumed to be the time of the last inspection, but it should be prior to the inspection, because the competing (censoring) event has happened within the inspection interval. The size of this bias is less significant than the bias due to interval censoring of the event of interest because right-censored observations have a smaller contribution to the total likelihood as compared to interval censoring. In the case of rutting additional bias may be present due to the incorrect assumption of a Weibull distribution. In the case of positive correlation (Figure 3.12(c)) the bias due to interval censoring and bias due to dependent competing risks are in the same direction leading to increased total bias (summation). In contrast, in the case of negative correlation (Figure 3.12(d)) the bias due to interval censoring is in the opposite direction of the bias due to competing risks leading to a reduction of the total bias (cancellation).

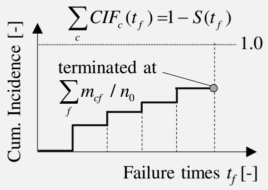
3.3.3 Cumulative incidence and conditional probability

Estimation of cumulative incidence functions (CIF) and conditional probability curves (CPC) is a common approach for competing risk data that does not rely on the independence assumption. For discussion and details see, for example, Pintilie (2006) or Kleinbaum and Klein (2012). The difference between KM and CIF is that CIF considers competing risk events as failures and not as censored observations as in the case of KM. Figure 3.13(a) shows the calculation of CIF in the case of negative correlated cracking and rutting failures. CIF provide information about the dominant failure cause, but they cannot be interpreted as survival probabilities. CIF are not equivalent to the marginal distributions and cannot be utilized in prediction models and the subsequent optimization of treatments.

Another concept in SA is the conditional survival defining the probability that a section survives the event of interest given that no competing failure has been experienced. Conditional probability curves (CPC) may be estimated based on the CIF estimates. In the case of independent failure causes the marginal and the conditional distributions are identical. In the case of negative correlation (Figure 3.13(b)) the conditional survival probabilities fall below the

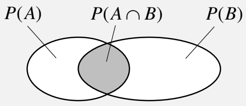
Cumulative Incidence Function (CIF):

- Does not require independence between competing risks
- Mutually exclusive failure causes
- Commonly used/easy to compute
- $1-CIF$ is not a survival curve → no meaningful interpretation for the prediction and modeling



Conditional Probability Curves (CPC):

- Estimation is unbiased only in the case of negative dependence
- Complete back-calculation of the marginal distribution is not possible in that case either



$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Independent events:

$$P(A \cap B) = P(A) \times P(B)$$

Empirical estimation (CI & CP):

Cause (distress)-specific hazard at ordered failure times t_f:

$$h_c(t_f) = m_{cf} / n_f$$

Incidence of failing from cause c:

$$\hat{I}_c(t_f) = S(t_{f-1}) \times h_c(t_f)$$

Cumulative incidence function (subdistribution function):

$$CIF_c(t_f) = Q_c(t_f) = \sum \hat{I}_c(t_f)$$

Conditional survival (estimated):

$$CP_c = \Pr(T_c \leq t | T \geq t)$$

$$CP_c(t) = CIF_c(t) / (1 - CIF_{c'})$$

Theoretical expressions (CP):

Bivariate joint (system/overall) survival function:

$$S(t_1, t_2) = \Pr(T_1 > t_1, T_2 > t_2)$$

Conditional survival function (true):

$$S_{1|2}(t_1 | T_2 > t_2) = \Pr(T_1 > t_1 | T_2 > t_2)$$

$$S_{1|2}(t_1 | T_2 > t_2) = \frac{S(t_1, t_2)}{S_2(t_2)} = 1 - CP_1$$

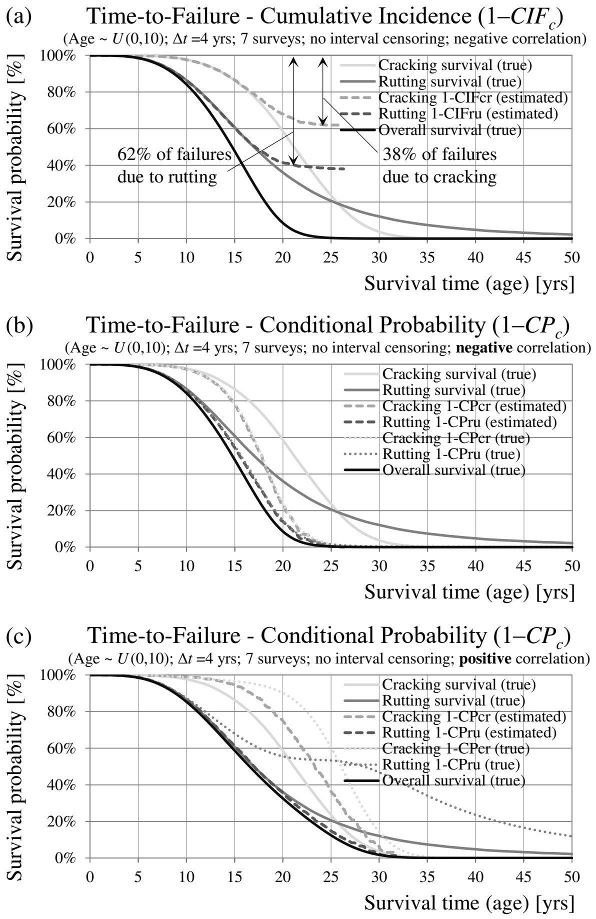
Independent failure causes:

$$S(t_1, t_2) = S_1(t_1) \times S_2(t_2) \quad \text{and}$$

$$S_{1|2}(t_1, T_2 > t_2) = S_1(t_1)$$

Legend

c…	failure cause of interest
c'…	all failure causes other than c
m_{cf}…	# failures of type c at time t_f
$S(t_{f-1})$…	overall prob. of surviving t_{f-1}
T_c…	type until failure c occurs
T…	time until any failure occurs
$S_i(t_i)$…	marginal distribution (cause i)


(a) Time-to-Failure - Cumulative Incidence ($1-CIF_c$)

(Age ~ $U(0,10)$; Δt =4 yrs; 7 surveys; no interval censoring; negative correlation)

(b) Time-to-Failure - Conditional Probability ($1-CP_c$)

(Age ~ $U(0,10)$; Δt =4 yrs; 7 surveys; no interval censoring; **negative** correlation)

(c) Time-to-Failure - Conditional Probability ($1-CP_c$)

(Age ~ $U(0,10)$; Δt =4 yrs; 7 surveys; no interval censoring; **positive** correlation)

FIGURE 3.13 Overview and estimation of cumulative incidence (a) and cumulative probability curves (b, c) based on competing risk data.

marginal survival probabilities, because surviving the competing event makes survival less likely. The opposite relationship can be observed in the case of positive correlation (Figure 3.13(c)). Even if CPC are estimated without bias, it is not possible to estimate the marginal distribution after the point where the overall survival falls down to 0%.

Yang and Kim (2015) provide the only study known to the authors in pavement management that considers different pavement distress types as competing risks. Unfortunately, the analysis was limited to an estimation of CIF and CPC without acknowledging the limitations of these approaches. Moreover, the nature of possible correlation between the three analyzed distresses was not investigated.

3.3.4 Performance history-based approach

In summary, traditional survival analysis methods lead to bias in the presence of dependent competing risks. Using information about previous conditions as well as distress progression characteristics has the potential to reduce this bias. A regression function with constant distress progression factor (power factor β_2) can be fitted at the section level by estimating the scale parameter β_1 provided that at least three condition observations are available. Since this is the exact same algorithm used for data generation, the performance function will fit perfectly with no deviations (residues). Therefore, a random measurement error is added (subtracted) to (from) each of the observations. Furthermore, condition probability distribution, confidence intervals and failure distribution can be estimated assuming a distribution (e.g. normal) for the error terms (Hoffmann and Donev 2016). The idea is to replace the censoring time for competing-risk censored observations (considered as right-censored) with unbiased upper and lower limits of the confidence interval for the service life derived with regression. Likelihood contributions for failure (censoring) times for the event of interest can be computed as usual without fitting of a regression function. The question how to choose the confidence level for the prediction interval has no correct answer in a mathematical sense. However, confidence levels producing average

Single-section model (SSC):

- Not applicable, if cracking has not yet initiated/not enough observations ($df \leq 1$)
- Considers deterioration history/ degradation process (power β_2)

Competing-risk SA:

- Independent competing risks → standard SA is sufficient
- Dependent competing risks → treating censored observations as right censored leads to bias
- SA does not consider previous conditions/degradation process

Combination of SA & SSC:

- Replacing censoring time due to competing risk with confidence interval for service life
- Uncertain predictions with small log-likelihood contributions
- Weighting/selection issues

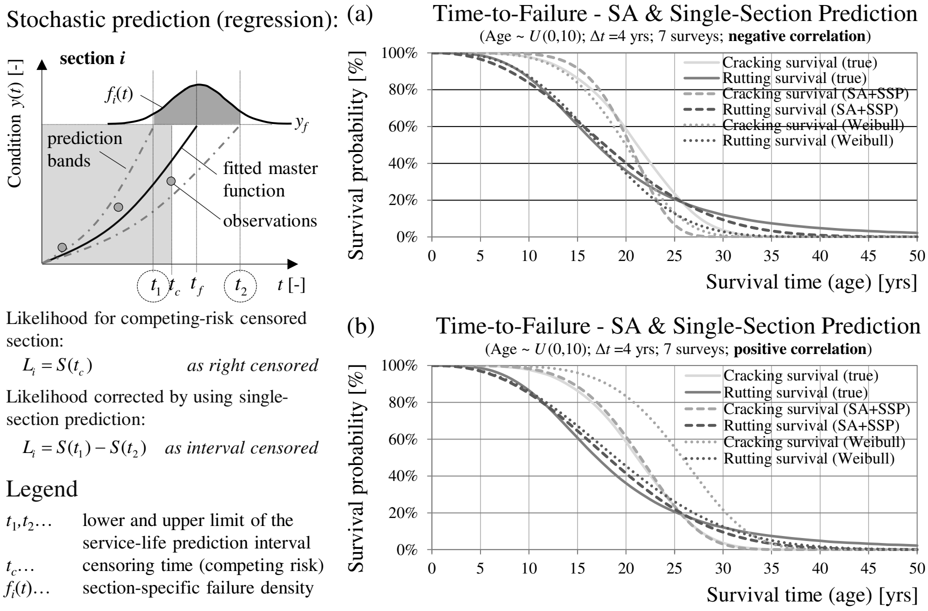


FIGURE 3.14 Improvement of survival estimates for dependent competing risks by using previous conditions and distress progression models at section level in combination with survival analysis in the case of negative (a) and positive (b) correlated distresses.

confidence intervals similar to the inspection interval will ensure that the competing-risk censored sections have no more or less weight in the likelihood than other sections. On the other hand, sections with an unreliable prediction and a wider prediction interval will exhibit a smaller log-likelihood contribution, which is a desirable property.

Figure 3.14(a,b) illustrates the application of this approach to the generated data using single-section regression with a confidence level of 68.2% (one standard deviation of the mean service life). In the case of rutting and cracking (positive correlation) the bias was substantially reduced in comparison to the parametric estimate from Figure 3.12. The residual bias for rutting is caused by the incorrect distribution assumption (Weibull instead of log-logistic). Nevertheless, in one of the four cases (cracking, negative correlation) the approach was not able to improve the estimates (Figure 3.14(a)). The reason for this is that the single-section prediction model could not be applied for roughly 1/3 of the competing-risk censored sections because they had less than three observations. Since this group is not random and not representative, the bias was even increased in the lower tail of the survival function in this case. In summary, using a regression prediction model at the section level in addition to survival analysis has the potential to eliminate or reduce bias due to dependent competing risks, if the model can be applied for the majority of the sections and the assumptions of the single-section prediction model are met.

3.3.5 Copula-based approach

As already shown, if competing risks are independent, the Kaplan-Meier method produces unbiased estimates, but the independence assumption is neither verifiable nor realistic for pave-

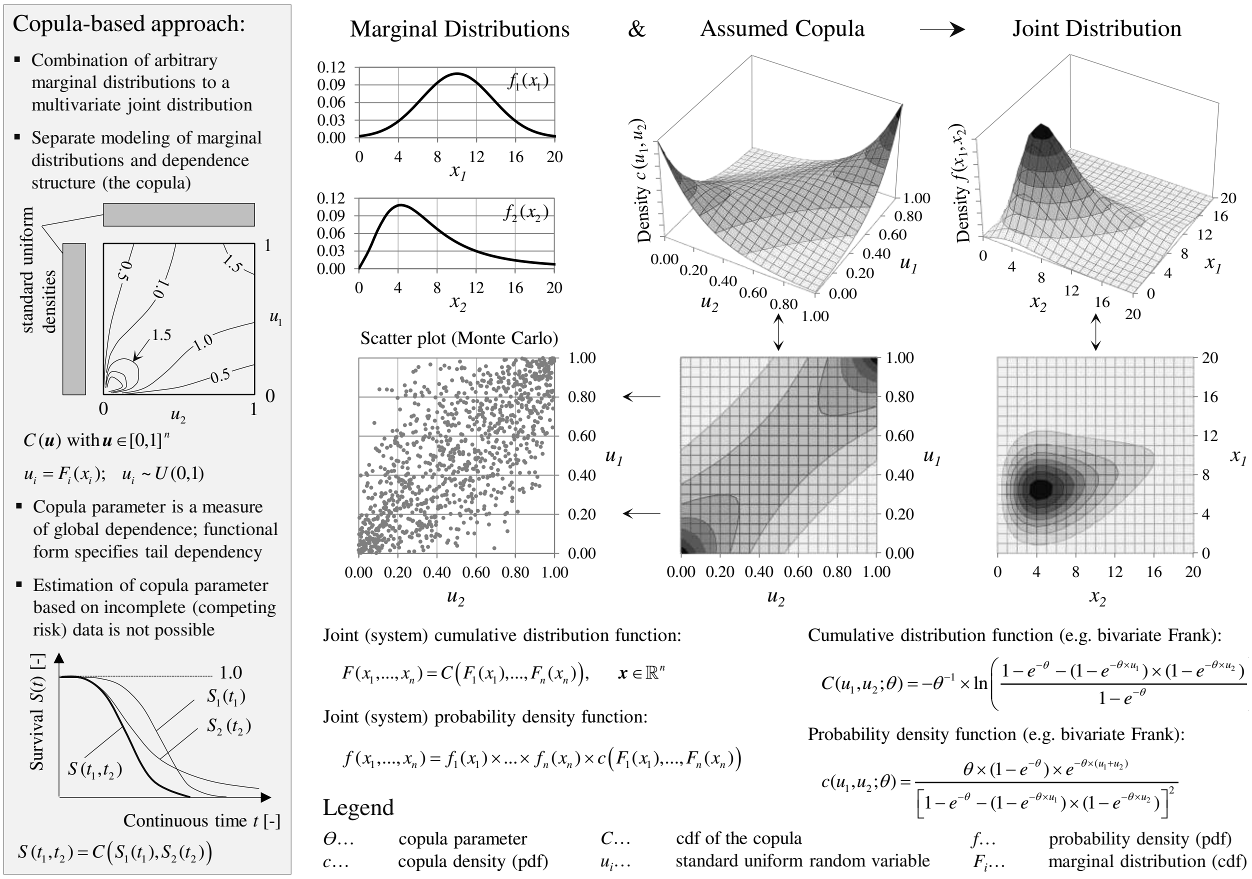


FIGURE 3.15 Combination of arbitrary marginal distributions to a multivariate joint distribution using copula (dependence) function.

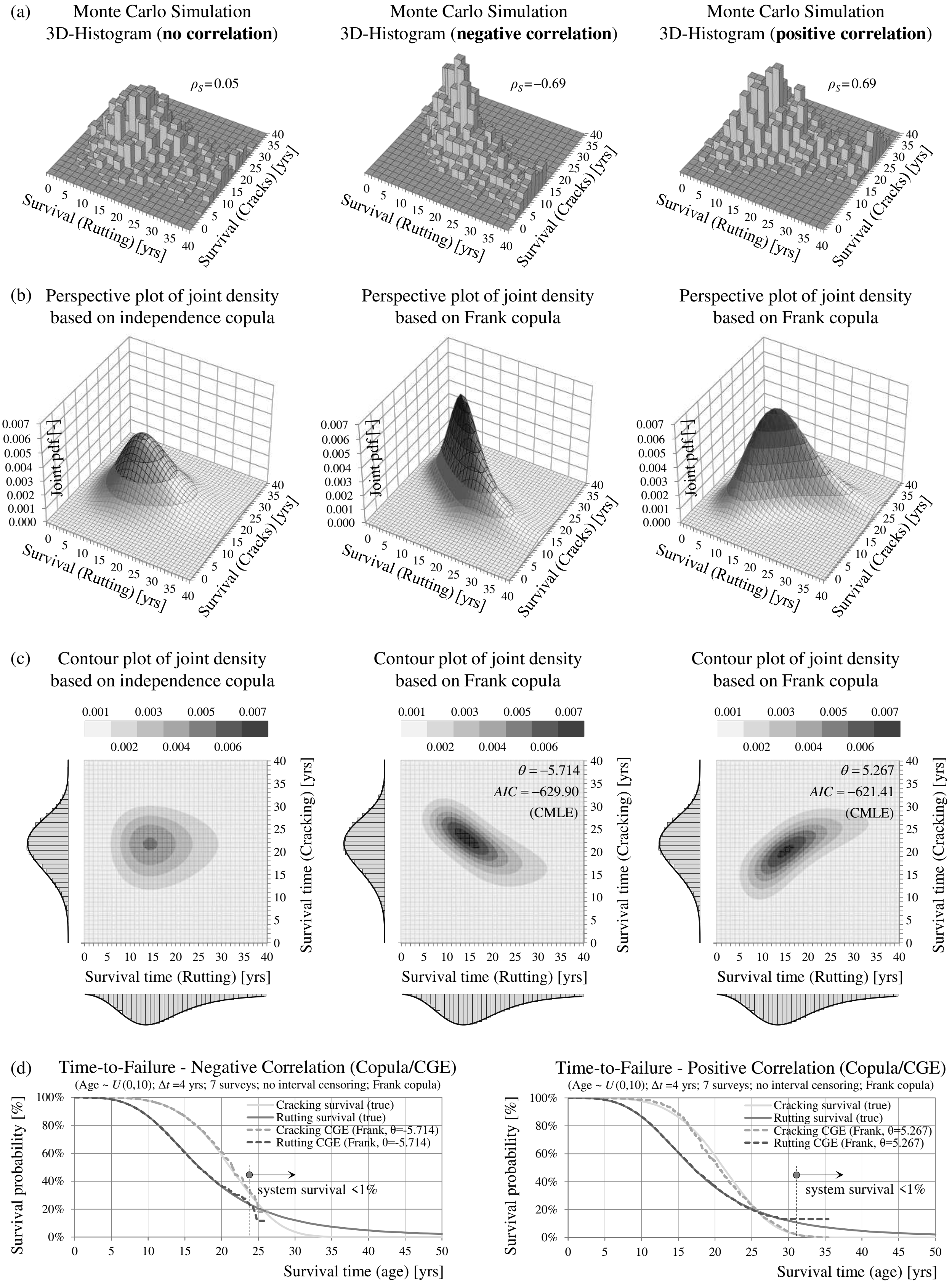


FIGURE 3.16 Histogram of the generated service life data (a), perspective plots (b) and contour plots (c) of the joint survival density based on a fitted Frank copula to the true data. Estimation of the marginal survival functions with the competing risk data and an assumed (fitted) copula using copula-graphic estimator (d) in the case of negative/positive correlation.

ment distress types. Thus, in the presence of dependent competing risks the marginal (latent) lifetime distributions are not identifiable and cannot be estimated without bias with common survival analysis models (Crowder 2012). Moreover, there exists a model with independent competing risks and one or many models with dependent different marginal distributions that produce the same joint (overall) survival function. However, if the dependence structure is known a priori, this identifiability issue can be resolved mathematically. One possibility for a characterization of dependencies are copula functions providing a way to simultaneously model marginal and joint distributions of random variables. A copula is defined as the multivariate (joint) distribution of standard uniform random variables. Thus, any arbitrary continuous marginal distribution can be uniquely linked to a specific joint distribution using probability integral transformation, if the functional form and parameters of the copula are specified (Figure 3.15). A description of the mathematical foundations and properties of copulas is beyond the scope of this paper and can be found in the literature (Kurowicka and Cooke 2006, Joe 2015).

The overall survival function and the cumulative incidence functions (CIF) can be estimated from the observed competing risk data. Given these functions and an assumed copula the marginal distributions can be identified with the copula-graphic estimator (CGE) of Zheng and Klein (1995). Figure 3.16(a) shows the structure of the generated multivariate data in a 3D histogram which can be used for the selection of an appropriate copula form. Furthermore, three different Archimedean copulas are fitted to the uncensored generated data using canonical maximum likelihood (Figure 3.16(b,c)). Finally, the marginal distributions are estimated non-parametrically based on a Frank copula using CGE. The estimates are largely identical to the true survival distributions in both cases of negative and positive dependence (Figure 3.16(d)).

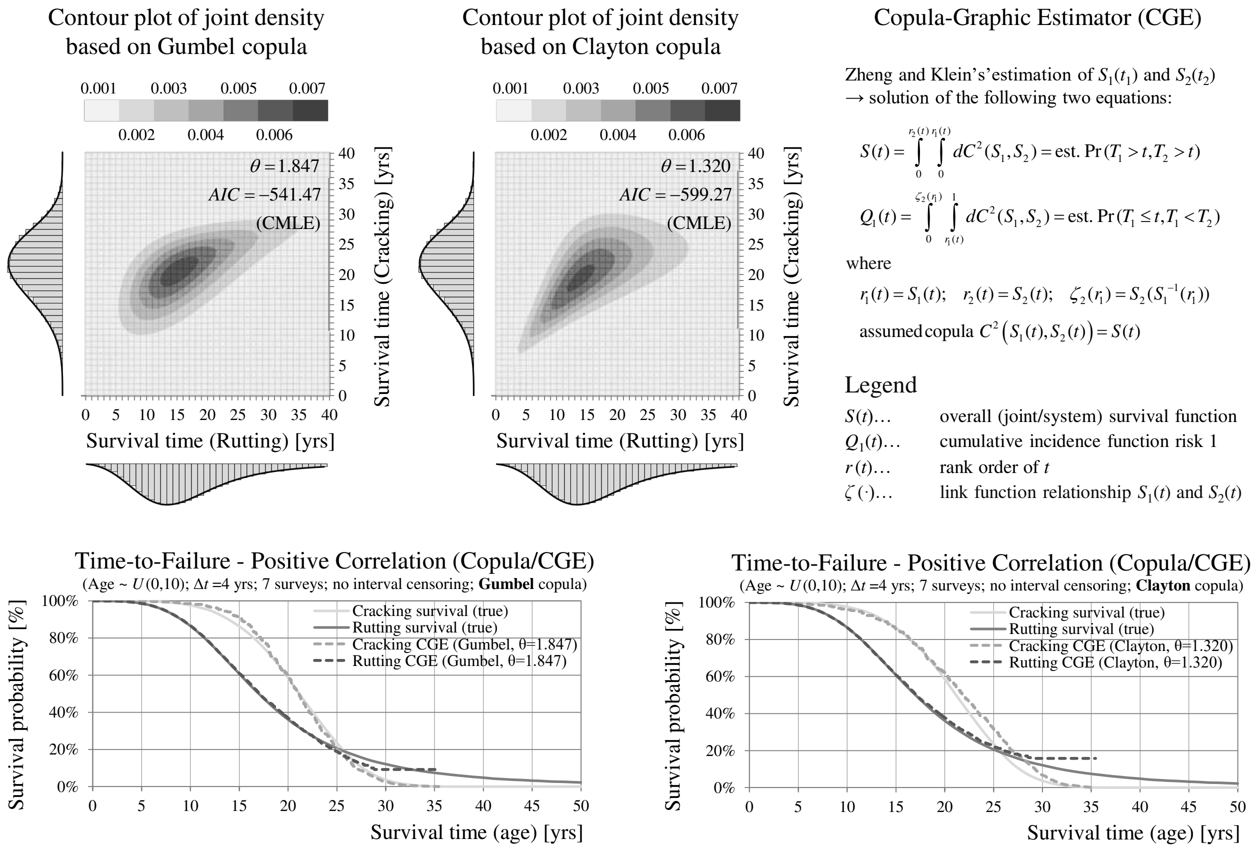


FIGURE 3.17 Definition of the copula-graphic estimator and sensitivity of the survival estimates to the functional form of the copula.

Figure 3.17 shows that marginal survival estimates with the flexible Frank copula exhibit almost no bias compared to other forms (Gumbel, Clayton). A critical factor for estimation accuracy is the copula parameter provided that a flexible functional form is selected. This approach may also be generalized to more than two competing risks and extended to a regression model with covariates (Lo 2015). However, in contrast to the performance history-based approach (Section 3.3.4), the complexity increases substantially when considering more than two failure types.

Another limitation of the copula-based competing risk approach is that the copula cannot be estimated from the censored data and must be assumed. Laboratory tests and accelerated pavement testing may provide insights about the sign and degree of correlation between different asphalt failure mechanisms. Another approach would be to estimate the survival times until lower condition thresholds (i.e. closer to the initial state) guaranteeing less censored observations. The distress correlation is invariant to this scaling of the survival distribution allowing for the copula to be fitted. In any case further research should be aimed at investigating the dependencies between dominant failure causes from real-world survey data.

3.4 CONCLUSIONS

Reliable pavement performance models are crucial for the planning of maintenance and rehabilitation activities and an efficient use of public funds. For a specific selection and optimization of treatments performance prediction models and service life estimates for each distress type are needed. The parameters of these models should be estimated with empirical data from periodic condition surveys. Due to the nature of pavement deterioration processes and practical limits of the inspection process, censoring of the data is unavoidable. There are different types of censoring, each showing a different impact on the estimation and a different frequency of occurrence. Approaches like regression analysis (OLS) or Markov chains do not account for censoring. Using them to develop distress progression models may result in substantial bias and more or less arbitrary estimates. Some researchers have used standard non-, semi- and full parametric survival analyses accounting for censoring, but their analyses were limited to the univariate case of only one possible failure type neglecting competing risk and correlation between failure causes. This study covers common cases for pavements exhibiting more than one distress type. Despite the fact that the presented models are in principle capable of accounting for covariates, the related aspects are not included in the paper. Using multiple explanatory variables leads to many issues (e.g. endogeneity bias, multicollinearity, ex-ante prediction, specification errors, etc.) that will be covered in another paper.

Distress modeling in the presence of dependent competing risks poses a challenge, but the reality of more than two and not necessarily mutually exclusive failure causes is even more complex. This paper provides evidence that models with better fit to real-world survey data (e.g. regression) are not necessarily models describing the underlying (unobserved) degradation process and service life accurately. As the “true” degradation process on the network level cannot be observed it is not possible to determine directly which method or model will perform better than others. Therefore, the analysis and conclusions in this paper are based on simulated data allowing a comprehensive and objective comparison of different models to the a priori specified true processes. Based on the results it can be shown that common survival analysis models show substantial bias in the simple case of two positive or negative dependent competing risks. Depending on the present types of censoring and violated assumptions related deviations may partly cancel each other out or add up. Moreover, competing risk approaches based on cumulative incidence reflect the observed data but not the underlying process (marginal distribution). It can be shown that further improvements and bias corrections are achieved by using prior conditions and degradation history at the section level. Another finding is that copulas are a promising alternative for multivariate simultaneous modeling of distress types allowing for an accurate estimation of marginal lifetime distributions with censored data and assumed (estimated) copula parameter. Future research will be concentrated on the comparative application of the presented models and methods to real-world data and the impact of dependence structures between failure causes on estimation accuracy.

REFERENCES

Aalen, O.O., Borgan, Ø. and Gjessing, H.K., 2008. *Survival and event history analysis. A process point of view*. New York, NY: Springer.

Abaza, K.A., 2017. Empirical approach for estimating the pavement transition probabilities used in non-homogenous Markov chains. *International Journal of Pavement Engineering*, 18 (2), 128-137.

Butt, A.A., Shahin, M.Y., Feighan, K.J. and Carpenter, S.H., 1987. Pavement performance prediction model using the Markov process. *Transportation Research Record: Journal of the Transportation Research Board*, 1123, 12-19.

Christofa, E.M. and Madanat, S.M., 2010. Development of pavement cracking initiation models by combining experimental and field data. *89th Annual Meeting of the Transportation Research Board*, 10-14 January, Washington, DC.

Crowder, M.J., 2012. *Multivariate survival analysis and competing risks*. Boca Raton, FL: Chapman & Hall/CRC Press.

Donev, V. and Hoffmann, M., 2016. Pavement performance prediction and estimation of service life in the presence of censored condition data. *Proceedings of the Eight International Conference on Maintenance and Rehabilitation of Pavements*, 27-29 July 2016 Singapore. Singapore: Research Publishing, 680-689.

Dong, Q. and Huang, B., 2015. Failure probability of resurfaced preventive maintenance treatments: investigation into long-term pavement performance program. *Transportation Research Record: Journal of the Transportation Research Board*, 2481, 65-74.

Joe, H., 2015. *Dependence Modeling with Copulas*. Boca Raton, FL: Chapman & Hall/CRC Press.

Hoffmann, M. and Donev, V., 2015. Probabilistic modeling of flexible pavement performance and LCC – based service lives with failure distributions. *In*: H. Furuta, D.M. Frangopol and M. Akiyama, eds. *Life-Cycle of Structural Systems: Design, Assessment, Maintenance and Management,* 16-19 November 2014 Tokio. London: Taylor & Francis, 1981-1988.

Hoffmann, M. and Donev, V., 2016. Introduction of a new continuous time and state space stochastic process in condition prediction. *International Journal of Pavement Engineering*. doi: 10.1080/10298436.2016.1162304.

Kalbfleisch, J.D. and Prentice, R.L., 2002. *The statistical analysis of failure time data*. 2nd ed. Hoboken, NJ: John Wiley & Sons.

Kargah-Ostadi, N. and Stoffels, S.M., 2015. Framework for development and comprehensive comparison of empirical pavement performance models. *Journal of Transportation Engineering*, 141(8), 04015012.

Kleinbaum, D.G. and Klein, M., 2012. *Survival analysis: a self-learning text*. 3rd ed. New York, NY: Springer.

Kurowicka, D. and Cooke, R, 2006. *Uncertainty analysis with high dimensional dependence modelling*. Chichester, UK: John Wiley & Sons.

Li, N., Xie, W. and Haas, R., 1996. Reliability-based processing of Markov chains for modeling pavement network deterioration. *Transportation Research Record: Journal of the Transportation Research Board*, 1524, 203-213.

Lo, S.M.S, 2015. *Competing risks model and copula*. Dissertation (PhD). University of Freiburg.

Loizos, A. and Karlaftis, M., 2005. Pavement crack initiation prediction from in-service pavements: a duration model approach. *Transportation Research Record: Journal of the Transportation Research Board*, 1940, 38-42.

Madanat, S., Mishalani, R. and Ibrahim, W.H.W., 1995. Estimation of infrastructure transition probabilities from condition rating data. *Journal of Infrastructure Systems*, 1(2), 120-125.

Mallick, R. and El-Korchi, T., 2013. *Pavement engineering: principles and practice*. 2nd ed. Boca Ranton, FL: CRC Press.

Molzer, C., Felsenstein, K., Viertl, R., Litzka, J. Vycudil, A., 2000. *Statistische Methoden zur Auswertung von Straßenzustandsdaten* [Statistics for pavement performance modeling]. Wien: BMVIT, Straßenforschung Heft 499 (in German).

Nakat, Z.S. and Madanat, S.M., 2008. Stochastic duration modeling of pavement overlay crack initiation. *Journal of Infrastructure Systems*, 14 (3), 185-192.

Norouzi, A., Kim, D. and Kim, Y. R., 2016. Numerical evaluation of pavement design parameters for the fatigue cracking and rutting performance of asphalt pavements. *Materials and Structures*, 49(9), 3619-3634.

Ozer, H., Al-Qadi, I.L., Singhvi, P., Khan, T., Rivera-Perez, J. and El-Khatib, A., 2016. Fracture characterization of asphalt mixtures with high recycled content using Illinois semicircular bending test method and flexibility index. *Transportation Research Record: Journal of the Transportation Research Board*, 2575, 130-137.

Paterson, W.D.O., 1987. *Road deterioration and maintenance effects: models for planning and management*. Baltimore, MD: The Johns Hopkins University Press.

Pintilie, M., 2006. *Competing risks: a practical perspective*. 1st ed. Chichester, UK: John Wiley & Sons.

Prozzi, J.A. and Madanat, S.M., 2000. Using duration models to analyze experimental pavement failure data. *Transportation Research Record: Journal of the Transportation Research Board*, 1699, 87-94.

Reger, D., Christofa, E., Guler, I. and Madanat, S., 2013. Estimation of pavement crack initiation models by combining experimental and field data. *Journal of Infrastructure Systems*, 19(4), 434-441.

Ross, S.M., 2014. *Introduction to probability models*. 11th ed. Waltham, MA: Academic Press.

Rübensam, J., Hellmann, L., Staroste, D. and Stoltz, J., 2005. *Untersuchungen zur Wirtschaftlichkeit und bautechnischen Bewährung von Fahrbahnbefestigungen aus Asphalt und Beton auf bestehenden Bundesautobahnen* [Investigations of the profitability and structural dependability of asphalt and concrete carriageway pavement on existent interstate highways]. Bonn: Forschung Straßenbau und Straßenverkehrstechnik, Heft 914 (in German).

Schram, S. and Abdelrahman, M., 2009. Mechanistic-empirical modeling in network-level pavement management. *Transportation Research Record: Journal of the Transportation Research Board*, 2093, 76-83.

Shin, H.C. and Madanat, S., 2003. Development of stochastic model of pavement distress initiation. *Journal of Infrastructure Planning and Management*, 744 (61), 61-67.

Smith, A.F., Prozzi, J., Tandon, V. and Mikhail, M., 2007. *TexSys: guide for selecting HMA for Texas flexible pavements*. El Paso, TX: Center for Transportation and Infrastructure Systems, University of Texas at El Paso.

Van Dam, T.J., Chesher, A.D. and Peshkin, D.G., 1997. Evaluation of LTTP data using HDM-III probabilistic failure-time models for crack initiation in bituminous pavements. *Transportation Research Record: Journal of the Transportation Research Board*, 1592, 125-133.

Wang, Y., Mahboub, K.C. and Hancher, D.E., 2005. Survival analysis of fatigue cracking for flexible pavements based on long-term pavement performance data. *Journal of Transportation Engineering*, 131 (8), 608-616.

Wooldridge, J., 2013. *Introductory econometrics: a modern approach*. 5th ed. Mason, OH: South-Western Cengage Learning.

Yang, J. and Kim, S.H., 2015. Investigating the performance of as-built and overlaid pavements: a competing risks approach. *International Journal of Pavement Engineering*, 16 (3), 191-197.

Zheng, M. and Klein, J.P., 1995. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82 (1), 127-138.

CHAPTER 4

Optimization of pavement maintenance and rehabilitation activities, timing and work zones for short survey sections and multiple distress types

by Valentin Donev & Markus Hoffmann

in *International Journal of Pavement Engineering*
Published online: 24 Jul 2018 (open access)
Published in print: Volume 21, Issue 5, 2020
DOI: 10.1080/10298436.2018.1502433

ABSTRACT

Existing pavement management systems (PMSs) aggregate short survey sections into long homogeneous sections (typically 1-5 km), aiming to reduce the amount of data and formulate a tractable optimization problem. However, assuming that sections in a similar condition will undergo a similar deterioration of this condition is incorrect, resulting in imprecise predictions and violations of thresholds. Thus, any maintenance and rehabilitation (M&R) optimization approach based on long road sections and a single condition indicator (aggregated or not) cannot provide applicable results at the project level. This paper provides a holistic model framework and consistent methodology for a new PMS, abolishing the need for composite condition indices, data aggregation and homogeneous sections in the M&R optimization process. The proposed bottom-up optimization is based on short survey sections (25-50 m) and solves for optimal treatment type, timing, work-zone length and layout, minimizing agency, road user and environmental life cycle costs. Instead of using just one condition indicator, the method models the effects of each treatment on multiple distress types and their combinations. Moreover, the short survey sections are grouped into larger work zones based on purely economic criteria (i.e. economies-of-scale costs). Treatment cost, temporary traffic control cost (TTCC), and work-zone user cost models are derived as a function of the project size and calibrated for actual prices in Austria. Genetic algorithms are employed as a part of the solution algorithm, being capable of handling an unlimited number of road sections. Furthermore, the approach can be extended to account for budget restrictions, risk and cross-asset allocation of funds for any aging repairable infrastructure system.

KEYWORDS: pavement management, bottom-up optimization, life-cycle costs, work-zone length, M&R strategies, economies of scale, genetic algorithms, treatment timing, freeways.

4.1 INTRODUCTION

Pavement management covers the processes of data collection, condition assessment and prediction, planning, optimization and application of maintenance and rehabilitation (M&R) treatments, as well as budgeting. The objective of M&R optimization is to maximize performance (e.g. network condition), or to minimize agency and user costs, or to minimize risk/environmental impacts by selecting appropriate treatments (decision variables), subject to constraints (e.g. budget). This problem can be formulated for a single road section (single facility), for a travel route consisting of multiple sections, or for the entire network (system) covering all sections and other road assets.

In the context of aging infrastructure, increasing traffic loads, demanding public expectations and limited budgets, road agencies in developed countries are faced with several requirements. They have to maintain the network condition, allocating their budget according to the needs by balancing trade-offs between agency and user costs in the long term, but also to reduce accidents, emissions and traffic flow interruptions in the short term. An ideal pavement management system (PMS) would therefore address these goals simultaneously, providing optimal budgets at the network level, as well as detailed location-specific treatment recommendations in terms of timing, type, work-zone length and layout that could be accepted by road engineers and applied with limited need for manual adaptation.

Common PMSs aggregate short survey sections to long homogeneous sections based on different algorithms, effectively resulting in an averaging of measured condition data. The purpose of this approach is to reduce the computing effort and the complexity of the optimization problem. However, there is a growing discrepancy between the on-going development of automated condition survey systems (laser technology), delivering more accurate measurements with higher resolution and collection frequency, sharply contrasting with the limited capabilities of current PMS, using only a fraction of the provided information in prediction and optimization.

Surprisingly, the validity of the approaches for sectioning and data aggregation is not questioned in the literature and the consequences of their use are not investigated. The assumption that a similar condition (at the time of the survey) implies similar future deterioration is proven wrong by empirical research. Thus, condition predictions based on homogeneous sections cannot be accurate. Moreover, the defined condition thresholds are systematically violated due to averaging effects, and it is not possible to select optimal work zones since the formation of homogeneous sections is conducted prior to prediction and optimization.

This paper provides a holistic framework and methodology for an end-to-end pavement management system (PMS) which does not rely on homogeneous sections and composite condition indices. The foundations for this system are laid by deriving calibrated cost, performance and duration models which are shortly presented in the paper as well. The core, however, is the new optimization approach which is based on short survey sections (e.g. 50 m), allowing for precise condition predictions and section-specific treatment recommendations. The developed optimization algorithm groups the short survey sections into longer work zones based on economies-of-scale costs. The application of this new methodology is demonstrated by using a case study of 1000 road sections, realistic condition data, as well as multiple distress types and treatments.

The following section (Section 4.2) gives an overview of the relevant literature by focusing on the shortcomings of common approaches. Section 4.3 introduces the case study, the calibration of the input parameters as well as the developed performance and cost models. The methodology of the optimization approach and the results of its application are presented in Section 4.4 and Section 4.5, respectively. The explicit mathematical formulation of the optimization model and the details of the solution algorithm are briefly explained, while an in-depth description will be the subject of another paper.

4.2 LITERATURE REVIEW

There is extensive literature on the topic of pavement M&R optimization. However, most authors approach the problem from different perspectives by focusing only on a single section at the project level, budget allocation at the network level or work-zone scheduling. The research on budget allocation is concerned with long-term performance and selection of M&R strategies, using either prioritization or optimization algorithms, but usually does not consider the network configuration or the planning of work zones. The literature on work-zone optimization considers trade-offs between one-time agency and work-zone user costs but does not optimize treatment type and timing within a comprehensive life cycle costing approach.

Typical prioritization methods, such as the marginal cost-effectiveness (MCE) approach, are widely used, despite the need for composite condition indices and being inferior to optimization (Haas et al. 1994, AASHTO 2012). The MCE approach, however, suffers from a few additional drawbacks. First, maximizing effectiveness as an objective favors reconstruction and heavy rehabilitation (larger area between the curves) and may produce significantly higher agency costs (only included as a constraint). Second, the computation of effectiveness based on an aggregated index is problematic because the results of the optimization depend on the aggregation rules and weights, and are therefore subjective.

4.2.1 Budget allocation and optimization based on life-cycle costs

In contrast to prioritization, mathematical optimization allows for an optimal (near-optimal) solution to be found in the case of exact (heuristic) methods. Moreover, multiple constraints (e.g. annual budget, condition, level of service, etc.) can be incorporated in the formulation of the problem. Two basic approaches are discussed in the literature: top-down and bottom-up. Top-down methods analyze road sections in groups (discrete classes) and treat all sections in a group as identical in terms of deterioration (e.g. Kuhn and Madanat 2005). In this way, these methods are independent of the number of road sections and thus very efficient for large-scale networks. The disadvantage is that all sections in a class receive the same treatment and section-specific recommendations cannot be given.

In contrast, bottom-up approaches consider section-specific distress information and deterioration rates. Bottom-up optimization can be further divided into one-stage and two-stage solution approaches. One-stage bottom-up approaches consider all possible combinations of road sections, M&R treatment types and application years simultaneously. Each combination is represented by a binary decision variable that takes the value “1” if the treatment is selected and “0” otherwise. The goal is to minimize the discounted total life-cycle costs, subject to budget and condition constraints. The problem cannot be solved exactly, even for a very small number of sections, due to the exponential complexity and inevitable nonlinearities in the objective function. Hence, many authors have adopted a heuristic technique (genetic algorithm) in their solution (Fwa et al. 1996, Ferreira et al. 2002, Santos et al. 2017). However, the major drawback of these approaches is that there is no guarantee and no way to know whether the obtained M&R program is even close to optimal (for large networks).

Two-stage bottom-up approaches consist of selecting a small set of optimal and sub-optimal M&R strategies for each section. Based on this selection, the objective is to find the optimal combination of strategies at the network level, satisfying the budget constraint in each year. Yeo et al. (2013) and Lee and Madanat (2015) formulate the network-level problem as a constrained combinatorial problem, which they solve with evolutionary algorithm or pattern search heuristic. These models represent pavement condition by using only one indicator, namely the pavement serviceability rating (PSR) or the international roughness index (IRI), respectively. In reality, however, pavements exhibit various types of distress that may require different treatments in each case. One advantage of the two-stage approaches is that the optimal strategy for each section provides the solution at the network level, if the budget constraint is not binding, which could be used as an orientation value for the constrained problem.

4.2.2 Optimization of work zones

The literature on work-zone scheduling is mostly concerned with operational decisions for short-term work zones by optimizing project length, duration, traffic control and starting time (e.g. Chien and Schonfeld 2001, Jiang and Adeli 2003). The optimization is based on the trade-off between economies-of-scale agency costs and work-zone impacts in terms of delays and accident costs. These costs only comprise one-time (initial) costs, with life-cycle costs and future treatments being neglected. Moreover, the analysis focuses only on single projects, where treatment type and timing have already been determined.

As a step forward, these models were extended to account for future costs with a very simplified life cycle (Chen and Schonfeld 2007, Yang 2010). The authors investigate trade-offs between longer service life (resurfacing interval) and higher one-time agency and user costs for thicker pavements. Thereby, resurfacing interval is a linear function of the decision variable overlay thickness. Then again, these studies focus only on a single project with no optimization of treatment timing or type and no true life cycle.

Hajdin and Lindenmann (2007) consider multiple projects based on a graph model of the network. M&R treatments on pavement sections and other road assets (bridges, tunnels) are grouped into larger work zones which minimize long-term costs, subject to budget constraints. The main limitation of this approach is its reliance on static input from external management systems which determine M&R strategies (optimal and alternatives) independently for each individual asset. Subsequently, the proposed network optimization considers only those treatments which are scheduled within a single budget period. However, a joint optimization of work-zone length and M&R timing for all assets (and not only for those with treatments in the budget period) over a longer period may lead to quite different results. But then again, this would enormously increase the complexity of the proposed model.

Medury and Madanat (2013) incorporate work-zone effects in a bottom-up network-level approach (budget allocation) by defining a network capacity threshold as a constraint. The applicability is demonstrated on the basis of a small case study of 11 road sections. Their model, however, neither accounts for economies-of-scale costs nor for different traffic demand on individual sections.

In summary, most of the approaches used in the literature are not scalable for large networks due to complexity issues. The models that are scalable, in turn, include too many simplifications to yield detailed applicable results at the project level. Apart from the mentioned limitations, almost all of the studies referred to above are based on road sections with lengths between 1 km and 5 km. Automated condition surveys, however, report data for much shorter survey sections (data segments) (e.g. 10 m in California, 50 m in Austria, 100 m in Germany, etc.). These survey sections are aggregated into homogeneous sections (management segments), with measurements being aggregated to condition variables. These long-used approaches were originally motivated by (i) the storage and processing capabilities of computers at the time, (ii) the exponential complexity of the optimization problem, and (iii) short sections not being suitable for practical application of treatments. However, the use of homogeneous sections leads to imprecise condition prediction, resulting in over- and underestimating service life on short sections. Imprecise predictions, in turn, lead to imprecise treatment recommendations at the section (project) level. Moreover, long homogeneous sections prevent the selection of work zones based on economies-of-scale costs, reducing the flexibility in the optimization.

Today, powerful database management systems which are capable of efficiently storing, managing and analyzing big data are available, so that the first issue (i) no longer applies. This paper contributes to resolving the second and third issue (ii and iii) by making use of short survey sections which will allow road agencies to benefit from high-quality data obtained by automated surveys.

4.3 CASE STUDY, PERFORMANCE AND COST MODELS

4.3.1 Overview of holistic PMS framework

A graphical overview of the new end-to-end PMS incorporating the novel optimization approach is provided in Figure 4.1. The figure aims to enhance the understanding of the optimization framework as a whole, illustrating the flow of information and the different computational steps. Based on a database containing section-specific information (structure, condition, age, traffic, etc.), the first crucial step is predicting the condition for each road section (and each distress type) and deriving a service life/condition distribution at the network level. If condition predictions are biased, the subsequent analysis, treatments and optimization results will deviate from reality, regardless of the methods applied. Modules 3, 5, 6 and 7 include the developed performance, duration and cost models described in this section. The first stage of optimization is conducted at the road section level (Section 4.4.1). The second stage is the optimization of work zones (Section 4.4.2), using the results from the first stage as initial values, also considering user costs and traffic control costs. The last stage is the optimization under budgetary restrictions, leading to less optimal results (as treatments are deferred) and the analysis of obtained results.

4.3.2 Service life with distress and spatial correlation

This section describes the development of the concept and data generation procedure for a case study with a simulated road network replicating real-world data characteristics (Donev and Hoffmann 2017b, Hoffmann and Donev 2017). Simulation studies are used quite often in the research literature on treatment optimization, especially when demonstrating the applicability of

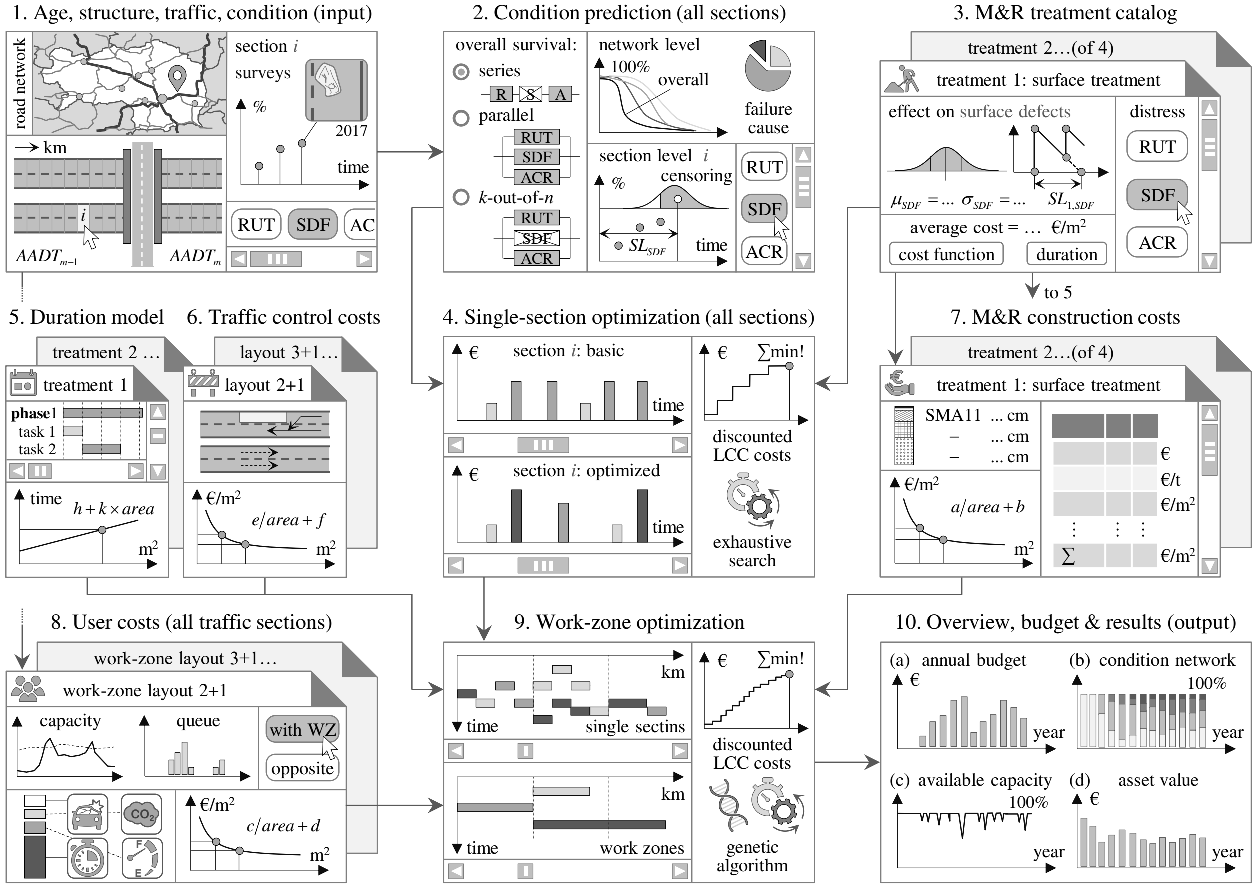


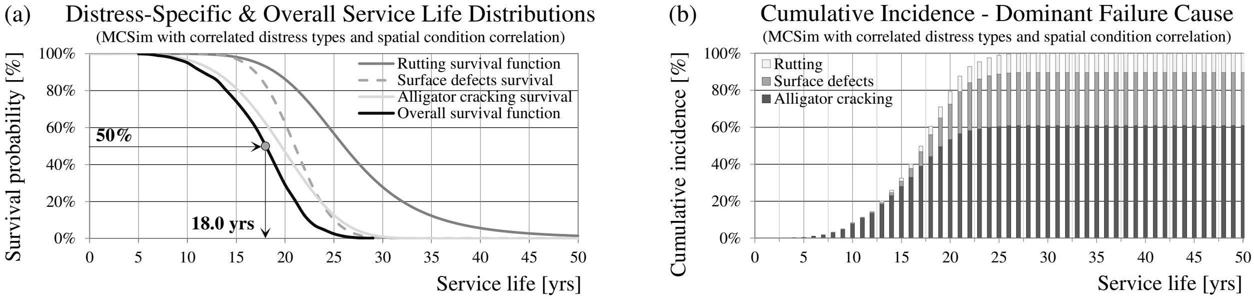
FIGURE 4.1 Overview of the developed performance, cost, duration and traffic models as modules of a holistic PMS, together with the sequence of computations from input to output.

new approaches (e.g. Fwa et al. 1996, Sathaye and Madanat 2012, Yeo et al. 2013). One of the reasons for this is that new approaches have to be tested for a variety of settings, and possible limitations have to be identified prior to any real-world data application. Sensitivity analysis and what-if scenarios help identify key factors, trade-offs and improvement potential of the overall system (e.g. design optimization, inspection intervals, etc.). Furthermore, a perfect knowledge of service life enables model comparison and objective quantification of the consequences of sectioning and aggregation in terms of threshold violations. This paper offers a general procedure for the simulation of multiple distress types and short survey sections, incorporating spatial correlation based on empirically calibrated input parameters.

The case study is based on 1000 road sections, each with a length of 50 m and located on one of the exterior lanes of a four-lane freeway. All input parameters are calibrated and conceptualized for flexible pavements on freeways in Austria. Other road classes characterized by different design or types of pavements may exhibit different distresses and failure causes. Pavement condition is being modeled by using multiple distress types, with different treatments being possible for different distress combinations. In contrast, many researchers only consider one distress type or aggregate individual distresses into one composite index which is used in the optimization. Roughness, as measured by the IRI, is perhaps the most common and often the only condition indicator included in optimization approaches. This is due to the fact that roughness is considered a primary determinant of user costs, and there are many roughness-based performance and user-costs models available in the literature. However, roughness as well as any composite index, is not in itself a failure mechanism but an indicator of several other distress types. Therefore, assigning treatments solely based on composite indices or roughness values fails to address the individual distress types appropriately.

For the purposes of a clear graphical representation of the results, this study considers only three main distress types for asphalt pavements: rutting, surface defects (raveling, polishing, bleeding, etc.) and alligator (bottom-up fatigue) cracking. However, the complexity of the approach is not affected in any way by the inclusion of more than three distress types, provided that the corresponding condition models are available. Service life with regard to each of these three distress types is assumed to follow its own distribution as shown in Figure 4.2(a,c). For more flexible modeling, three different distributions are used: the normal, the Weibull and the log-logistic. From a theoretical point of view, the service lives of the three distress types have to be correlated (i.e. correlated random variables). The used correlation coefficients shown in Figure 4.2(d) are estimated in the course of a preliminary analysis of data based on 1400 km regional roads in Austria (approx. 28,500 road sections). Without any claim to general validity, the assumed matrix is considered a realistic possibility for the purposes of the case study. As expected, the service lives are positively correlated, because most of the distress types share common influencing factors (traffic, climate, etc.). However, the estimated correlation coefficients show that the degree of correlation is not very strong.

Service life for each distress is defined as the physical time from construction until exceeding a specified condition threshold. The individual distress types are combined in a series system representing the service life of the whole pavement. Thus, on a given section, the individual distresses may be considered as competing risks with the first distress that exceeds its thresholds determining the failure cause and the overall service life. Figure 4.2(a) shows that the overall service life is on average shorter as the service life of the most critical distress. This follows from the fact that, for a series system, the overall survival probability at any time is the product of the survival probabilities for the individual distress types (in the case of uncorrelated distresses). Details about the estimation of service life with censored data and correlated competing risks can be found in Donev and Hoffmann (2017a).



(c) Distress-Specific Service Life Distributions
(Based on literature and preliminary analysis of real data in Austria)

Distress types

	Distribution type, parameters and expectation
Rutting	Shifted log-logistic: $\mu = 3.8$ (shift); $p = 5.8$ (shape); $\lambda = 0.045$ (scale); $E(X) = 27.1$ yrs
Surface defects	Normal distribution: $\mu = 21.0$ (mean); $\sigma^2 = 9.923$ (variance)
Alligator cracking	Weibull distribution: $p = 4.5$ (shape); $\lambda = 0.047$ (scale); $E(X) = 19.4$ yrs

(d) Distress Correlation Matrix Definition
(Based on preliminary analysis of regional-roads condition data in Austria)

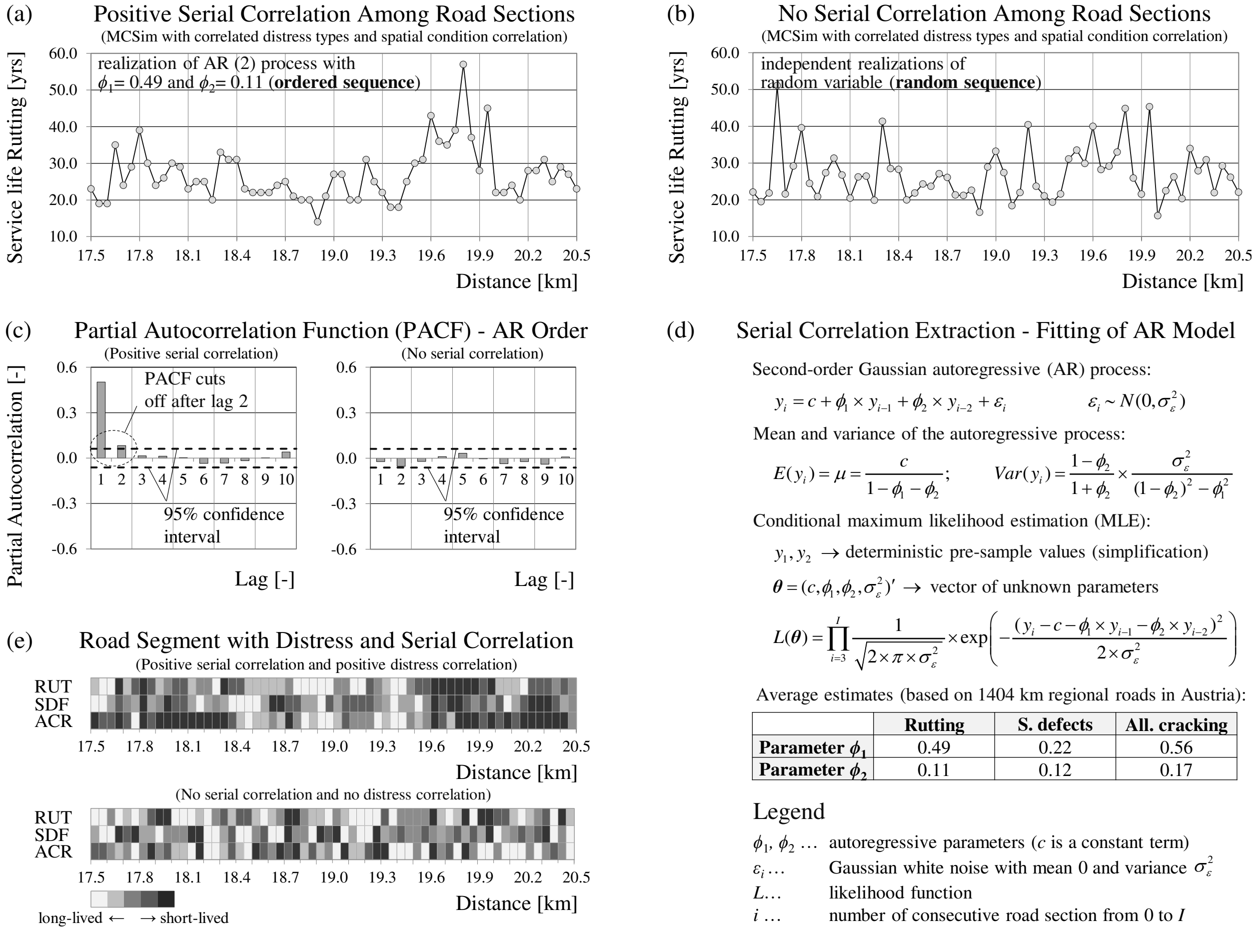
Pearson correlation

	Rutting	S. defects	All. cracking
Rutting	1.00	0.16	0.33
Surface defects	0.16	1.00	0.53
Alligator cracking	0.33	0.53	1.00

FIGURE 4.2 Overall service life (a) and dominant failure cause (b) resulting from the distress-specific service life distributions (c) and the correlations between these distributions (d).

The combination of marginal (distress-specific) distributions and distress-correlation matrix not only defines the overall survival function but also the dominant failure cause given by the cumulative incidence function, as shown on Figure 4.2(b). In particular, alligator cracking causes 60% of all failures, followed by surface defects with 30% and rutting with 10%. These values are assumed on the basis of the available literature and on the results of network-wide surveys of freeways in Austria (Fromm et al. 2012, unpublished). The parameters of service-life distributions are reverse-engineered based on the assumed failure causes, the estimated distress correlations and a structural life (cracking) of approximately 20 years (FSV, 2016).

So far, the spatial arrangement of the road sections has not been taken into account. The assumption of independent condition developments on adjacent sections is unrealistic, as they often exhibit quite similar condition. This study employs a times-series model to account for a positive correlation between adjacent road sections (spatial, serial or autocorrelation). The focus, however, is not on a sequence of observations in time, as it is usual for time series, but on series of ordered measurements at intervals of equal length (survey sections). Thus, a second-order autoregressive (AR) process was fitted to real-world data by using conditional maximum likelihood estimation. The basic formulas are shown in Figure 4.3(d) and for more details on time series, reference to the literature is made (e.g. Hamilton 1994, Montgomery et al. 2015). Condition data of regional roads from two Austrian states was used for this analysis with models



Serial Correlation Extraction - Fitting of AR Model

Second-order Gaussian autoregressive (AR) process:

$$y_i = c + \phi_1 \times y_{i-1} + \phi_2 \times y_{i-2} + \varepsilon_i \qquad \varepsilon_i \sim N(0,\sigma_\varepsilon^2)$$

Mean and variance of the autoregressive process:

$$E(y_i) = \mu = \frac{c}{1-\phi_1-\phi_2}; \qquad Var(y_i) = \frac{1-\phi_2}{1+\phi_2} \times \frac{\sigma_\varepsilon^2}{(1-\phi_2)^2-\phi_1^2}$$

Conditional maximum likelihood estimation (MLE):

$y_1, y_2 \rightarrow$ deterministic pre-sample values (simplification)

$\boldsymbol{\theta} = (c, \phi_1, \phi_2, \sigma_\varepsilon^2)' \rightarrow$ vector of unknown parameters

$$L(\boldsymbol{\theta}) = \prod_{i=3}^{I} \frac{1}{\sqrt{2 \times \pi \times \sigma_\varepsilon^2}} \times \exp\left(-\frac{(y_i - c - \phi_1 \times y_{i-1} - \phi_2 \times y_{i-2})^2}{2 \times \sigma_\varepsilon^2}\right)$$

Average estimates (based on 1404 km regional roads in Austria):

	Rutting	**S. defects**	**All. cracking**
Parameter ϕ_1	0.49	0.22	0.56
Parameter ϕ_2	0.11	0.12	0.17

Legend

ϕ_1, ϕ_2 …	autoregressive parameters (c is a constant term)
ε_i …	Gaussian white noise with mean 0 and variance σ_ε^2
L…	likelihood function
i …	number of consecutive road section from 0 to I

FIGURE 4.3 Positive serial correlation (a) vs. no serial correlation (b) of rutting service life for a 3-km road segment with partial autocorrelation function for the two cases (c). Fitting of AR (2) process to real-world data with the estimated autoregressive parameters (d) and graphical representation of the road segment simultaneously exhibiting serial correlation and distress correlation (e).

being fitted to individual roads. While the mean condition may vary between roads, the estimated autoregressive parameters are relatively consistent with the average values shown in Figure 4.3(d). The order (i.e. number of lags) of the autoregressive models is determined based on the partial autocorrelogram (Figure 4.3(c)). For the vast majority of sections, the correlation is significant over a length of one to four sections (50 m length) with two lags being the most common. Furthermore, the results showed that the number of lags and the degree of the correlation depend on the distress type and length of the survey sections (25 m and 50 m were examined). Figure 4.3(a,b) illustrates the difference between ordered and random sequences of service lives for rutting.

Incorporating spatial correlation will not change the marginal distributions in Figure 4.2, but will affect the arrangement of the simulated service lives. Generating a data set that simultaneously accounts for distress and spatial correlation, as shown in Figure 4.3(e), is not straightforward, but a detailed description is beyond the scope of this paper. In short, the procedure consists of the following two steps. First, a set of correlated random numbers is obtained using Cholesky decomposition of the distress correlation matrix (Meissner 2014). In a second step, the spatial correlation is introduced to the same set of correlated random numbers based on the estimated autoregressive parameters. Random variates (service lives) are then obtained by using the inverse-transform method (Rossetti 2016).

The incorporation of spatial correlation is of critical importance as it affects the grouping of adjacent short-survey sections to longer work zones based on economies of scale. The goal here was not to develop a general methodology for modeling of spatial dependence but to generate realistic data set for the case study, also allowing for a ceteris paribus comparison with common approaches. A more accurate and flexible way to account for spatial variability is based on the theory of random fields (Lea and Harvey 2015).

In summary, the simulated road network could be considered as a realistic case study due to the extensive empirical calibration. This not only assigns more weight to the delivered proof of application (see Section 4.5), but also allows for some practical conclusions to be drawn, which are subject to further investigation.

4.3.3 M&R treatments: definition and service life

This work assumes a flexible pavement structure which is based on the Austrian design catalogue (FSV 2016) and consists of three asphalt layers (surface, binder and base) as well as two unbound layers (road base and subbase). Accordingly, four M&R treatment types are considered, with each treatment allowing more extensive (deeper) repair of the pavement (Figure 4.4). Each treatment is represented by a color code (for visualization) and an integer code (mathematical model). Since the case study only considers the exterior lane, there is a vertical restriction and the existing pavement height must be maintained. Therefore, all treatments include the removal of the corresponding existing layer(s).

As will be seen in Section 4.4.4 of this work, the number of treatment types has a great impact on the complexity of the optimization problem. Therefore, only a moderate number of M&R alternatives should be considered. Treatments with effect on the same distresses, exhibiting similar service life and costs, can be represented as a group or by the most cost-effective. For example, there are many types of surface treatments: microsurfacing, slurry seal, fog seal, chip seal, ultrathin friction course, and so forth. If the optimization recommends a surface treatment,

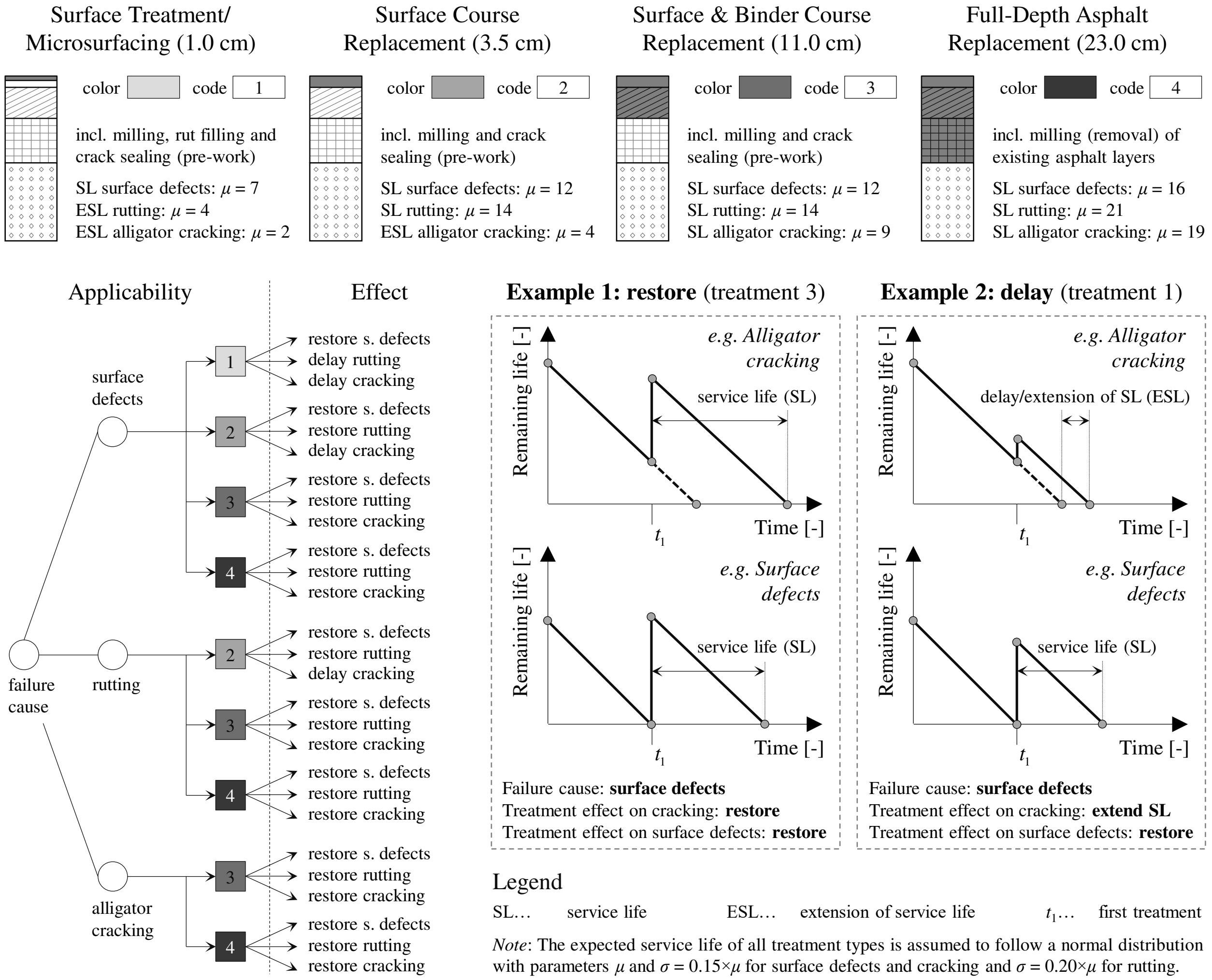


FIGURE 4.4 M&R treatment alternatives, characterized by depth of repair, effect (restore/delay) on the different distress types and technical applicability (decision tree).

then the exact type could be specified at the planning stage based on other factors (e.g. experience, available technology, bidding results). Furthermore, the selection of treatments here represents the most common treatments applied on Austrian freeways. Bearing-capacity problems of the unbound layers, which would require a full-depth pavement replacement, are the exception. Crack sealing and pothole patching are usually considered in the budget for routine (corrective) maintenance and applied on an annual basis or as soon as possible, respectively.

The effect of each treatment on the three modeled distress types is described with a distinct service-life distribution (see Figure 4.4). Due to the lack of evidence for other functional forms, and for simplicity, it is assumed that the treatment life follows a normal distribution with the mean value based on literature review (e.g. Hall et al. 2001, Cuelho et al. 2006, Anastasopoulos and Mannering 2015, Birbaum 2016, Nobakht et al. 2016). Since the literature rarely distinguishes between effects on different distress types, the values were supplemented using engineering judgment and surveys. The standard deviation is computed as a percentage of the mean value (20% for rutting and 15% otherwise). These values were selected in accordance with the literature and in order to achieve similar distribution of failure causes as compared to the initial service life.

Besides the generated service lives, no continuous distress or improvement model is employed in the current case study. This property ensures a highly robust model which is not dependent on additional assumptions about performance functions. Therefore, service life, defining the latest possible timing of treatment application, is the only information needed. Since all considered treatments apply to the whole area of the lane and no local repairs (patching) are considered, there is no particular benefit in knowing the exact distress extent prior to the threshold. However, if the condition thresholds cannot be satisfied (e.g. due to budgetary restrictions) and the service life is exceeded, a performance model would be necessary to measure the extent of threshold violation. Furthermore, the presented approach can incorporate any mechanistic, empirical or stochastic condition model for the description of intermediate states, with no repercussions as to the complexity of the problem.

In the absence of a performance model, the deterioration is illustrated with the remaining life function, which equals the service life at the time of construction/treatment application and decreases linear to zero at the time of failure (see Figure 4.4). The model also considers the technical applicability of treatments depending on the failure cause. As illustrated by the decision tree, if the failure cause is alligator cracking, only two treatments can be applied. For the purposes of the case study, it is assumed that rutting is confined within the surface course.

Furthermore, two types of treatment effect are considered: restoration/reset (Example 1) and delay/service life extension (Example 2). The restoration effect is the standard one, and it resets the remaining service life (if any) to the treatment life. For example, if microsurfacing is applied, the service life for surface defects will be reset, but the service lives for rutting and alligator cracking will continue to elapse unaffected. The delay effect, however, assumes that the remaining life for rutting and alligator cracking will be extended by a small amount, shifting the remaining life function horizontally. This assumption is justified by the inclusion of pre-treatment works (crack sealing, rut leveling) in the description of the treatment (microsurfacing).

The treatment life is correlated with the initial service life as well as with the treatment lives on the adjacent road sections. The proposed approach allows for modeling of stochastic treatment lives for the time period after the first treatment when no condition information is yet available (see Appendix 4.A.1 for details). In such cases, deterministic models cannot be calibrated, as customary, using shift or scale factors. Deterministic models will thus predict equal service life, same failure causes and equal treatment intervals for large groups of sections leading to underestimation of M&R needs.

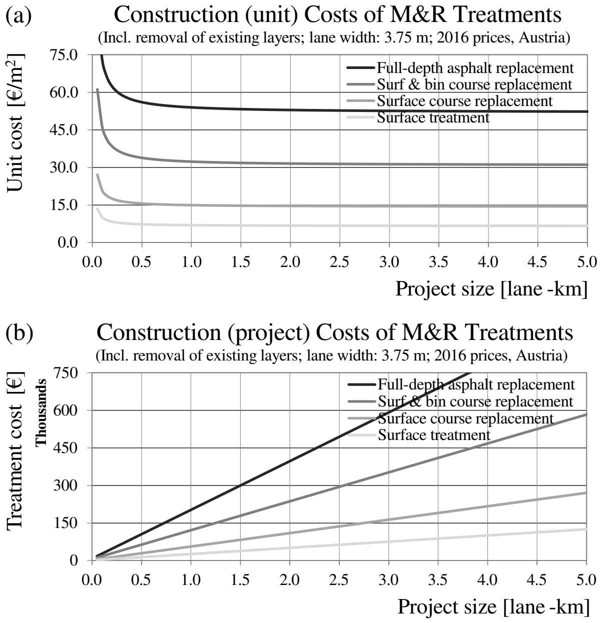
4.3.4 Agency costs for M&R treatments

The estimation of agency costs for M&R treatments is based on a linear model with fixed and variable cost components. The parameter estimates for the four treatment types are provided in Figure 4.5 (bottom right) with graphical representation as unit costs and total project costs shown in Figure 4.5(a,b). The unit costs decrease with increasing project size due to the spreading of fixed costs over larger work-zone area (economies-of-scale effect). It is not possible to capture these effects with average costs used by many PMS leading to an over- or underestimation of treatment costs for a given project.

The cost functions are based on detailed bottom-up calculation, a simplified version of which is also shown in Figure 4.5. The calculation includes the necessary work activities for each treatment, the selection of asphalt mix and thickness, as well as unit prices for each item.

The prices for each item are then divided into fixed and variable components and summed up accordingly. The advantage of this bottom-up calculation is its facile adjustment to other regions or countries by using current prices for material, labor, fuel, etc.

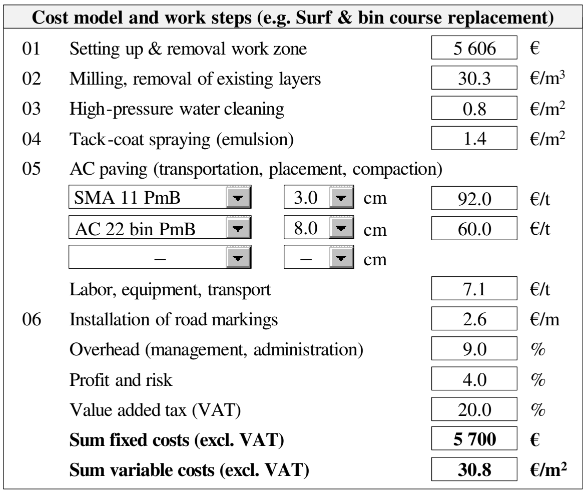
The unit cost estimates are based on a comprehensive analysis of contractor's bid prices for 60 M&R projects in Austria, price lists from asphalt mixing plants, data from road planning offices and available literature (Wießmayer 2006, Lindenmann 2008, Birbaum 2016). All prices were converted to 2016 constant euros using construction price indices for Austria. The costs for setting up and removal of the work zone (often given as a lump sum in bids) cover the following activities: site clearance, office facility setup, power/water supply, material storage, transportation of tools, equipment, machinery, vehicles, scaffolding, and others. The analysis has shown that a large portion of these costs are fixed costs. However, the setup costs not only depend on the treatment type, but on the size and complexity of the project. The cost model accounts for this variable proportion by using empirical relationships, with the project duration as a proxy for size and complexity (Birbaum 2016). If pre-treatment works (crack sealing, rut filling) are considered for a given treatment, their costs are included as well.


(a) Construction (unit) Costs of M&R Treatments

(Incl. removal of existing layers; lane width: 3.75 m; 2016 prices, Austria)

(b) Construction (project) Costs of M&R Treatments

(Incl. removal of existing layers; lane width: 3.75 m; 2016 prices, Austria)


Cost model and work steps (e.g. Surf & bin course replacement)

01	Setting up & removal work zone		5 606	€
02	Milling, removal of existing layers		30.3	€/m³
03	High-pressure water cleaning		0.8	€/m²
04	Tack-coat spraying (emulsion)		1.4	€/m²
05	AC paving (transportation, placement, compaction)			
	SMA 11 PmB	3.0 cm	92.0	€/t
	AC 22 bin PmB	8.0 cm	60.0	€/t
	–	– cm		
	Labor, equipment, transport		7.1	€/t
06	Installation of road markings		2.6	€/m
	Overhead (management, administration)		9.0	%
	Profit and risk		4.0	%
	Value added tax (VAT)		20.0	%
	Sum fixed costs (excl. VAT)		**5 700**	**€**
	Sum variable costs (excl. VAT)		**30.8**	**€/m²**

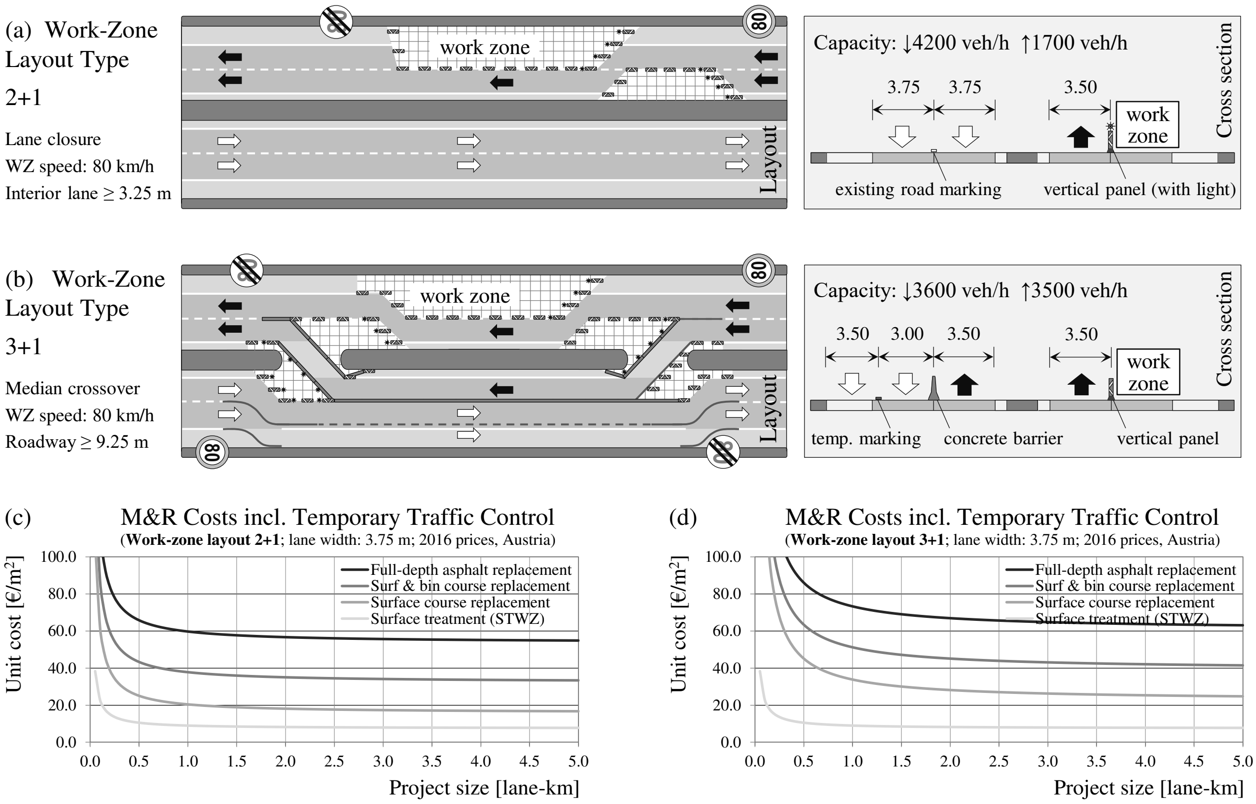
M&R treatment type (depth)	Fixed [€]	Variable [€/m²]
Surface treatment (1 cm)	1 300	6.6
Surface course replacement (3.5 cm)	2 400	14.3
Surf & bin course replacement (11.0 cm)	5 700	30.8
Full-depth asphalt replacement (23.0 cm)	7 800	51.9

FIGURE 4.5 Bottom-up M&R construction cost calculation based on price lists and contractor's bid prices from Austria. Unit costs (a) and total project costs (b) as a function of the project area for the considered four treatment types.

4.3.5 Agency costs for temporary traffic control

Due to extensive work-zone safety measures on freeways with heavy traffic, temporary traffic control costs (TTCC) represent a large proportion of the total agency costs for M&R projects in Austria and Germany. Naturally, TTCC depend on the work-zone layout and the functional class of the road. Figure 4.6(a,b) shows two typical configurations for a work zone in the exterior lane of four-lane freeways with the corresponding lane widths, work-zone capacity, speed limit and channelizing devices. These layouts are simplified versions of the exemplary traffic management plans for long-term work zones provided in the Austrian standards (FSV 2012). For the estimation of work-zone capacity a base capacity of 2100 veh/h per lane is used, together with capacity reduction factors derived from the literature (Hellmann et al. 2008, Fischer 2009). If the work zone extends over the entire roadway, the construction can be staged with lane closures in sequence shifting traffic from the existing to the newly repaired lane. However, technological and quality considerations may require both lanes to be repaired in a single pass with full-width asphalt paver. In that case, a full roadway closure will be necessary (e.g. layout 3+0 or 4+0). Each alternative is associated with different agency costs, available roadway capacity and accident risk.

TTCC are often neglected in the literature on budget allocation (see Section 4.2.1), as there is no way to attribute these costs without work zones. Moreover, temporary traffic control is sometimes an in-house service, which the agency performs with its own staff. Thus, there is not much literature on the estimation of TTCC, apart from some estimates as a percentage of the project's total cost. The cost functions for layout 2+1 and 3+1 given in Figure 4.6(c,d) are based on data provided in the German literature (Fischer 2009, Birbaum 2016), with adjustment for Austria. The cost functions are based on a more detailed bottom-up cost calculation with unit prices for the different items and services. The fixed component consists of costs for the setup of the work zone approach and transition area, asphalt works for unpaved median strips (where relevant) and installation of lighting devices in crossover areas. Length-dependent variable costs include the delivery/setup of vertical panels separating the worksite from vehicular traffic as well



Cost components	STWZ	2+1	3+1
Fixed costs [€]	4 500	15 000	40 000
Variable costs (length) [€/m]	3.0	5.0	30.0
Variable costs (length & time) [€/m×day]	-	0.07	0.11

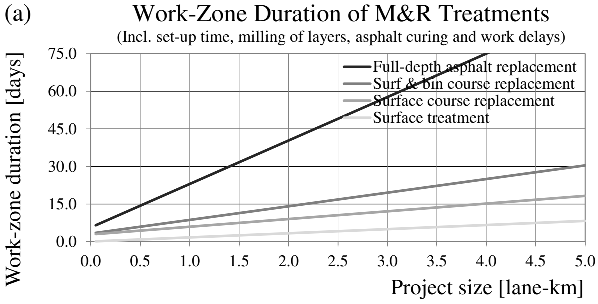
Note: Short-term work zone (STWZ) is the default WZ layout for surface treatment. The layouts for long-term work zones comply with the Austrian standards (RVS). Cost and capacity calculations are based on literature and assessment of experts.

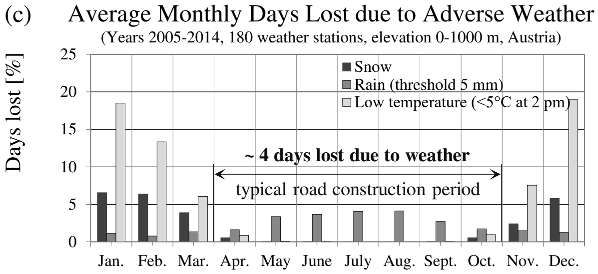
FIGURE 4.6 Common exterior-lane work-zone layouts on Austrian freeways (simplified) with corresponding speed limits, channelizing devices and roadway capacity values (a, b). Temporary traffic control costs derived by bottom-up approach as a function of the project length (c, d).

as the installation of temporary markings and traffic signs. Time-dependent costs comprise running costs (e.g. electricity), rental costs (for traffic control devices), surveillance, daily inspections, maintenance, and cleaning. Microsurfacing is applied by a specialized truck- or trailer-mounted mixing and paving machine in a continuous way, which requires a mobile short-term work zone (STWZ) with significantly lower TTCC. In summary, TTCC are an important component in the agency budget with a high proportion of fixed costs, which is essential for the grouping of work zones.

4.3.6 Work-zone duration model

To account for the impact of construction activities on road users as well as time-dependent traffic control costs a treatment duration model is needed. The simplified treatment duration model shown in Figure 4.7(b) is based on paving and milling production rates and accounts for setup time, delays and adverse weather days. Production rates for an eight-hour workday are selected using literature (Lindenmann 2008, Brozek et al. 2009) and data from manufacturers' specifications (e.g. Wirtgen 2013). A fixed time for work-zone setup and removal (factor *c*) is included for each treatment type. Especially for new construction, the setup time will also depend on the size and complexity of the project. The productivity delay factor *r* is assumed to be equal to 1.2 for M&R works (i.e. 20% increase of project duration), accounting for material- and equipment-related idle times, design changes or errors, poor management, and so forth (Anastasopoulos et al. 2012). Factor *f* considers days lost due to adverse weather conditions which have an influence on asphalt paving activities (snow, rainfall, low or high temperatures). In order to determine this factor, observations from 180 weather stations in Austria for nine consecutive years were analyzed, with the average results shown in Figure 4.7(c). Region- and month-specific factors could be used in a more detailed model, although the individual analysis of each of the nine Austrian states for two elevation groups (0-500 m and 500-1000 m) showed differences primary in the winter period, when no M&R activities take place. The thresholds





(b) Work-zone duration *D* in calendar days (discontinuous model):

$$D = \left(c + \frac{Area}{Paving\ output\ \ [\mathrm{m}^2/\mathrm{day}]} + \frac{Area \times Milling\ depth}{Milling\ output\ \ [\mathrm{m}^3/\mathrm{day}]} \right) \times r \times f$$

Legend

c…	fixed time for setup and removal of work zone, temporary traffic control and asphalt curing
r…	factor productivity delays (e.g. breakdown, out-of-sequence)
f…	factor lost days due to weather, holidays and weekends

Production rates (daily output):

M&R treatment type	**Paving [m²/work day]**	**Milling [m³/work day]**
Surface treatment (1 cm)	4 000	800
Surface course replacement (3.5 cm)	3 000	800
Surf & bin course replacement (11.0 cm)	2 000	500
Full-depth asphalt replacement (23.0 cm)	800	200

Continuous work-zone duration model (linear approximation):

M&R treatment type	**Fixed [days]**	**Variable [days/1000 m²]**
Surface treatment (1 cm)	0.0	0.439
Surface course replacement (3.5 cm)	2.8	0.823
Surf & bin course replacement (11.0 cm)	3.2	1.451
Full-depth asphalt replacement (23.0 cm)	5.6	4.630

FIGURE 4.7 Derivation of the work-zone duration model (a, b) using production rates from the literature and accounting for days lost due to adverse weather conditions based on an extensive analysis of 180 weather stations over nine years in Austria (c).

defining lost days are based on the Austrian specifications for asphalt paving and literature (Williams et al. 2009). The resulting work-zone duration as a function of the project size for each treatment is graphically represented in Figure 4.7(a). The linear models show a very close approximation to the formula, which exhibits discontinuities for a duration of less than five days.

4.3.7 Work-zone user and environmental costs

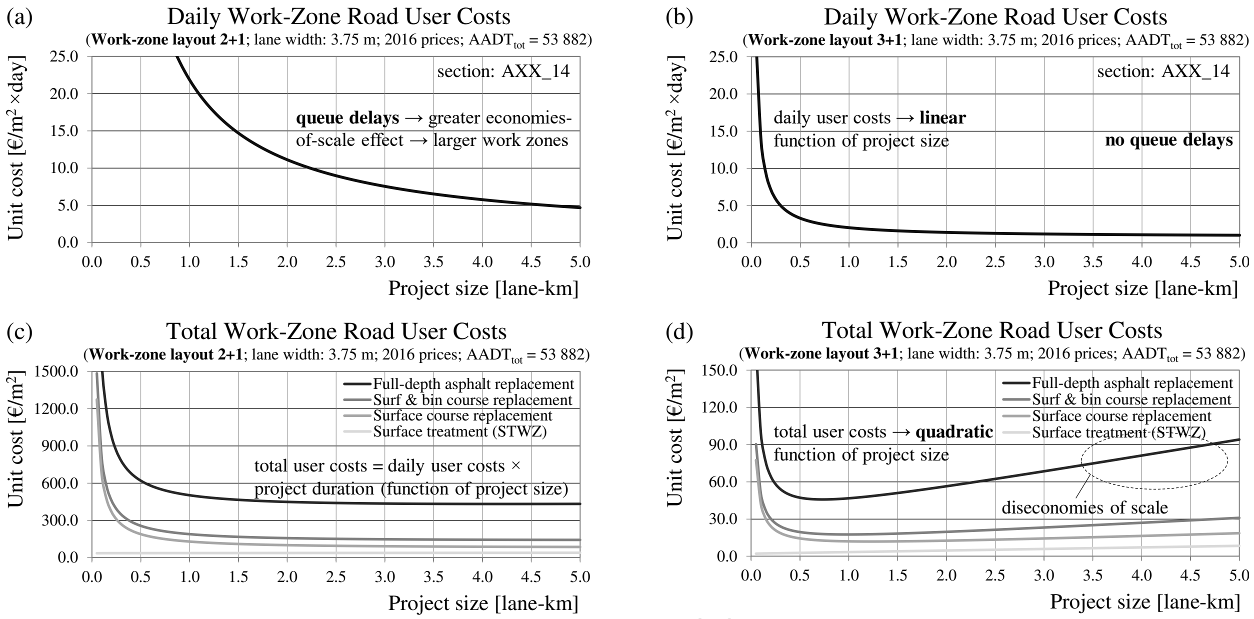
User costs include traffic delay costs (TDC), vehicle operating costs (VOC), emissions costs (EC), and crash costs (CC) according to EU regulations (Eurovignette Directive). In general, a distinction can be made between condition-related and work-zone related user costs. Condition-related user costs are a function of the road condition, with the majority of models linking user costs to roughness (IRI). However, measured IRI values are often not directly comparable between agencies due to differences in equipment, methods, and reporting (Ogwang 2016). Thus, the existing IRI-based models are not always applicable and, in addition, not every agency disposes of sufficient resources in order to develop its own user-cost models. Furthermore, a situation where a poor pavement condition (represented by IRI) will lead to speed reductions and ultimately to delay times is not realistic for freeways in Austria. Condition surveys performed on Austrian freeways show that only 0.3% of the road sections are in the worst roughness class (Fromm et al. 2012, unpublished).

Therefore, condition-related user costs are not considered currently as their effect within the given thresholds is very limited. Instead, the model focuses on the additional work-zone induced user costs, including TDC, VOC, EC, and CC. The employed model parameters and unit costs are representative for the Austrian freeways as they are a result of extensive literature analysis and calibration (e.g. Bennett and Greenwood 2001, Brozek et al. 2009, FSV 2010, BMVIT 2013, Ntziachristos and Samaras 2016). Furthermore, all components of the model are adjusted to 2016 constant euros with the consumer price index. Travel delay costs are estimated for each traffic section, direction of travel and work-zone layout using a macroscopic demand-capacity model based on Walls III and Smith (1998) as well as Mallela and Sadasivam (2011). The hourly traffic demand is the product of the annual average daily traffic (AADT) and typical hourly distribution factors obtained from historical traffic counting data. The vehicle operating costs are computed as a product of vehicles number (by type), the additional fuel consumption (gasoline/diesel), and fuel unit costs. The emission costs are estimated on the basis of emission factors (e.g. kg CO_2/kg fuel) and unit costs, assuming proportionality to the additional fuel consumption. The crash costs per day are calculated using the expected additional number of crashes per day and the average costs per crash, assuming increased risk in the work zone. Additional details on the employed model parameters, as well as on the estimation of TDC and CC are provided in Appendix 4.A.2.

With these parameters, daily user-cost functions are derived for each combination of traffic segment and work-zone configuration. The daily user costs apply to all treatment types, being computed as a sum of TDC, VOC, EC, and CC for both directional roadways. The estimated user costs for a given traffic section (AXX_14) and two work-zone layouts (2+1 and 3+1) are shown in Figure 4.8(a,b). The function parameters for the other traffic segments are given in Figure 4.8 (bottom right). The daily user costs show a linear increase with project size, having a large intercept value (fixed costs). The high proportion of fixed costs results from components independent of work-zone length, namely: speed change delay, stopping delay, queue delay, and fixed lengths for approach, transition and buffer areas. Layout 2+1 provides lower capacity due to the closed lane, and the traffic demand on this section exceeds the capacity, leading to the formation of queue. In contrast, layout 3+1 provides sufficient capacity and no queue delays occur. Thus, in the case of congestion (layout 2+1), fixed costs become extremely high, as the number of work zones increases.

The additional work-zone user costs for the whole project are obtained by multiplying daily user costs by treatment duration in days. Thus, the total (project) user costs are a quadratic function of the project's size as both daily user costs and project duration are a linear function of the project's size. In the case of queue formation (layout 2+1), the total user costs exhibit economies-of-scale characteristics (Figure 4.8(c)). If the demand does not exceed the available capacity (layout 3+1), the total user costs are lower by an order of magnitude (see Figure 4.8(d); note the different scale of the *y*-axis). However, with the increase of project size (> 1.0 km), diseconomies of scale occur. The reason for this is that the daily costs converge to a constant value for work zones larger than 1.0 km, while the project duration increases further. In other words, for work-zone lengths which exceed an optimal value (e.g. 1.0 km), it will be more beneficial to have more work zones with shorter durations than to have fewer work zones with longer durations.

In this work, the network-level pavement management is represented by a single freeway or a route (50 lane-km). For a network-wide extension, a network model (e.g. graph) will be needed, taking into account possible detours and the redistribution of traffic. Appendix 4.A.3 includes a summary table of all calibrated service life, performance and cost model parameters.



Traffic volume - annual average daily traffic and % trucks:

Traffic section	From km	To km	Direction ↓		Direction ↑WZ	
			AADT	Trucks	AADT	Trucks
AXX_10	0.000	11.500	27 348	15%	28 066	14%
AXX_14	11.500	20.500	26 498	12%	**27 384**	**12%**
AXX_23	20.500	25.150	18 778	16%	19 462	14%
AXX_35	25.150	40.400	19 779	14%	20 509	14%
AXX_50	40.400	50.000	20 407	13%	18 680	14%

Derived user cost functions for the given traffic segments:

Traffic section	User cost (layout 2+1)		User cost (layout 3+1)	
	€/day	€/m×day	€/day	€/m×day
AXX_10	101485	1.48	4924	2.93
AXX_14	80513	1.45	4757	2.87
AXX_23	1555	1.03	3402	2.02
AXX_35	1634	1.09	3573	2.13
AXX_50	1488	0.99	3462	2.07

FIGURE 4.8 User costs incorporating TDC, VOC, EC and CC derived by bottom-up approach as a function of the project length for a single day (a, b) and for the total project duration by treatment type (c, d). Different traffic volumes and work-zone layouts lead to different parameters of user-cost functions.

4.4 SOLUTION METHODOLOGY

This section describes a new formulation of the M&R optimization problem based on short survey sections and multiple distress types. The solution is obtained in multiple stages: section-level, work-zone and budget. However, the results from a previous stage are not binding and are only used as initial values, subject to change.

4.4.1 Section-level optimization

At the section level, the M&R optimization problem consists of finding the temporal sequence of treatments (timing and type) that minimizes the discounted total agency life-cycle costs over a sufficiently long planning period. In general, there are two principal approaches for determining treatment timing on a single section: threshold-based and/or consideration of user costs. As discussed in Section 4.3.7, this work does not consider condition-related user costs as their effect within the considered condition thresholds remains limited.

Condition thresholds are employed by many authors as a simplification in the optimization (e.g. Chu and Chen 2012, Sathaye and Madanat 2012). Although the proposed approach also relies on thresholds at the section level, thresholds are only used as an upper limit for the optimization of work zones (see Section 4.4.2). Nevertheless, condition thresholds limit the solution space, with the final solution being optimal only within these limits. This underlines the importance of thresholds in general and their influence on service lives and costs. Furthermore, the fixed thresholds can be relaxed for some distress types (e.g. cracking) if the treatment effectiveness is estimated as a function of the current condition.

The methodology for M&R optimization at the section level is illustrated in Figure 4.9. For simplicity, the figure only shows two distress types and the first two treatment applications. As a reminder, service life defines the latest possible timing of treatment application (hard constraint), and at the end of the remaining life a treatment is triggered. The remaining life function equals service life at the time of construction/treatment and decreases linear to zero at the time of failure. For example, in Figure 4.9, the first failure occurs at t_1 due to surface defects. Alternative 1 includes a surface treatment which delays alligator cracking propagation, but a structural (second) treatment is still needed at time t_2. Alternative 2 consists of bringing forward the structural treatment to the time t_1 resetting both distress types. Depending on the discount rate, if the increase of present value of structural treatment ($t_2 \rightarrow t_1$) is less than the present value of surface treatment, then the second alternative will be optimal. Although the first distress type that reaches its threshold defines the timing of the first treatment, the type of the selected treatment in the algorithm depends on the effect of the treatment on all distress types and the resulting timing for the second and all subsequent treatments.

This approach considers all feasible temporal sequences of M&R treatments. In contrast, most models assume repetitive applications of the same treatment or treatment sequence and constant treatment intervals. Furthermore, the selection of a specific treatment determines the failure cause for the next treatment depending on the generated stochastic treatment lives. Thus, failure causes are not known a priori, and the applicability constraints (see Section 4.3.3) are not fixed but change dynamically during the optimization.

In this paper, the planning horizon for M&R activities is set to 100 years accounting for residual value at the end of this period. The residual value is computed as a ratio of remaining life

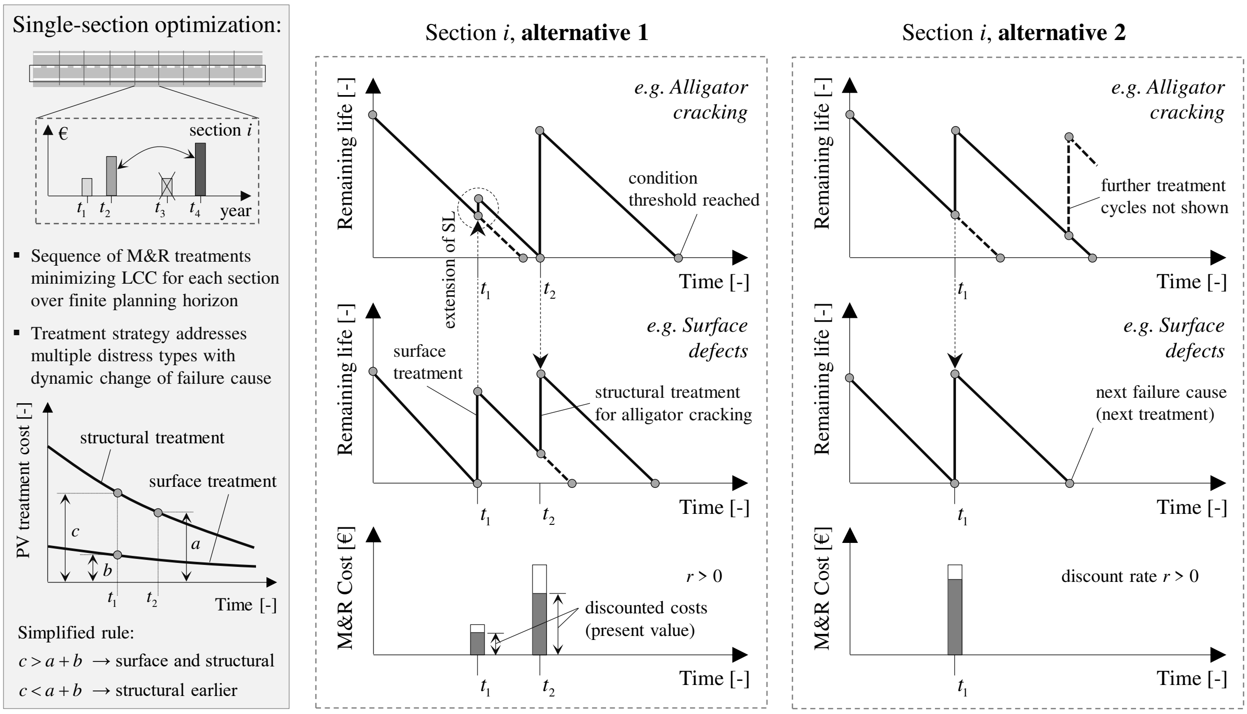


FIGURE 4.9 Methodology for M&R optimization on a single road section for multiple distress types based on average costs, condition thresholds and treatment applicability considerations.

to service life of the last planned treatment multiplied by the cost of this treatment (Stone 2013). This value is discounted and subtracted from the total life-cycle costs. In contrast, having a short planning period without accounting for residual value will push the more cost-intensive repairs right after the end of the period, creating an M&R backlog. As the computation of residual value (with no real market) is somewhat subjective, a combination of long planning period and residual value is preferred. Moreover, a planning horizon of 100 years is adequate with regard to cross-asset optimization and typical service lives of bridges and tunnels (80-100 years).

The optimization at the section level uses average treatment costs (without traffic control) and no economies-of-scale cost functions, as the costs would be too high for 50-m sections. This is not a drawback, as both timing and treatment type could be changed at the next stage (work zones). Nevertheless, the single-section optimization is still needed to optimize the full sequence of treatments up to 100 years. Then, at the second stage, the first three treatments are optimized with regard to work zones, and the remaining sequence of treatments is adopted from the first stage (section level).

4.4.2 Work-zone optimization

The work-zone optimization for short survey sections is, in essence, the main step that distinguishes the proposed methodology from any other approach known to the authors. In contrast to the literature, the focus is not only on operational aspects and short-term impacts of work zones on the road users. Instead, treatment type and timing on short survey sections (50 m) are optimized and coordinated in order to build larger work zones, based on economies-of-scale cost functions (treatment and traffic control costs) and a full life cycle. At the same time, the pavement condition is modeled by using multiple distress types. The present distress or distress combination determines the technical applicability of treatments. Thus, the length and location of a work zone is a pure economic decision, which is based on more accurate predictions with detailed untransformed survey data. The objective for the work-zone optimization is to minimize the sum of discounted total costs for all considered single sections. The decision variables consist of treatment timing and type with initial (and feasible) values provided by the single-section optimization.

The proposed approach can only allow the formation of a work zone if the constraints on all involved single sections remain satisfied. Thus, the section with the earliest failure time in a work zone determines the treatment timing for the entire work zone (due to condition thresholds). If failure causes are also different, the treatment selection is based on the most severe (deeper) failure cause (applicability constraint). The methodology for work-zone optimization is illustrated in Figure 4.10. For simplicity, only two road sections and one distress type are shown. The treatment on section *i* corresponds to Alternative 2 in Figure 4.9. The contiguous section *i*+1 has the same first treatment, but at a later time (t_1'). The cost for two separate work zones simply represents the sum of the discounted total costs for both sections. In the case of a combined work zone, both sections share the fixed costs, resulting in lower treatment costs per section. The treatment on section *i*+1 is brought forward in this case, resulting in an increased present value and a loss of service life. The optimal alternative is the one with a lower present value of life-cycle costs for both road sections.

In summary, the work-zone optimization is based on the trade-off between loss of service life and benefits gained by scale economies. The loss of service life is expressed in monetary terms as the increased present value of the required treatment's cost. The increase of present value is calculated for all subsequent treatments in the life cycle, as not only the first treatment, but the entire cycle is brought forward. The average work-zone length depends on the degree of spatial correlation (see Section 4.3.2), the share of fixed costs on total costs and the discount rate. Furthermore, at this stage, work-zone user costs may be considered as well (see Section 4.5.4).

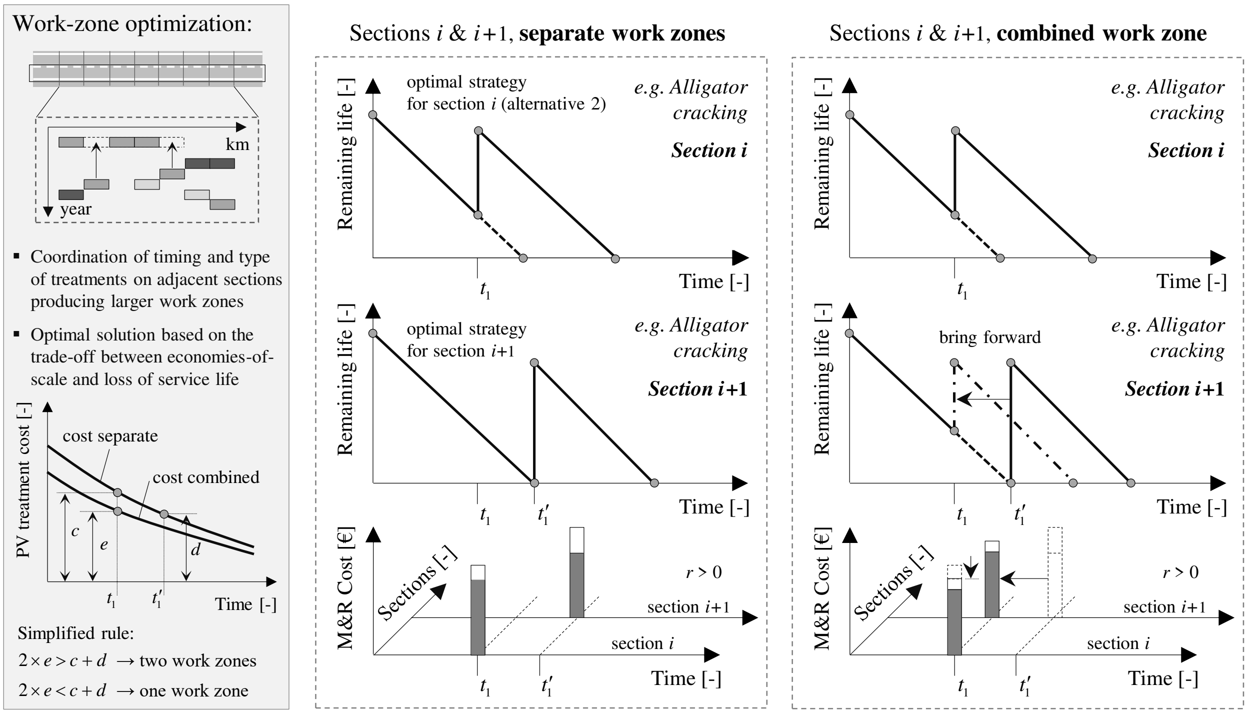


FIGURE 4.10 Optimization of M&R treatment type, timing and work zones for multiple distress types based on economies-of-scale costs, condition thresholds and treatment applicability considerations.

4.4.3 Soft constraints and treatment of outliers

In contrast to long homogeneous sections the proposed model guarantees the satisfaction of all condition thresholds. However, when grouping a number of road sections together, the timing of treatments in resulting work zones equals the timing of the most short-lived section. This may lead to early treatments and an inefficient use of pavement service lives. If the goal is to always satisfy existing condition thresholds, either higher investments will be needed or the agency may choose to adjust these thresholds. In contrast, the current practice is to form long homogeneous sections representing the condition with an average value, which effectively masks current and future condition threshold violations.

A possible way to solve this issue is to model condition thresholds as soft constraints. Thus, the model will allow for the exceedance of thresholds by a specific amount which is determined by a penalty function. For each distress type, one of the two options (hard or soft constraints) has to be chosen. It is reasonable to select hard constraints for distress types related to road safety (e.g. skid resistance, rutting). Soft constraints can be used for other distress types, where the threshold is more arbitrary (e.g. alligator cracking). The use of soft constraints, however, imposes the integration of a performance model which measures the extent of exceedance. The penalty function adds additional costs to the treatment costs, depending on the degree of exceedance. For a road section with more gradual condition deterioration a longer treatment delay will thus be allowed, ceteris paribus, in comparison to a section showing a more rapid deterioration. The penalty cost may be thought of as the additional cost for small repairs maintaining the condition until the timing of the delayed repair is optimal. Alternatively, penalty values may represent condition-related effects from deferring treatments (e.g. water ingress → deeper damage) which are requiring the replacement of more material. In any case, the penalty function must be calibrated in such a way that a postponement of treatments will lead to longer work zones. A first application of this concept is demonstrated in Section 4.5.2.

Even if soft constraints are used, some sections are not grouped after the optimization, because it is more economical to pay high fixed costs for a single section then to form a combined work zone. The presence of such outliers (short-lived or long-lived) is not a problem, since these sections could be repaired locally in short-term (daytime) work zones, as long as their number remains limited.

4.4.4 Complexity and solution approach

A more detailed and accurate modeling (short sections, multiple distress types and treatments) leads to an optimization problem that is more difficult to solve compared to common approaches. At the section level the number of solutions is limited by applicability and threshold constraints, which allows for an exact solution to be found. An exhaustive search algorithm is automated to perform optimization for all sections in a sequence. However, the problem of the work-zone optimization is far more complex (see Figure 4.11). The decision variables include the type (x_i) and timing (y_i) of treatment on section i. In the literature (e.g. Lee and Madanat 2015), timing is often considered a discrete variable with step size of one year, regardless of the use of a continuous or discrete time deterioration model. In this paper, timing is also an integer. If continuous time is used instead, the solution needs to be rounded to an integer anyway in order for scale economies to apply. Treatment type is also a general integer variable with a domain of possible values {1, 2, 3, 4}.

Furthermore, the problem is non-linear as the discounting function and the deterioration models (in general) are not linear. However, the most unpleasant feature of the problem is the non-convex and non-smooth objective function, resulting from economies-of-scale effects and the nominal scale of the treatment type variable. This is another reason why continuous time and mixed-integer formulation are not beneficial, since methods for mixed integer non-linear programming (e.g. outer approximation, generalized Benders decomposition) are suited for solving convex problems.

An evaluation of the objective function (fitness) landscape for only two variables (timing of sections 5 and 6) sheds light on the complexity of the search space (Figure 4.11, bottom right). The structure of the real multi-dimensional fitness landscape includes multiple local traps and is difficult to exploit, as there is no specific direction in which the function is increasing or decreasing. Furthermore, the number of possible solutions grows exponentially with the size of the problem (see bubble chart). So, even a relatively small number of road sections (20-30) results in an astronomical number of solutions. Due to the fact that the problem cannot be solved by exact methods, a heuristic method based on stochastic principles is chosen and applied.

The main emphasis of this paper is on the development of a holistic methodology for M&R optimization, as even the best solution algorithm will be of little use with a poorly developed methodology. Nevertheless, it is important that the formulated problem is tractable and that high-quality solutions can be found within a reasonable time. The results presented in the following section are obtained by making use of a unique algorithm which is tailored to fit the problem at hand. The algorithm is currently being improved, and it will be described in a separate paper due to its complexity. However, some details are already discussed below.

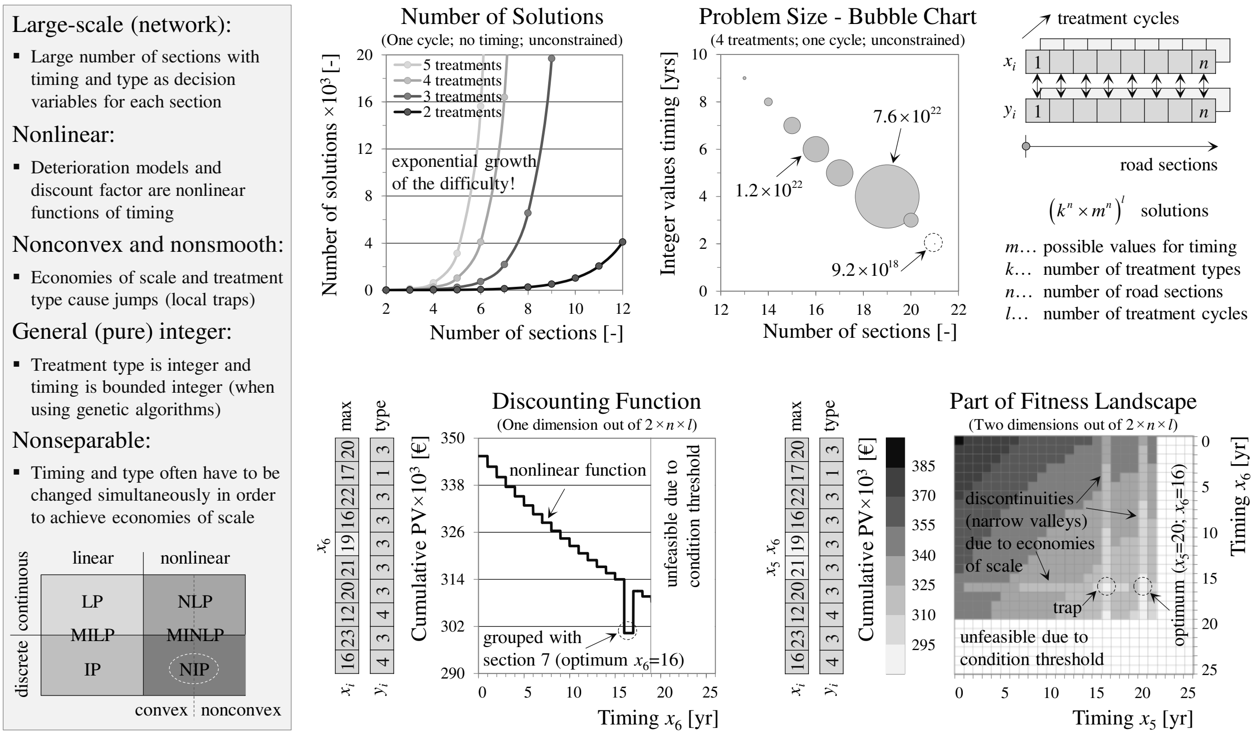
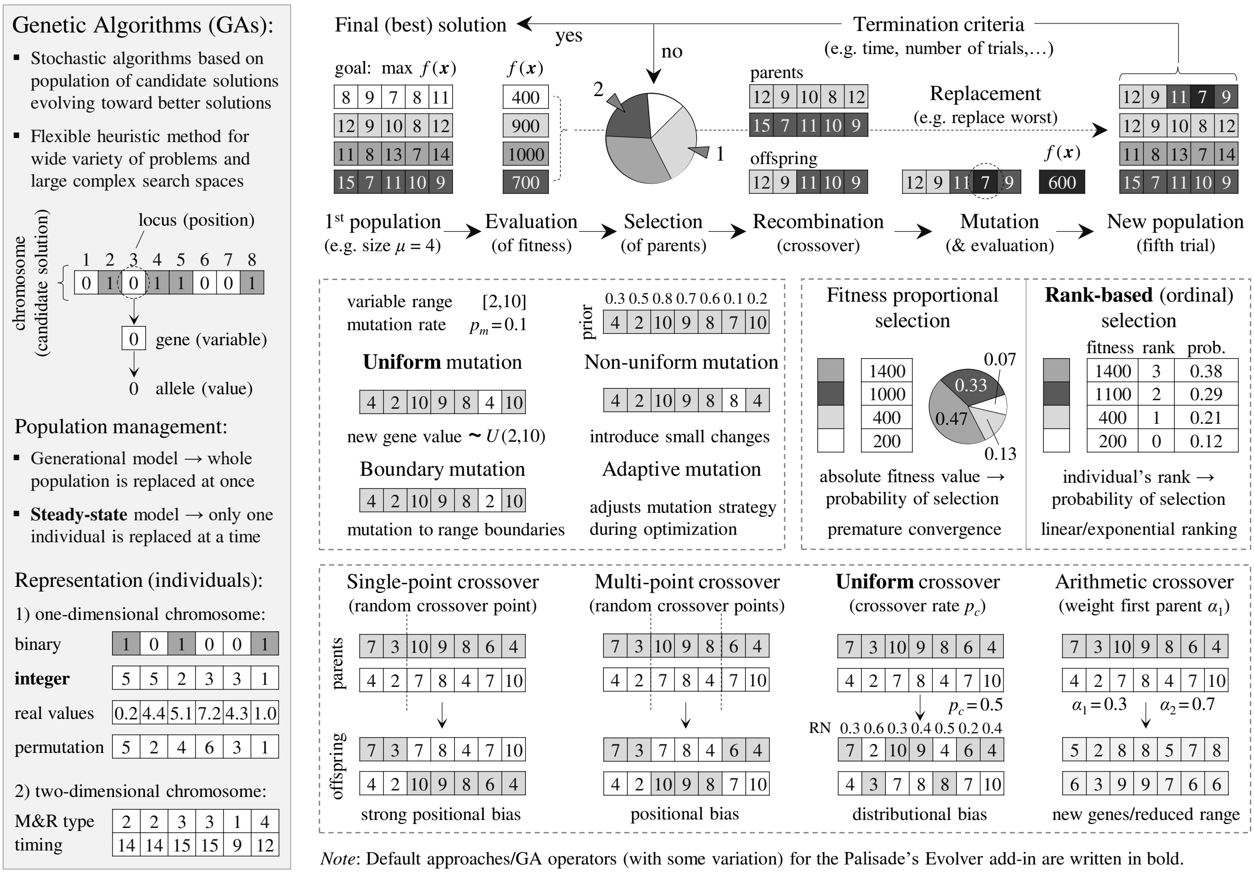


FIGURE 4.11 Size, characteristics and complexity of the formulated optimization problem. Simple example with 10 sections and evaluation of the objective function for just one decision variable (timing for section 6) and for two decision variables (sections 5 and 6).

The developed algorithm, currently labeled “expand, merge & solve” (EMS), is based on multiple iterations and resembles features of “divide and conquer” and greedy algorithms (Cormen et al. 2009). Despite the incorporation of a divide and conquer technique, the solution quality is not affected by the decomposition (overlapping is not needed). Thus, EMS is capable of handling an unlimited number of sections, with the running time being proportional to the number of sections. The algorithm code is written using the Visual Basic for Applications (VBA) programming language. For the “solve” part, the commercial software @Evolver 7.5.1, an add-in for Microsoft Excel, is used (Palisade 2015). Evolver’s genetic algorithm is integrated in the EMS algorithm (VBA) using Evolver’s Excel Developer Kit (XDK).

Genetic algorithms (GAs) constitute a metaheuristic search method which is based on the principles of “natural selection” and “survival of the fittest” (Goldberg 1989, Eiben and Smith 2015). GAs are applied in many research areas to solve difficult optimization problems with complex search spaces. A general overview of GAs is provided in Figure 4.12. The idea behind GAs is to initially generate and then iteratively enhance a population of solutions with the help of variation (crossover, mutation) and selection (parents, survivor) operators. Eventually, a near-optimal solution will be obtained by examining a relatively small number of possible solutions. Evolver employs an integer representation of solutions, a steady-state population management model, rank-based selection and multiple crossover and mutation operators (see Figure 4.12). Population size, crossover and mutation rate, as well as stopping time for the algorithm are estimated on the basis of comparative studies. The analysis showed, however, that different combinations of parameters may achieve similar results. Thus, as optimal parameters also depend on the data sample, the importance of parameter tuning should not be overestimated.



Note: Default approaches/GA operators (with some variation) for the Palisade’s Evolver add-in are written in bold.

FIGURE 4.12 General overview of genetic algorithms: simplified example of steady-state GA, as well as most common operators for mutation, parent selection and crossover.

4.5 APPLICATION AND RESULTS

This section presents the results derived from the application of the proposed methodology (Section 4.4) to the case study (Section 4.3.2), with an emphasis on graphical representation which is more suitable for the discussion of the findings on optimal work zones.

4.5.1 Section-level results

The single-section optimization is conducted for all 1000 road sections. A (real) discount rate of 4% is used as a standard case (Stone 2013). Figure 4.13(a,c,e) shows the remaining service life with regard to the three distress types before (basic) and after the optimization for a randomly chosen section (section 466). The basic strategy results from selecting the cheapest feasible treatment for the given failure cause. For example, if the failure cause is rutting, the basic strategy will be to select a surface course replacement. The expenditure stream diagrams (Figure 4.13(b,d)) shows that—as a result from life cycle optimization—three treatments (instead of four) are necessary in the first 50 years, decreasing the associated present value by 9%, as shown in Figure 4.13(f). Summed up for 1000 sections, the total reduction of discounted agency costs amounts to 2% (3% undiscounted). Many PMSs, however, do not jointly optimize maintenance, rehabilitation and reconstruction activities and do not consider different temporal treatment sequences. Compared to such strategies, like, for example, “treatment 3 only” (surf & bin course) and “treatment 4 only” (full-depth asphalt), the LCC-optimized strategy results in a total reduction in present value of 9% and 18%, respectively (12% and 17% undiscounted).

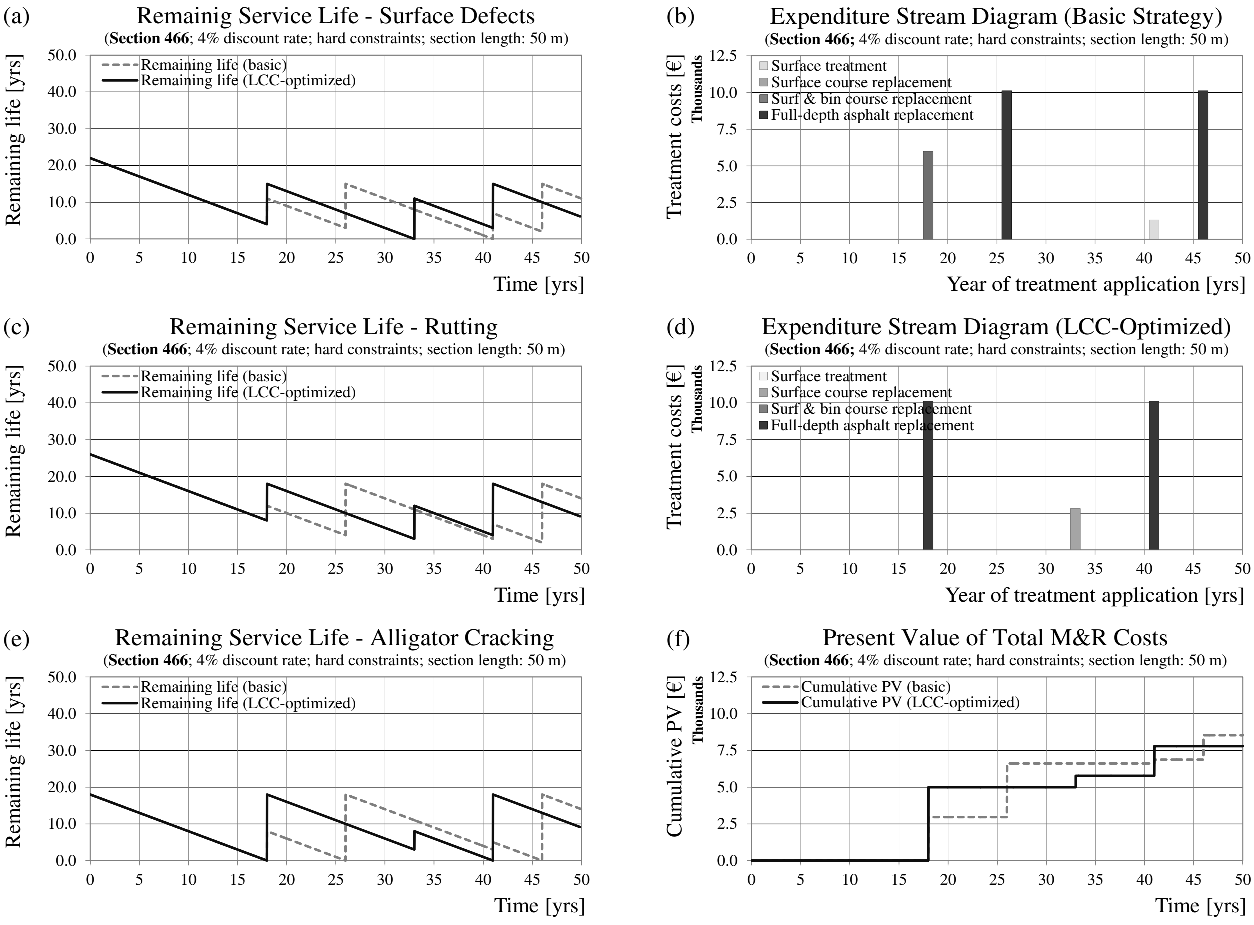


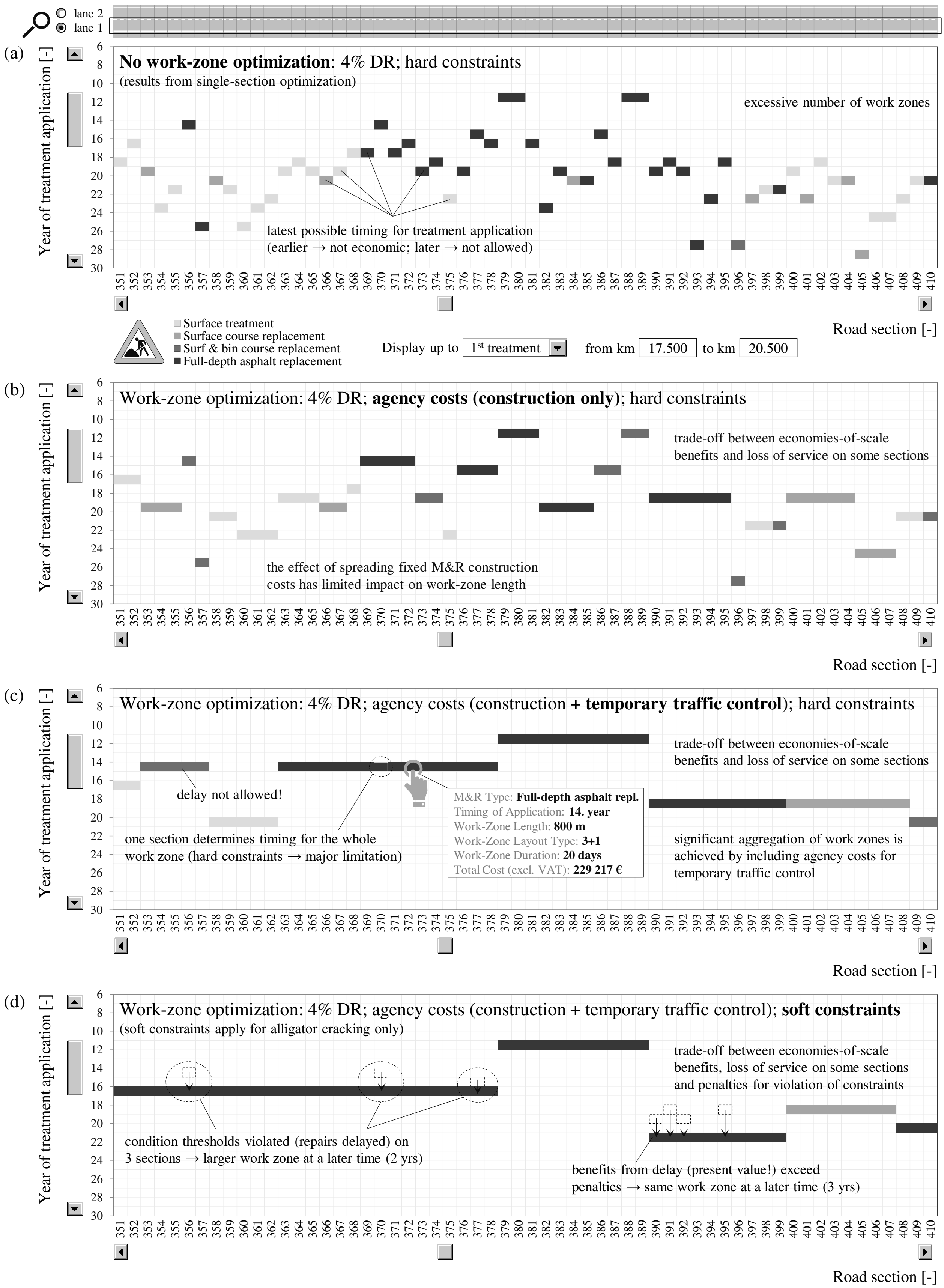
FIGURE 4.13 Results of the optimization for a single road section (randomly chosen) with remaining life (a, c, e), expenditure stream (b, d) and present value (f) prior (basic) and after the optimization.

4.5.2 Work-zone optimization

The results from the single-section optimization for 60 contiguous road sections (chosen at random) are shown in Figure 4.14(a). For simplicity, only the first treatment in the planning period is shown, depicting the positive autocorrelation pattern of service lives. For each section treatment timing is the latest possible. However, such an M&R strategy is not practical due to the excessive number of work zones. For the formation of work zones, scale-economies cost functions must be employed. Minimizing the total discounted costs for all sections based on M&R treatment costs generates the work zones shown in Figure 4.14(b). Significantly larger work zones are obtained from the optimization when traffic control costs are taken into account according to Figure 4.14(c). For each work zone, treatment type, timing, work-zone length, layout, duration and costs are computed, together with more a detailed cost calculation and workflow schedule. Despite the fact that, at this point, some work zones already exhibit practical lengths, there is still potential for improvement. For example, a work zone consisting of sections 351-362 may be economically beneficial. The consideration of such an alternative in the optimization process is not possible due to the condition thresholds. Further investigations show that only a few sections limit the timing of the obtained work zones. For example, the threshold on section 370 is solely responsible for determining the timing (14th year) of the corresponding work zone. Given the subjective nature of condition thresholds, this is a major limitation in any optimization method.

A possible solution to this problem is to model condition thresholds as soft constraints instead of hard constraints (see Section 4.4.3), with the results shown in Figure 4.14(d). This approach leads to larger work zones, a better utilization of service life and longer treatment intervals. In this case, only the threshold for alligator cracking is modeled as a soft constraint. On a general note, in some cases the monetary difference between two separate work zones and one joint work zone is very small (a few euros), so it might be practical to include a small cost tolerance in favor of larger work zones, also benefitting road users.

The soft-constraint method requires a penalty function, with penalty costs being added to the objective function (agency costs), taking the amount of threshold violation as a basis. Performance models are needed for estimating threshold violation. Although only alligator cracking is modeled as a soft constraint in the case study, power performance functions ($\beta_0 + \beta_1 \times SL^{\beta_2} = y_f$) are defined for all distress types. The power parameter implies slightly progressive deterioration ($\beta_2 = 1.20$) for surface defects, degressive deterioration ($\beta_2 = 0.80$) for rutting and progressive development ($\beta_2 = 2.90$) for alligator cracking. The values are taken from previous research (Donev and Hoffmann 2017a), having no impact on the results presented here (except in the case of soft constraints for alligator cracking). Figure 4.15(a,c,e) shows the condition distribution at the network level (1000 sections) for the three distress types without optimization of the work zone (see Figure 4.14(a)). Figure 4.15(b,d,f) shows the corresponding distributions after the optimization of work zones (the case of Figure 4.14(c)). Since both cases employ hard constraints, there is no road section in condition class 5 (very poor). Moreover, the network is in a better average condition as a result of the work-zone optimization and brought-forward treatments. This secondary effect is more noticeable for the primary and secondary failure cause (alligator cracking and surface defects). At the section level, brought-forward M&R increase the probability of surviving until the treatment timing.



Display up to 1st treatment from km 17.500 to km 20.500

M&R Type: **Full-depth asphalt repl.**
Timing of Application: **14. year**
Work-Zone Length: **800 m**
Work-Zone Layout Type: **3+1**
Work-Zone Duration: **20 days**
Total Cost (excl. VAT): **229 217 €**

FIGURE 4.14 Results of the work-zone optimization for a segment of 60 road sections (randomly chosen) and 4% discount rate displaying only the first treatment of the life cycle for the cases of no optimization (a), minimum M&R costs (b), minimum total agency costs (c) and minimum total agency costs with soft constraints for alligator cracking (d).

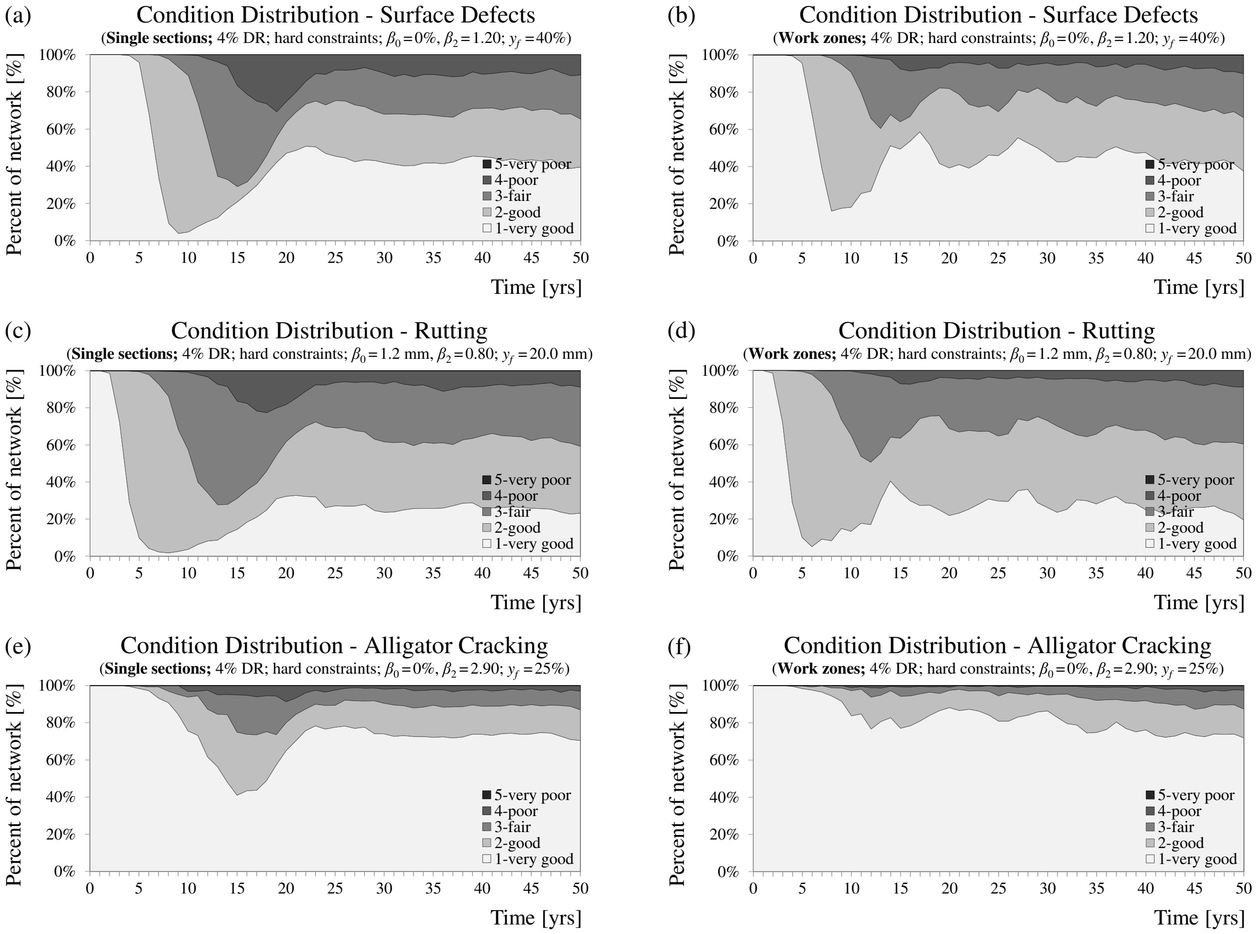
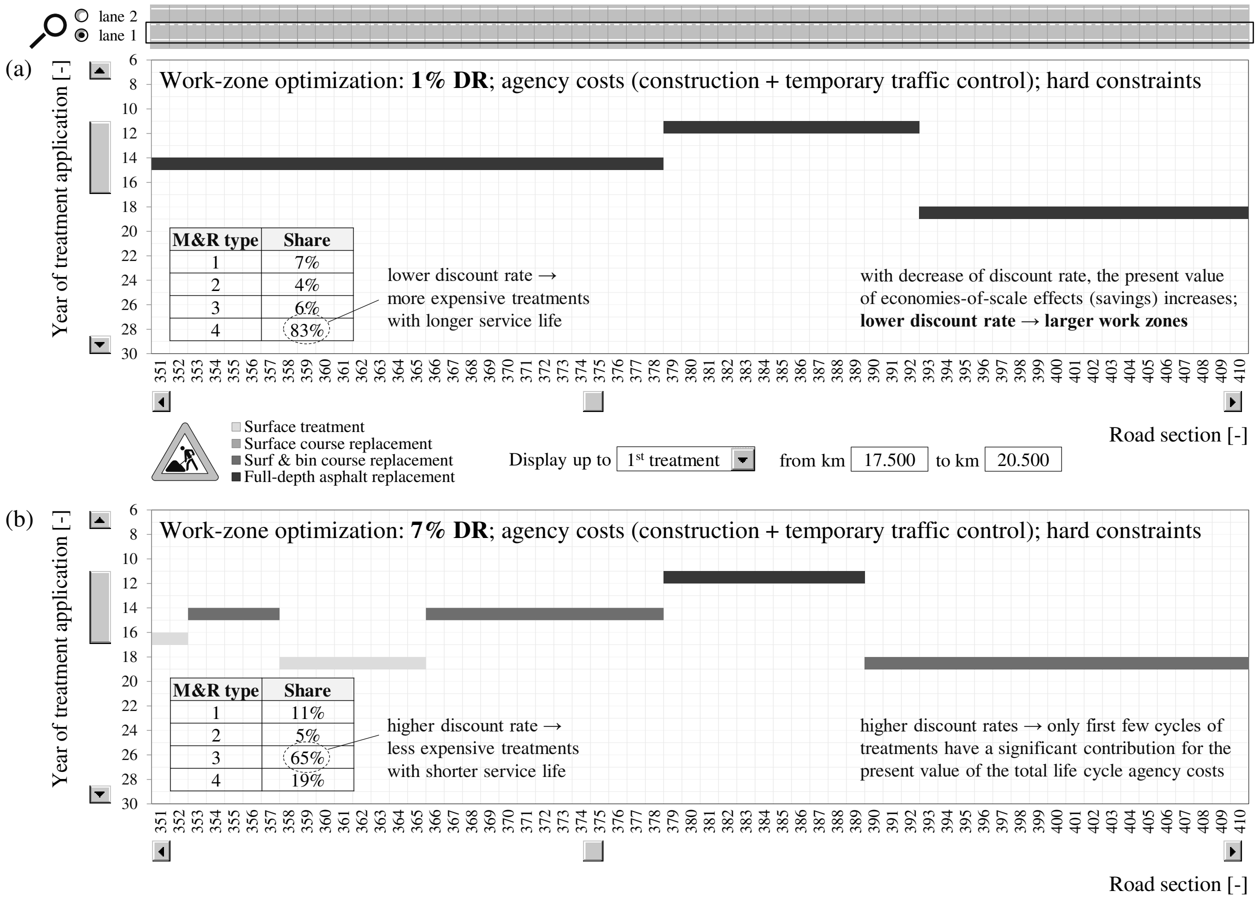


FIGURE 4.15 Distress-specific condition distributions for all sections in the cases of no-work zone optimization (a, c, e) and optimal work zones (b, d, f) using power functions for intermediate conditions.

4.5.3 Sensitivity to different discount rates

In the following, the sensitivity of the results should be investigated on the basis of discount rates of 1% and 7%, as presented in Figure 4.16(a,b). The results show that lower discount rates lead to greater work-zone lengths. As already mentioned, the work-zone optimization is based on the trade-off between loss of service life (increase of present value) and economies-of-scale benefits. For the discount rate of 7%, the increase of present value is high, but only for the first treatment, as subsequent treatments show almost no present-value contribution. In the case of 1% discount rate, the increase of present value is not as high, but a greater number or follow-up treatments exhibit a significant increase. In total, the increase of present value for both cases (1% and 7%) is similar. The difference results from discounting the benefit of scale economies. This benefit has much higher present value with lower rates (1%) leading to a greater length of work zones.

In addition to the length of the work zone, the applied discount rate shows a major impact on the type of the treatment. For example, a higher discount rate leads to preference of partial-depth replacement (treatment 3) over full-depth replacement (treatment 4). High discount rates attach more importance to short-term decisions favoring less costly treatments, despite the shorter service lives. On the other hand, given that future pavement performance and associated treatment costs are largely uncertain, low discount rates (implying a high level of certainty for the type and timing of future treatments) are not very appropriate.



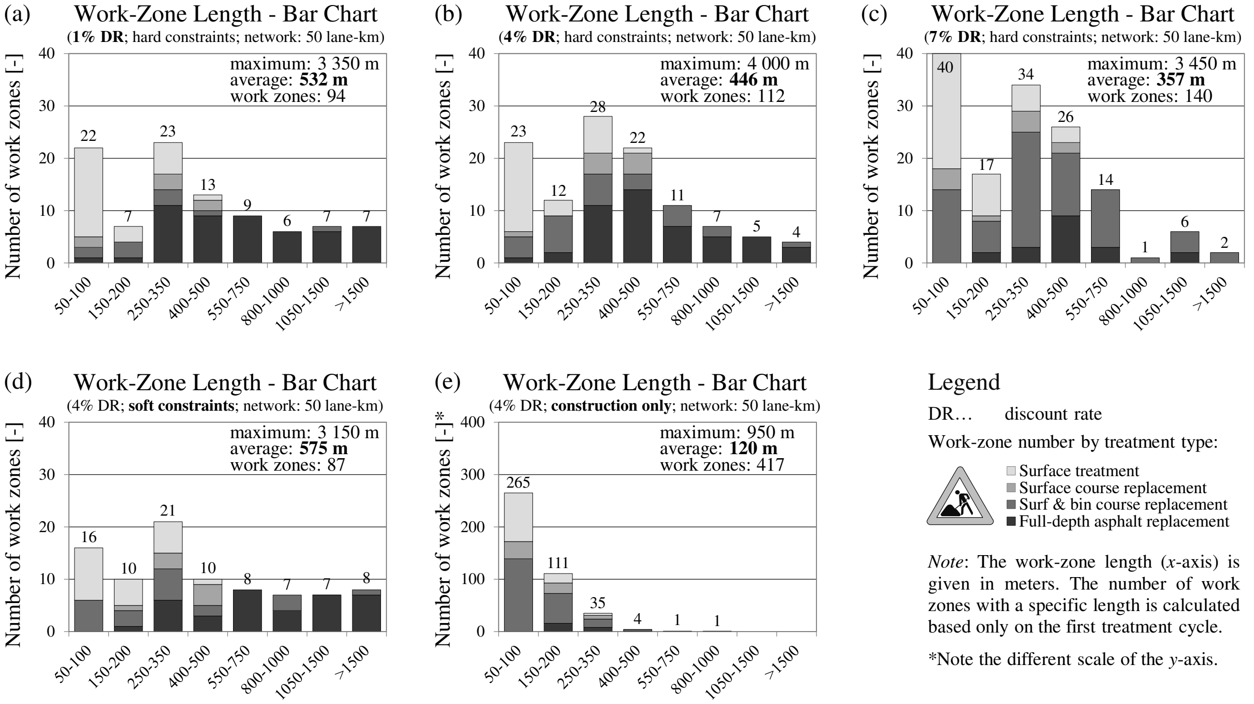
lower discount rate → more expensive treatments with longer service life

with decrease of discount rate, the present value of economies-of-scale effects (savings) increases; **lower discount rate → larger work zones**

higher discount rate → less expensive treatments with shorter service life

higher discount rates → only first few cycles of treatments have a significant contribution for the present value of the total life cycle agency costs

FIGURE 4.16 Effect of lower (a) and higher (b) than 4% discount rates on optimal work zones, treatment types and timings for the case of minimum total agency costs and hard constraints.



Note: The work-zone length (*x*-axis) is given in meters. The number of work zones with a specific length is calculated based only on the first treatment cycle.

*Note the different scale of the *y*-axis.

FIGURE 4.17 Number of work zones in different length classes by type of treatment for cases of different discount rates (a, b, c), soft constraints (d) and minimum M&R costs (e).

The number of work zones by length and treatment type for all 1000 road sections is presented in Figure 4.17. Neglecting traffic control costs results in average work zones of 120 m, as shown in Figure 4.17(e). The average length of a work zone increases with the decrease of discount rates. The greatest average work-zone length (575 m) is achieved in the case of soft constraints (Figure 4.17(d)). Short work zones (< 200 m) mainly include surface treatments which are applied as short-term work zones and may not be representative for the average work zone. In some cases, work zones of up to 4000 m are proposed.

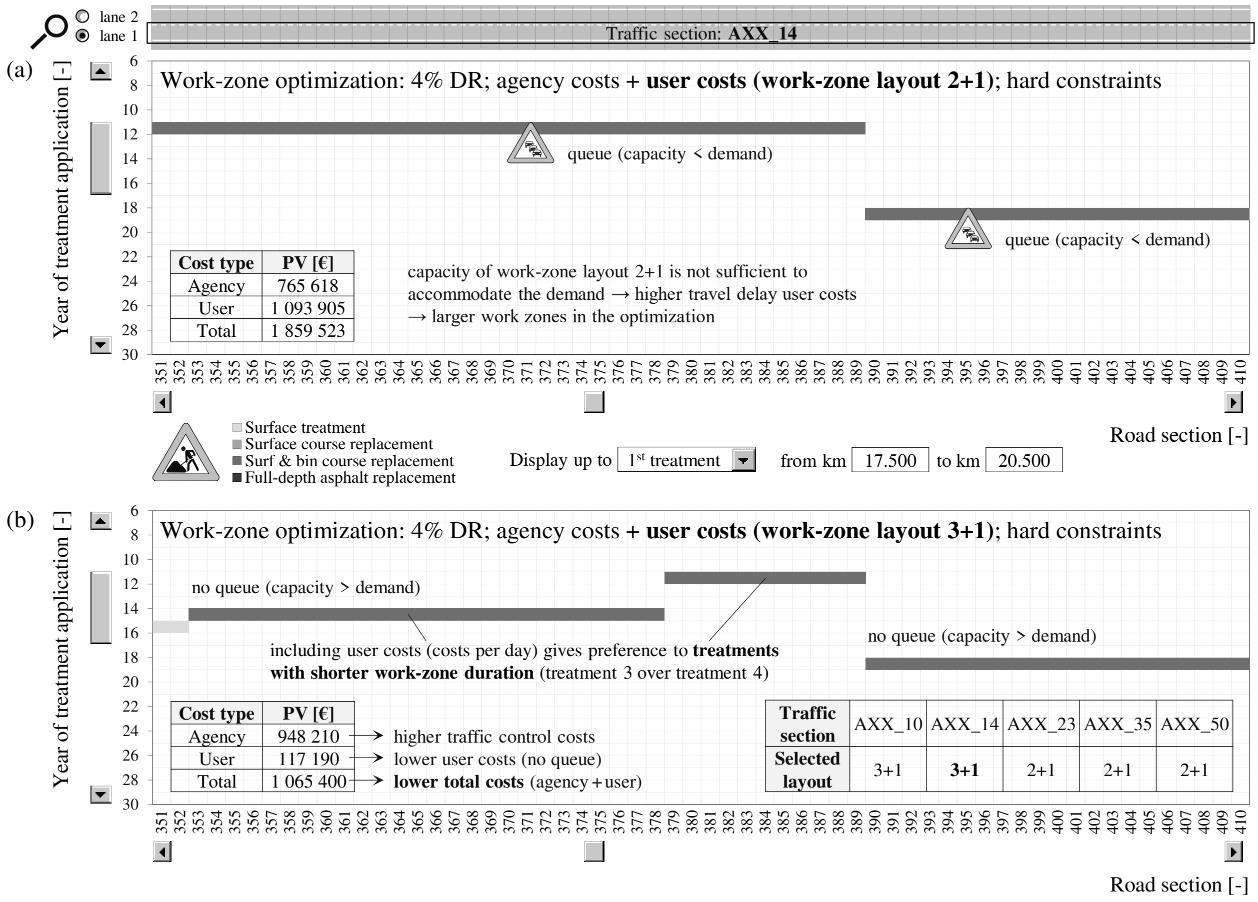
The average length of a work zone may seem too short as compared to typical lengths known from practice. On the one hand, there is no proof that the work zones in practice are optimal from an economic point of view. On the other hand, the optimization does not yet account for a few factors that will significantly increase the obtained work-zone lengths. First, an aggregated work zone may include multiple M&R activities (see Figure 4.14(c), sections 390-408). These sections may share the fixed costs for traffic control and part of the fixed costs for construction. Two work zones with a small gap in between (i.e. sections where no construction activities take place, but the same traffic control applies) can also be aggregated in one joint work zone. For this purpose, some additional constraints reflecting the approach in practice have to be defined (e.g. technically reasonable activity combinations, maximum gap length, maximal number of different activities, etc.). These situations are currently not considered, but the model can easily be adjusted to account for them.

Second, the employed cost models currently consider fixed costs related to construction and temporary traffic control. Other (in-house) activities with a high proportion of fixed costs are not included yet (e.g. costs for supervision and planning, road safety audits, carrying out of call-for-tender procedures, quality assurance and acceptance testing). Taking into account these costs will further reduce the number of work zones and increase their length. On the other hand, considering the second roadway lane will lead to a shorter average work-zone length (due to economies-of-scale in transverse direction). With less than 5% of the truck traffic on the interior lane in Austria, however, it is to be expected that the interior lane will require substantially less frequent structural repairs.

4.5.4 Selection of a work-zone layout

In the following, additional road user costs which are caused by work zones should be considered. The corresponding objective is to minimize the sum of agency (including traffic control) and user costs. The optimization is conducted for two work-zone layouts: 2+1 and 3+1, as shown in Figure 4.18(a,b). The results are proof that the taking into account of user costs leads, in general, to larger work zones. For the given traffic section (AXX_14), the available capacity of layout 2+1 is not sufficient to accommodate the traffic demand, resulting in a queue and very high user costs. In contrast, the configuration 3+1 does not cause congestion and leads to user costs which are approximately nine times lower (60 sections). However, maintaining the number of open lanes by using median crossover leads to higher agency costs (20%). In the case of 2+1, the forced-flow conditions lead to much higher user costs, resulting in a higher level of aggregation (only two work zones) in comparison to 3+1.

Besides greater work-zone lengths, considering user costs leads to different optimal treatment types (see Figure 4.14(c) for comparison). Work-zone user costs are computed as the product of daily user costs and project duration. Thus, it is not surprising that treatments with a shorter duration (treatment 3) are preferred over treatments with a longer duration (treatment 4). User costs (non-cash expense) are subjective and primarily represent short-term impacts. Therefore, they should not be mixed with agency costs which are of a more objective nature. Instead, it is recommended to utilize user costs in making short-term decisions (e.g. selection of work-zone configurations). In the present case, layout 3+1 will be selected for minimizing total costs. The layouts for the other traffic segments are provided in the table of Figure 4.18(b). Subsequently, work-zone optimization can be conducted with the determined layouts based solely on the agency costs. By doing so, the interests of road users will be taken into account (even if this adds up to more expensive traffic control), but other decisions with long-term impacts like treatment type and work-zone length will be more stable as they are based on agency costs.



(a)

Cost type	PV [€]
Agency	765 618
User	1 093 905
Total	1 859 523

capacity of work-zone layout 2+1 is not sufficient to accommodate the demand → higher travel delay user costs → larger work zones in the optimization

(b)

including user costs (costs per day) gives preference to **treatments with shorter work-zone duration** (treatment 3 over treatment 4)

Cost type	PV [€]	
Agency	948 210	→ higher traffic control costs
User	117 190	→ lower user costs (no queue)
Total	1 065 400	→ **lower total costs** (agency + user)

Traffic section	AXX_10	AXX_14	AXX_23	AXX_35	AXX_50
Selected layout	3+1	**3+1**	2+1	2+1	2+1

FIGURE 4.18 Effect of considering user costs together with agency costs in the work-zone optimization for layout 2+1 (a) and layout 3+1 (b) and a given traffic section (AXX_14). Selection of a work-zone layout based on total (agency plus user) costs.

4.5.5 Special topic: pavement design optimization

The main idea of long-life (perpetual) pavements is to design thicker pavements in order to avoid premature structural failure (Timm and Newcomb 2006). Such pavements mainly exhibit surface-related distress and require minor (light) rehabilitation which results in lower life cycle costs despite higher initial construction costs. In the context of the case study (with alligator cracking as a dominant failure cause), the effects of extending structural life by thicker design is investigated. For this purpose, the expectation of the service-life distribution for alligator cracking is extended from 19.4 to 24.2 years (by 4.8 years). The service life distributions for surface defects and rutting remain unchanged, as they are not affected by thicker design.

Figure 4.19(a) shows the optimal work zones minimizing agency life cycle costs (including traffic control) for the standard design case (19.4 years). This figure supplements Figure 4.14(c) by displaying the second and the third treatment, not previously shown, as the focus of the discussion was on work zones. It is also shown that the employed (EMS) algorithm evaluates work zones consisting of, for example, the second treatment (on one section) and the third treatment (on the neighboring section). Figure 4.19(c) includes the corresponding work zones with enhanced design (24.2 years). Extending the life for alligator cracking by 4.8 years on average leads to a new dominant (first) failure cause (surface defects with 55%). The first visible consequences based on the segment of 60 sections are an increased share of minor treatments (treatments 1 and 2), as well as shorter work zones and treatment intervals.

For the entire network (50 lane-km), the annual M&R expenditures and length-dependent treatment type distributions (treatment demand) for both cases of standard and enhanced design are shown in Figure 4.19(b,d). The annual costs are provided in 2016 constant-value euros (inflation free) and constitute net costs excluding value-added tax (VAT). For simplicity, only the first 30 years are shown, although the planning horizon is 100 years. The comparison of average annual M&R costs (years 6-30) shows that extending design life leads to a reduction in costs from €0.65 million to €0.57 million per year (12% reduction), despite shorter work zones and treatment intervals. The treatment-demand diagrams show that the share of minor treatments is significantly increased in the case of enhanced design. Furthermore, the treatment-demand peak has shifted from years 13-14 to years 17-18, but the total repaired network area is almost the same in both cases.

The difference in annual expenditures and present value is highlighted in Figure 4.20(a,b). The strategy with extended structural life leads to an M&R cost reduction of €2.08 million over 30 years. The majority of these M&R savings are incurred in the first 15 years, with the enhanced design being more expensive in the second half of the shown period. The total reduction in present value amounts to €1.73 million (4% discount rate). In order to determine whether the enhanced design reduces total life cycle costs or not, the initial construction costs have to be considered. These costs are not taken into account here due to the absence of a reliable empirical model which links pavement thickness and service life. However, for €1.73 million, approximately 7 cm of asphalt base course could be added to the pavement structure (for all 1000 sections). The singular aim of this example is to show that structural design life has a major impact on life-cycle costs, and pavement thickness should therefore be a subject of optimization. This conclusion is in contrast to the fixed design life of 20 years which is embedded in the standard procedures of many countries, but is in accordance with similar findings in scientific literature (McDonald and Madanat 2012, Lee and Madanat 2014).

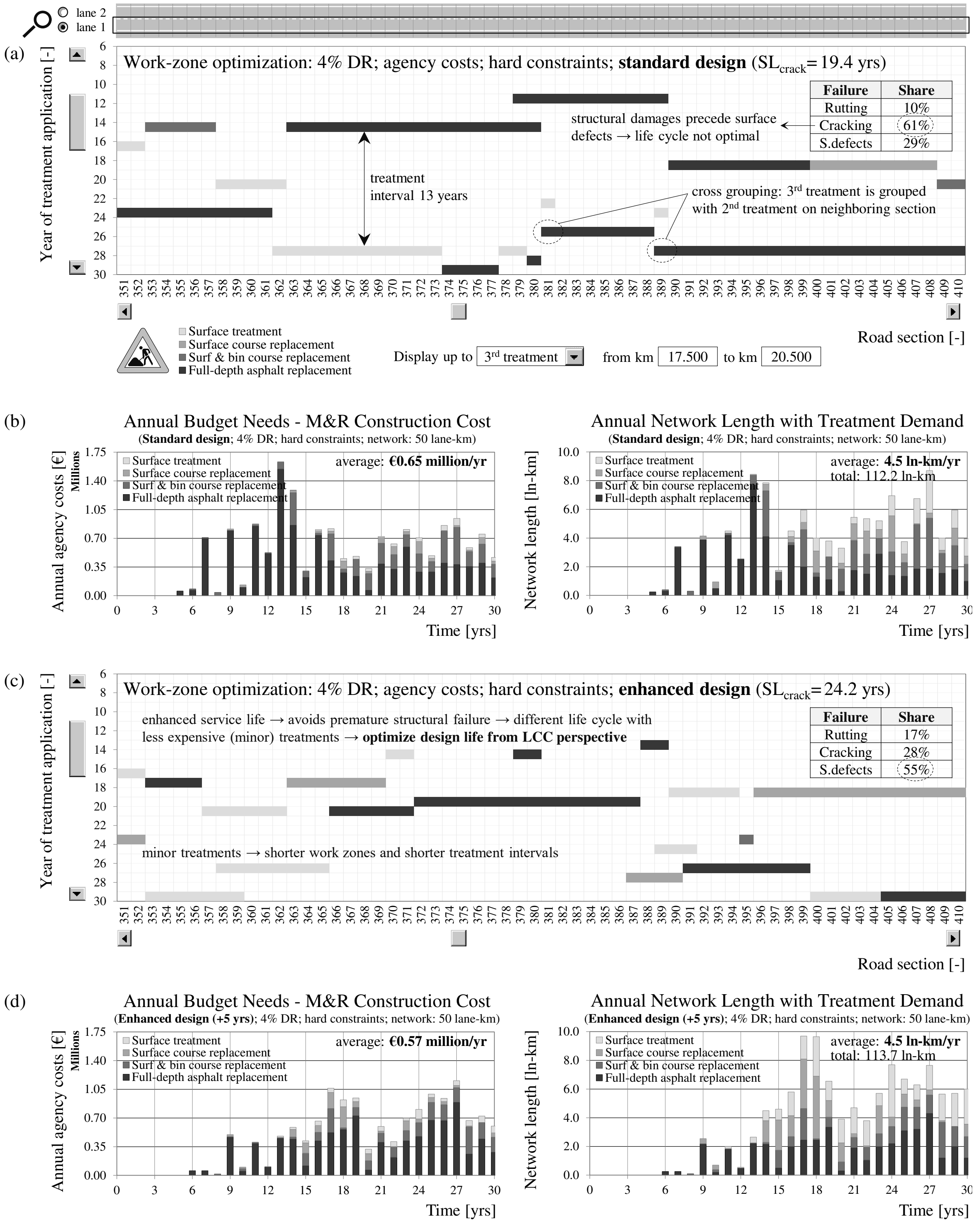


FIGURE 4.19 Comparison of work zones, annual budget and distribution of treatment types for the standard case of minimum total agency costs, hard constraints and 4% discount rate (a, b) and for the case of enhanced design life for alligator cracking under the same conditions (c, d) with display of the first three treatment cycles.

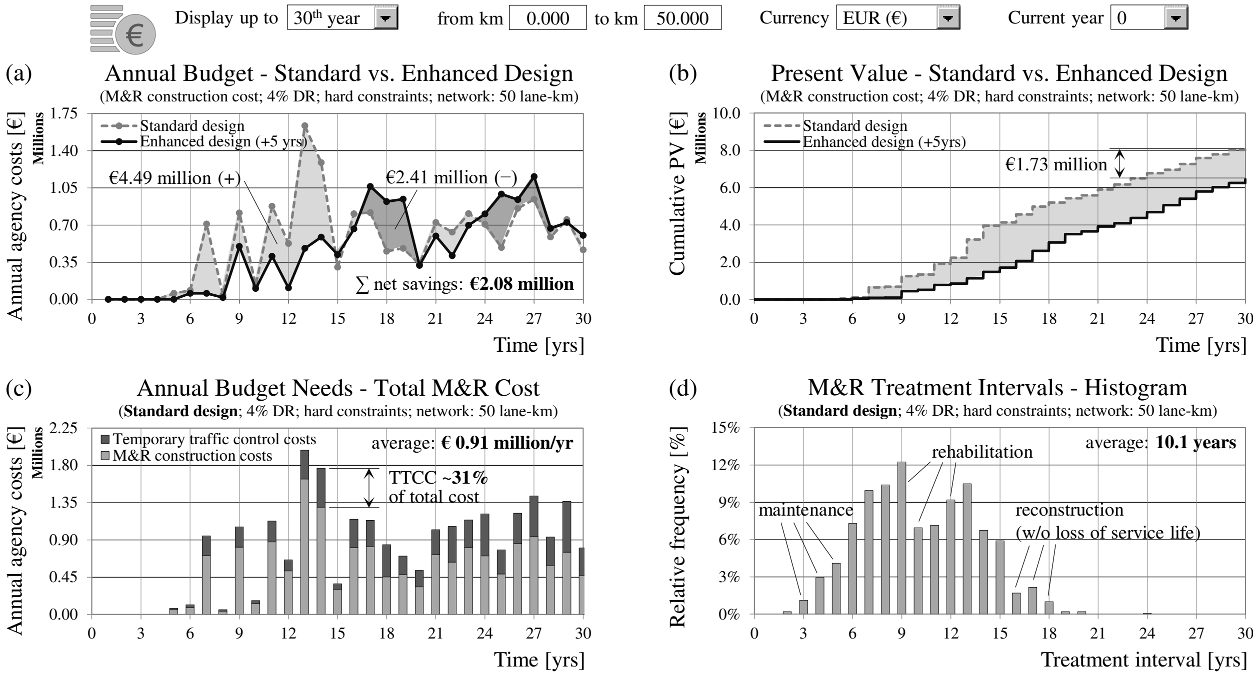


FIGURE 4.20 Potential savings as a result of avoiding premature structural failures by extending design life for alligator cracking in terms of annual budget (a) and present value (b). Total agency costs including TTCC (c) with histogram of treatment intervals (d) in the standard design case.

The total agency costs are shown in Figure 4.20(c), with the costs for temporary traffic control amounting to 31% of total costs. This percentage can be reduced with the ongoing development of the approach and the potential increase of the length of work zones (see Section 4.5.3). The probability distribution of treatment intervals shows a wide dispersion which is attributed to the different service lives of the employed treatments (see Figure 4.20(d)), with the average interval being 10.1 years.

If budget constraints are not binding, the solution derived by unconstrained work-zone optimization is final. The network-level M&R optimization under annual agency budget constraints will be the subject of a forthcoming paper. The solution of the unconstrained problem, however, is the best possible solution for the agency. Any other M&R program resulting from budgetary restrictions is suboptimal, assuming that condition thresholds are not violated.

4.6 CONCLUSIONS

The primary goal of any PMS is to determine optimal M&R strategies for each section of a road network, subject to an annual budget. State-of-the-art PMS aggregate short survey sections into longer homogeneous sections, prior to any optimization. Due to averaging effects, the estimated service lives are not accurate, condition thresholds are systematically violated and a determination of optimal work zones is no longer possible. Instead of forming long homogeneous sections, the proposed approach is based on short survey sections, allowing for more accurate condition predictions for each failure type and stricter threshold control. Survey sections are combined into longer work zones based on economies-of-scale effects, optimizing agency M&R and temporary traffic control costs. In addition, work-zone related user and environmental costs are taken into account in order to select appropriate work-zone layouts.

Common approaches are based on a single condition index which represents pavement condition and on predefined M&R strategies (temporal sequence of treatment types). In contrast, the suggested method considers every possible sequence of treatments, depending on the predicted distress types or a combination thereof. The proposed approach is unique in its capabilities to simultaneously account for multiple treatments, multiple distress types, stochastic treatment lives, scale economies, traffic control, work-zone layouts, capacity and road user impacts. The developed cost and performance models are an integral part of the optimization approach. All input information is estimated on the basis of real data or an extensive literature comparison and calibrated for flexible pavements on freeways in Austria. The formulated optimization problem exhibits exponential complexity and is solved by a unique algorithm which is specifically designed for this class of problems.

The proposed methodology is successfully applied to a case study with a realistic road network consisting of 1000 consecutive sections. The results show the superiority of the proposed approach when compared to a predefined sequence of treatments at the single section level. According to the results, traffic control costs are the primary driver for the formation of work zones. The use of soft constraints increases the flexibility of the approach, reducing the dependence of the results on condition thresholds and outliers. Furthermore, the conducted sensitivity analysis shows that discount rates have a huge impact on the length of work zones and the type of the treatment. Higher discount rates attach more weight to short-term decisions and favor less costly treatments, despite shorter service lives. The consideration of additional user costs due to work zones leads to larger work zones and treatments with a shorter duration. In the light of the European Commission's strategy towards internalization of transport accident and environmental costs, the model is capable of addressing these challenges, providing the means for an estimation of future infrastructure charges.

Furthermore, the designed structural life has a huge impact on the extent of subsequent treatments and life cycle costs. Extended-life pavements have the potential to reduce total agency costs and road user impacts with less costly surface treatments, shorter work zone durations and a reduction of lane closures, especially in the case of low discount rates. Future developments include the extension of the model to simultaneous modeling of multiple lanes, estimation of optimal inspection intervals, application to rigid pavements and cross-asset optimization. A follow-up paper that is currently under preparation will compare the developed holistic model framework and methodology to existing PMS approaches, with and without budget constraints, based on the presented case study.

REFERENCES

AASHTO, 2012. *Pavement management guide*. 2nd ed. Washington, DC: American Association of State Highway and Transportation Officials.

Anastasopoulos, P.C., Labi, S., Bhargava, A. and Mannering, F.L., 2012. Empirical assessment of the likelihood and duration of highway project time delays. *Journal of Construction Engineering and Management*, 138 (3), 390-398.

Anastasopoulos, P.C. and Mannering, F.L., 2015. Analysis of pavement overlay and replacement performance using random parameters hazard-based duration models. *Journal of Infrastructure Systems*, 21 (1), 04014024.

Bennett, C.R. and Greenwood, I.D., 2001. *Modeling road user and environmental effects in HDM-4*. Highway Development and Management Series, vol.7. Paris: World Road Association (PIARC).

Birbaum, J., 2016. *Entwicklung eines Verfahrens zur Ermittlung der ebenszykluskosten von Straßen in Asphaltbauweise* [Development of a procedure to determine the life-cycle-costs of asphalt roads]. Dissertation (PhD). Universität Siegen (in German).

BMVIT, 2013. *Transport in figures: Austria - edition 2011*. Vienna: Federal ministry for transport, innovation and technology (BMVIT).

Brozek, B., Litzka, J. and Weninger-Vycudil, A., 2009. *Entwicklung eines Nutzerkostenmoduls im österreichischen PMS* [Development of a user cost module in the Austrian pavement management system]. Wien: BMVIT, Straßenforschung Heft 585 (in German).

Chen, C.H. and Schonfeld, P., 2007. Pavement resurfacing intervals in work zone optimization. *86th Annual Meeting of the Transportation Research Board*, 21-25 January, Washington, DC.

Chien, S. and Schonfeld, P., 2001. Optimal work zone lengths for four-lane highways. *Journal of Transportation Engineering*, 127 (2), 124-131.

Chu, J.C. and Chen, Y.-J., 2012. Optimal threshold-based network-level transportation infrastructure life-cycle management with heterogeneous maintenance actions. *Transportation Research Part B*, 46 (9), 1123-1143.

Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C., 2009. *Introduction to algorithms*. 3rd ed. Cambridge, MA: MIT Press.

Cuelho, E., Mokwa, R. and Akin, M., 2006. *Preventive maintenance treatments of flexible pavements: A synthesis of highway practice*. Report No. FHWA/MT-06-009/8117-26. Helena, MT: Montana Department of Transportation.

Donev, V. and Hoffmann, M., 2017a. Condition prediction and estimation of service life in the presence of data censoring and dependent competing risks. *International Journal of Pavement Engineering*. doi: 10.1080/10298436.2017.1293264.

Donev, V. and Hoffmann, M., 2017b. Network-level optimization of pavement M&R activities, timing and work-zone length for multiple distress types. *World Conference on Pavement and Asset Management*, 12-16 June, Milan, Italy.

Eiben, A.E. and Smith, J.E., 2015. *Introduction to evolutionary computing*. 2nd ed. Berlin: Springer.

Ferreira, A., Picado-Santos, L. and Antunes, A., 2002. A segment-linked optimization model for deterministic pavement management systems. *International Journal of Pavement Engineering*, 3 (2), 95-105.

Fischer, L., 2009. *Wirtschaftlichkeitsuntersuchung von Behelfsverkehrsführungen an Autobahnarbeitsstellen unter Berücksichtigung der Querschnittsabmessungen* [Economic feasbility study of work-zone layout types and traffic control on freeways with regard to the dimensions of the cross section]. Bauhaus-Universität Weimar (in German).

Fromm et al., 2012. *ASFINAG: Erhaltungsstrategie* [ASFINAG: Asset management strategy]. Wien: ASFINAG Service GmbH und ASFINAG Alpenstraßen (unpublished, in German).

FSV, 2010. RVS 02.01.22: *Nutzen-Kosten-Untersuchungen im Verkehrswesen* [Transportation cost and benefit analysis]. Wien: Forschungsgesellschaft Straße-Schiene-Verkehr (in German).

FSV, 2012. RVS 05.05.42: *Straßen mit getrennten Richtungsfahrbahnen* [Roads with separate directional carriageway]. Wien: Forschungsgesellschaft Straße-Schiene-Verkehr (in German).

FSV, 2016. RVS 03.08.63: *Oberbaubemessung* [Pavement design]. Wien: Forschungsgesellschaft Straße-Schiene-Verkehr (in German).

Fwa, T.F., Chan, W.T. and Tan, C.Y., 1996. Genetic-algorithm programming of road maintenance and rehabilitation. *Journal of Transportation Engineering*, 122 (3), 246-253.

Goldberg, D.E., 1989. Genetic algorithms in search, optimization & machine learning. Reading, MA: Addison-Wesley.

Haas, R., Hudson, W.R., Zaniewski, J., 1994. *Modern pavement management*. Malabar, FL: Krieger.

Hajdin, R. and Lindenmann, H.-P., 2007. Algorithm for the Planning of Optimum Highway Work Zones. *Journal of Infrastructure Systems*, 13 (3), 202-214.

Hall, K.T., Correa, C.E., Carpenter, S.H. and Elliot, R.P., 2001. *Rehabilitation strategies for highway pavements*. NCHRP Project C1-38: Conctractor's final report. Washington, DC: Transportation Research Board.

Hamilton, J.D., 1994. *Time series analysis*. Princeton, NJ: Princeton University Press.

Hellmann, L., Rübensam, J. and Schwiethal, S., 2008. *Entwicklung von Verfahrenshilfen zur Berücksichtigung baustellenbedingter Nutzerkosten im Erhaltungsmanagement* [Development of process procedures in the maintenance management, considering construction site-specific user costs]. Bonn: Forschung Straßenbau und Straßenverkehrstechnik, Heft 988 (in German).

Hoffmann, M. and Donev, V., 2017. Optimization of pavement M&R activities and work zones with user costs for multiple distress types. *World Conference on Pavement and Asset Management*, 12-16 June, Milan, Italy.

Jiang, X. and Adeli, H., 2003. Freeway work zone traffic delay and cost optimization model. *Journal of Transportation Engineering*, 129 (3), 230-241.

Kuhn, K.D. and Madanat, S.M., 2005. Model uncertainty and management of a system of infrastructure facilities. *Transportation Research Part C*, 13 (5), 391-404.

Lea, J.D. and Harvey, J.T., 2015. Using spatial statistics to characterise pavement properties. *International Journal of Pavement Engineering*, 16 (3), 239-255.

Lee, J. and Madanat, S., 2014. Joint optimization of pavement design, resurfacing and maintenance strategies with history-dependent deterioration models. *Transportation Research Part B*, 68, 141-153.

Lee, J. and Madanat, S., 2015. A joint bottom-up solution methodology for system-level pavement rehabilitation and reconstruction. *Transportation Research Part B*, 78, 106-122.

Lindenmann, H.P., 2008. *Massnahmenplanung im Erhaltungsmanagement von Fahrbahnen; Synthesebericht* [Action planning within the management of road maintenance; Synthesis report]. UVEK-Schriftenreihe Nr. 1259. Zürich: Schweizerischer Verband der Strassen- und Verkehrsfachleute (in German).

Mallela, J. and Sadasivam, S., 2011. *Work zone road user costs: concepts and applications*. Report No. FHWA-HOP-12-005. Washington, DC: Federal Highway Administration.

McDonald, M. and Madanat, S., 2012. Life-cycle cost minimization and sensitivity analysis for mechanistic-empiric pavement design. *Journal of Transportation Engineering*, 138 (6), 706-713.

Medury, A. and Madanat, S., 2013. Incorporating network considerations into pavement management systems: A case for approximate dynamic programming. *Transportation Research Part C*, 33, 134–150.

Meissner, G., 2014. *Correlation risk modeling and management: An applied guide including the Basel III correlation framework- with interactive correlation models in Excel/VBA*. 1st ed. Singapore: Wiley.

Montgomery, D.C., Jennings, C.L. and Kulahci, M., 2015. *Introduction to time series analysis and forecasting*. 2nd ed. Hoboken, NJ: Wiley.

Nobakht, M., Sakhaeifar, M.S., Newcomb, D. and Underwood, S., 2016. Mechanistic-empirical methodology for the selection of cost-effective rehabilitation strategy for flexible pavements. *International Journal of Pavement Engineering*. doi: 10.1080/10298436.2016.1199878.

Ntziachristos, L. and Samaras, Z., 2016. Exhaust emissions from road transport. *In*: *EMEP/EEA air pollutant emission inventory guidebook 2016*. EEA Report No 21/2016. Luxembourg: European Environment Agency.

Ogwang, A., 2016. *Threshold-based resurfacing policies in pavement management to minimize costs and greenhouse gas emissions*. Dissertation (PhD). University of California, Berkeley

Palisade, 2015. *Evolver: the genetic algorithm solver for Microsoft Excel*. User's guide, version 7. Ithaca, NY: Palisade Corporation.

Rossetti, M.D., 2016. *Simulation modeling and Arena*. 2nd ed. Hoboken, NJ: Wiley.

Santos, J., Ferreira, A. and Flintsch, G., 2017. An adaptive hybrid genetic algorithm for pavement management. *International Journal of Pavement Engineering.* doi: 10.1080/10298436.2017.1293260.

Sathaye, N. and Madanat, S., 2012. A bottom-up optimal pavement resurfacing solution approach for large-scale networks. *Transportation Research Part B*, 46 (4), 520-528.

Stone, M.L.C., 2013. *Development of unit cost estimating models with respect to scale economies and material price volatility for use in probabilistic life cycle cost analysis*. Dissertation (PhD). University of Alabama.

Timm, D.H. and Newcomb, D.E., 2006. Perpetual pavement design for flexible pavements in the US. *International Journal of Pavement Engineering*, 7 (2), 111–119.

Walls III, J. and Smith, M.R., 1998. *Life-cycle cost analysis in pavement design: Interim technical bulletin*. Report No. FHWA-SA-98-079. Washington, DC: Federal Highway Administration.

Wießmayer, W., 2006. *B1 Wiener Straße Teststrecke: Abschlussbericht* [B1 Wiener Straße test track: final report]. Linz: Amt der O.ö. Landesregierung – Landesbaudirektion, Abteilung Straßenerhaltung und –betrieb (in German).

Williams, R.C., Hildreth, J.C. and Vorster, M.C., 2009. Highway construction data collection and treatment in preparation for statistical regression analysis. *Journal of Construction Engineering and Management*, 135 (12), 1299-1306.

Wirtgen, 2013. *Wirtgen cold milling manual: Technology and application*. Windhagen, Germany: Wirtgen GmbH.

Yang, N., 2010. *Optimization of highway work zone decisions considering short-term and long-term impacts*. Dissertation (PhD). University of Maryland.

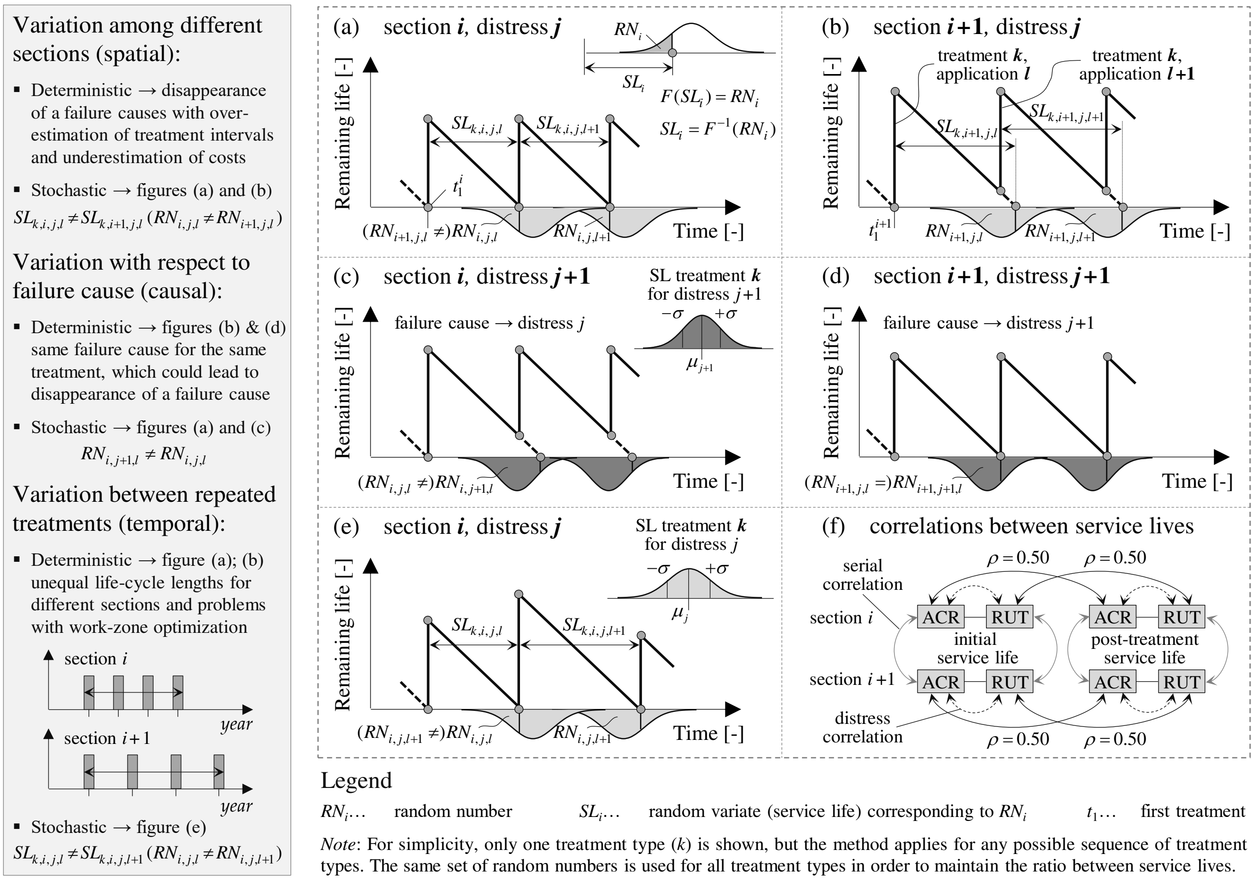
Yeo, H., Yoon, Y. and Madanat, S., 2013. Algorithms for bottom-up maintenance optimisation for heterogeneous infrastructure systems. *Structure and Infrastructure Engineering*, 9 (4), 317-328.

APPENDIX 4.A* DETAILS ON TREATMENT LIFE AND USER COSTS

4.A.1 Stochastic treatment life

Figure 4.A.1 aims to shed some light on the generation of treatment lives based on the normal distributions from Figure 4.4. Three different types of variability are possible, with the first being variation between different sections. The second type is variation of the treatment effect between distress types (i.e. variation of the failure cause). The third type of variability is between repeated applications of the same treatment on the same section.

In this work, the first two types of variability are considered which results in a semi-stochastic treatment life. The variability between applications is neglected, as the inclusion results in a disproportional impact on Excel's performance in comparison to the benefits. Although all evidence suggests that treatment life is stochastic, it is not realistic that the variability is completely random. Thus, the same serial and distress correlations as in Section 4.3.2 are incorporated in the simulated treatment lives (Figure 4.A.1(f)). Furthermore, it is assumed that pre-treatment and post-treatment service lives are correlated as well, with a correlation coefficient equal to 0.5 (conservative estimate). The generated data set comprises an array with 1000 rows (sections) and 12 columns (4 treatments 3 distress types).



Legend

RN_i… random number SL_i… random variate (service life) corresponding to RN_i t_1… first treatment

Note: For simplicity, only one treatment type (k) is shown, but the method applies for any possible sequence of treatment types. The same set of random numbers is used for all treatment types in order to maintain the ratio between service lives.

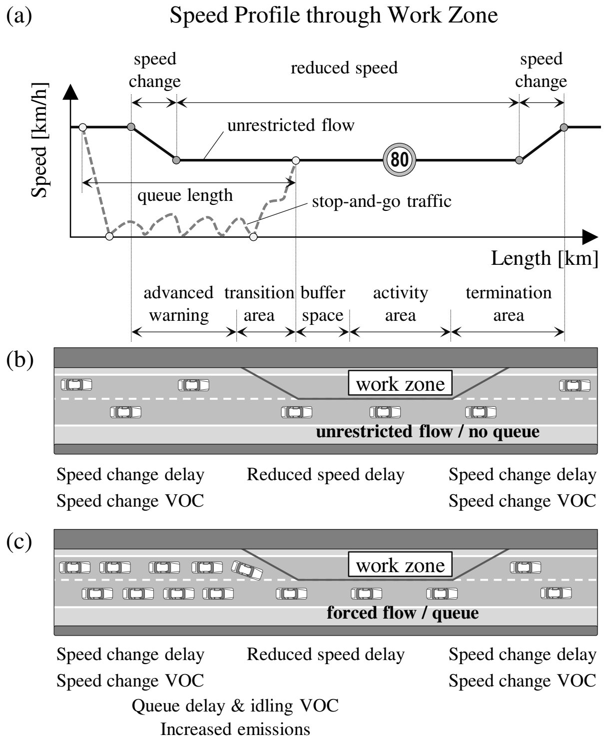
FIGURE 4.A.1 Stochastic treatment-life modeling with three types of variation: among different sections (a, b), with respect to the failure cause (a, c vs. b, d) and between repeated treatment applications (e). Consideration of three types of correlation: among neighboring road sections, between distress types and between pre- and post-treatment service lives (f).

---

[*] *This appendix is not part of the published journal article.*

4.A.2 Work-zone user costs

Figure 4.A.2 provides an overview of selected user-cost model parameters and unit costs, together with a typical speed profile and associated user costs in each area of the work zone. The parameters and unit costs (adjusted to 2016 constant euros) are representative for the Austrian freeways as they are a result of extensive literature analysis and calibration.



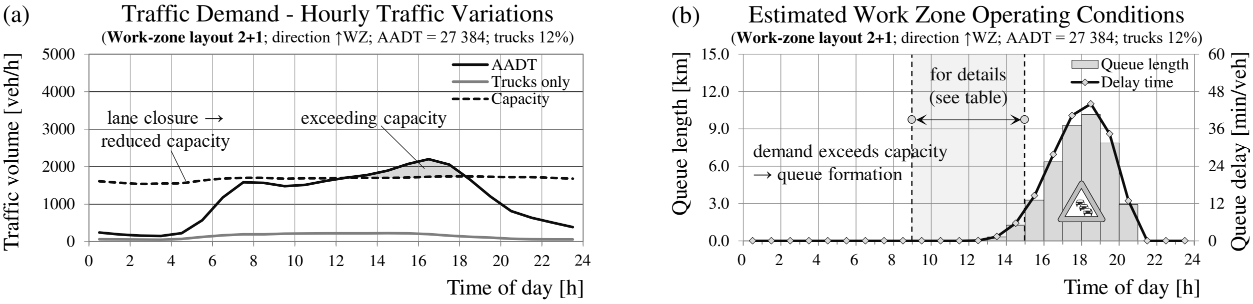
User costs (selected parameters and unit costs, 2016 prices, Austria)		
Travel delay costs (TDC)		
Hourly value of time travel (personal)	9.1	€/pers-h
Hourly value of time travel (business)	34.1	€/pers-h
Percent of business travel (freeway)	5.0	%
Average vehicle occupancy for personal travel	1.3	pers/veh
Percent of empty trucks (e.g. backhaul)	15.0	%
Average payload	15.0	t/truck
Hourly time value of trucks	29.3	€/truck-h
Vehicle operating costs (VOC)		
Passenger cars with diesel	57.0	%
Net cost diesel (Nov. 2016)	0.48	€/l
Net cost gasoline (Nov. 2016)	0.44	€/l
Emissions costs (EC)		
Greenhouse gases (CO_2)	57.0	€/t
Nitrogen oxide (NO_x)	11 604	€/t
Non-methane volatile organic compound	2 275	€/t
Particular matter (PM)	91 014	€/t
Crash costs (CC)		
Fatality (comprehensive costs)	3 261 543	€
Serious injury (comprehensive costs)	412 511	€
Minor injury (comprehensive costs)	29 082	€
Property damage	28 358	€/crash
Increase of crash rate in a work zone	50.0	%
Average crash rate (freeways)	0.090	crashes/ 10^6veh-km

FIGURE 4.A.2 Typical speed profile (a) and user-cost components associated with the different work-zone areas for unrestricted flow (b) and forced flow (c). Selected model parameters and unit costs calibrated for Austria (2016 prices).

Travel delay costs are estimated using a macroscopic demand-capacity model. Figure 4.A.3(a) and (c) show the analysis for a given traffic section (AXX_14), direction of travel (with work zone) and two work-zone layouts: 2+1 and 3+1, respectively. The work zone capacity is shown in Figure 4.6 and adjusted for the percentage of trucks in every hour using passenger car equivalents (upgrade ≤ 2%). Layout 2+1 provides lower capacity due to the closed lane, and exceeding of the capacity (in the afternoon) leads to queue formation. The resulting queue delay time and length are shown in Figure 4.A.3(b), together with a detailed spreadsheet for a time window from 9 am to 3 pm. Queued vehicles include the cumulative number of vehicles in queue (demand minus capacity for each hour). Queue length and delay are determined using average queue speed, which, in turn, is given as function of the ratio volume through queue to unrestricted capacity (Walls III and Smith 1998). In contrast, layout 3+1 provides sufficient capacity and no queue delays occur (Figure 4.A.3(d)). Delay time includes speed change, reduced speed, stopping and queue delays. Reduced speed delay, for example, is regarded as the difference between the time to cross the entire segment at the reduced work zone speed (80 km/h) and the time to cross the work zone at free-flow speed (130 km/h for passenger cars and 100 km/h for trucks).

The hourly traffic demand in Figure 4.A.3 is the product of the annual average daily traffic (AADT) and typical hourly distribution factors. These inputs can be obtained from historical traffic counting data at the work-zone site. In the case study, the traffic demand information is obtained from an analysis of 10 permanent traffic counters on two Austrian freeways. Based on data for seven consecutive years (2008-2014), the following input for the delay model was computed: average hourly distributions (Tuesday to Thursday) for each travel direction and separately for trucks and for all vehicles, annual average daily traffic/truck traffic (see Figure 4.8, bottom left) and growth rates (all vehicles/trucks). Traffic counting is based on mounted overhead triple-tech detectors (radar, ultrasound and passive infrared) and embedded in pavement inductive loops. Such historical measurements are available for the entire Austrian freeway network, comprising traffic data for each lane, vehicle classification (nine categories) and hourly values for each day of the year.

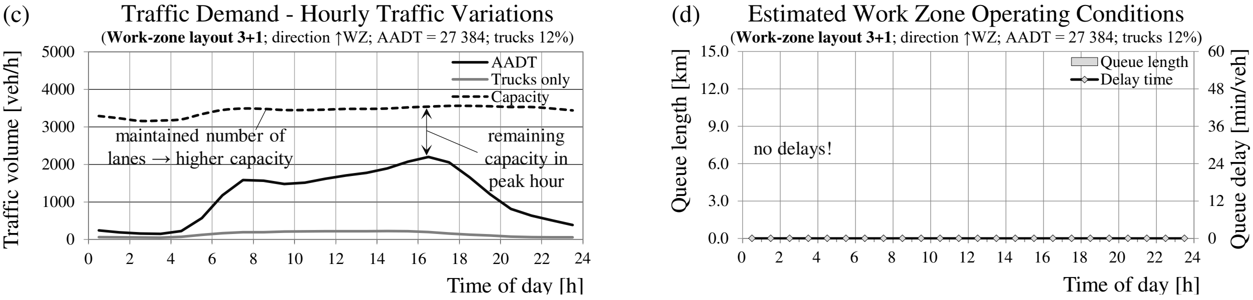
The crash costs per day are calculated using the expected additional number of crashes per day and the average costs per crash, assuming increased risk in the work zone. The average costs per crash are estimated based on unit costs (by severity) and average accident damages (e.g. fatalities/crash, serious injuries/crash, etc.), derived from statistical analysis of accident data for Austrian freeways (2013-2017). The expected number of crashes, in turn, is estimated as a product of the increase in crash rate and the exposure per day (work-zone length AADT).


(a) Traffic Demand - Hourly Traffic Variations
(**Work-zone layout 2+1**; direction ↑WZ; AADT = 27 384; trucks 12%)

(b) Estimated Work Zone Operating Conditions
(**Work-zone layout 2+1**; direction ↑WZ; AADT = 27 384; trucks 12%)

Traffic section AXX_14 ▾ from km 11.500 to km 20.500 Direction with WZ ▾ WZ Layout 2+1 ▾

Time	Hourly traffic [%]	Hourly traffic trucks [%]	Traffic volume [veh/h]	Volume trucks [trucks/h]	Adjusted capacity [veh/h]	Queued vehicles	Number of vehicles that				Operating conditions
							Traverse WZ	Traverse queue	Stop for queue	Slow down for WZ	
…											…
09-10	5.40%	6.35%	1479	213	1680	0	1479	0	0	1479	Free flow
10-11	5.53%	6.44%	1515	216	1690	0	1515	0	0	1515	Free flow
11-12	5.91%	6.61%	1617	221	1690	0	1617	0	0	1617	Free flow
12-13	6.23%	6.51%	1705	218	1700	5	1700	1700	1705	0	Forced/queue
13-14	6.48%	6.59%	1775	221	1700	80	1700	1700	1775	0	Forced/queue
14-15	6.91%	6.72%	1893	225	1700	273	1700	1700	1893	0	Forced/queue
…											…


(c) Traffic Demand - Hourly Traffic Variations
(**Work-zone layout 3+1**; direction ↑WZ; AADT = 27 384; trucks 12%)

(d) Estimated Work Zone Operating Conditions
(**Work-zone layout 3+1**; direction ↑WZ; AADT = 27 384; trucks 12%)

FIGURE 4.A.3 Example of demand-capacity analysis based on estimated hourly traffic variations and work-zone capacity, together with resulting queue length and delay time for a given traffic section, direction of travel (with work zone) and the two considered work-zone layouts 2+1 (a, b) and 3+1 (c, d).

4.A.3 Summary of parameters used in the case study

Figure 4.A.4 provides an overview of the distress types, service life distributions, distress and spatial correlations employed in the case study, as well as the cost and performance model parameter values (see Section 4.3). The case study considers three major distress types: rutting, surface defects and alligator cracking. However, including additional distress types will not affect the principal methodology or complexity of the formulated optimization problem. The service lives for the three distress types are positively correlated (distress correlation), as they are influenced by common factors (e.g. traffic loading). Furthermore, the service lives of neighboring road sections are also positively correlated (spatial correlation), which is modeled by a second-order autoregressive model with parameters ϕ_1 and ϕ_2 (see Figure 4.3, bottom right).

Cost and performance:

Distress types:

- Service life for alligator cracking, surface defects, rutting, as well as overall service life (series system)
- Correlation between distresses & among adjacent sections (spatial)

M&R treatment alternatives:

- Four most common treatments on Austrian freeways with delay/reset effect and stochastic life
- M&R cost based on contractor's bid prices and other sources

Temporary traffic control costs:

- Depend on the work zone layout; high proportion of fixed costs

Duration model:

- Based on daily production rates; accounting for setup time, delays and adverse weather

Work-zone user costs:

- Travel delay costs based on demand-capacity analysis
- Vehicle operating costs, emission costs and crash costs
- Total (project) user costs → product of daily user costs and project duration in days

Distress-specific service life distributions:

Distress type	Distribution	Parameters	$E(X)$ [yrs]
Rutting	Log-logistic	$\mu = 3.8; p = 5.8; \lambda = 0.045$	27.1
Surface defects	Normal	$\mu = 21.0; \sigma^2 = 9.923$	21.0
Alligator cracking	Weibull	$p = 4.5; \lambda = 0.047$	19.4

Distress and spatial correlation (ϕ_1; ϕ_2):

Rutting	Surface	Cracks	ϕ_1	ϕ_2
1.00	0.16	0.33	0.49	0.11
0.16	1.00	0.53	0.22	0.12
0.33	0.53	1.00	0.56	0.17

Treatment life (normally distributed with mean μ):

M&R treatment type (depth)	Rutting	Surface	Cracks
Surface treatment (1 cm)	$\mu = 4$*	$\mu = 7$	$\mu = 2$*
Surface course replacement (3.5 cm)	$\mu = 14$	$\mu = 12$	$\mu = 4$*
Surf & bin course replacement (11.0 cm)	$\mu = 14$	$\mu = 12$	$\mu = 9$
Full-depth asphalt replacement (23.0 cm)	$\mu = 21$	$\mu = 16$	$\mu = 19$

Temporary traffic control costs:

Cost components	STWZ	2+1	3+1
Fixed costs [€]	4 500	15 000	40 000
Variable costs [€/m]	3.0	5.0	30.0
Variable costs [€/m×day]	-	0.07	0.11

Agency cost functions (M&R treatment costs):

M&R treatment type (depth)	Fixed [€]	Variable [€/m²]
Surface treatment (1 cm)	1 300	6.6
Surface course replacement (3.5 cm)	2 400	14.3
Surf & bin course replacement (11.0 cm)	5 700	30.8
Full-depth asphalt replacement (23.0 cm)	7 800	51.9

Work-zone duration model:

Fixed [days]	Variable [days/1000 m²]
0.0	0.439
2.8	0.823
3.2	1.451
5.6	4.630

User cost functions for the given traffic segments:

Traffic section	User cost (layout 2+1)		User cost (layout 3+1)	
	€/day	€/m×day	€/day	€/m×day
AXX_10	101485	1.48	4924	2.93
AXX_14	80513	1.45	4757	2.87
AXX_23	1555	1.03	3402	2.02
AXX_35	1634	1.09	3573	2.13
AXX_50	1488	0.99	3462	2.07

Traffic by direction of travel:

From km	To km	↓ no WZ AADT	↑ with WZ AADT
0.000	11.500	27 348	28 066
11.500	20.500	26 498	27 384
20.500	25.150	18 778	19 462
25.150	40.400	19 779	20 509
40.400	50.000	20 407	18 680

Note: All prices are in 2016 constant euros and calibrated for Austria. The treatment life follows a normal distribution with parameters μ and $\sigma = 0.15\times\mu$ for surface defects/cracking and $\sigma = 0.20\times\mu$ for rutting. STWZ stands for short term work zone.

*Delay effect; does not apply, if the distress in question is the cause of failure (no effect).

FIGURE 4.A.4 Summary of estimated service life, performance and cost model parameter values.

CHAPTER 5

Aggregation of condition survey data in pavement management: shortcomings of a homogeneous sections approach and how to avoid them

by Valentin Donev, Markus Hoffmann & Ronald Blab

in *Structure and Infrastructure Engineering*
Published online: 24 Feb 2020 (open access)
DOI: 10.1080/15732479.2020.1730409

ABSTRACT

A common feature of most pavement management systems is the aggregation of condition data collected from periodic surveys to appropriate long homogeneous sections, aiming to reduce the amount of data to be analyzed. It can be shown that regardless of the methods used for sectioning and prediction, homogeneous sections cannot describe the true variation in condition and failure time, leading to either over- or underestimation of service life and hence suboptimal selection of timing and type of maintenance and rehabilitation (M&R) treatments. Furthermore, the generation of homogeneous sections hinders the derivation of project-level performance models from historical condition surveys due to shifting of sections and greatly reduces the number of possible work-zone solutions in the optimization. This paper illustrates the deviations resulting from different levels of data aggregation based on a case study of 1000 road sections. The results of common sectioning algorithms are then compared to those derived from a LCC-optimization approach based on accurate condition predictions on short survey sections and work-zones based on economies of scale. The conclusions of this work may be useful to road agencies in avoiding substantial drawbacks of data aggregation and improving project-level M&R treatment recommendations based on the same input condition data.

KEYWORDS: pavement management, life cycle costs, data processing, optimization, roads & highways, service life, rehabilitation, predictions, infrastructure.

5.1 INTRODUCTION

Pavement management systems (PMSs) rely on information about past and current condition to identify road sections in need of immediate repair and develop multi-year maintenance and rehabilitation (M&R) strategies. For this purpose, condition surveys of distress extent and severity are conducted periodically. Automated high-speed condition surveys produce an ever-increasing amount of data due to advancements in measuring technology, allowing more accurate and more detailed surveys. However, the first operational network-level PMSs have been introduced in the early 1980s in the USA and around 2000 in German-speaking countries. At that time, the available personal computers were not capable of handling large amounts of data. The solution was to aggregate the measured data to long condition-homogeneous sections and management sections facilitating the subsequent processing and analysis (Weninger-Vycudil et al. 2009, Lea et al. 2014). Today, the aggregation of condition data to homogeneous and/or management sections is still a part of every common PMS, despite the drastic improvements in computing power and efficiency of mathematical algorithms. However, the consequences of data aggregation and the associated loss of information are hardly ever discussed in the literature and almost unknown in the practice (Hudson et al. 2011, p.15).

The objective of this research is to systematically investigate the impact of aggregating survey data to homogeneous sections on the accuracy of service-life predictions and determination of M&R work zones. The selection of the topic is motivated by its practical importance and huge impact on the results in any PMS. In addition, the segmentation of condition data produced by automated surveys of linear assets is a common research issue concerning e.g. pipelines and railways as well (e.g. Alfelor and McNeil 1992, Amaya-Gómez et al. 2019). In this paper, different methods for generation of homogeneous sections are applied to a parametric case study and compared to a new life cycle cost (LCC) approach proposed by Donev and Hoffmann (2018b), which eliminates the need for an aggregation of survey data. An overview of other shortcomings in existing PMSs concerning condition assessment and condition prediction can be found in Hoffmann (2006) and Hoffmann and Donev (2018). A critical review of M&R programming using maximization of benefits based on aggregated condition index as well as possible alternatives are covered in Donev et al. (2020).

The reminder of this paper is organized as follows. An overview of selected literature on the topic of road segmentation and homogeneous sections is provided in Section 5.2. Using a parametric case study (Section 5.3) as a basis for comparison, Section 5.4 presents the consequences from the aggregation of survey data for the estimation of service life vis-à-vis the new LCC approach. The conclusions in Section 5.5 put the findings into perspective and provide an outlook into further applications and research potential.

5.2 INVESTIGATED METHODS

Automated condition survey systems provide pavement data with high accuracy and spatial resolution. Distress data is collected practically continuously or at very short distress-specific longitudinal sampling intervals (e.g. 0.10-0.15 m). As these measurements contain white noise, the data is averaged “online” and stored for longer recording intervals (e.g. 1-5 m). Pavement managers are usually not provided with the raw data but with data averaged over short survey (reporting) sections (e.g. 25-100 m). All common PMSs aggregate these survey sections to much longer condition-homogeneous and/or management sections (e.g. 1-5 km) using some algorithms or specific decision rules. A condition-homogeneous section usually consists of several survey sections exhibiting similar condition characteristics, using as a basis the most recent survey. The main goals of this aggregation are a reduction of the amount of data to be processed and analyzed, and obtaining practical lengths for planning of M&R projects in the PMS.

The typical aggregation process consists of determination of section borders and aggregation of the survey data. Alternatively, the aggregation of short survey sections may be viewed as segmentation of data series or road sectioning. The distress data associated with survey sections and homogeneous sections are referred to as condition values and representative values, respectively. The borders of the aggregated sections and, by extension, their lengths can be determined using fixed or dynamic sectioning (Bennett and Paterson 2000, Bennett 2004). Fixed sections do not change with time (Caltrans 2015). The goal of dynamic sectioning is to produce homogeneous (uniform) sections in terms of condition, structure, etc., limiting the deviations from the original data and the associated loss of information. Thus, each condition survey results in different condition-homogeneous sections. As a result, at the road section level, the performance history cannot be replicated, the trends in distress progression cannot be assessed and the information from previous surveys cannot be included in condition predictions. Therefore, in German-speaking countries, condition predictions are based on information only from the last survey, leading to inaccurate predictions without consideration of uncertainty (Hoffmann 2006). On the other hand, if fixed management sections are used, more information will be lost due to averaging and the predicted condition will deviate more from the actual variation within the section. Subsequently, condition-homogeneous sections for each distress type are further aggregated to management sections, considering pavement structure, traffic volume, etc. Once the sectioning is completed, the data from the survey sections must be transformed to these homogeneous or management sections by using distress-specific aggregation rules (e.g. weighted average, maximum value, minimum value).

There are many different approaches for road segmentation. Three of the most common methods are briefly described below and applied to the case-study data. The absolute difference approach (ADA), which has been used in Germany, transforms the original measurement series to a series of smoothed values using a symmetric moving average filter (Rübensam and Schulze 1996). Subsequently, the absolute differences between each two neighboring values that follow each other in a user-defined interval are computed. Section borders are inserted each time the graph of absolute differences exceeds a defined threshold value (see Figure 5.1(a)). The threshold value, the window length for the calculation of moving average and the window length for absolute differences determine the number of resulting sections and the degree of segmentation. The selection of these control parameters depends on the distress type and variability of the data, but there are no objective technical criteria indicated for determination of their values. The formation of homogeneous sections considering all distress types is based on the "lowest common denominator" principle (Bennett 2004). A second algorithm aggregates the resulting sections to ensure that all sections satisfy a specified minimum-length constraint. In Germany, a separate algorithm is used for generating homogeneous sections with regard to the pavement structure (type, layers, etc.). Another algorithm aggregates these two sets of homogeneous sections to final management sections with lengths of approximately 5 km, being considered as practical for freeways (Kunze and Rübensam 2007).

Absolute difference in moving averages (ADA):

- Smoothing of survey data series (condition values) using moving average (MA) filter
- Calculation of absolute differences between neighboring values of the smoothed series
- Section borders are inserted at the maxima of each proportion of the absolute-differences series exceeding a defined threshold

Cumulative difference approach (CDA):

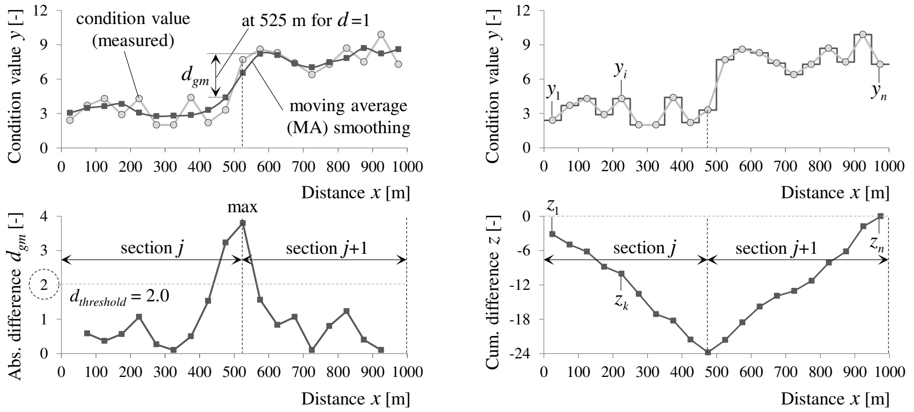
- Cumulative differences → adding the difference between current value (y_i) and the series average ($\bar{y}$) to the previous sum (z_{i-1})
- Section borders are inserted at the places where the slope of the cumulative-differences curve changes algebraic signs

Minimization of the sum of squared errors (SSE):

- Section borders minimize the sum of squared errors resulting from the replacement of survey data with the representative values (averaged data) for the homogeneous sections
- The number of sections depends on additional requirements (minimum section length, etc.)

(a) Absolute Difference Approach (ADA)

(b) Cumulative Difference Approach (CDA)



Symmetric moving average of window length $(2\times q+1)$:

$$MA(y_i)=\frac{1}{2\times q+1}\sum_{r=-q}^{q} y_{i+r} \qquad q \text{ is integer} \qquad r=-q,\ldots,+q$$

Absolute difference of MAs for a window $(2\times d+1)$:

$$d_{gm,i}=\left|MA(y_{i-d})-MA(y_{i+d})\right| \qquad d \text{ is integer}$$

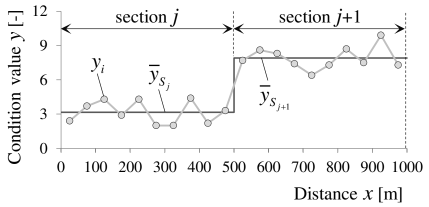
Cumulative difference values (equal survey sections):

$$z_i=(y_i-\bar{y})+z_{i-1} \qquad \bar{y}=\frac{1}{n}\times\sum_{i=1}^{n} y_i \qquad i=1,\ldots,n$$

$y_1, y_2, \ldots, y_n$ → series of condition values

$\bar{y}$ → average of all condition values y_i

(c) Minimization of the Sum of Squared Errors (SSE)



Sum of squared errors between original data (condition values) and section averages (homogeneous sections):

$$SSE=\sum_{j=1}^{k}\sum_{i\in S_j}(y_i-\bar{y}_{S_j})^2 \qquad k \rightarrow \text{number of homogeneous sections}$$

Minimum length of a homogeneous section (constraint):

$$L_{S_j}=\sum_{i\in S_j}(x_i-x_{i-1})\geq L_{\min} \qquad \forall S_j=S_1,S_2,\ldots,S_k$$

FIGURE 5.1 Common methods for generation of homogeneous road sections and segmentation of condition survey data: absolute difference approach (a), cumulative difference approach (b) and minimization of the sum of squared errors (c).

The cumulative difference approach (CDA) is the most well-known and widely cited method for road segmentation in the literature (AASHTO 1993). For the case of equally long survey sections, a series of cumulative differences is constructed by calculating the cumulative sum of differences between each current measurement and the average of the entire series. The cumulative differences are plotted against distance, and the points where the graph changes its direction from decreasing to increasing and vice versa are identified as section borders (see Figure 5.1(b)). If different parts of the measurement series are evaluated with the algorithm, this may result in a substantial change in number and borders of homogeneous sections, especially for series with structural breaks. Moreover, the method does not allow control over the number of generated sections. Hence, several authors have modified the CDA by incorporating additional constraints like minimum length of homogeneous sections and minimum difference in averages between two adjacent sections (Ping et al. 1999, Kennedy et al. 2000, Cafiso and Di Graziano 2012, Haider and Varma 2016).

Another group of methods for the identification of homogeneous sections is based on the minimum sum-of-squared-errors (SSE) criterion (Misra and Das 2003, Cafiso and Di Graziano 2012). The problem may be formulated as a nonlinear binary programming model, where the goal is to minimize the sum of squared deviations between the original data (survey sections) and the aggregated data (averages for homogeneous sections), as illustrated in Figure 5.1(c). A binary decision variable, indicating the presence of a section border or a change point is assigned to each condition value of the series, where the variable equals "1" if section border should be inserted and "0" otherwise. Since the solution of the unconstrained problem is the survey sections, additional constraints are necessary (e.g. minimum length of homogeneous sections). Due to the complexity, the problem cannot be solved exactly. Therefore, heuristic methods are applied, providing no information about the quality of the solution.

In Austria, a method based on Bayesian principles is used for the segmentation of condition survey data, assuming that measurements can be described by a first-order autoregressive process (Thomas 2005). In the Austrian PMS, however, homogeneous sections are only used to reduce the amount of data (Weninger-Vycudil et al. 2009). The final management sections are determined based solely on inventory data (pavement structure, traffic, other assets), completely disregarding the condition-homogeneous sections. An exception is made for sections where the safety-relevant distress types rutting and skid resistance exceed their condition thresholds. Such sections are considered as separate management sections. In all other cases, the homogeneous-sections data is transformed to the management sections using length-weighted averaging.

In contrast to other works (cf. Latimer et al. 2004, Thomas 2004, El Gendy and Shalaby 2008, Cafiso and Di Graziano 2012), the goal of this paper and case study is not to compare different algorithms for road segmentation and determine the most appropriate among them. Instead, the goal here is to show that any aggregation of condition data to homogeneous and management sections has significant consequences for the estimation of service life and determination of M&R timing leading to suboptimal results regardless of the used method.

However, the compromise between data accuracy and practical project lengths can be avoided. The solution proposed by Donev and Hoffmann (2018b) is to use short survey sections (e.g. 25-50 m) for condition prediction. Survey sections can then be combined to longer work zones based on more accurate condition predictions and economies-of-scale cost functions for each M&R treatment type. The length and location of a work zone, as well as treatment's timing and type, result directly from solving an optimization problem, minimizing the total LCC. A detailed description of this approach is provided in the cited paper. The current paper presents a comparison of work zones resulting from homogenous sections with different level of aggregation (based on current condition) and work zones determined by using LCC optimization (based on M&R costs).

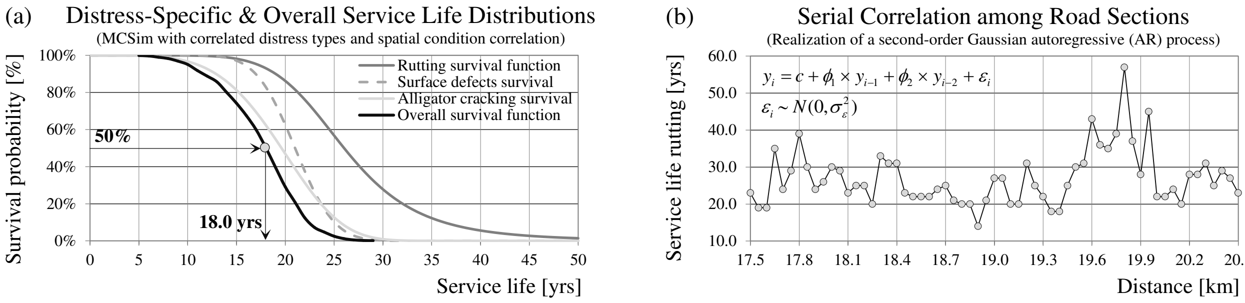
5.3 CASE STUDY

This paper presents an in-depth analysis of the consequences from aggregation of condition survey data to homogeneous sections in existing PMSs. Selected commonly-used data-segmentation algorithms are applied to a case study of a simulated road network and compared to an innovative LCC approach that does not require data aggregation. The advantage of a simulation approach can be found in the complete knowledge of service life and condition distribution at any point in time, avoiding the limitations of empirical data like censoring and ex ante prediction. Therefore, synthetic data allows researchers to isolate, objectively quantify and compare the deviations resulting from the application of a specific approach (e.g. prediction bias, threshold violations).

The case study is based on 1000 road sections with flexible pavement located on one of the exterior lanes of a four-lane freeway. Each section has a length of 50 m which is the standard survey length in the Austrian PMS and a width of 3.75 m, resulting in a route with a total length of 50 lane-km. Figure 5.2(a,c) gives an overview of the three considered distress types, namely rutting, surface defects and alligator cracking, together with their corresponding service life distributions. Service life is defined as the time from initial construction or application of M&R activity until the exceedance of a certain distress-specific condition threshold. For a given road section, the first distress exceeding its threshold determines the overall service life (series system), the cause for failure and also the latest possible timing for M&R measures.

At the network level, the service lives for the three distress types are positively correlated (distress correlation), as they are influenced by common factors like traffic loading. Furthermore, the service lives of neighboring road sections are also positively correlated (spatial correlation), which is modelled by a second-order autoregressive model with parameters ϕ_1 and ϕ_2 (see Figure 5.2(b,d)). A realistic estimate of the spatial correlation is of a critical importance, as it directly affects the formation of homogeneous sections. Ultimately, a set of three distress-specific service lives is generated for each of the 1000 road sections. Power functions with fixed power parameter (β_2) are fitted to the individual service lives by varying the scale parameter (β_1) in order to describe intermediate conditions. The employed performance functions together with the corresponding rating scales for computation of an aggregated condition index (see Section 5.4.2) are presented in Figure 5.3(a,b,c).

A comparison between generated homogeneous sections and the LCC approach proposed by Donev & Hoffmann (2018b) requires the definition of M&R treatment alternatives. While homogeneous sections may be used as basis for determination of work zones, the former are not linked to specific M&R treatments. Thus, in order to ensure comparability, only one possible treatment is considered, eliminating “treatment type” as a decision variable in the LCC approach. Full-depth asphalt replacement is chosen in the case study, since it has effect on all three distress types. Figure 5.3(d) presents economies-of-scale cost functions for this treatment, including temporary traffic control costs (TTCC) for the two work-zone layouts depicted in Figure 5.3(e). The effect of full-depth asphalt replacement on the different distress types is shown in Figure 5.3(f). A more detailed discussion on the selected service life distributions, cost models and the data-generation procedure is offered in Donev and Hoffmann (2018a, 2018b). Different parametric values or distribution functions may more or less change the specific results in terms of value and possibly sign but will not change the essence of the conclusions of this paper.


(a) Distress-Specific & Overall Service Life Distributions
(MCSim with correlated distress types and spatial condition correlation)

(b) Serial Correlation among Road Sections
(Realization of a second-order Gaussian autoregressive (AR) process)

(c) Distress-Specific Service Life Distributions
(Based on literature and preliminary analysis of real data in Austria)

Distress types

	Distribution type, parameters and expectation
Rutting	Shifted log-logistic: $\mu = 3.8$ (shift); $p = 5.8$ (shape); $\lambda = 0.045$ (scale); $E(X) = 27.1$ yrs
Surface defects	Normal distribution: $\mu = 21.0$ (mean); $\sigma^2 = 9.923$ (variance)
Alligator cracking	Weibull distribution: $p = 4.5$ (shape); $\lambda = 0.047$ (scale); $E(X) = 19.4$ yrs

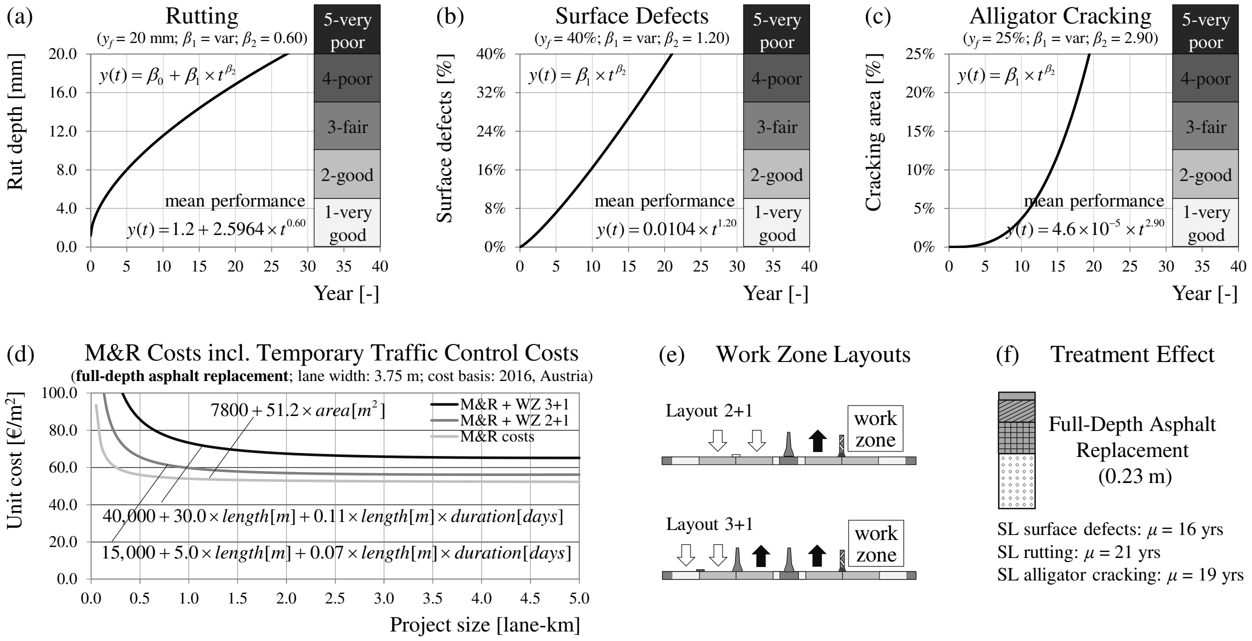
(d) Serial Correlation Estimation – AR(2) Model
(Average estimates based on 1400 km regional roads in Austria)

	Rutting	S. defects	All. cracking
Parameter ϕ_1	0.49	0.22	0.56
Parameter ϕ_2	0.11	0.12	0.17

Legend

ϕ_1, ϕ_2 … autoregressive parameters (c is a constant term)
ε_i … Gaussian white noise with mean 0 and variance σ_ε^2

FIGURE 5.2 Overall and distress-specific service life distributions (a, c) as well as correlation between neighboring road sections (b, d) employed in the case study.



Note: M&R costs and TTCC are based on analysis of empirical data from Austria, available literature and expert opinions. The expected service life for full-depth asphalt replacement is assumed to follow a normal distribution with parameters μ and $\sigma = 0.15\times\mu$ for surface defects and cracking and $\sigma = 0.20\times\mu$ for rutting.

FIGURE 5.3 Distress-specific performance functions (a, b, c), economies-of-scale cost functions (d), considered work-zone layouts (e) and service life for full-depth asphalt replacement (f).

5.4 CONSEQUENCES OF USING CONDITION-HOMOGENEOUS SECTIONS

5.4.1 Service life prediction

In this section, the three algorithms for road segmentation ADA, CDA and SSE described in Section 5.2 are applied to the case study road network (50 lane-km), separately for each distress type. Homogeneous sections based on a minimization of SSE are generated for three different minimum lengths (200, 600 and 1000 m) to investigate different levels of aggregation. Figure 5.4(a) shows for example a 3-km road segment, consisting of 60 survey sections and cracking measurements associated with each 50-m survey segment, serving as a current condition and basis for segmentation. Thus, the 3-km segment is divided into four homogeneous sections using the SSE approach with a minimum-length constraint of 600 m. The representative values for the generated homogeneous sections are obtained by taking the average. The first obvious consequence from the aggregation is that peak condition values will be lost or smoothed out.

Even if an algorithm singles out sections with critical condition values, this does not solve the problem with masking of critical sections when predicting future conditions (see Figure 5.4(b)). Condition prediction is performed once using the values of the survey sections and once using the aggregated values of the homogenous sections. Detailed predictions for four randomly selected survey sections are provided in Figure 5.4(c). The average performance function for cracking (see Figure 5.3(c)) is scaled through both the average value for the homogeneous section and the single condition values, reflecting the current practice for section-specific predictions in the German-speaking countries (e.g. Weninger-Vycudil et al. 2009). Service life equals the time until reaching the defined condition threshold (25% alligator

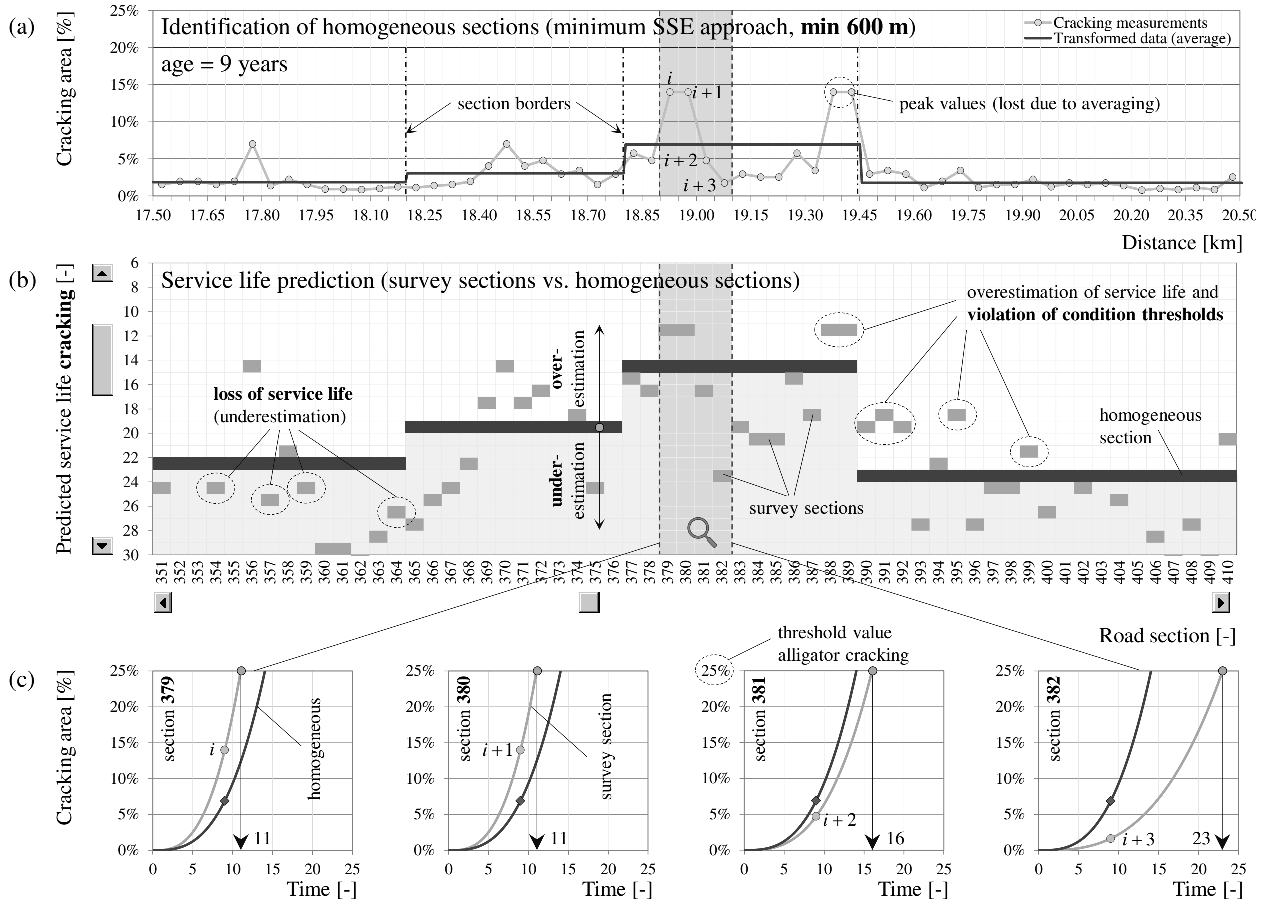
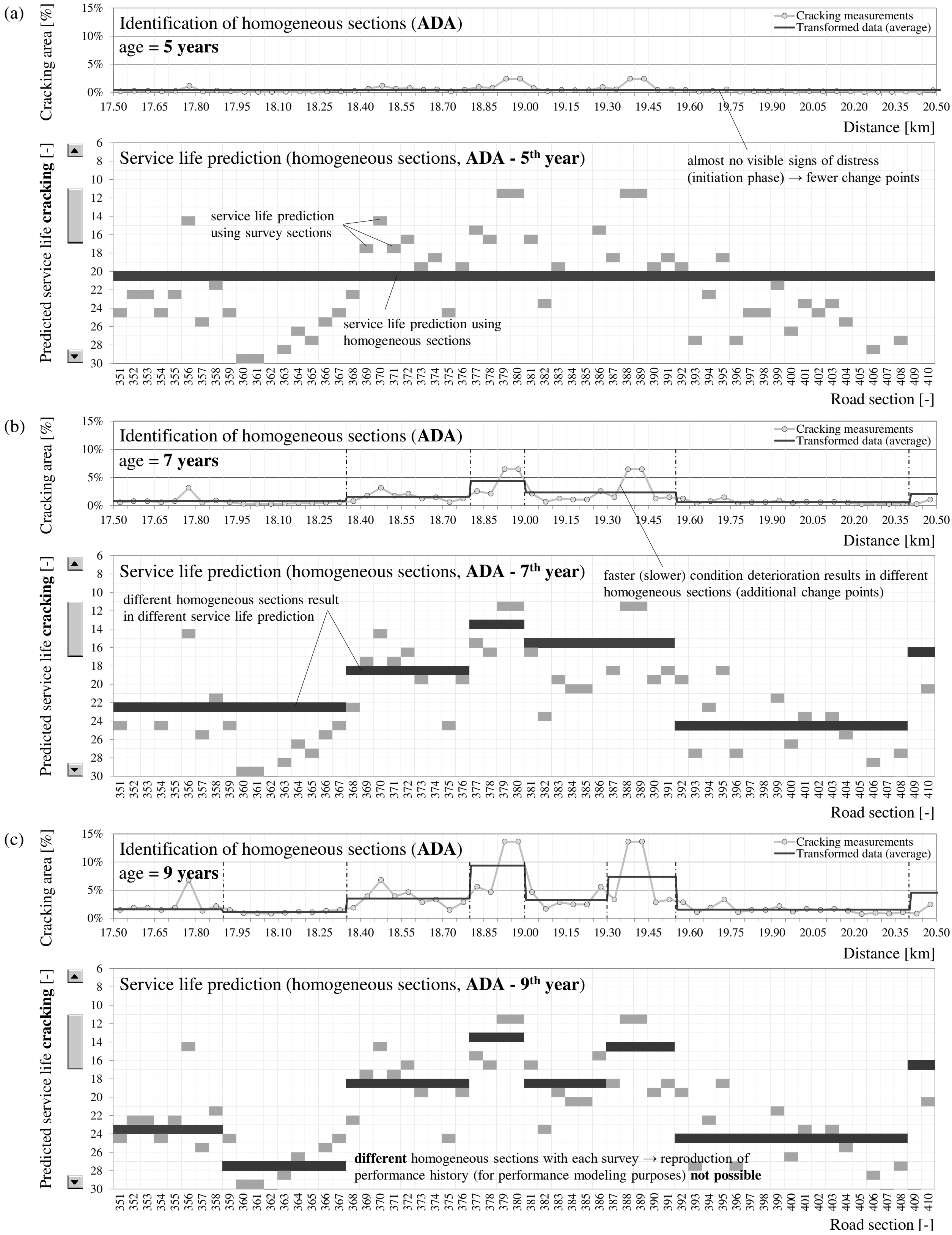


FIGURE 5.4 Identification of homogeneous sections and data transformation for a segment of 60 randomly chosen road sections (a). Service-life prediction errors resulting from the use of homogeneous sections in comparison to prediction based on short survey sections (b, c).

cracking). Figure 5.4(b) shows that it is not possible to predict the true variation in failure time with the aggregated sections. Thus, some short sections will fail prior to their predicted service life (prediction based on homogeneous sections), resulting in an overestimation of service life. Instead of recognizing the need for local repairs in a shorter term, the practitioners will be unaware of the potential threshold violations, as the detailed data is not analyzed in existing PMSs. In other cases, homogeneous sections will underestimate the true service life, leading to unnecessary loss of service life. This problem will persist in the case of fixed management sections and/or more sophisticated methods for prediction due to the underlying averaging.

Next, the generation of homogenous sections is repeated with the ADA approach without minimum-length constraint using the condition in different years as basis for the segmentation. Figure 5.5(a) presents cracking area at pavement age of 5 years, together with the resulting homogeneous section(s). The majority of sections do not exhibit any cracks, which can be explained by the progressive performance function and typical initiation phase for cracking. Thus, there are no significant changes in the condition, resulting in only one homogeneous section with service life equal to 20 years. Only two years later, cracking occurs on multiple sections which produces five homogeneous sections (see Figure 5.5(b)). The updated condition prediction based on the new averaged condition values is quite different from the uniform prediction in Figure 5.5(a). Applying again the same algorithm at the age of 9 years leads to further segmentation and different prediction results (Figure 5.5(c)). Thus, instead of improving condition predictions, the data from each consecutive condition survey will lead to entirely new road segmentation, resulting in different averaged condition values and different service-life predictions. Last but not least, the number and position of homogeneous sections depend heavily on the applied segmentation algorithm (compare Figure 5.4(a) with Figure 5.5(c)).



almost no visible signs of distress (initiation phase) → fewer change points

service life prediction using survey sections

service life prediction using homogeneous sections

faster (slower) condition deterioration results in different homogeneous sections (additional change points)

different homogeneous sections result in different service life prediction

different homogeneous sections with each survey → reproduction of performance history (for performance modeling purposes) **not possible**

FIGURE 5.5 Effect of using consecutive condition surveys as a basis for homogeneous sections based on cracking measurements at different points of time – pavement age of 5 (a), 7 (b) and 9 years (c).

5.4.2 Combined effect of homogenous sections and aggregated indices

The deviations in Figure 5.4 and Figure 5.5 result from an aggregation of survey sections for just one distress type. In the presence of multiple distress types, the overall service life for a given section can be determined as time until the first failure (Figure 5.6(a)). However, if a composite condition index like overall condition index (OCI) or present serviceability index (PSI) is used as a trigger of M&R activities, overall service life will always be overestimated due to averaging and weighting. In addition, this effect is overlaid by the formation of homogeneous sections at the network level, as demonstrated in Figure 5.6(b). The dashed line shows the survival function based on the survey sections (no data aggregation) and the first-failure principle. In the present case, the time until the first failure (series system) is equivalent to taking the maximum/worst of the three individual distress indices. The solid black line represents the survival function based on homogeneous sections (e.g. ADA) and a series system. The area difference between the solid black and the dashed line displays the ceteris paribus effect of homogeneous sections on service life. The remaining three lines describe the combined effect of homogeneous sections and OCI, resulting in overestimation of service life and survival probability. For example, according to the true distribution (dashed line) 60% of all sections survive until year 17, while the corresponding percentage in the case of OCI (weights 0.20, 0.50, 0.30) amounts to 96%.

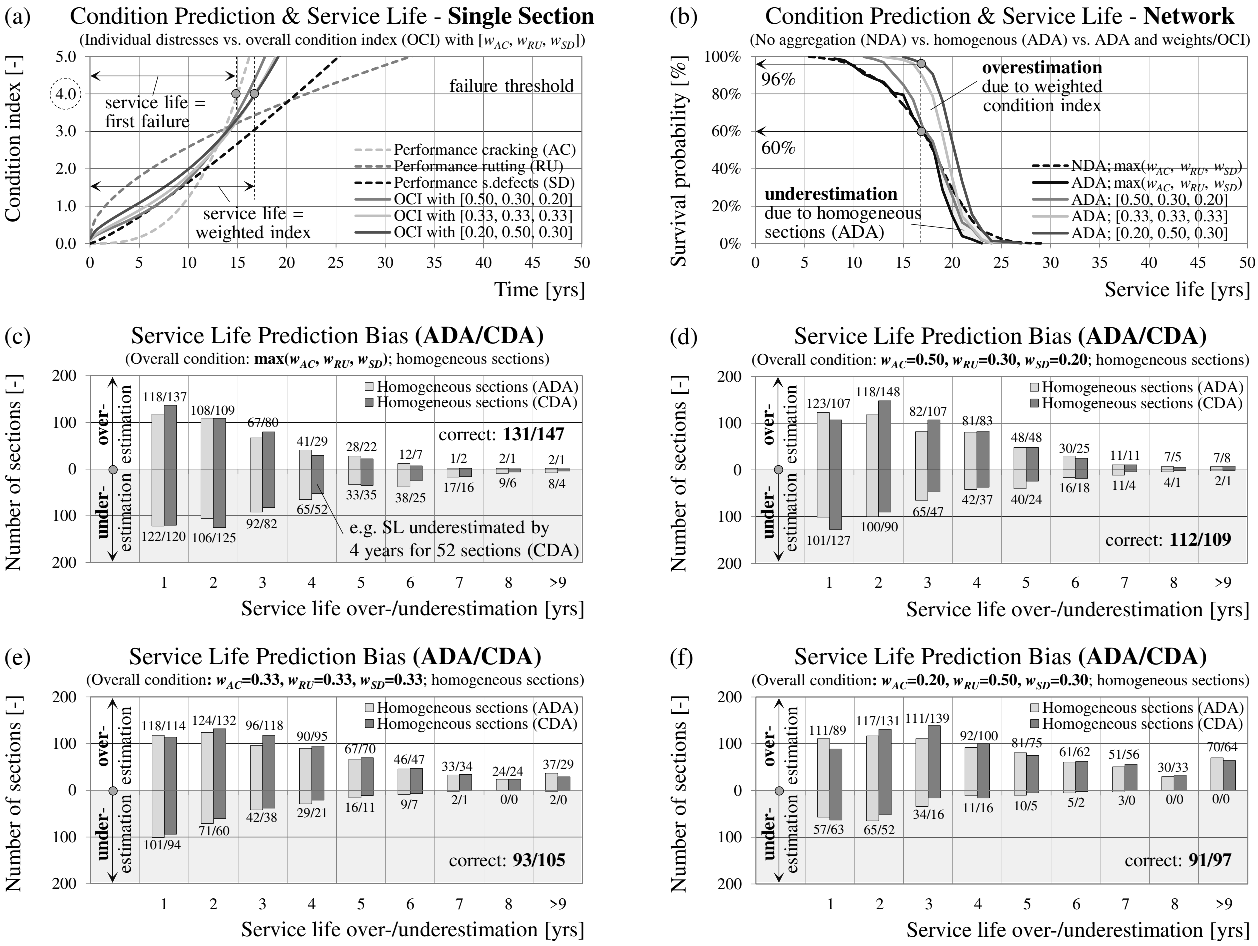


FIGURE 5.6 Overestimation of service life for a road section with multiple distress types by using OCI (a). Over-/underestimation of service life in years resulting from aggregation to homogeneous sections (ADA/CDA), where the overall service life is based on a series system (c) or on an OCI with different weights (b, d, e, f).

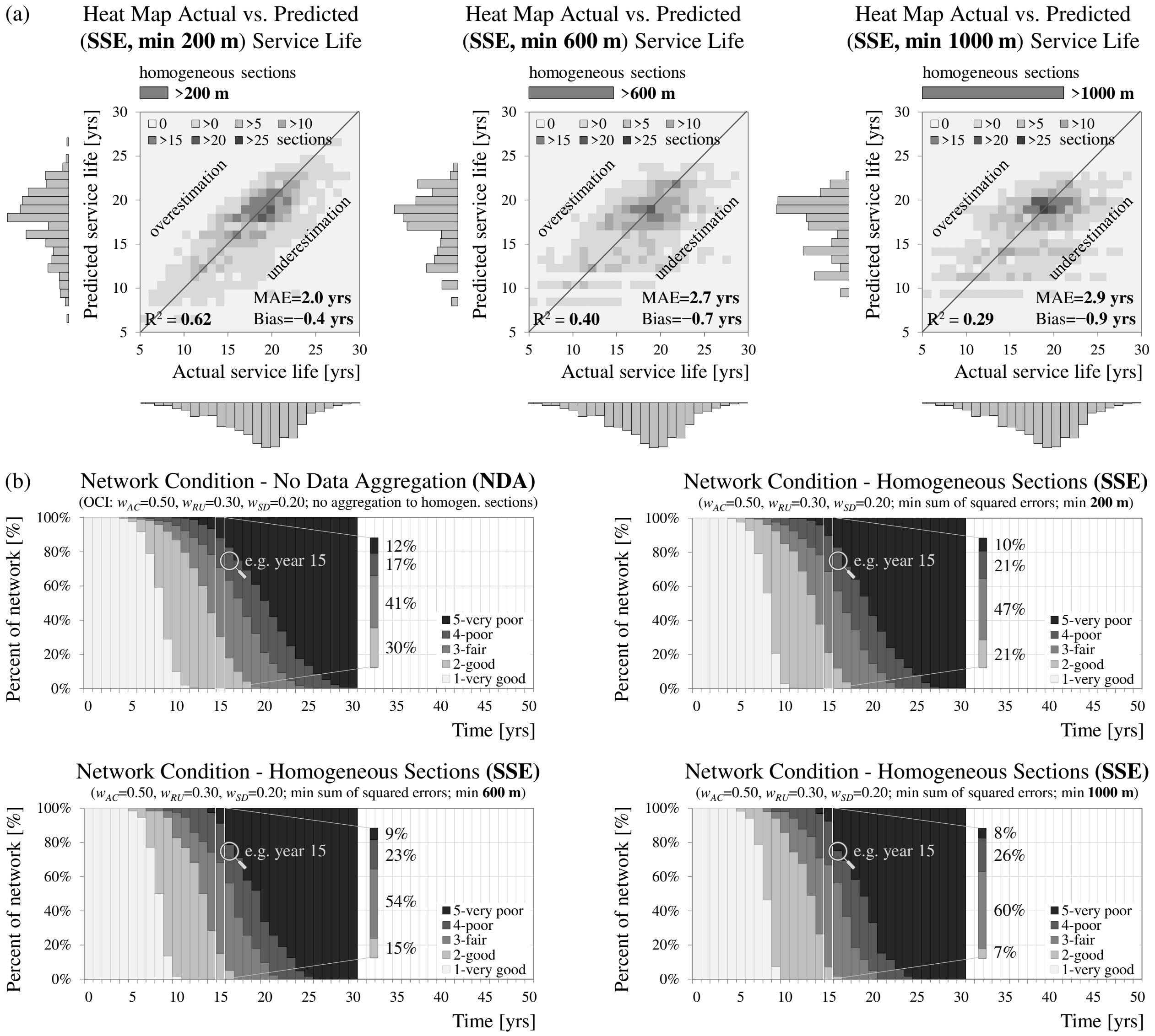
Figure 5.6(c) shows a histogram (tornado chart) of the bias in service life prediction, resulting from the formation of homogeneous sections alone (ADA/CDA). The columns provide the number of sections (out of 1000), for which the service life is over- and underestimated sorted by the error in years on the horizontal axis. The columns are slightly shifted downwards, indicating systematic underestimation. Figure 5.6(d,e,f) provides corresponding analysis for the cases of homogeneous sections and three different sets of weights for OCI. Based on the calibrated model parameters, alligator cracking is the dominant failure cause in the case study. Consequently, if lower weights are attached to cracking, the systematic overestimation becomes more obvious (see Figure 5.6(e,f)). The use of a composite condition index as a trigger for treatments increases the overestimation of service life for short-lived sections caused by homogeneous sections (overlay effect) and partially reduces, on average, the underestimation for long-lived sections.

5.4.3 Influence of aggregation level

The deviations between actual (survey sections) and predicted (homogeneous sections) service life for three distinct levels of aggregation are visualized using heat maps in Figure 5.7(a). The aggregation is based on the SSE approach constraining the minimum length of a homogeneous section, respectively, by 200 m, 600 m and 1000 m. The pixels of the heat map represent the number of sections for each combination of actual and predicted service life, whereby darker color indicates higher frequency. In these three cases overall service life is based on the first-failure principle (i.e. without using OCI). The corresponding histograms of the service life distributions based on survey sections and homogeneous sections are shown below and to the left of each heat map, respectively. It can be seen that the aggregation of survey sections distorts the service life distribution in a way that the variation is reduced and the tails are lost. Furthermore, with increasing level of aggregation the predicted values are more dispersed, indicating larger prediction errors. More road sections fall below the diagonal line, suggesting systematic underestimation of service life, which is confirmed by the estimated mean error (bias).

The consequences of the aggregation to homogeneous section for the condition distribution are revealed in Figure 5.7(b). The figure shows the distribution of the OCI, which is computed from the distributions of the individual distress types. The case of no data aggregation (NDA) displays the true network condition. At first glance the condition distributions may seem similar, but a closer look, for example, at year 15 reveals that the percentages of the network in different condition classes change substantially.

The table at the bottom of Figure 5.7 summarizes the results for the individual distress types and overall service life. The mean error (bias) and the mean absolute error (MAE) are the average difference and the average absolute difference, respectively, between the actual and predicted service life. The MAE of overall service life is already 2.0 years for a length of 200 m and increases to 2.9 years for 1000 m. The mean error shows systematic underestimation of service life in all cases. The errors when predicting individual distress types are in general larger (e.g. 3.7 years for alligator cracking and 1000 m). As the different distresses "compete" to determine the overall service life, a large error for an individual distress type will not be relevant, if the overall life is determined by a competing event. Moreover, a larger error in service life for one individual distress may coincidentally compensate the error in overall service life by changing the failure cause. The errors of the individual distress types depend on the variance of the corresponding distress-specific service life or condition distributions. Thus, a distribution with larger variance, like the one for rutting (see Figure 5.2(a)), results in larger errors for the same minimum length of homogeneous sections. Another consequence of the aggregation is the change of the dominant failure cause (i.e. first distress that reaches its threshold). For example, in the case of 1000-m length, the failure cause changes for 32% of the sections and the share of alligator cracking as leading failure cause increases from 63% to 86%. This will ultimately result in M&R programs with different selection of treatments – in this case, a program with more structural repairs. The above-mentioned consequences will be more severe for the cases of fixed sections and dynamic sections with greater length, which are more common in practice.



Prediction quality	Service life prediction - SSE (200 m)					Service life prediction - SSE (600 m)					Service life prediction - SSE (1000 m)				
	RU	AC	SD	Overall	Failure cause	RU	AC	SD	Overall	Failure cause	RU	AC	SD	Overall	Failure cause
Mean abs. error (MAE) [yrs]	3.9	2.2	1.8	2.0		4.5	3.2	2.2	2.7		4.8	3.7	2.4	2.9	
Bias (mean error) [yrs]	−0.7	−0.8	−0.3	−0.4		−1.0	−1.5	−0.4	−0.7		−1.1	−1.9	−0.6	−0.9	
Correct predictions	8%	17%	17%	18%	78%	10%	10%	15%	13%	71%	8%	8%	13%	12%	68%

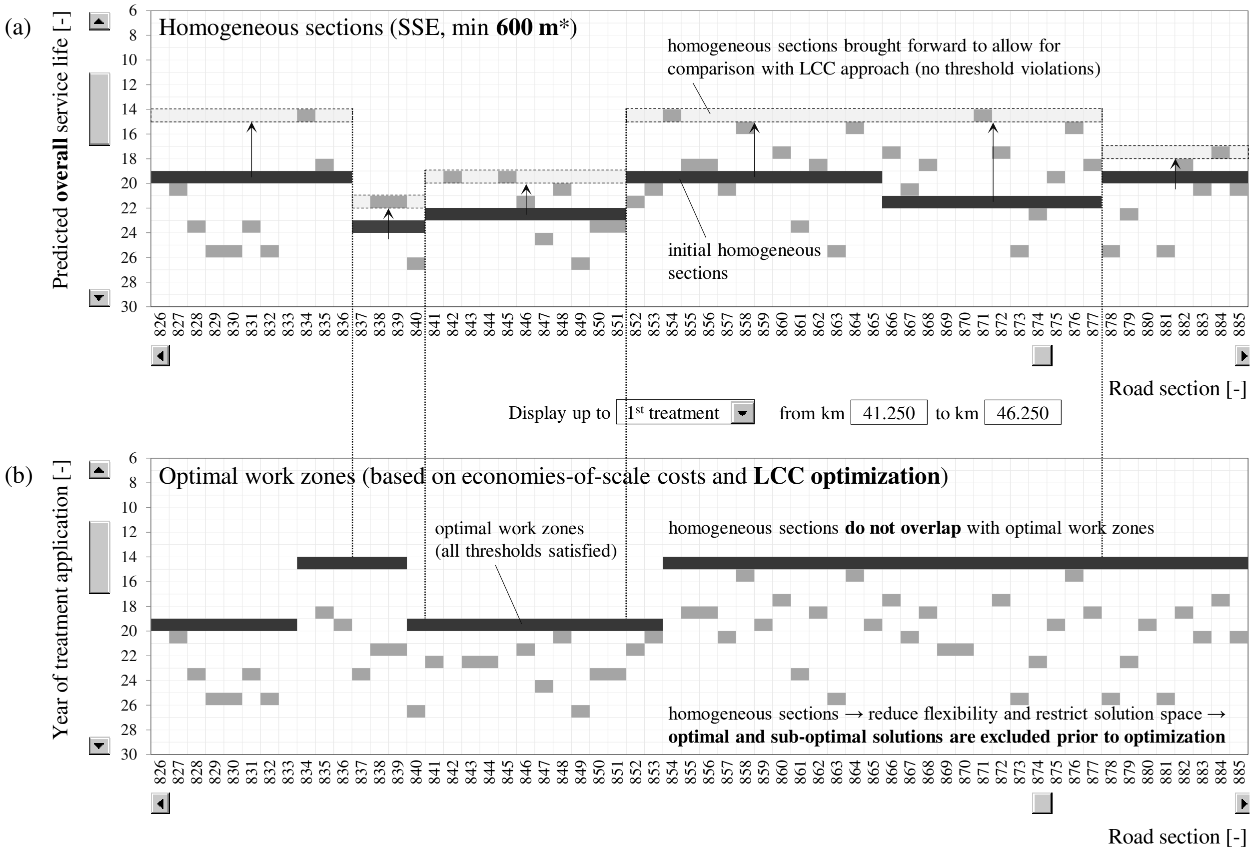
Note: The network condition distribution is computed from the distributions of the individual distress types, while the overall condition index is only used as a means to consolidate individual distributions and facilitate comparison. In all cases the overall service life is estimated as time to first failure (series system).

FIGURE 5.7 Heat maps of actual (survey sections) vs. predicted service life based on homogeneous sections (SSE) with different level of aggregation (a). Distortions of the condition distribution at the network level in each case (b).

5.4.4 Comparison to LCC approach

As homogeneous sections are not directly related to specific M&R treatment types, an objective comparison with the LCC approach proposed by Donev and Hoffmann (2018b) is only possible for one preselected treatment. In this case, full-depth asphalt replacement is chosen, as it has impact on all three considered distress types. Furthermore, the timing of work zones based on homogeneous sections may be set to the corresponding overall service life prediction. However, homogeneous sections lead to a violation of condition thresholds, whereas the work zones based on the LCC approach do not allow threshold violations, unless soft constraints are being used. The violation of thresholds is not straightforward to quantify objectively in monetary terms. Nonetheless, the timing of work zones based on homogeneous sections can be shifted upwards to earliest failure time of the short sections forming them, as illustrated in Figure 5.8(a). Thus, the work zones based on homogeneous sections will also satisfy all condition constraints with timing determined by first failing short survey section and the only possible treatment type in this case study – a full-depth asphalt replacement.

The generation of homogeneous sections does not take into account predicted future condition and possible M&R actions. In contrast, the work zones resulting from the LCC optimization are based on more accurate condition predictions and economies-of-scale cost functions for each M&R treatment type and work-zone layout. Figure 5.8(b) shows the results of the optimization



Note: A different set of 60 survey sections was used to highlight the differences, as the results obtained using the previous example were similar (by chance).

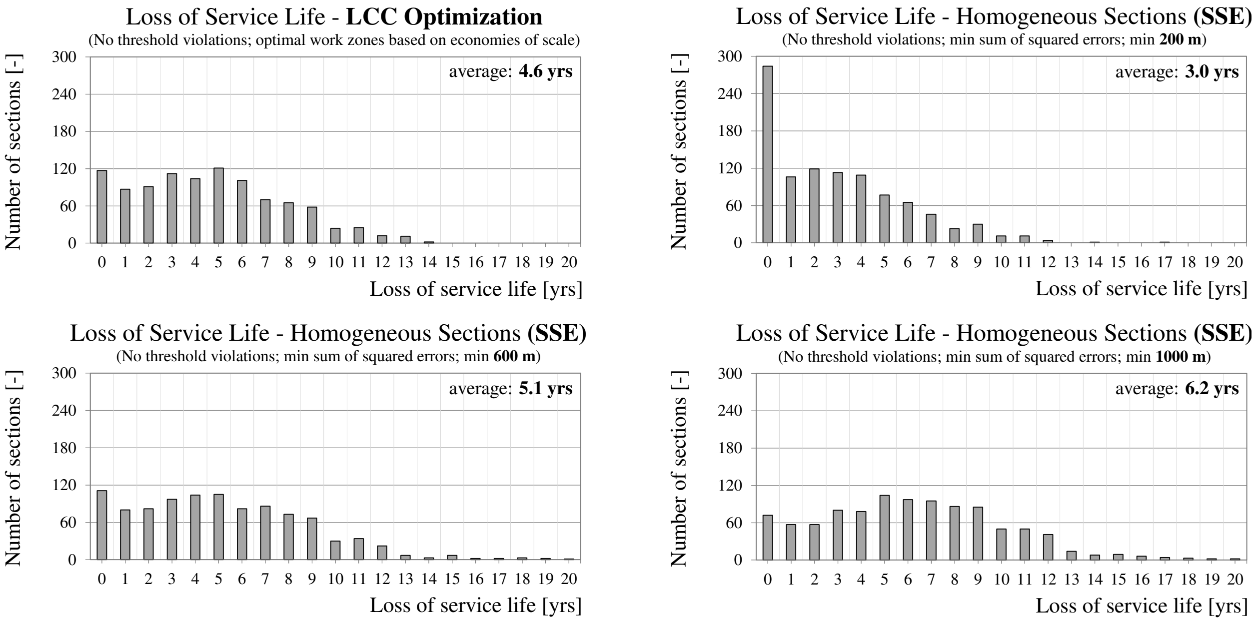
*Overall homogeneous sections are determined using homogeneous sections for the individual distress types (based on min SSE, 600 m), whereby the change points of the first failing distress type are relevant. As a result, some overall homogeneous sections exhibit length shorter than the minimum (600 m).

FIGURE 5.8 Comparison of work zones resulting from condition-homogeneous sections with modified timing (no threshold violations) and work zones resulting from an LCC optimization considering economies-of-scale effects.

for a discount rate of 4%, with details on the solution algorithm being given in Donev and Hoffmann (2018b). The results show that work zones, accounting for economic criteria and M&R treatments, will not overlap with condition-homogeneous sections in almost any case. Therefore, any aggregation of sections prior to condition prediction and M&R optimization significantly reduces the number of possible work-zone solutions and the potential for cost optimization. Nevertheless, the work zones generated by the LCC approach are still too short as compared to typical project lengths known from practice. The reason for this is that the algorithm currently does not account for a few practical considerations like for example combining two work zones with a short gap in between. However, a significant increase of work-zone length is to be expected, once the approach is fully developed. A detailed discussion on this issue is offered in Donev and Hoffmann (2018b).

Furthermore, for a thorough comparison of the two approaches in terms of costs, different levels of aggregation for homogeneous sections have to be examined (200 m, 600 m and 1000 m). Longer homogeneous sections will benefit from economies-of-scale effects, which will be reflected in the future costs, but will also cause greater loss of service life on individual sections, which will increase the discounted costs. Consequently, the discount rate has a considerable effect on the work-zone length in the LCC optimization. Figure 5.9 shows the distribution of lost service life for all 1000 road sections, resulting from the formation of work zones for the first M&R treatment. As expected, the average service-life loss increases with increasing length of homogeneous sections. As shown later, the average length of work zones determined by the LCC approach amounts to 714 m. However, the LCC approach is accountable for a smaller average service-life loss (4.6 years) compared to homogeneous sections with 600-m length (5.1 years). This indicates that the work zones based on homogeneous sections are not optimally located.

The annual agency costs, comprising M&R treatment costs and TTCC, are presented in Figure 5.10. The LCC approach results in the lowest annual costs with €0.79 million per year. The costs for 200-m homogenous sections are significantly higher (€1.21 million/year) in


Loss of Service Life - **LCC Optimization**
(No threshold violations; optimal work zones based on economies of scale)

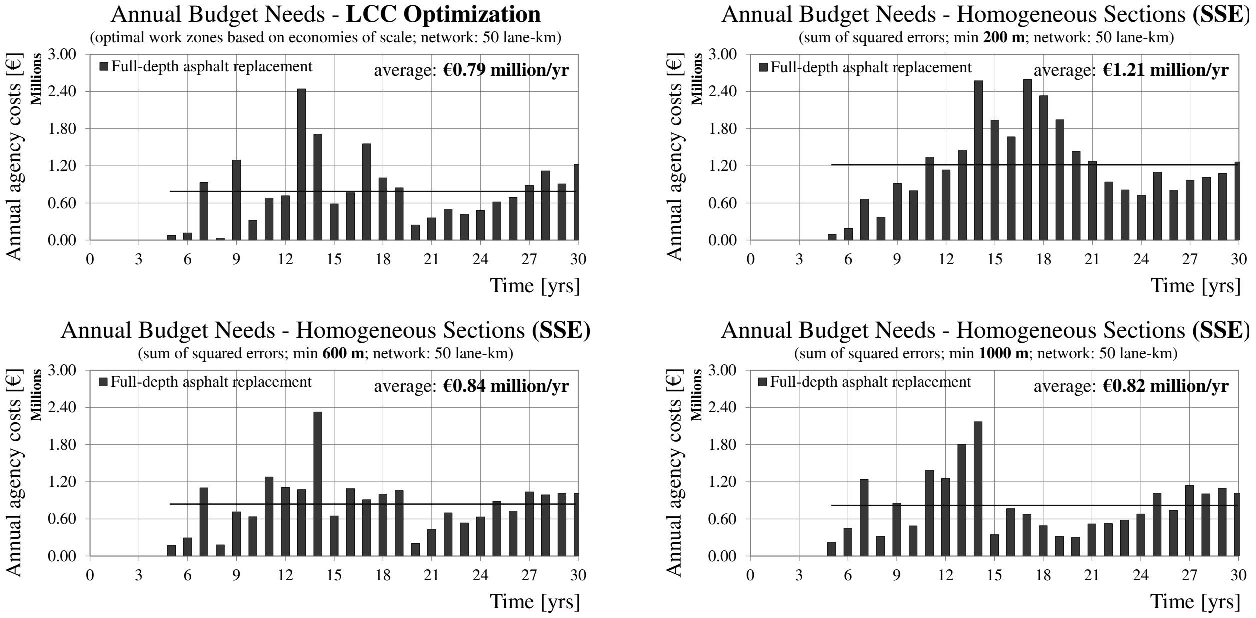
Loss of Service Life - Homogeneous Sections **(SSE)**
(No threshold violations; min sum of squared errors; min **200 m**)

Loss of Service Life - Homogeneous Sections **(SSE)**
(No threshold violations; min sum of squared errors; min **600 m**)

Loss of Service Life - Homogeneous Sections **(SSE)**
(No threshold violations; min sum of squared errors; min **1000 m**)

Note: Loss of service life is computed as a difference between service life for a work zone (based on homogeneous sections/LCC optimization) and service life for the short sections in the work zone. Short homogeneous sections (200 m) lead to a smaller loss of service life, but also result in very high fixed costs.

FIGURE 5.9 Comparison of loss of service life associated with work zones resulting from homogeneous sections with different level of aggregation (condition based) and work zones resulting from an LCC optimization (M&R based).

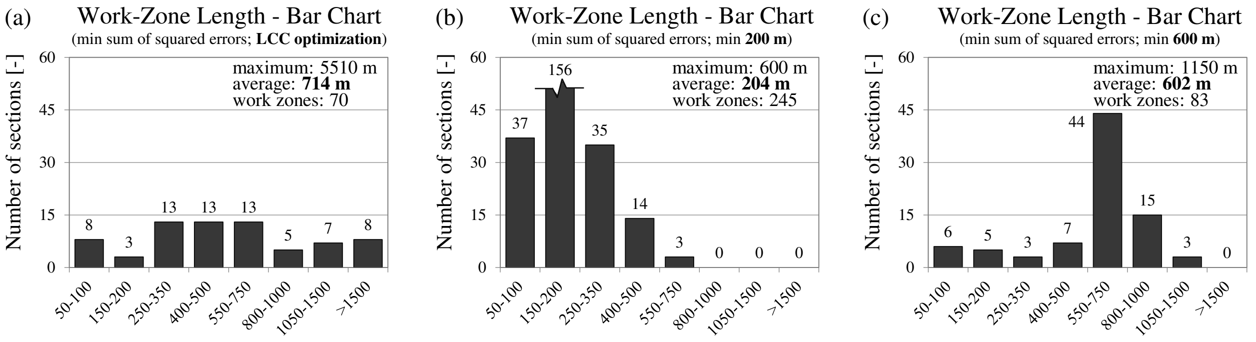


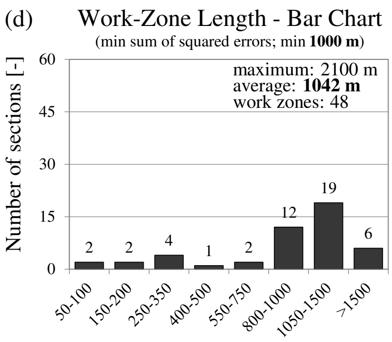
Note: Annual agency costs include M&R costs and TTCC (layout 2+1 or layout 3+1) for the first and all subsequent M&R treatments (asphalt replacement only) in the planning period of 30 years (no threshold violations allowed). The average annual costs are computed for the period from the 5th year to the 30th year.

FIGURE 5.10 Comparison of agency costs associated with work zones resulting from homogeneous sections with different level of aggregation and work zones resulting from an LCC optimization.

comparison to the other three cases due to higher fixed costs for short work zones. However, the difference between the LCC approach and homogeneous sections of 600 m and 1000 m turns out to be not very large. The reason for this is that the optimization objective is to minimize the total discounted costs. The summary table in Figure 5.11 reveals that the present value of the total costs in the case of LCC optimization is approximately 0.58 million (6%) lower than the second-best solution – 600-m homogeneous sections. Moreover, this difference can be attributed primarily to the TTCC, as these include high fixed costs, translating into more distinctive economies-of-scale effects. The present value in the case of 600 m is lower than the present value for 1000 m, suggesting that the optimum for homogeneous sections might be found between these two lengths if the relationship is convex. Figure 5.11 provides the distribution of the resulting work-zones lengths in each case.

The comparison between work zones based on homogeneous sections (condition) and work zones based on the new LCC approach (M&R costs) provides another perspective on the limitations of data-aggregation approaches. However, the full potential of the LCC approach lies in considering all possible combinations of M&R treatments in the life cycle. So, the cost savings are likely underestimated in this case study, as considering only one treatment alternative is a limiting factor for the selection of optimal treatment strategies. Moreover, for the purposes of this work, inaccurate predictions and restricted solution space are treated as independent effects from the generation of homogeneous section. In reality, these effects are superimposed with the accuracy of available performance models and the huge variation of the results depending on the applied sectioning algorithm and used condition survey if data from more than one campaign are available. Therefore, the overall effect – whether it is an overestimation or underestimation of service life and/or costs – will vary from case to case.





Summary - LCC Optimization vs. Homogeneous Sections

	LCC optimization (work zones)	Homogeneous sections - SSE (200 m)	Homogeneous sections - SSE (600 m)	Homogeneous sections - SSE (1000 m)
Average M&R costs [€]	663,277	690,027	688,982	712,837
Average total costs [€]	789,034	1,208,206	836,991	822,336
Σ PV M&R costs [€]	7,015,067	7,016621	7,388,307	7,769,997
Σ PV total costs [€]	8,533,426	13,014,968	9,116,603	9,124,680

Note: Total costs include M&R costs & TTCC. The computation of present values (PV) is based on 4% discount rate, considering any residual value at the end of the planning period (30 years).

FIGURE 5.11 Distribution of the length of work zones resulting from homogeneous sections with different level of aggregation and work zones resulting from an LCC optimization. Summary of average annual costs and total discounted costs with and without TTCC for each case.

5.5 CONCLUSIONS

The presented paper focuses on the aggregation of condition data to homogeneous sections as a major pitfall in most PMSs. Commonly used methods for generation of homogeneous sections are described and the consequences of their use are analyzed based on a parametric case study.

The aggregation of short survey sections to longer condition-homogeneous sections was historically motivated by the limited storage and processing capabilities of the computers at the time of the development of the first PMSs. Today, every common PMS still employs some form of section aggregation, disregarding the serious consequences of this approach. Even more surprising is the fact that these consequences are rarely discussed and not systematically investigated in literature. The results of the application of three common approaches for road segmentation provide proof that the aggregation to homogeneous sections leads to a substantial loss of information and smoothing of peak condition values. Moreover, homogeneous sections produce inaccurate predictions of service life and leading failure causes, impeding the selection of optimal timing and type of M&R treatments. In addition, the aggregation of sections hinders the development of section-level performance models based on historical condition data, as homogeneous sections shift with each survey. Since the aggregation is conducted prior to condition prediction and M&R optimization, the flexibility of selecting work zones is greatly reduced, as it is only possible to combine or partition existing homogeneous sections.

A possible solution to this problem is to use short survey sections for more accurate predictions and then to determine work zones based on economic criteria. In essence, the work-zone/LCC optimization is based on the trade-off between loss of service life and benefits gained by scale economies. The applicability of this LCC approach to large-scale networks has already been proven by previous research. This paper compares work zones resulting from homogeneous sections (condition based) and work zones resulting from LCC optimization (M&R based). Under the same boundary conditions of no threshold violations, the new LCC approach outperforms homogenous sections with different level of aggregation (200 m, 600 m and 1000 m) in terms of loss of service life and total discounted costs. Furthermore, the LCC approach can be extended to account for risk and consider other road assets like bridges and tunnels, minimizing total LCC, traffic-flow interruptions and other negative impacts of work zones. The extension of this new LCC approach, the validation of its perceived benefits and the evaluation of the practical applicability of the results calls for a future research based on real-world cost, traffic and condition data.

The conclusions of this research may be of use for road agencies in avoiding the described drawbacks of homogeneous sections and data aggregation. Applying the proposed LCC approach will improve the quality of condition prediction and the quality of project-level M&R treatment recommendations by making better use of the collected condition survey data. In summary, this will also allow for considerable efficiency gains regarding availability of the road network as well as substantial cost savings compared to common approaches.

REFERENCES

Alfelor, R.M. and McNeil, S., 1992. Definition of homogeneous segments on the basis of condition data: general approaches and specific application to rail maintenance. *Transportation Research Record: Journal of the Transportation Research Board*, 1341, 63-69.

Amaya-Gómez, R., Sánchez-Silva, M. and Muñoza, F., 2019. Integrity assessment of corroded pipelines using dynamic segmentation and clustering. *Process Safety and Environmental Protection*, 128, 284-294.

AASHTO, 1993. *AASHTO guide for design of pavement structures*. III-19–III-24, and Appendix J. Washington, D.C.: American Association of State Highway and Transportation Officials.

Bennett, C.R., 2004. Sectioning of road data for pavement management. *6th International Conference on Managing Pavements*, 19-24 October, Brisbane, Queensland, Australia.

Bennett, C.R. and Paterson, W.D.O., 2000. *A guide to calibration and adaptation* (The Highway Development and Management Series, Volume 5). Paris, France: World Road Association (PIARC).

Cafiso, S. and Di Graziano, A., 2012. Definition of homogenous sections in road pavement measurements. *Procedia - Social and Behavioral Sciences*, 53, 1069-1079.

Caltrans, 2015. *Automated pavement condition survey manual*. Sacramento, CA: California Department of Transportation.

Donev, V. and Hoffmann, M., 2018a. Comparison of optimization approaches for pavement maintenance and rehabilitation policies on road section and network level. *6th International Symposium on Life - Cycle Civil Engineering* (IALCCE2018), 28-31 October, Ghent, Belgium.

Donev, V. and Hoffmann, M., 2018b. Optimisation of pavement maintenance and rehabilitation activities, timing and work zones for short survey sections and multiple distress types. *International Journal of Pavement Engineering*. Advance online publication. doi: 10.1080/10298436.2018.1502433.

Donev, V., Hoffmann, M. and Blab, R., 2020. Benefit maximization based on aggregated condition indices: drawbacks for selection of pavement treatments. *International Journal of Pavement Engineering*. Advance online publication. doi: 10.1080/10298436.2020.1728533.

El Gendy, A. and Shalaby, A., 2008. Using quality control charts to segment road surface condition data. *7th International Conference on Managing Pavement Assets*, 23-28 June, Calgary, Alberta, Canada.

Haider, S.W. and Varma, S., 2016. Another look at delineation of uniform pavement sections based on falling weight deflectometer deflections data. *Canadian Journal of Civil Engineering*, 43 (1), 40-50.

Hoffmann, M., 2006. *Instandsetzung von Strassen* [Rehabilitation of roads]. Dissertation (PhD). Technische Universität Graz (in German).

Hoffmann, M. and Donev, V., 2018. Introduction of a new continuous time and state space stochastic process in condition prediction. *International Journal of Pavement Engineering*, 19 (4), 339-354.

Hudson, W.R., Rada, G.R. and Martin, T., 2011. Balancing pavement management, software requirements, and data collection details. *8th International Conference on Managing Pavement Assets* (ICMPA8), 15-19 November, Santiago, Chile.

Kennedy, J., Shalaby, A. and Van Cauwenberghe, R., 2000. Dynamic segmentation of pavement surface condition data. *3rd Transportation Specialty Conference*, 7-10 June, London, Ontario, Canada.

Kunze, A. and Rübensam, J., 2007. *Erarbeitung eines Verfahrens zur Bildung von Erhaltungsabschnitten für das Erhaltungsmanagement (PMS) auf Basis von Zustands- und Aufbaudaten* [Development of a process to generate maintenance sections for the maintenance management based on condition data and pavement structures] (Forschung Straßenbau und Straßenverkehrstechnik, Heft 972). Bonn: Bundesministerium für Verkehr, Bau und Stadtentwicklung (in German).

Latimer, D.P., Manoharan, S.R., Robertson, N.F. and Nata-atmadja, A., 2004. A study of the impact of road segmentation schemas on predicted maintenance investment outcomes. *6th International Conference on Managing Pavements*, 19-24 October, Brisbane, Queensland, Australia.

Lea, J., Harvey, J. and Tseng, E., 2014. Aggregating and modeling automated pavement condition survey data for flexible pavements for use in pavement management. *Transportation Research Record: Journal of the Transportation Research Board*, 2455, 89-97.

Misra, R. and Das, A., 2003. Identification of homogeneous sections from road data. *International Journal of Pavement Engineering*, 4 (4), 229-233.

Ping, W.V., Yang, Z., Gan, L. and Dietrich, B., 1999. Development of procedure for automated segmentation of pavement rut data. *Transportation Research Record: Journal of the Transportation Research Board*, 1655, 65-73.

Rübensam, J. and Schulze, F., 1996. *Entwicklung einer Methodik zur zweckmäßigen Zusammenfassung maßnahmebedürftiger Abschnitte der BAB-Betriebsstrecken auf der Grundlage von Zustands- und Bestandsdaten* [Development of a method for appropriate merging of sections of the federal highways based on condition and inventory data] (Forschung Straßenbau und Straßenverkehrstechnik, Heft 736). Bonn: Bundesministerium für Verkehr (in German).

Thomas, F., 2004. Generating homogeneous road sections based on surface measurements: available methods. *2nd European Pavement and Asset Management Conference*, Berlin, Germany.

Thomas, F., 2005. Automated road segmentation using a Bayesian algorithm. *Journal of Transportation Engineering*, 131 (8), 591-598.

Weninger-Vycudil, A., Simanek, P., Haberl, J. and Rohringer, T., 2009. *Handbuch Pavement Management in Österreich 2009* [Manual pavement management in Austria 2009] (Straßenforschung, Heft 584). Wien: Bundesministerium für Verkehr, Innovation und Technologie (in German).

CHAPTER 6

Benefit maximization based on aggregated condition indices: drawbacks for selection of pavement treatments

by Valentin Donev, Markus Hoffmann & Ronald Blab

in *International Journal of Pavement Engineering*
Published online: 24 Feb 2020 (open access)
DOI: 10.1080/10298436.2020.1728533

ABSTRACT

Road agencies employ different criteria and methods to compare treatment alternatives and develop maintenance and rehabilitation (M&R) programs for their networks. The maximization of benefits (effectiveness) has been integrated in several leading pavement management systems and used for network-level programming since the late 1980s. The approach defines benefit as the area between the performance curves with and without M&R actions based on an aggregated index, representing the overall pavement condition. Further simplifications like the screening of solutions using an efficiency frontier and the incremental benefit-cost technique have made it possible to apply the method with the available computing power at the time. Based on a case study of 1000 road sections (50 km route), this paper analyses the effect of benefit maximization on annual budget, condition and treatment type distributions, trigger values and remaining life. M&R programs minimizing costs and maximizing benefits are compared at the project and network level for various budget scenarios. The results show that the maximization of benefits based on an aggregated condition index leads to substantially higher agency costs and favors the selection of expensive treatments with earlier timing, irrespective of actual failure causes. The conclusions of this work may be useful to road agencies in developing more efficient budget-allocation polices at the network level.

KEYWORDS: pavement management, life cycle costs, optimization methods, incremental benefit-cost ratio, area between the curves, benefit maximization, composite condition index.

6.1 INTRODUCTION

Network-level pavement management systems (PMSs) have been in operation since early 1980s in the USA and around 2000 in German-speaking countries, assisting road agencies with the development of maintenance and rehabilitation (M&R) programs, cost estimations and budgeting. An overview of typical modules and workflow in a widely used PMS together with the different stages of data processing and analysis are illustrated in Figure 6.1. The entire process is based on aggregating data from periodic condition surveys with treatment impact and costs into a comprehensive decision-making criterion on the foundation of cost-effectiveness analysis. The results of the surveys are reported for short survey sections (50 m) which are then aggregated to condition-homogeneous sections and management sections to facilitate the processing and analysis of the data (Weninger-Vycudil et al. 2009, Lea et al. 2014, Haas et al. 2015).

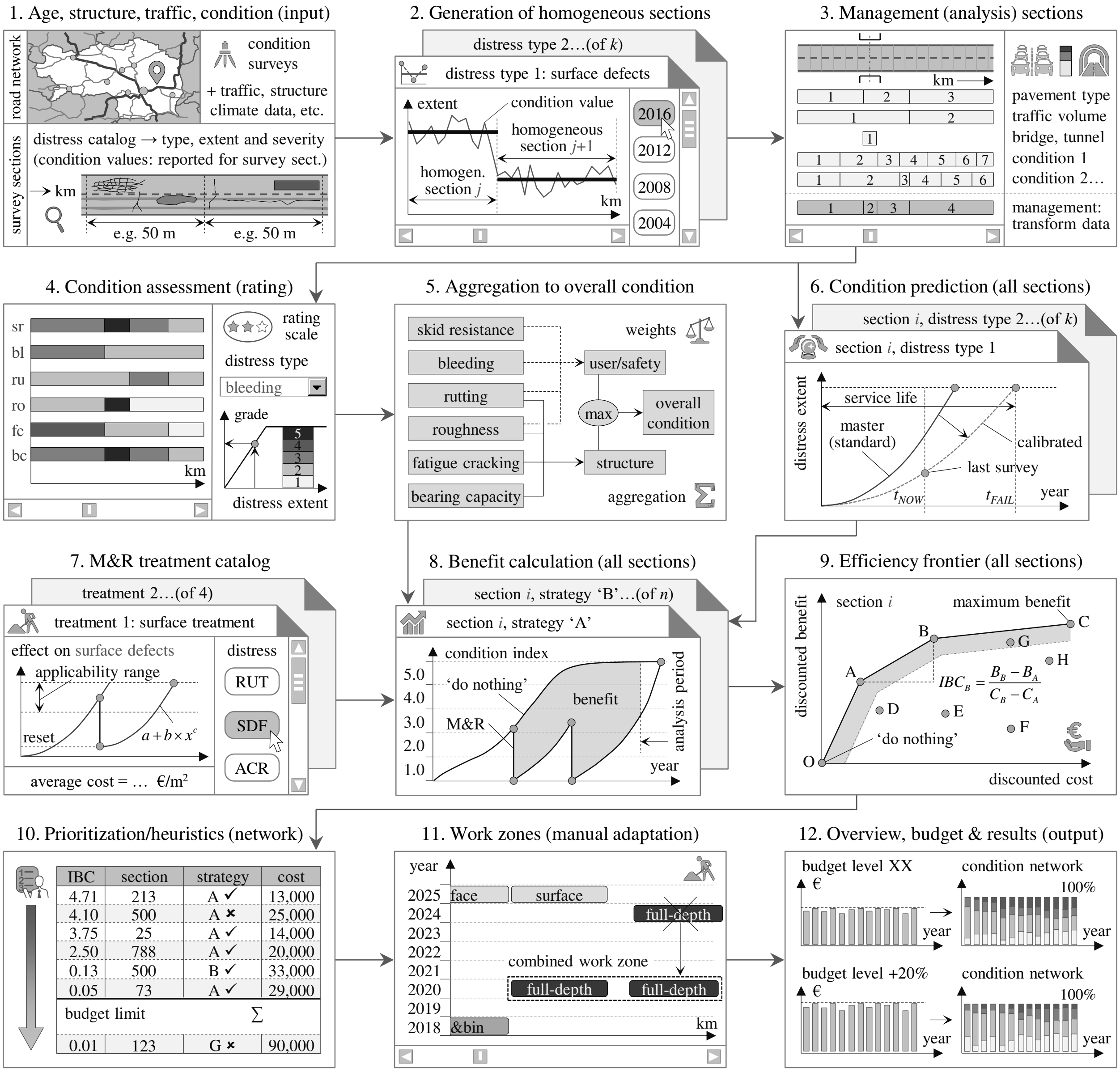


FIGURE 6.1 Overview of data-processing flow and sequence of computations from input to output in a common PMS.

For these management sections, different M&R strategies are evaluated and compared in terms of costs and benefits. Benefit (or effectiveness) is defined as the area enclosed between the do-nothing performance curve and the post-treatment curve. Sometimes this area is multiplied by traffic volume and section length. To obtain a one-dimensional measure of benefits, the different distress types have to be aggregated into a composite condition index like, e.g. the overall condition index (OCI) or the present serviceability index (PSI). At the network level, heuristic methods such as the incremental benefit-cost (IBC) technique, ranking approaches (worst first) or mathematical programming are usually employed for the final selection of M&R strategies (Haas et al. 2015).

This paper examines the effect of setting different network-level objectives (minimize life cycle costs vs. maximize benefits) on the resulting budget needs, condition distribution, remaining life and asset value for various budget scenarios. It is also investigated, if the different objectives favor specific M&R treatment types or trigger values (early vs. late timing). The comparison is based on numerical examples under ceteris paribus conditions, allowing to separate and quantify the impact of distinct computational steps in PMSs on the results. The paper focuses on the benefit-maximization approach due to its wide application in existing PMSs used by road agencies and the lack of comprehensive discussion in the literature. A comparison to other (more advanced) optimization methods mainly from the literature is beyond the scope of this paper.

The following section (Section 6.2) provides an overview of selected literature on benefit maximization with description of the method. The parametric case study, performance and cost models in Section 6.3 define the basis for application and comparison of the investigated optimization approaches. Section 6.4 presents network-level M&R programs developed for different optimization goals and various budget scenarios, together with the resulting condition distributions and annual budget needs. The last section contains the conclusions.

6.2 LITERATURE AND METHODS

The pavement management optimization problem can be formulated for single sections (project level) and at the network level. At the project level, the goal is to determine the temporal sequence of M&R treatments (timing and type) that minimizes agency costs or maximizes benefits for a given time period. At the network level, however, it may not always be possible to implement the optimal strategy for each section due to budget restrictions. A sophisticated network-level approach should also consider interdependencies between different road sections and other assets in terms of economies-of-scale costs, capacity utilization and work-zone effects. Hence, the network level problem is a very complex one, with three categories of approaches being available: ranking, heuristic techniques (e.g. IBC) and optimization.

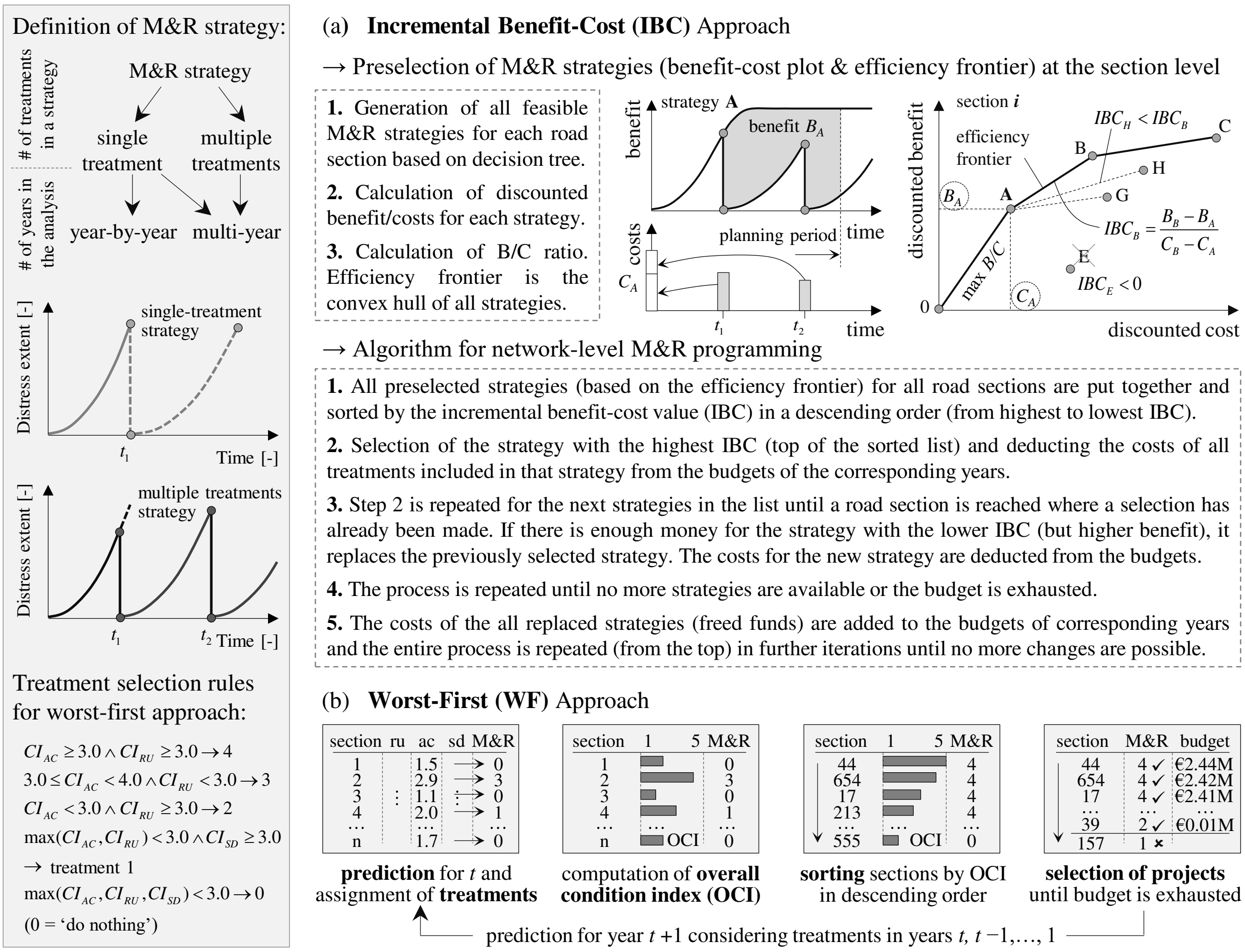
Ranking is very simple and intuitive but also the most rudimentary approach for development of M&R programs at the network level which can be applied to a small road network even without a computer. Apart from their historical significance for pavement management, ranking methods are still being used by some agencies at regional level and in developing countries. Ranking of road sections by current condition, also known as worst-first (WF) approach, is the most common ranking method. The decision which road sections should be repaired in any given year is made according to the current condition, assigning treatments first to the worst-performing section and subsequently moving down in the sorted list until the budget is exhausted

(FHWA 1998, Javed 2011, Menendez et al. 2013). To include other decision criteria, several authors have developed composite priority indices, considering multiple factors like condition, traffic volume, functional class, drainage, etc. (e.g. Shah et al. 2014). The WF approach does not require condition prediction and does not consider treatment effects. Moreover, the method is suited to compare different projects (sections) but not different treatment alternatives for the same project (only "where", not "what/when"). Therefore, the M&R type is usually determined based on trigger rules or decision trees. In this work, the WF approach is used to develop a multi-year M&R program on a year-by-year basis, providing a basis for comparison with other methods (see Figure 6.2(b)). Using the predicted individual distress types in a given year, an M&R strategy, including the do-nothing alternative, is determined for each section. The sections are then sorted according to OCI and the highest-ranked projects that would still fit the budget are selected. The procedure is repeated sequentially for each year in the planning period, whereby the planning period is the time period in which M&R treatments can be planned.

More advanced incremental benefit-cost (IBC) or marginal cost-effectiveness (MCE) methods have been implemented in a few commercial off-the-shelf software products and used by many road agencies for decades. The history of the IBC approach goes back to the early years of pavement management (Juster and Pecknold 1976, Haas et al. 1985, Shahin et al. 1985). The method is integrated in pavement management systems like dTIMS - Deighton total infrastructure management system (Keleman et al. 2008, Zavitski et al. 2008), the Austrian VIAPMS (Weninger-Vycudil 2001, Weninger-Vycudil et al. 2009) and the PMS software in Germany (Maerschalk and Socina 2008), with the latter two being based on dTIMS. The highway development and management model (HDM-IV) uses the IBC concept too, but relies on monetary benefits and user costs (Kerali and Mannisto 1999, Odoki and Kerali 2000). According to MDOT (2019), the highway pavement management application (HPMA) also employs the MCE approach for selection of treatments. Furthermore, the IBC method is used in bridge management as well (e.g. Farid et al. 1994, Thomson et al. 1998).

The general idea of the IBC method is to consider a predefined set of M&R strategies for each road section and select the ones that maximize the sum of individual benefits over the planning period (typically 10 years), subject to annual budget constraints (Haas et al. 1994, AASHTO 2012). Thus, the method is superior to simple ranking, since it considers multiple treatment alternatives for the same section and future benefits in the network-level analysis. Multi-year M&R programs can be developed using either year-by-year or multi-year analysis. If a strategy consists of multiple treatments, however, only multi-year analysis can be conducted, as such strategies have impact on multiple budget constraints.

Figure 6.2(a) presents a rather verbal description of the methodology of the IBC approach. For each section different treatment strategies are generated using trigger rules. Although the effect of a treatment on each distress type is usually modelled and predicted separately, the individual distress types are ultimately combined into one composite condition index (like OCI), allowing for a single-objective formulation of the optimization problem. The number of generated M&R strategies varies between sections and may become considerably large depending on the number of different treatments, trigger rules and the length of the planning period. Hence, before conducting the network-level evaluation, the number of possible strategies for each section has to be limited with the intention to reduce computing effort. This is done by using the efficiency frontier, that is, the convex envelope of all points (strategies) in the benefit-cost plot. IBC value is computed for each strategy as the ratio of difference in benefits to difference in



1. Generation of all feasible M&R strategies for each road section based on decision tree.

2. Calculation of discounted benefit/costs for each strategy.

3. Calculation of B/C ratio. Efficiency frontier is the convex hull of all strategies.

1. All preselected strategies (based on the efficiency frontier) for all road sections are put together and sorted by the incremental benefit-cost value (IBC) in a descending order (from highest to lowest IBC).

2. Selection of the strategy with the highest IBC (top of the sorted list) and deducting the costs of all treatments included in that strategy from the budgets of the corresponding years.

3. Step 2 is repeated for the next strategies in the list until a road section is reached where a selection has already been made. If there is enough money for the strategy with the lower IBC (but higher benefit), it replaces the previously selected strategy. The costs for the new strategy are deducted from the budgets.

4. The process is repeated until no more strategies are available or the budget is exhausted.

5. The costs of the all replaced strategies (freed funds) are added to the budgets of corresponding years and the entire process is repeated (from the top) in further iterations until no more changes are possible.

Treatment selection rules for worst-first approach:

$CI_{AC} \geq 3.0 \wedge CI_{RU} \geq 3.0 \rightarrow 4$

$3.0 \leq CI_{AC} < 4.0 \wedge CI_{RU} < 3.0 \rightarrow 3$

$CI_{AC} < 3.0 \wedge CI_{RU} \geq 3.0 \rightarrow 2$

$\max(CI_{AC}, CI_{RU}) < 3.0 \wedge CI_{SD} \geq 3.0$ → treatment 1

$\max(CI_{AC}, CI_{RU}, CI_{SD}) < 3.0 \rightarrow 0$

(0 = 'do nothing')

FIGURE 6.2 Common methods for network-level project prioritization and budget allocation in PMS: incremental benefit-cost approach (a) and worst-first approach (b).

costs between the strategy in question and a reference strategy. Only strategies on the efficiency frontier or within a tolerance margin are ranked as cost-effective and kept for further analysis.

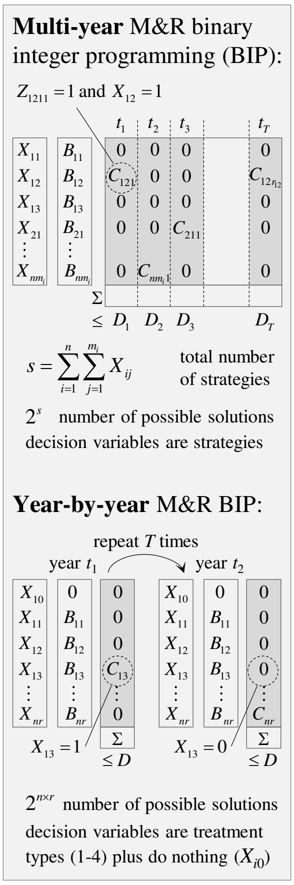
At the network level, all strategies for all road sections are pooled together and ranked according to their IBC value. In contrast to ranking methods, however, the final program usually does not consist of the sections on top of the list. In fact, the strategies at the bottom of the list provide the highest benefits, but their selection depends on the available budget(s). Due to the presence of an optimization objective (maximization of benefits), the IBC method may be classified as a heuristic technique for solving an optimization problem. The algorithm for network-level prioritization given in Figure 6.2(a) can be found with some minor variations in the literature (Haas et al. 1994, FHWA 1998, Weninger-Vycudil 2001, Peng and Ouyang 2010, AASHTO 2012).

The IBC formulation can be solved exactly using binary integer mathematical programming (BIP). Integer formulations of the M&R optimization problem are common in pavement management (Haas et al. 1994, Li et al. 1998, Ferreira et al. 2002, Wang et al. 2003, Scheinberg and Anastasopoulos 2010). A multi-year binary integer program, minimizing life cycle costs or maximizing benefits is presented in Figure 6.3(a). Each strategy is represented by a decision variable, taking the value of "1" if selected and "0" otherwise. The sum of the decision variables for all strategies on a given section must equal "1", which guarantees that only one strategy can be selected per section. The resulting total costs in each year are limited by the budget constraint.

In contrast to heuristic methods, BIP considers all possible combinations of strategies. If the cost and benefits of the optimal and suboptimal M&R strategies for each section are determined prior to the network level optimization, then the optimization problem is linear. Linear BIP can be solved efficiently for a relatively large number of decision variables (e.g. hundreds of thousands), using, for example, branch-and-cut algorithms (Hillier and Lieberman 2015).

Multi-year programs can also be developed on a year-by-year basis, using a single-year BIP formulation, as shown in Figure 6.3(b). This approach is based on single-treatment strategies (see Figure 6.2), with the different treatment types, including "do nothing", serving as decision variables. Prior to each consecutive optimization run, feasible M&R treatments and benefits (condition prediction) for each alternative must be calculated based on the current condition. The single-year problem will be, in general, easier to solve, but the procedure must be repeated for each year in the planning period. Moreover, the solution of a single-year program will be inferior to the solution of multi-year programming, since it does not simultaneously consider all possible combinations of treatments and potential trade-offs in the life cycle.

In this paper, the analysis focuses on the maximization of only non-monetary benefits defined as the area between the curves. The reason for this is that common PMSs usually do not account for economic benefits in terms of user and environmental costs savings (Wu and Flintsch 2009). However, instead of maximizing the area between the curves (condition), standard approaches can be extended to maximize economic benefits (cost savings). A possible formulation is to maximize user costs savings due to improved condition in comparison to a do-nothing or a do-minimum (base) strategy minus work-zone related user costs (Brozek et al.



(a) **Multi-year** M&R Programming (multiple years simultaneously):

Maximize $\sum_{i=1}^{n}\sum_{j=1}^{m_i} X_{ij} \times B_{ij}$ (objective function maximum total benefits)

Minimize $\sum_{i=1}^{n}\sum_{j=1}^{m_i} X_{ij} \times G_{ij}$ (alternative objective minimum discounted costs)

subject to

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} X_{ij} \times (\sum_{k=1}^{r_{ij}} Z_{ijkt} \times C_{ijk}) \leq D_t \qquad \forall t = 1,2,...,T$$

(budget constraint)

$$\sum_{j=1}^{m_i} X_{ij} = 1 \qquad \forall i = 1,2,...,n$$

(one strategy per section)

and X_{ij} is binary (decision variable)

Z_{ijkt} is binary (auxiliary variable)

where

$$X_{ij} = \begin{cases} 1, & \text{if M\&R strategy } j \text{ on section } i \text{ is selected} \\ 0, & \text{otherwise} \end{cases}$$

$$Z_{ijkt} = \begin{cases} 1, & \text{if } k^{\text{th}} \text{ treatment of strategy } j \text{ on section } i \text{ is in year } t \\ 0, & \text{otherwise} \end{cases}$$

Legend

B_{ij}...	discounted benefit resulting from strategy j on section i
G_{ij}...	total discounted costs of strategy j on section i
m_i...	number of feasible strategies for section i
r_{ij}...	number of treatments included in strategy j for section i
C_{ijk}...	actual cost of the k^{th} treatment of strategy j for section i

n...	total number of sections
D_t...	budget for year t
i...	road section $i = 1,\ldots,n$
j...	strategy $j = 1,\ldots,m_i$
k...	treatment $k = 1,\ldots,r_{ij}$

(b) **Year-by-year** M&R Programming (each consecutive year separately):

Maximize $\sum_{i=1}^{n}\sum_{j=1}^{r} X_{ij} \times B_{ij}$ (objective maximum benefit)

subject to

$$\sum_{i=1}^{n}\sum_{j=1}^{r} X_{ij} \times C_{ij} \leq D$$

(budget constraint)

$$\sum_{j=1}^{r} X_{ij} = 1 \qquad \forall i = 1,2,...,n$$

(one strategy per section)

and X_{ij} is binary (decision variable)

Feasible treatment alternatives - trigger rules (based on individual distress types):

treatment 3/4: $CI_{SD} \geq 3.0 \vee CI_{RU} \geq 3.0 \vee CI_{AC} \geq 3.0$

treatment 2: $(CI_{SD} \geq 3.0 \vee CI_{RU} \geq 3.0) \wedge CI_{AC} < 3.0$

treatment 1: $CI_{SD} \geq 3.0 \wedge CI_{RU} < 3.0 \wedge CI_{AC} < 3.0$

do nothing: $CI_{SD} < 3.0 \wedge CI_{RU} < 3.0 \wedge CI_{AC} < 3.0$

Legend

X_{ij}...	treatment j on section i applied in the year of analysis
B_{ij}...	benefit of application of treatment j on section i

r...	total number of M&R treatments
i...	road section $i = 1,\ldots,n$

FIGURE 6.3 Formulation of the network-level pavement management optimization problem as a binary integer program, simultaneously considering multiple budget years (a) or only a single year (b).

2009). In addition, also the increased agency cost in comparison to base alternative can be considered in the objective function (Odoki and Kerali 2000). Both formulations are only possible as alternative to the benefit maximization based on aggregated condition indices. It is possible, however to conduct a multi-objective optimization with two or more objectives, one including user costs and one being the maximization of benefits based on area between the curves and select a solution based on Pareto frontier (e.g. Yu et al. 2015).

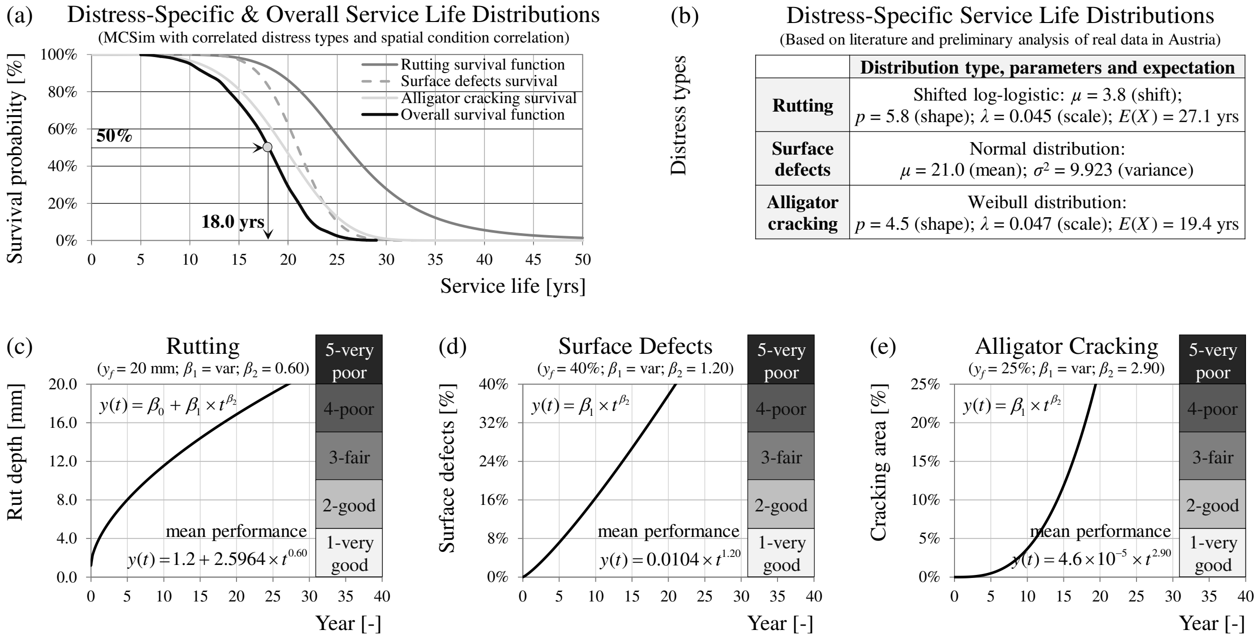
In general, there is a gap in the literature regarding the systematic comparison and discussion of the approaches for condition rating, condition prediction and M&R optimization employed in PMSs. Kuhn (2012) compared optimal M&R policies developed using individual distress types versus a composite condition index and criticized any other use of aggregated indices in pavement management other than for communication with high-level decision makers. Gharaibeh et al. (2010) compared six different aggregated indices based on distress data and concluded that even seemingly similar indices lead to significantly different rating of the same pavement sections. In the literature there are also several studies that evaluated different solution algorithms for the benefit-maximization problem (e.g. Yoo and Garcia-Diaz 2008, Peng and Ouyang 2010, Patidar et al. 2011). In contrast, this paper questions the suitability of maximizing benefits based on the area between the curves as an objective in pavement management.

6.3 CASE STUDY

6.3.1 Service life and distress correlation

This paper presents a comprehensive analysis of the consequences arising from using benefit maximization strategies in pavement management. Selected commonly-used optimization approaches are applied to a case study of a simulated road network. The individual computational steps in the following analysis, the cost and performance models resemble the typical process in common PMSs. Moreover, each parameter of the case study is estimated based on real-world data, with more details on the calibration being given in Donev and Hoffmann (2018a, 2018b). A simulation approach provides exact knowledge of service life and condition distribution at any point in time, avoiding limitations of empirical data like censoring and ex ante prediction. Therefore, complete data allows researchers to isolate, objectively quantify and compare the deviations resulting from the application of a specific approach.

The case study is based on 1000 road sections, each with a length of 50 m, which is the standard for survey sections in Austria and a width of 3.75 m, resulting in a total length of 50 lane-km. All inputs are calibrated and conceptualized for flexible pavements on freeways in Austria, even though the generated sections are not tied to a specific location or a specific Austrian freeway. Figure 6.4(a,b) provides an overview of the three considered distress types (rutting, surface defects and alligator cracking), together with their respective service life distributions. Initial service life is defined as the time from construction until exceedance of a given distress-specific condition threshold. At the section level, the first distress exceeding its threshold determines the overall service life, the cause for failure (series system) and the latest possible timing of treatment application (boundary condition). In this work, a trigger value is defined as the condition or condition index at the time of the treatment application. If pavement condition is constrained only to non-failing states, the trigger value can be less than or equal to the threshold value, where a lower value corresponds to a better condition. It must be noted that



	Distribution type, parameters and expectation
Rutting	Shifted log-logistic: $\mu = 3.8$ (shift); $p = 5.8$ (shape); $\lambda = 0.045$ (scale); $E(X) = 27.1$ yrs
Surface defects	Normal distribution: $\mu = 21.0$ (mean); $\sigma^2 = 9.923$ (variance)
Alligator cracking	Weibull distribution: $p = 4.5$ (shape); $\lambda = 0.047$ (scale); $E(X) = 19.4$ yrs

FIGURE 6.4 Overall and distress-specific service life distributions (a, b) and performance functions (c, d, e) employed in the case study.

the selection of thresholds and trigger values directly impacts the observed service life in practice and restricts the solution space in the optimization (e.g. Ford et al. 2012). Much like the development of aggregated indices, the definition of condition threshold is often based solely on subjective expert opinions. This drawback can be mitigated, for example by using soft constraints, user costs or condition-dependent treatment life (Donev and Hoffmann 2018b).

The service lives for these three distress types are positively correlated (distress correlation), as they are influenced by common factors like traffic loading. A set of three service lives (for each distress) is generated for each of the 1000 road sections. Performance functions with fixed power parameter (β_2) are fitted to the individual service lives by varying the scale parameter (β_1) to describe intermediate conditions. The employed power functions, as well as the corresponding rating scales for the computation of OCI are presented in Figure 6.5(c,d,e). The scale for OCI ranges from 1.0 (very good) to 5.0 (very poor).

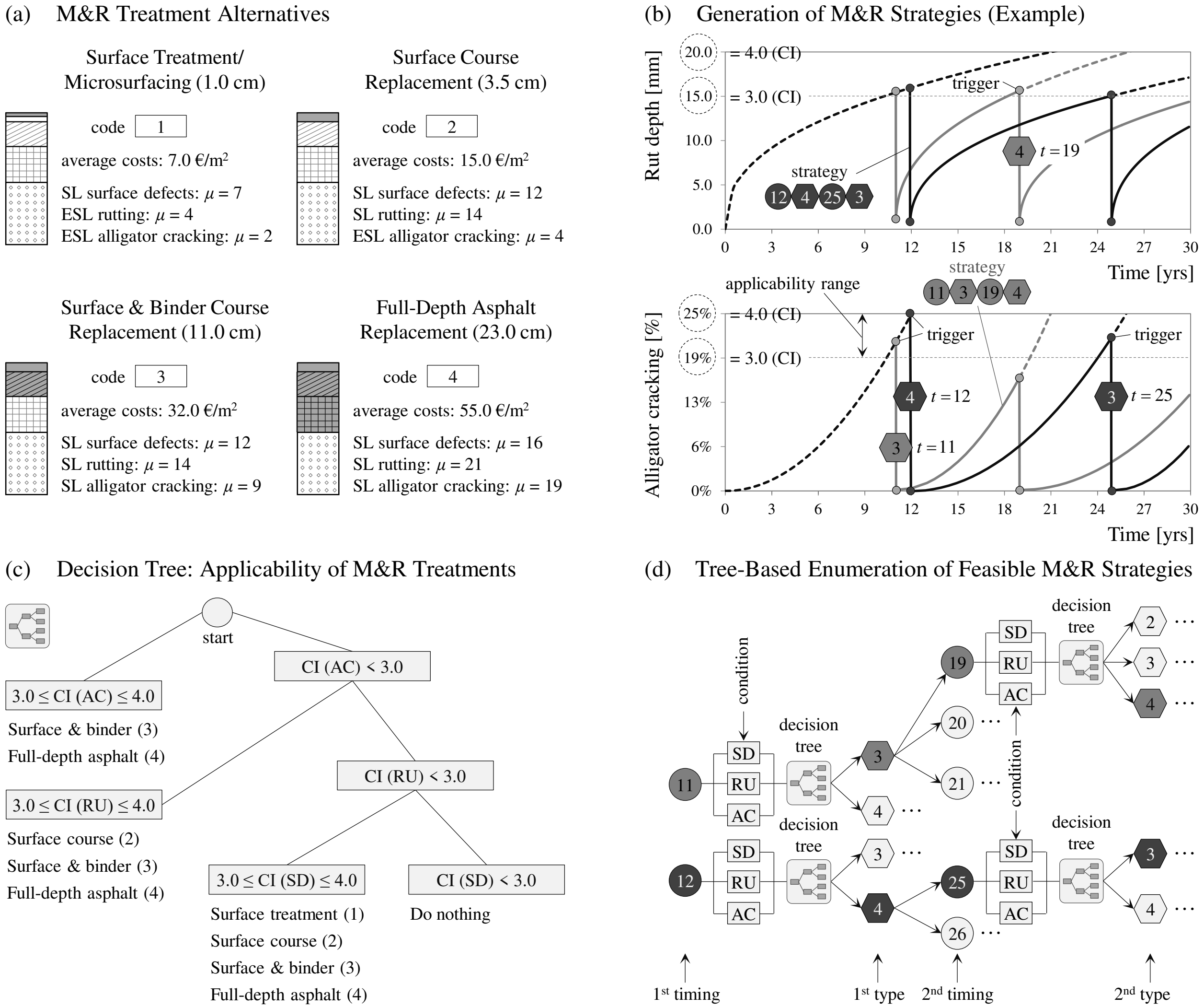
More details on initial and post-treatment (see Section 6.3.2) service life distributions, on the estimated correlations and on the data generation procedure are offered in Donev and Hoffmann (2018b). The focus here is not on the derivation of cost and performance models but on the definition of a realistic case study as basis for a comparison of different treatment selection and optimization approaches. Different parameter values or distribution forms may change the exact numeric results more or less, but will not change the nature of the conclusions of this work, regarding the methodological drawbacks of the analyzed approaches (Section 6.4).

6.3.2 Cost and performance modelling

The application of methods for M&R prioritization and optimization to the case study data also requires the definition of treatment types, post-treatment service lives and treatment costs. Four different treatment types for asphalt pavements are considered, ranging from preventive maintenance (surface treatment) to full-depth asphalt replacement (see Figure 6.5(a)). Each treatment is represented by an average unit cost, a distress-specific service life and an integer code from 1 to

4 for modelling purposes. The average unit cost estimates are based on analysis of empirical data from Austria and provided in 2016 constant euros. Furthermore, the average unit costs are computed for typical M&R project lengths. Economies-of-scale costs are not considered here to allow more clear interpretation of the results (cf. Irfan et al. 2012, Donev and Hoffmann 2018b, Qiao et al. 2019). The effect of a treatment on each distress type is modelled with distinct distribution. For the case study, it is assumed that treatment life follows a normal distribution with a mean value (μ) based on literature analysis (e.g. Hall et al. 2001, Cuelho et al. 2006, Anastasopoulos and Mannering 2015, Birbaum 2016, Nobakht et al. 2016). Similar to the initial service life, treatment life is defined as the time from the application of the treatment until the exceedance of the same distress-specific threshold. The effect “extension of service life” (ESL) models a treatment effect which delays distress progression for a short period of time, without correcting the underlying problem.

The technical applicability of treatments is determined using the decision tree depicted in Figure 6.5(c). The decision tree determines one or more treatment alternatives for any possible



Legend

ESL… extension of service life (SL) CI… condition index AC… alligator cracking RU… rutting SD… surface defects

Note: Service life of all treatments follows a normal distribution with μ and $\sigma = 0.15\times\mu$ for surface defects and cracking and $\sigma = 0.20\times\mu$ for rutting.

FIGURE 6.5 M&R treatment alternatives with average costs and service life (a), technical applicability based on decision tree (c) and generation of feasible combinations of treatment type and timing (M&R strategies) for a given road section (b, d).

combination of distresses and their extent or severity. For consistency reasons, the extent is represented by a condition index (CI), which is a linear transformation to an uniform scale (see Figure 6.4(c,d,e) for employed rating scales). For example, if the CI for alligator cracking is within the range from 3.0 to 4.0, two structural treatments (treatment 3 and 4) are feasible. The decision tree does not allow exceedance of condition thresholds for the individual distress types ($CI \leq 4.0$), being an important boundary condition in the optimization. If the condition indices of all distress types fall below 3.0, no M&R treatments are planned.

At the road section level, different temporal sequences of treatments (M&R strategies) are possible. Each feasible strategy defined by timings (integer) and types (integer) has to be evaluated in terms of cost and benefits. The generation of treatment strategies is based on an enumeration algorithm to ensure that all possible combinations are evaluated. A simplified example is presented in Figure 6.5(b,d). The first year in which a treatment can be planed is year 11, as this is the first year with a CI exceeding 3.0. Feasible treatment alternatives are then determined for the predicted combination of distresses in year 11 using the decision tree. In the present example, only treatments 3 and 4 are viable alternatives, as the CI for alligator cracking exceeds 3.0. The selected treatment (in this case treatment 3) determines the performance after year 11 and the timing range for the next treatment (from year 19 to 21). If the next treatment is planned in year 19, three treatment alternatives become feasible according to the decision tree. Thus, the ordered sequence of pairs of treatment's timing and type (11, 3; 19, 4) completely defines the strategy (in dark grey). Figure 6.5(b,d) traces a second possible strategy (in black), starting from year 12 (i.e. 12, 4; 25, 3).

6.4 CONSEQUENCES OF APPLYING BENEFIT-MAXIMIZATION STRATEGIES

This section presents the results from the application of the methods described in Section 6.2 to the case study data. First, all possible M&R strategies are generated for each of the 1000 road sections based on a decision tree, trigger rules and service lives (see Figure 6.5). The generation of strategies, as well as the computation of benefits and costs for each strategy is conducted automatically using a small add-in for Microsoft Excel written in Visual Basic for Applications (VBA) programming language. The number of single treatments in a strategy results from the requirement to always keep the condition within acceptable limits during the planning period. The planning period is set to 30 years as a compromise between typical periods for IBC analysis of 10-20 years and the aim of gaining some insights about the long-term effects of the M&R strategies. In any case, a true life cycle cost analysis in PMS will require a longer planning horizon (e.g. at least 50 years). A discount rate of 4% was selected for consistency with previous research. The case-study data produces 300 M&R strategies per section on average or approximately 300,000 strategies in total. Figure 6.6 provides a general overview of the computational steps in the case study, with the comparison taking place at the section level, at the network level with unlimited budget and, finally, at the network level with budget constraints. A comparison of M&R programs without budgetary restrictions has little practical relevance with regard to the reality of limited agency budgets. Nevertheless, it is important from a theoretical point of view as it allows conclusions about the suitability of the defined optimization goals.

In the case study, four different objectives found in the literature are evaluated: maximization of benefits with and without residual benefits and minimization of costs with and without residual value. Residual benefit (RB) is the additional benefit area associated with a specific period (in this case 20 years) beyond the end of the planning period of 30 years (Weninger-Vycudil et al. 2009). The residual value (RV) at the end of the planning period is computed as a ratio of remaining life to service life of the last planned treatment multiplied by the cost of this treatment. RV is discounted and subtracted from the total life-cycle costs (Al-Qadi et al. 2009, Zuniga-Garcia et al. 2018).

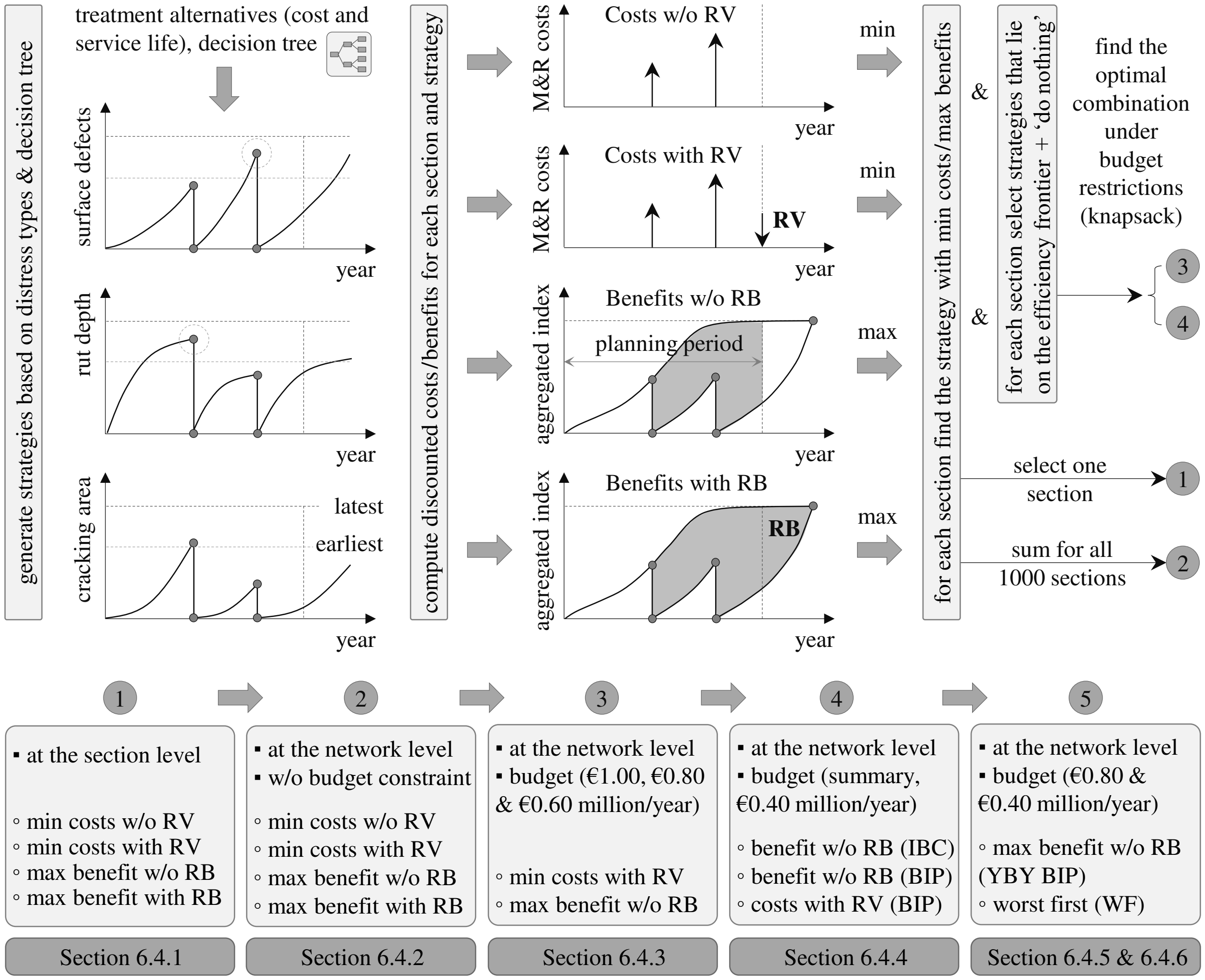


FIGURE 6.6 Overview of major computational steps in the analysis as well as different levels of comparison and associated settings (1-5).

6.4.1 Maximizing benefits/minimizing costs at the section level

The optimal strategies at the section level are found by a simple lookup in the tables of the generated strategies. If the budget constraint is not binding, the solution at the network level consists of the optimal solutions for each individual section. The solution of the unconstrained problem is also the overall best solution that can be achieved for the defined objective. The optimal strategies for a road section randomly-selected for illustration purposes (e.g. section 55) are displayed in Figure 6.7. Figure 6.7(a) shows the expenditure stream and benefit of the strategy that minimizes total discounted costs without RV. This strategy minimizes the costs strictly in the planning period without accounting for the network condition afterwards. Hence, such strategies may be of interest to investors in public-private partnership (PPP) projects. In this case, a small maintenance treatment in year 29 ensures that condition thresholds are not violated and the pavement holds until the end of the planning period. Figure 6.7(b) depicts the strategy that minimizes costs with RV. Such strategies will normally be desired by long-term oriented road agencies, as they do not push more expensive (structural) treatments behind the end of a considered or contracted planning period with substantial M&R backlogs thereafter.

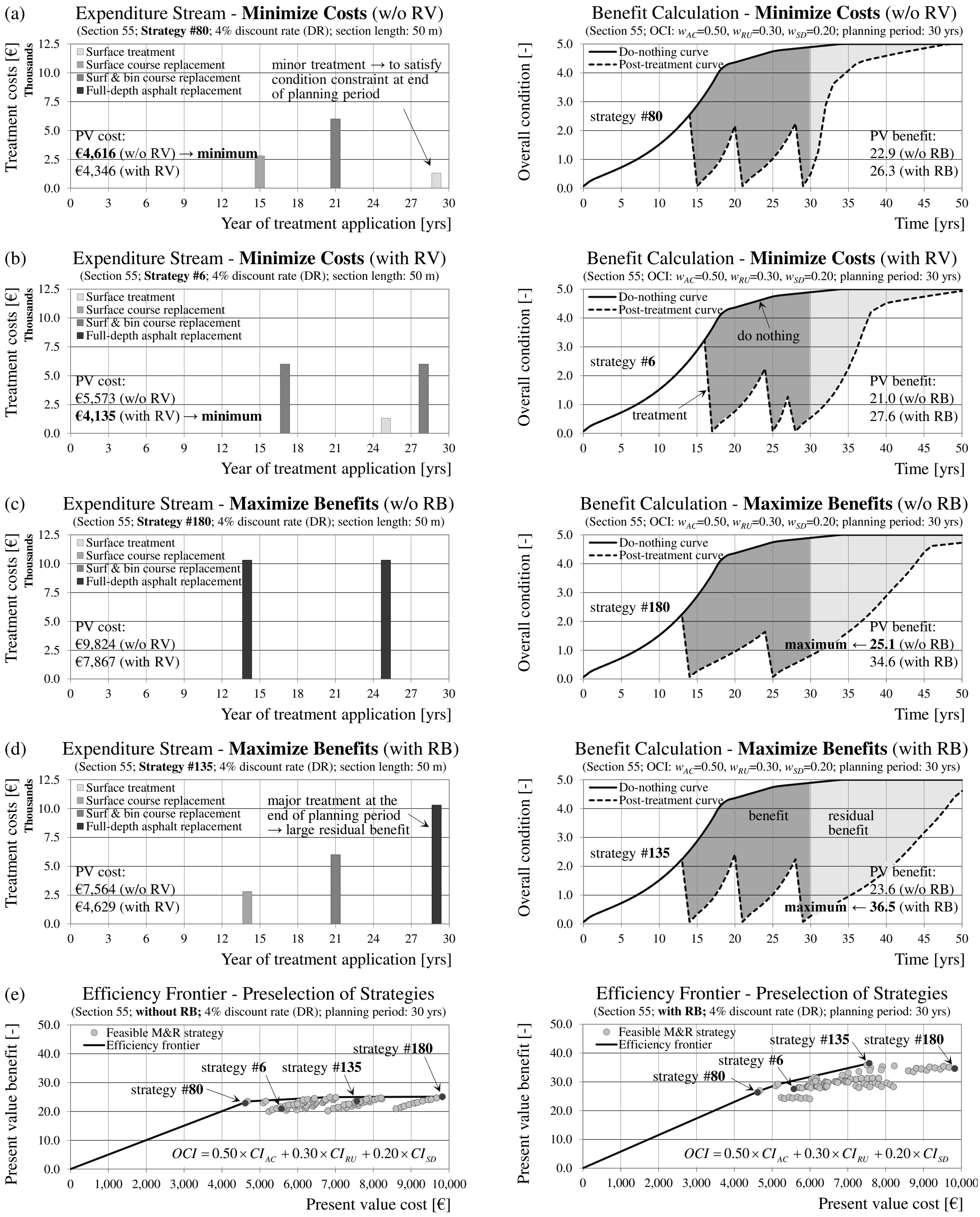


FIGURE 6.7 Expenditure stream diagram and benefit calculation for strategies that minimize costs with/without residual value (a, b) and maximize benefits with/without residual benefits (c, d) for a given road section, together with benefit-cost plot and efficiency frontier as a basis for selection of strategies for network-level programming (e).

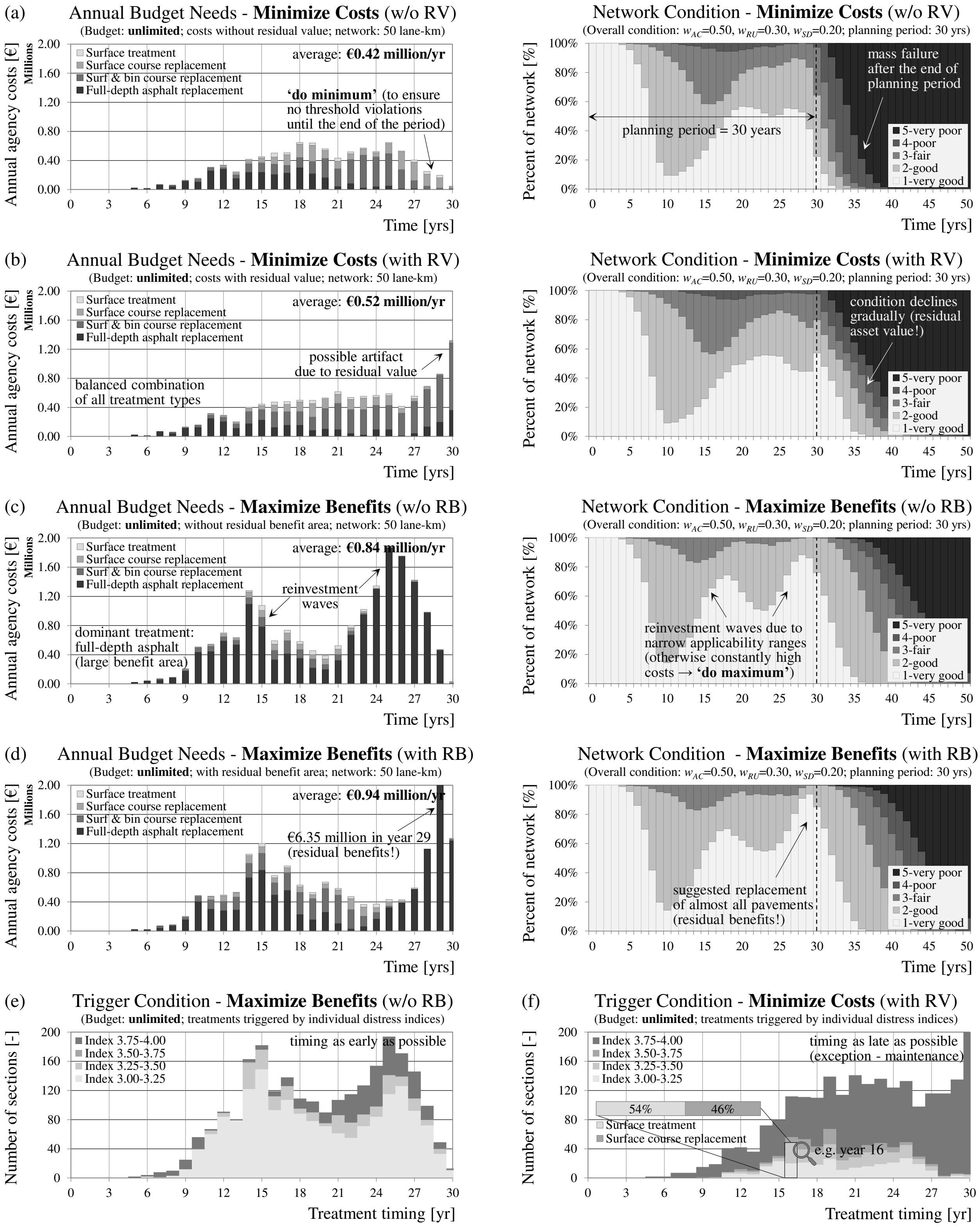
The strategy that maximizes benefits without RB is shown in Figure 6.7(c). In contrast to the previous two strategies, this strategy is composed solely of the most extensive (also the most expensive) treatment, as it produces a larger area between the curves. The objective of maximizing benefits with RB (see Figure 6.7(d)) has a reverse effect to the investments near the end of the planning period in comparison to the strategy of minimum costs without RV. The reason for this is that an additional treatment in the last years of the planning period generates a substantially larger RB area.

The post-treatment performance curves in Figure 6.7 reveal that treatments are applied before the overall condition reaches its threshold value of 4.0. This is due to the fact that treatments are triggered by the individual distress types, and the overall condition index is only used for the calculation of benefits. Figure 6.7(e) presents the efficiency frontier in the cases of benefits with and without RB. The efficiency frontier is used to select strategies for network-level M&R programming under budget restrictions (see Section 6.4.3). The figure shows also the equation for a hypothetical OCI used for all computations presented in this paper. For the purpose of the case study, more weight is attributed to the structural condition indicator (alligator cracking), reflecting the practice in many PMSs. Thus, the following weights have been assigned: 0.50 to alligator cracking, 0.30 to rutting and 0.20 to surface defects.

6.4.2 Maximizing benefits/minimizing costs without budgetary restrictions

The insights from the comparison of strategies at the section level become more obvious by examining the resulting annual costs and condition distribution for the entire network (50 lane-km). Figure 6.8(a) shows the strategy that minimizes costs without RV, resulting in average agency costs of €0.42 million per year. The cost diagram exhibits decreasing investments in the last years (no structural treatments), which leads to a rapidly worsening condition right after the end of the planning period. The strategy that minimizes costs with RV results in average annual expenditures of €0.52 million, as depicted in Figure 6.8(b). This strategy produces steady cash flows and a more balanced mix of maintenance and rehabilitation treatments. As the remaining asset value at the end of the planning period is implicitly considered, the network condition does not decline as rapidly as in Figure 6.8(a).

The maximization of benefits without RB yields average agency costs of €0.84 million per year (see Figure 6.8(c)), which is 62% higher than the corresponding costs of the strategy minimizing costs with RV. Moreover, the resulting M&R program is dominated by one particular treatment type: full-depth asphalt replacement. This treatment is suggested in almost all cases, irrespective of the actual distress combination and the leading failure cause. Furthermore, the two reinvestment waves are caused by the lower limit of the applicability range (CI = 3.0). The lower limit of the applicability range translates into minimum time interval between treatments. Without such a limitation, the maximization of benefits will push follow-up treatments forward and more treatments will be included in the planning period, resulting in larger benefit area but also in constantly high expenditures. Thus, the case of unlimited budget clearly shows that the maximization of benefits based on a subjective composite condition index is not an appropriate objective in pavement management. Moreover, additional analysis has revealed that using equal weights for the distresses in OCI will change the optimal strategy at the section level for 21% of all sections. Assigning weights of 0.20, 0.50 and 0.30 to alligator cracking, rutting and surface defects, respectively, leads to change of the optimal strategy on 36% of all sections. The resulting



Note: The average annual agency costs are computed for the period from the 10th year to the 30th year. All costs are provided in 2016 constant-value euros (inflation free) and constitute net costs excluding value-added tax (VAT). Temporary traffic control costs (TTCC) are not considered in the analysis.

FIGURE 6.8 Annual budget needs and network condition without budgetary restrictions for the cases of minimum costs with/without residual value (a, b) and maximum benefits with/without residual benefits (c, d), as well as corresponding distributions of trigger values in each year of the planning period (e).

average annual agency costs, however, remain insensitive to any changes of weights. The reason for this is that, relative to other treatments, a major treatment produces larger area between the curves not only for the overall condition but also for all individual distress types, making the choice of the individual weights more or less irrelevant (see Appendix 6.A.1). Furthermore, the strategy that maximizes benefits with RB results in even higher annual costs (81% higher than "minimum costs with RV"). The reason for this is the concept of residual benefits that leads to a selection of strategies with a major treatment in the last three years to gain the additional benefit.

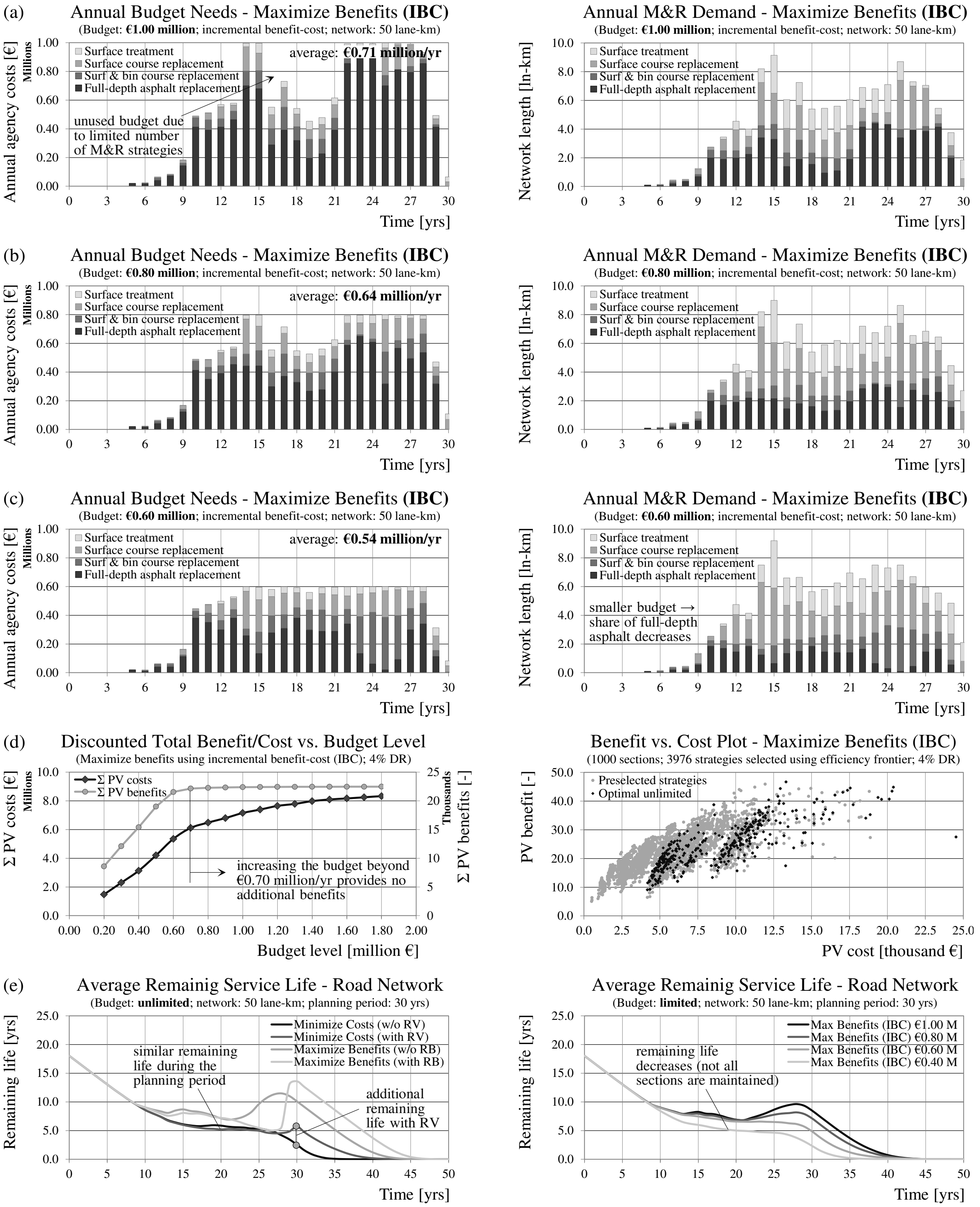
The distribution of trigger values, depending on the application year of the treatment (*x*-axis) is illustrated in Figure 6.8(e,f). The majority of treatments for the strategy of maximum benefits without RB are selected as early as possible (i.e. CI = 3.0 for any individual distress). In the case of minimization of costs with RV, the majority of treatments are planned as late as possible to minimize the loss of service life but without exceeding the thresholds. The magnified detail from Figure 6.8(f) shows that treatments with earlier timing are mostly preventive maintenance treatments. The following section will focus only on strategies of maximum benefits without RB and minimum costs with RV, as these objectives are more appropriate than their respective alternatives.

6.4.3 Maximizing benefits/minimizing costs under restricted budget (IBC)

The next objective is to find the combination of M&R strategies that maximizes the total benefit under limited budgets using the IBC algorithm (see Figure 6.2(a)). Only strategies that lie on the efficiency frontier are selected for the network-level analysis, amounting to 3976 M&R strategies in total (out of 300,000), not counting the "do nothing" strategies for each section. Figure 6.9(a,b,c) depicts the resulting annual agency costs and treatment demand (in lane-km) for annual budget levels of €1.00 million, €0.80 million and €0.60 million. In some of the cases the budget is not fully utilized due to the relatively small number of sections and strategies in the case study. The flexibility is also reduced by the fact that individual treatments composing a strategy impact budget constraints in multiple years. Nevertheless, in the case of maximum annual budget of €1.00 million, the resulting average agency costs (€0.71 million per year) are still 37% higher compared to the strategy minimizing costs with RV. For the other two annual budget scenarios, €0.80 million and €0.60 million, the expenditures are respectively 23% and 4% higher.

If the budget constraint for IBC is too tight, the resulting M&R program becomes more similar to the solution of the cost-minimization problem. If the budget constraint is less binding, the maximization of benefits will lead to substantially higher costs as corresponding cost-minimization strategies for the same reasons as those explained in Section 6.4.2. The graphs of the annual treatment demand show that with more severe budget restrictions, the share of full-depth asphalt replacement decreases in favor of surface and partial-depth repairs. If other performance and cost models are used, the alternative possibility cannot be ruled out that the share of repaired sections decreases and the shift in treatment type distribution is less significant.

The higher costs are further explained by the fact that the IBC and MCE approaches do not actually balance costs and benefits. The objective is to maximize benefits, while costs are only included in the budget constraint (see Peng and Ouyang 2010). Benefit-cost (B/C) or IBC values are only used for a pre-selection of strategies and to determine the order in which the strategies will be evaluated (i.e. to aid the heuristic procedure). Figure 6.9(d) displays the total discounted



Note: The average annual agency costs are computed for the period from the 10th year to the 30th year. All costs are provided in 2016 constant-value euros (inflation free) and constitute net costs excluding value-added tax (VAT). Note the different scale of the agency-costs axis (in comparison to Figure 11).

FIGURE 6.9 Annual budget needs and distribution of treatment types based on IBC approach for different funding levels (a, b, c). Total discounted benefits, agency costs and average remaining life depending on the budget (d, e).

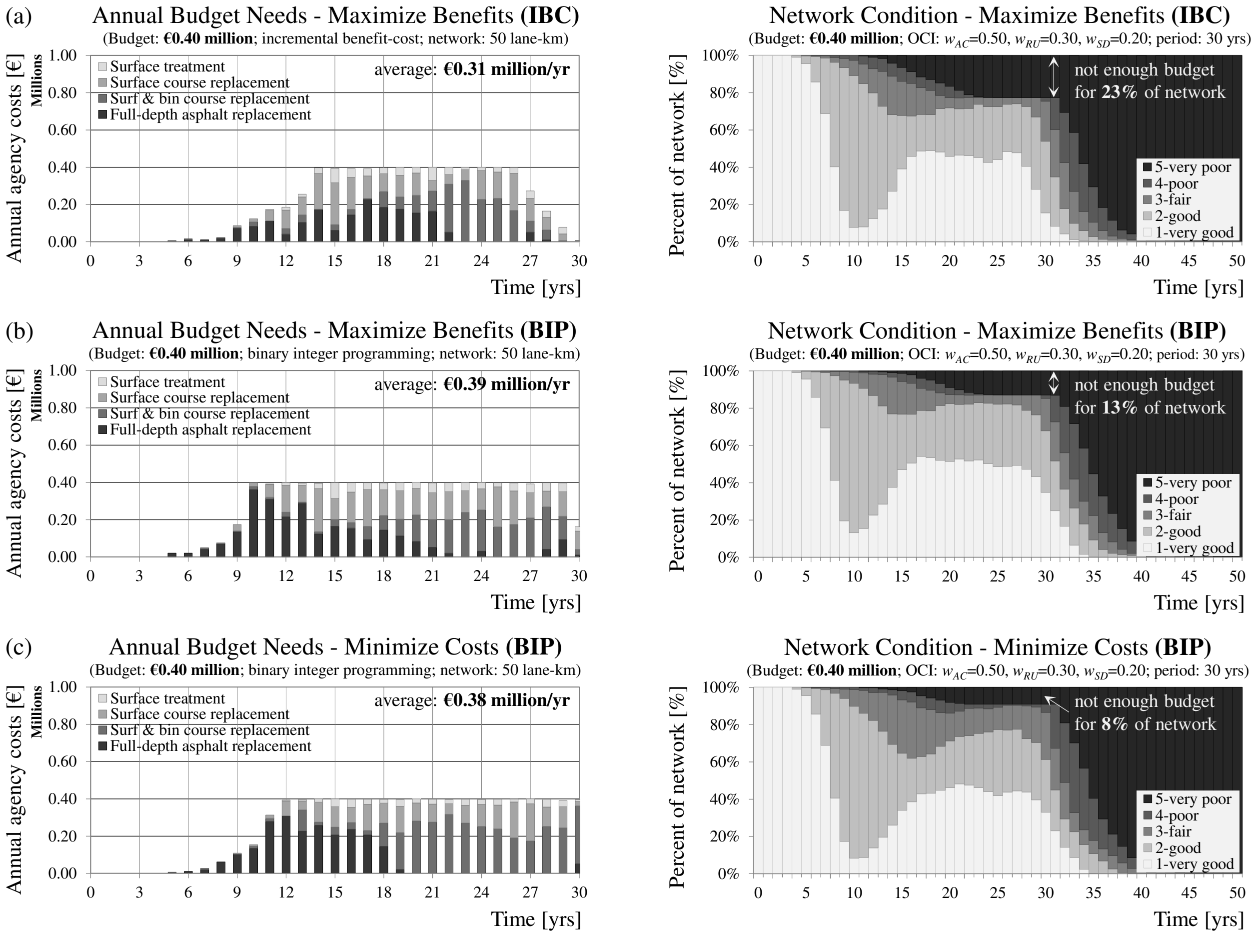
costs and benefits for different budget situations. As can be seen, increasing the budget level beyond €0.70 million per year only leads to higher total costs but does not lead to a further increase of benefits. This follows from the fact that each subsequent more expensive strategy on the efficiency frontier provides smaller benefit improvements (law of diminishing marginal returns). Some authors suggest using a threshold for the IBC ratio (e.g. Farid et al. 1994). However, it is not possible to objectively determine, what constitutes a sufficient improvement in benefit (i.e. justifiable investment), as the benefit value is based on an overall condition with no monetary value. Figure 6.9(e) illustrates the impact of the different objectives and the different budget levels on the resulting average remaining service life (RSL). Thus, the close link between condition distribution, remaining life, asset value and investment level is revealed, providing a complete picture of the implications of specific M&R strategies for the road agency.

6.4.4 Comparison to binary integer programming (BIP)

Next, the solution quality achieved by the heuristic algorithm IBC is evaluated. Figure 6.10(a) displays the annual costs and network condition for an annual budget of €0.40 million. As a result of the tight budget, the do-nothing strategy must be selected for 23% of all sections, implying that these sections will not receive any treatments in the entire planning period. Still, the available budget in the last years of the planning period is not fully utilized. An optimization approach can evaluate all possible combinations of strategies, while IBC evaluates the strategies in a sequence according to a specific order. So, a multi-year BIP (see Figure 6.3(a)) model is formulated, where each of the 3976 strategies is represented by a binary variable. The problem is solved using the branch-and-cut solver (COIN-OR-CBC) supported by @OpenSolver 2.9.0, an open-source add-in for Microsoft Excel (Forrest and Lougee-Heimer 2005, Mason 2012). Figure 6.10(b) presents the results, revealing that the budget is better utilized with this solution and only 13% of the network remains without funding (10% less!). In the literature it is argued that the IBC approach produces close to optimal results. In light of the already mentioned difficulties with the interpretation of benefits based on OCI, the finding here suggests that the difference between IBC and true optimization might be considerable (depending on the situation), measured in terms of reducing the backlog demand (cf. Haas et al. 1985).

A BIP model minimizing costs was also developed by considering the top ten strategies with the lowest life cycle costs for each section (9985 strategies). Minimizing costs requires an additional constraint on the condition (e.g. no failures), otherwise the “do nothing” strategy will always be selected. This approach leads to selection of strategies for additional 5% of the network, with only 8% remaining without M&R, as illustrated in Figure 6.10(c). As can be seen, the pre-selection of M&R alternatives for the network-level optimization is fundamentally different for the two objectives: benefit maximization and cost minimization. For the minimization of costs is sufficient to sort the strategies by total discounted costs in ascending order and select the top alternatives. For maximization of benefits, the efficiency frontier is needed, as the top strategies maximizing benefits have all high agency costs and cannot provide the necessary flexibility for the satisfaction of the budget constraints.

A summary of the results for different budget levels is provided at the bottom of Figure 6.10, while additional graphical representations are given in Appendix 6.A.2. It is clear that the BIP model leads to slightly higher total costs in comparison to IBC, but the two approaches should be compared with regard to their goal (i.e. maximization of benefits). In summary, BIP outperforms IBC, especially for tight budgets. Moreover, the solution of BIP allows an additional 10% (7%) of the network to be repaired for the budget scenario of €0.40 million per year (€0.50 million per year). The strategy of maximizing benefits with additional strategies includes strategies in the efficiency frontier tolerance range (16,229 strategies). This larger problem was solved using BIP, as the consideration of these strategies with IBC would require a modification of the algorithm. Anyhow, this provided only a little improvement compared to the original BIP formulation (3976 strategies). For example, in the case of €40 million annual budget, the amount of sections without treatments was reduced from 13% to 11% of the network.



Budget level	Maximize Benefits (a) **(Incremental Benefit-Cost)** 3,976 strategies		Maximize Benefits (b) **(Binary Integer Programming)** 3,976 strategies		Maximize Benefits (not shown) **(Additional Strategies - BIP)** 16,229 strategies		Minimize Costs (c) **(Binary Integer Programming)** 9,985 strategies	
	Σ PV benefits	Σ PV costs	Σ PV benefits	Σ PV costs	Σ PV benefits	Σ PV costs	Σ PV benefits	Σ PV costs
unlimited	22,476 (1000)	8,359,103	22,476 (1000)	8,359,103	22,476 (1000)	8,359,103	18,593 (1000)	4,950,230
€1.00 million	22,388 (1000)	7,189,750	22,407 (1000)	7,441,584	22,414 (1000)	7,533,996	18,599 (1000)	4,902,904
€0.80 million	22,272 (1000)	6,499,004	22,324 (1000)	6,769,672	22,344 (1000)	6,989,235	18,609 (1000)	4,845,767
€0.60 million	21,811 (998)	5,568,279	22,100 (1000)	5,910,076	22,145 (1000)	5,991,477	18,637 (1000)	4,724,359
€0.50 million	19,058 (**914**)	4,218,466	21,266 (**984**)	5,038,140	21,633 (996)	5,183,750	18,773 (**1000**)	4,581,188
€0.40 million	15,394 (**774**)	3,129,109	18,570 (**871**)	4,093,348	18,962 (894)	4,202,957	17,021 (**917**)	3,852,920

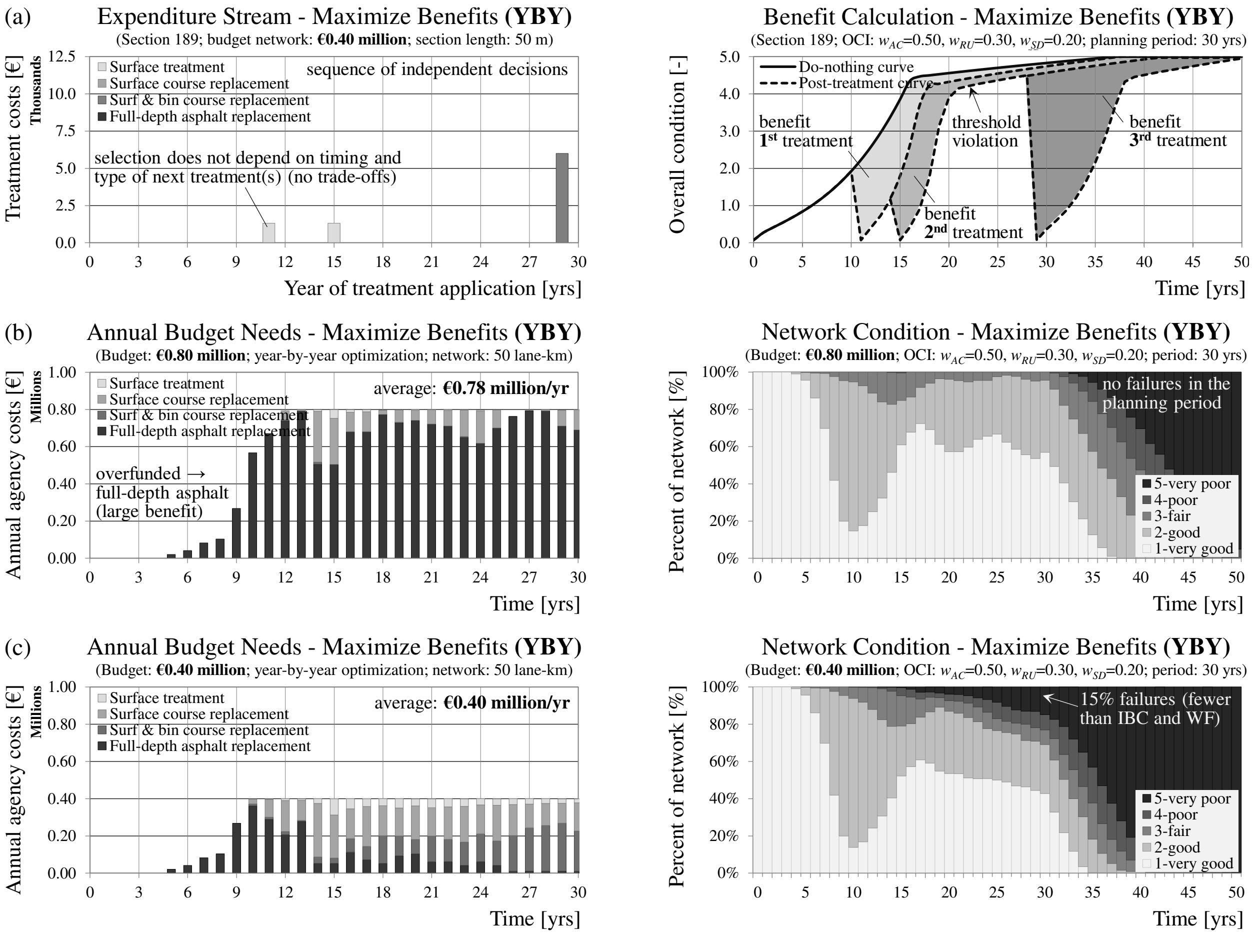
Note: Values in parentheses indicate the number of sections (of a total of 1000 sections) that can be maintained without violating the corresponding budget constraint. The total discounted benefits (Σ PV benefits) and the total discounted life cycle costs (Σ PV costs) are computed using a discount rate of 4% without consideration of residual benefit area and residual value, respectively. The average annual needs are computed for the period from the 10th year to the 30th year.

FIGURE 6.10 Comparison of annual budget needs and overall network condition resulting from maximizing benefits using IBC approach (a), maximizing benefits using BIP (b) and minimizing costs using BIP (c), with summary of the results for different budget scenarios.

6.4.5 Year-by-year binary programming

Alternatively, the benefit maximization problem can be solved using a BIP model on a year-by-year (YBY) basis (see Figure 6.3(b)). The decision variables for this formulation are the individual treatments (single-treatment strategies), resulting in a total of 5000 decision variables (1000 sections × 5 treatments). Thus, the benefits of all feasible treatments for all sections must be predicted prior to the optimization in each year. Furthermore, the decision tree is slightly modified so that upper limits for the individual distresses are removed (i.e. threshold violations are allowed). Otherwise, the problem may become infeasible for tight budget scenarios.

Figure 6.11(a) shows the optimal strategy for a selected road section and an annual budget level of €0.40 million. In contrast to a multi-year program, the selection of a specific treatment in a given year does not depend on the type and timing of the subsequent treatments. Therefore, an M&R optimization on a year-by-year basis does not constitute a life-cycle approach. A closer look at the performance curves reveals that the overall condition exceeds the threshold value in year 21 (individual distress types earlier). Due to a tight budget, this section is not repaired until year 29. Figure 6.11(b) reveals that for a funding level of €0.80 million per year, the budget is almost fully utilized (compare with Figure 6.9(b)). As treatment 4 produces the largest benefit area, for a larger budget, the problem is reduced to maximizing the number of sections where



Note: The average annual agency costs are computed for the period from the 10th year to the 30th year (2016 constant euros; net costs). The total discounted agency costs (Σ PV costs) for annual budget limits of €0.80 million and €0.40 million amount respectively to €7,988,496 and €4,310,741 (4% discount rate).

FIGURE 6.11 Maximization of benefits for a single road section (a) and for the entire network with annual budget limits of €0.80 million (b) and 0.40 million (c) using BIP separately for each year of the planning period (year-by-year).

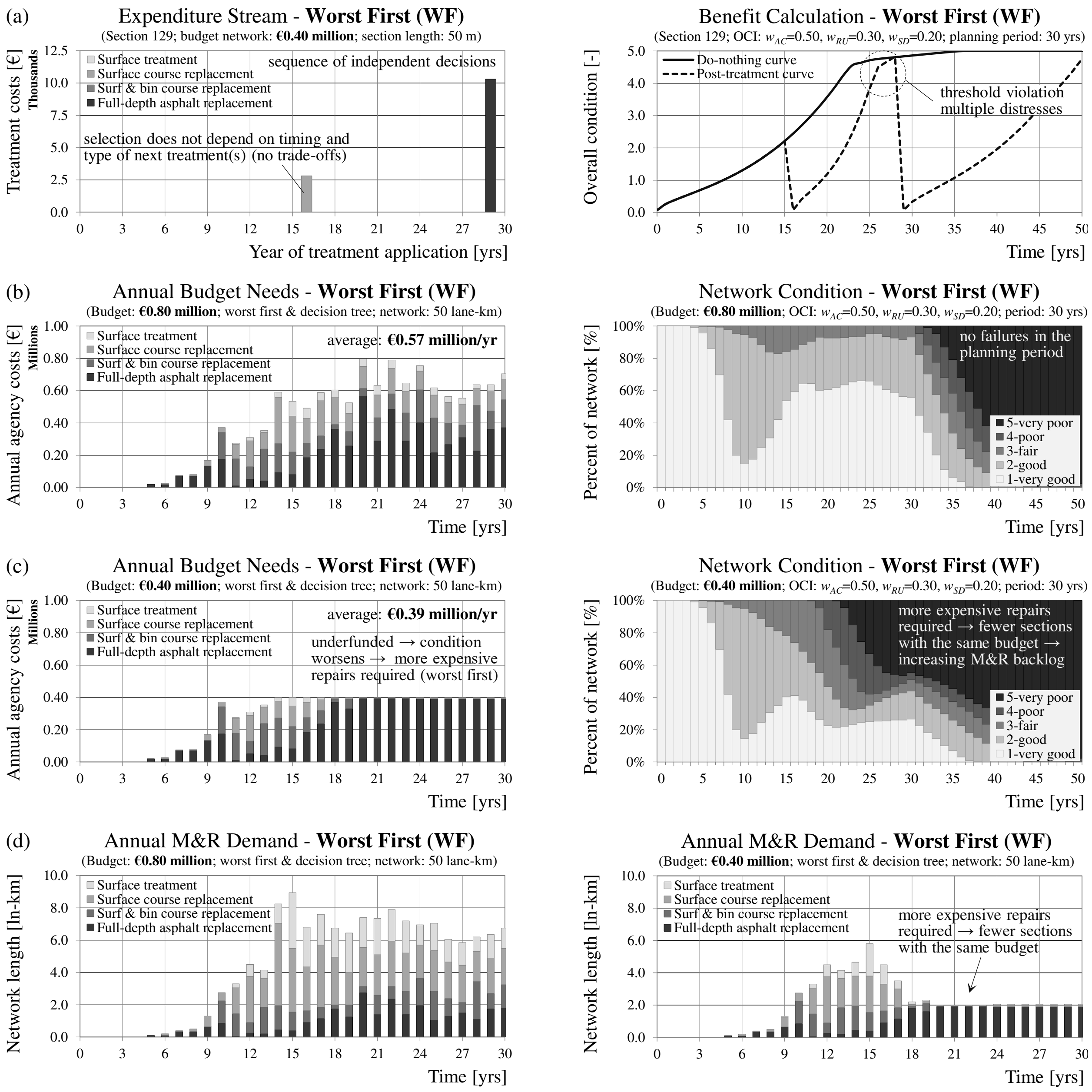
this treatment can be applied (“do maximum”). However, if the budget is insufficient, the optimization will try to minimize the sections without treatment (zero benefit) by selecting less expensive treatment types, as illustrated in Figure 6.11(c). The corresponding network condition shows 15% failures at the end of the planning period, which is still less than the failures with the IBC technique and a little more than the failures with the multi-year BIP (compare with Figure 6.10(a,b)). However, in the latter case the violations of condition thresholds occur earlier in the planning period. The reason for this is that the YBY model may consider repair of an already failed section, while the “do nothing” strategy for the multi-year model does not allow this option.

A limitation of the analysis in favor of the YBY approach follows from neglecting the negative effects of applying treatments by conditions worse than the threshold of 4.0. Such negative effects may include the need for more extensive treatment (higher costs) or shortening of post-treatment service life. This limitation relates mostly to the application of treatment 3 after year 15 in the case of €0.40 million annual budget (Figure 6.10(c)).

6.4.6 Worst-first approach

Finally, the WF ranking approach is applied to the case study data on a year-by-year basis. The current pavement condition is used to prioritize road sections and assign treatments based on a decision tree. The modification to the decision tree is given in Figure 6.2, defining a specific treatment without alternatives for each distress combination. Similar to the YBY BIP model, the selection of a treatment with the WF approach is independent from the timing and type of the subsequent treatments and temporary threshold violations are possible, as illustrated in Figure 6.12(a). However, neither treatment costs nor benefits are used as a decision-making criterion. In the case of a budget limit of €0.80 million per year, the average annual expenditures amount to €0.57 million, as shown in Figure 6.12(b). Thus, a simple WF approach results in respectively 11% and 27% lower costs compared to the multi-year IBC method and the BIP model on a year-by-year basis. The success of the WF approach, however, is highly dependent on the development of an appropriate decision tree.

Nevertheless, the shortcomings of this method are evident for tight budgets, as demonstrated in Figure 6.12(c). Since the available funding is not sufficient to cover all M&R needs, only the worst-performing sections will be repaired. A pavement in poor condition requires replacement (treatment 4), and thus fewer sections can be repaired with the same budget. This argument is confirmed by a decreasing annual network length with M&R actions in Figure 6.12(d) (right). The ultimate result of WF ranking approaches under tight budget constraints is a large M&R backlog and 45% failed sections in the last year of the planning period. M&R programs and resulting costs for annual budget levels of €1.00 million and €0.60 million based on YBY programming and WF approach are provided in Appendix 6.A.2.



Note: The average annual agency costs are computed for the period from the 10th year to the 30th year (2016 constant euros; net costs). The total discounted agency costs (Σ PV costs) for annual budget limits of €0.80 million and €0.40 million amount respectively to €5,649,363 and €4,023,941 (4% discount rate).

FIGURE 6.12 M&R program for a single section (a), annual budget needs and network condition (b, c), as well as distribution of treatment types (d) for budget levels of €0.80 million and 0.40 million, resulting from the application of the worst-first approach.

6.5 CONCLUSIONS

The present paper focuses on one major shortcoming in several existing PMSs, namely the development of network-level M&R policies by maximizing benefits (effectiveness) based on a composite condition index. The consequences of following benefit-maximization strategies for condition distribution, agency costs and remaining life are analyzed based on a parametric case study. Strategies maximizing benefits and minimizing costs are compared for the theoretical case of unlimited budget and for different funding scenarios using the IBC technique and true optimization (BIP).

The results show that the maximization of benefits gives preference to the most expensive replacement treatment with timing as early as possible, irrespective of actual service life and failure causes. However, the selection of treatments providing a larger area between the curves (benefit) based on an overall condition is neither an objective nor an economically justified decision-making approach. In the case of unlimited budgets, the benefit-maximization strategy results in 62% higher costs for the agency compared to a strategy that minimizes total life cycle costs. For limited budgets this percentage ranges between 4% and 37%. For road agencies with a tight budget, the IBC approach will lead to an inefficient use of funds, violations of thresholds and unnecessary M&R backlogs. Thus, the IBC objective of maximum benefits is a misleading, as these benefits have very little to do with actual monetary and non-monetary benefits resulting from following a specific M&R policy. In addition, the analysis reveals that using mathematical programming instead of heuristic approaches (IBC) may provide a substantial improvement, resulting in higher benefits and fewer sections in a failed state. Nevertheless, the results suggest that the selection of an appropriate optimization objective (benefits vs. costs) is far more important than finding the best solution algorithm (prioritization vs. optimization).

A comparison with a simple ranking approach (worst first) suggests that WF may perform very well in the case of less binding budget constraints, conditional on the appropriate definition of a decision tree for selection of M&R treatments. For a less sufficient funding level, however, the agency budget is quickly consumed by expensive replacements for an increasing number of sections in a failed condition, leading to a quickly increasing gap between M&R needs and available budget.

The conclusions of this research contribute to both theory and practice of pavement management. In the context of aging infrastructure, increasing M&R needs and limited public budgets, a better budget allocation at the network level and substantial savings can be achieved by minimizing agency or total (incl. user) costs instead of maximizing the area between curves based on subjective weighting. In any case, road agencies should be aware of the short- and long-term effects of pursuing a specific optimization objective on the average network condition and annual investment needs. The presented simulation approach avoids the limitations of empirical data providing researchers and agencies with the means to compare the effects of different parameters and optimization objectives on the results.

REFERENCES

AASHTO, 2012. *Pavement management guide*. 2nd ed. Washington, DC: American Association of State Highway and Transportation Officials.

Al-Qadi, I.L., Buttlar, W.G. and Baek, J., 2009. *Cost-effectiveness and performance of overlay systems in Illinois. Volume 2: Guidelines for interlayer system selection decision when used in HMA overlays*. Report no. FHWA-ICT-09-045. Urbana, IL: Illinois Center for Transportation.

Anastasopoulos, P.C. and Mannering, F.L., 2015. Analysis of pavement overlay and replacement performance using random parameters hazard-based duration models. *Journal of Infrastructure Systems*, 21 (1), 04014024.

Birbaum, J., 2016. *Entwicklung eines Verfahrens zur Ermittlung der Lebenszykluskosten von Straßen in Asphaltbauweise* [Development of a procedure to determine the life-cycle-costs of asphalt roads]. Dissertation (PhD). Universität Siegen (in German).

Brozek, B., Litzka, J. and Weninger-Vycudil, A., 2009. *Entwicklung eines Nutzerkostenmoduls im österreichischen PMS* [Development of a user cost module in the Austrian pavement management system]. Wien: BMVIT, Straßenforschung Heft 585 (in German).

Cuelho, E., Mokwa, R. and Akin, M., 2006. *Preventive maintenance treatments of flexible pavements: A synthesis of highway practice*. Report no. FHWA/MT-06-009/8117-26. Helena, MT: Montana Department of Transportation.

Donev, V. and Hoffmann, M., 2018a. Comparison of optimization approaches for pavement maintenance and rehabilitation policies on road section and network level. *6th International Symposium on Life - Cycle Civil Engineering (IALCCE2018)*, 28-31 October, Ghent, Belgium.

Donev, V. and Hoffmann, M., 2018b. Optimisation of pavement maintenance and rehabilitation activities, timing and work zones for short survey sections and multiple distress types. *International Journal of Pavement Engineering*. doi: 10.1080/10298436.2018.1502433.

Farid, F., Johnston, D.W., Rihani, B.S. and Chen, C.-J., 1994. Feasibility of incremental benefit-cost analysis for optimal budget allocation in bridge management systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1442, 77-87.

Ferreira, A., Picado-Santos, L. and Antunes, A., 2002. A Segment-linked optimization model for deterministic pavement management systems. *International Journal of Pavement Engineering*, 3 (2), 95-105.

FHWA, 1998. *Pavement Management Systems*. FHWA Report HI-97-024. Washington, DC: Federal Highway Administration.

Ford, K.M, Arman, M.H.R., Labi, S., Sinha, K.C., Thompson, P.D., Shirole, A.M. and Li, Z., 2012. *Estimating life expectancies of highway assets, volume 2: Final report*. NCHRP Report no. 713. Washington, DC: Transportation Research Board.

Forrest, J. and Lougee-Heimer, R., 2005. CBC User Guide. *INFORMS TutORials in Operations Research*, 257-277.

Gharaibeh, N. G., Zou, Y. and Saliminejad, S., 2010. Assessing the agreement among pavement condition indices. *Journal of Transportation Engineering*, 136 (8), 765–772.

Haas, R., Hudson, W.R. and Falls, L.C., 2015. Birth and teen years of pavement management (1967–1987). *In*: R. Haas, W.R. Hudson and L.C. Falls, eds. *Pavement asset management*. Salem, MA: Scrivener, 5-13.

Haas, R., Hudson, W.R. and Zaniewski, J., 1994. *Modern pavement management*. Malabar, FL: Krieger.

Haas, R., Karan, M.A., Cheetham, A. and Khalil, S., 1985. Pavement rehabilitation programming: a range of options. *1st North American Pavement Management Conference*, 18-21 March, Toronto, Ontario, Canada.

Hall, K.T., Correa, C.E., Carpenter, S.H. and Elliot, R.P., 2001. *Rehabilitation strategies for highway pavements*. NCHRP Project C1-38: Conctractor's final report. Washington, DC: Transportation Research Board.

Hillier, F.S. and Lieberman, G.J., 2015. *Introduction to operations research*. 10th ed. New York, NY: McGraw-Hill Education.

Irfan, M., Khurshid, M.B., Ahmed, A. and Labi, S., 2012. Scale and condition economies in asset preservation cost functions: case study involving flexible pavement treatments. *Journal of Transportation Engineering*, 138 (2), 218-228.

Javed, F., 2011. *Integrated prioritization and optimization approach for pavement management*. Dissertation (PhD). University of Singapore.

Juster, R.D. and Pecknold, W.M., 1976. Improving the process of programming transportation investments. *Transportation Research Record: Journal of the Transportation Research Board*, 599, 19-24.

Keleman, M., Henry, S., Farrokhyar, A., Mero, B., Strome, G. and Olson, S., 2008. *Pavement management manual*. Denver, CO: Colorado Department of Transportation.

Kerali, H.R. and Mannisto, V., 1999. Prioritization methods for strategic planning and road work programming in a new highway development and management tool. *Transportation Research Record: Journal of the Transportation Research Board*, 1655, 49-54.

Kuhn, K.D., 2012. Pavement network maintenance optimization considering multidimensional condition data. *Journal of Infrastructure Systems*, 18 (4), 220-277.

Lea, J., Harvey, J. and Tseng, E., 2014. Aggregating and modeling automated pavement condition survey data for flexible pavements for use in pavement management. *Transportation Research Record: Journal of the Transportation Research Board*, 2455, 89-97.

Li, N., Haas, R. and Huot, M., 1998. Integer programming of maintenance and rehabilitation treatments for pavement networks. *Transportation Research Record: Journal of the Transportation Research Board*, 1629, 242-248.

Maerschalk, G. and Socina, M., 2008. *Weiterentwicklung der Bewertung des Pavement Management Systems (PMS) um ein Verfahren für die Umsetzung von Qualitätszielen* [Extension of the evaluation functions of the pavement management system (PMS) by a system for the implementation of quality goals]. Forschung Straßenbau und Straßenverkehrstechnik, Heft 1001. Bonn: Bundesministerium für Verkehr, Bau und Stadtentwicklung (in German).

Mason, A.J., 2012. OpenSolver - An open source add-in to solve linear and integer progammes in Excel. *In*: D. Klatte, H.-J. Lüthi and K. Schmedders, eds. *Operations Research Proceedings 2011*. Berlin Heidelberg: Springer, 401-406.

MDOT, 2019. *Transportation asset management plan*. Saint Paul, MN: Minnesota Department of Transportation.

Menendez, J.R., Siabil, S.Z., Narciso, P. and Gharaibeh, N.G., 2013. Prioritizing infrastructure maintenance and rehabilitation activities under various budgetary scenarios: Evaluation of worst-first and benefit-cost analysis approaches. *Transportation Research Record: Journal of the Transportation Research Board*, 2361, 56-62.

Nobakht, M., Sakhaeifar, M.S., Newcomb, D. and Underwood, S., 2016. Mechanistic-empirical methodology for the selection of cost-effective rehabilitation strategy for flexible pavements. *International Journal of Pavement Engineering*. doi: 10.1080/10298436.2016.1199878.

Odoki, J.B. and Kerali, H.G.R., 2000. *Analytical framework and model descriptions*. The Highway Development and Management Series, Volume 4. Paris, France: World Road Association (PIARC).

Patidar, V., Labi, S., Morin, T., Thompson, P.D. and Sinha, K.C., 2011. Evaluating methods and algorithms for multicriteria bridge management at the network level. *Transportation Research Record: Journal of the Transportation Research Board*, 2220, 38-47.

Peng, F. and Ouyang, Y., 2010. *Pavement program planning based on multi-year cost-effectiveness analysis*. Report no. FHWA-ICT-10-067. Springfield, IL: Illinois Department of Transportation.

Qiao, Y., Labi, S., Fricker, J. and Sinha, K.C., 2019. Costs and effectiveness of standard treatments applied to flexible and rigid pavements: case study in Indiana, USA. *Infrastructure Asset Management*, 6 (1), 15-29.

Scheinberg, T. and Anastasopoulos, P.C., 2010. Pavement preservation programming: A multi-year multi-constraint optimization methodology. *89th Annual Meeting of the Transportation Research Board*, 10-14 January, Washington, DC.

Shah, Y.U., Jain, S.S. and Parida, M., 2014. Evaluation of prioritization methods for effective pavement maintenance of urban roads. *International Journal of Pavement Engineering*, 15 (3), 238-250.

Shahin, M.Y., Kohn, S.D., Lytton, R.L. and McFarland, M., 1985. Pavement M&R budget optimization using the incremental benefit-cost technique. *1st North American Pavement Management Conference*, 18-21 March, Toronto, Ontario, Canada.

Thompson, P.D., Small, E.P., Johnson, M. and Marshall, A.R., 1998. The Pontis bridge management system. *Structural Engineering International*, 8 (4), 303-308.

Wang, F., Zhang, Z. and Machemehl, R.B., 2003. Decision-making problem for managing pavement maintenance and rehabilitation projects. *Transportation Research Record: Journal of the Transportation Research Board*, 1853, 21-27.

Weninger-Vycudil, A., 2001. *Entwicklung von Systemelementen für ein österreichisches Pavement Management System* [Development of system elements for the Austrian pavement management system]. Dissertation (PhD). Technische Universität Wien (in German).

Weninger-Vycudil, A., Simanek, P., Haberl, J. and Rohringer, T., 2009. *Handbuch Pavement Management in Österreich 2009* [Manual pavement management in Austria 2009]. Straßenforschung, Heft 584. Wien: Bundesministerium für Verkehr, Innovation und Technologie (in German).

Wu, Z. and Flintsch, G.W., 2009. Pavement preservation optimization considering multiple objectives and budget variability. *Journal of Transportation Engineering*, 135 (5), 305-315.

Yoo, J. and Garcia-Diaz, A., 2008. Cost-effective selection and multi-period scheduling of pavement maintenance and rehabilitation strategies. *Engineering Optimization*, 40 (3), 205-222.

Yu, B, Gu, X., Ni, F. and Guo, R., 2015. Multi-objective optimization for asphalt pavement maintenance plans at project level: Integrating performance, cost and environment. *Transportation Research Part D*, 41, 64-74.

Zavitski, J.L., Rose, R.T. and Kuhl, G., 2008. Integrating pavement management into a comprehensive strategic asset management system for the state of Utah Department of Transportation. *7th International Conference on Managing Pavement Assets*, 23-28 June, Calgary, Alberta, Canada.

Zuniga-Garcia, N., Martinez-Alonso, W., de Fortier Smit, A., Hong, F. and Prozzi, J.A., 2018. Economic Analysis of Pavement Preservation Techniques. *Transportation Research Record: Journal of the Transportation Research Board*, 2672 (12), 10-19.

APPENDIX 6.A[†] SUPPLEMENTS AND INFLUENCE OF WEIGHTS

6.A.1 Influence of different weights in OCI on resulting optimal M&R strategies

This appendix includes supplemental figures to Section 6.4 regarding the influence of the distress weights in OCI on the optimal M&R strategies resulting from the benefit maximization approach. Figure 6.A.1 shows the strategies maximizing benefits without RB for four different combinations of weights for a given road section (section 254). It can be seen that changing the weights may lead to the different optimal strategies at the section level. Nevertheless, in this case, two combinations (Figure 6.A.1(b,c)) resulted in the same optimal strategy (strategy #75).

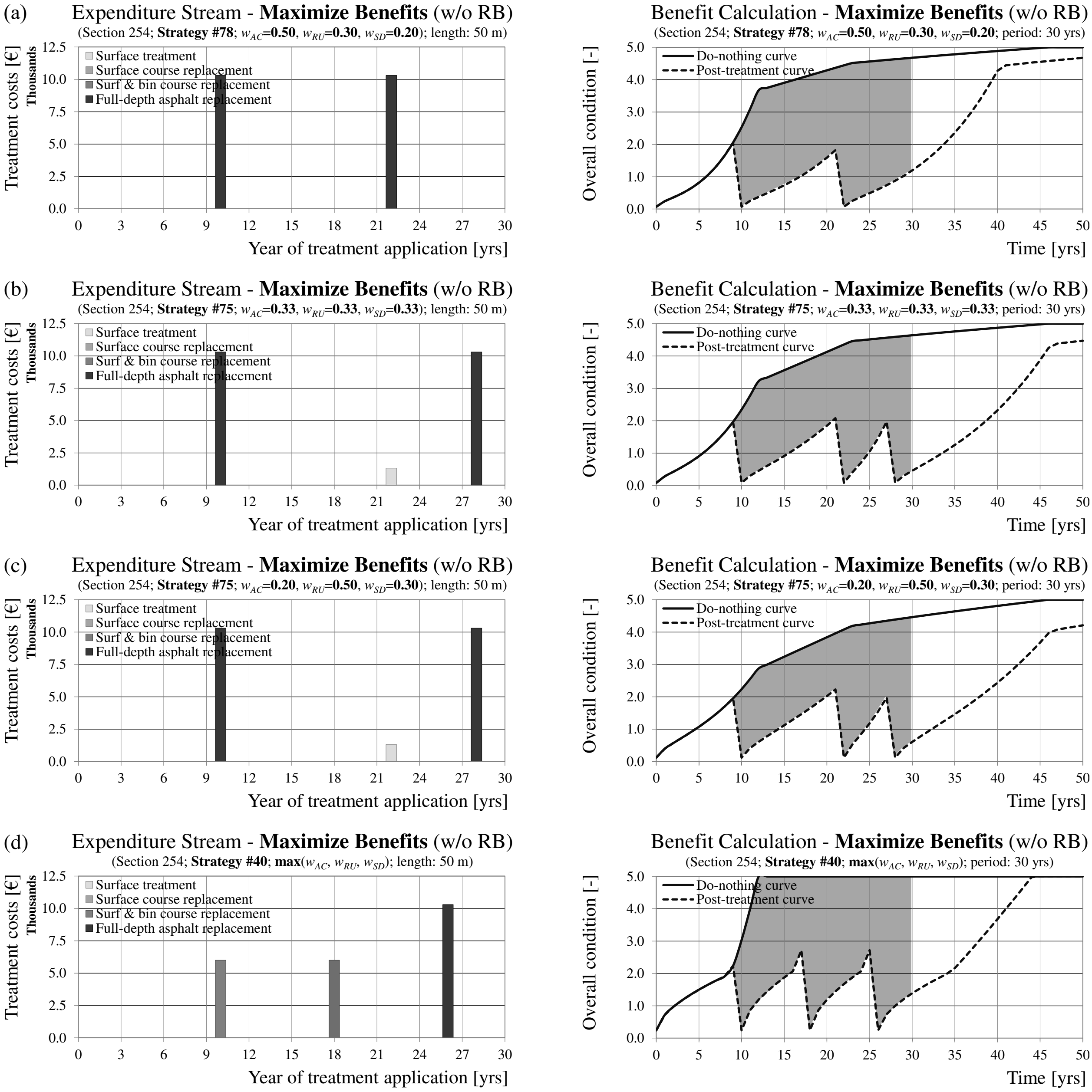


FIGURE 6.A.1 Expenditure stream diagram and benefit calculation for strategies that maximize benefits without RB at the road section level for different weights of the individual distress types in OCI.

[†] *This appendix is not part of the submitted article.*

The corresponding results at the network level are presented in Figure 6.A.2, whereby Figure 6.A.2(a) is identical to Figure 6.8(c). Although, the optimal strategy may change for an individual section, in this case study there are no significant differences in the annual costs and the treatment type distribution at the network level. Moreover, the average annual agency costs amount to €0.84 million in all four cases. Nonetheless, the share of treatment 3 (surf & bin course replacement) in budget years 10 to 21 is higher in three cases (see Figure 6.A.2(b,c,d)) in comparison to the case in Figure 6.A.2(a). Furthermore, there are seemingly substantial differences in the network condition distribution. However, this does not mean that the actual distributions resulting from the four M&R programs will be very different. Similarly, it cannot be concluded that the M&R program in Figure 6.A.2(d) results in the worst network condition. The second worse rating (Figure 6.A.2(c)), for example, can be explained by assigning larger weight to rutting together with its condition function and rating scale (see Figure 6.4 in Section 6.3.1).

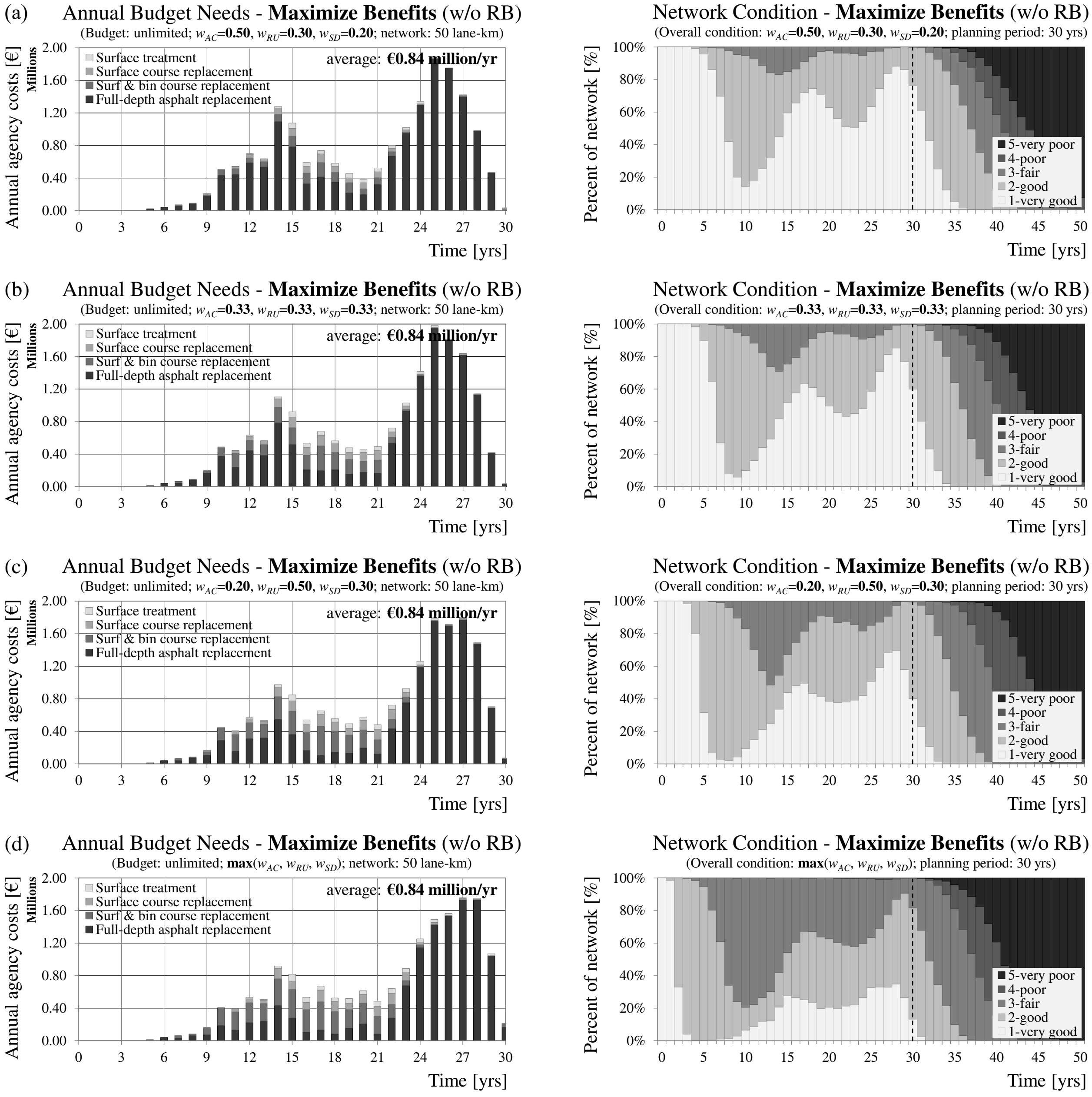


FIGURE 6.A.2 Annual budget needs and network condition for strategies maximizing benefits (without RB) without budgetary restrictions for different weights of the individual distress types in OCI.

All figures in Section 6.4 and Appendix 6.A show condition development (at section level) and condition distribution (at network level) using the OCI. However, it should be recalled that the do-nothing performance and the M&R treatment modeling are always based on the individual distress types, as shown in Figure 6.A.3 for the case of maximum benefits with weights w_{AC}=0.50, w_{RU}=0.30 and w_{SD}=0.20. Figure 6.A.3(a) (left) is identical to Figure 6.A.1(a) (right) and Figure 6.A.3(a) (right) is identical to Figure 6.A.2(a) (right). The computations in all other cases are similar, but the performance prediction and network distributions of the individual distress types are not shown for reasons of easier comparability and due to length limits. Thus, OCI is used only for the calculation of benefits and as a compact way to show the results. The averaging condition with regard to rutting is worse than the other two distress types. The reason for this lies in the degressive condition function and the equally-divided rating scale, leading to poor condition states being reached relatively early in comparison to the other distresses.

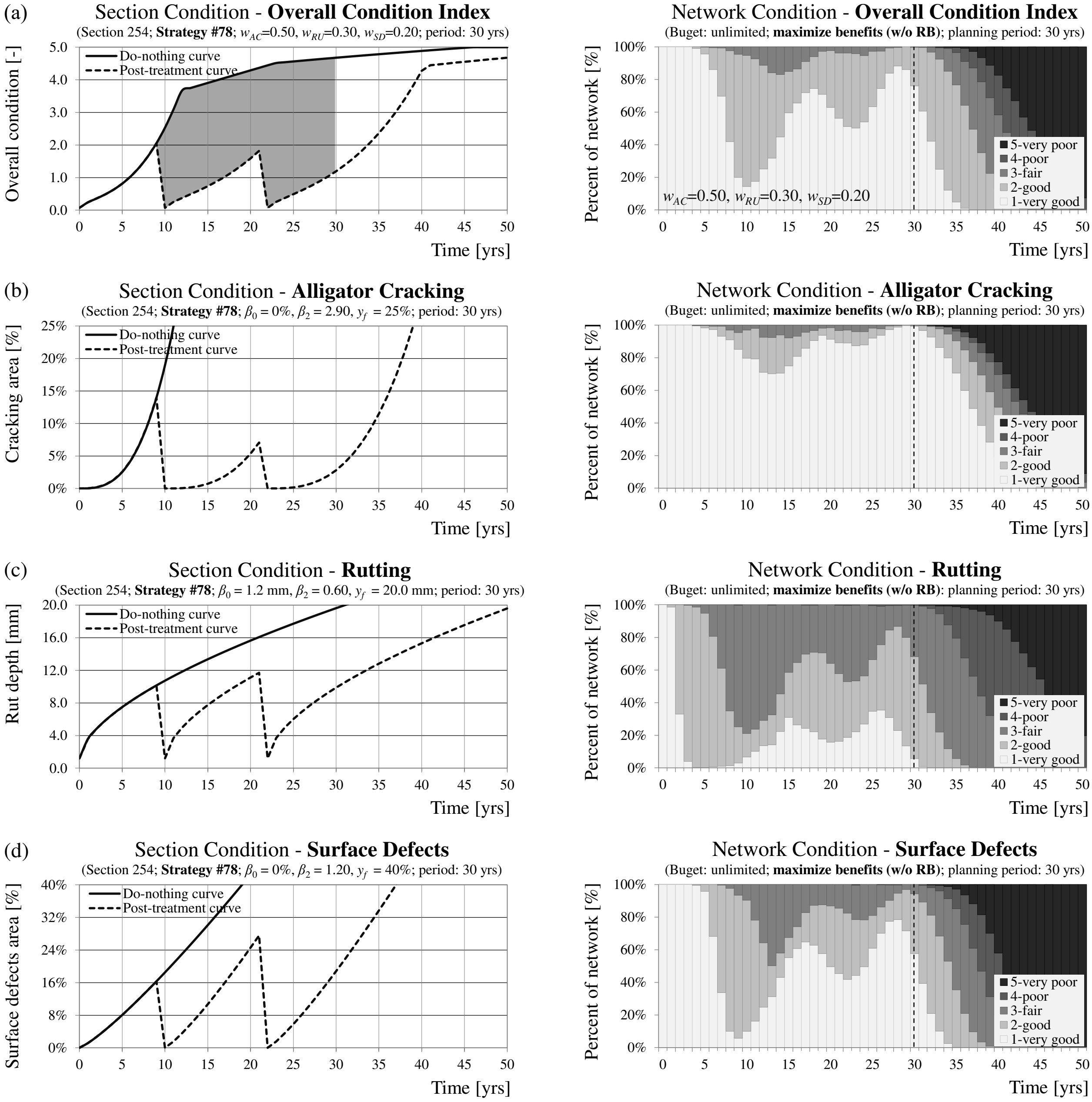


FIGURE 6.A.3 Performance prediction at the section level and condition distribution at the network level for OCI (a), resulting from condition prediction and M&R modeling for individual distresses (b,c,d).

In contrast to Figure 6.A.1 where for the same section the optimal strategies for different distress weights in OCI are presented, Figure 6.A.4 shows the same section and strategy, only using different weights for the computation of the condition development. Thus, strategy #78 is optimal only in the case of Figure 6.A.4(a) for weights w_{AC}=0.50, w_{RU}=0.30 and w_{SD}=0.20. The figure shows that changing the distress weights alters only slightly the condition development with and without M&R treatments based on OCI. The exception is Figure 6.A.4(d), where the overall condition at any time equals the worst of the individual distress indices. In the Austrian PMS, OCI is based on a similar formula with some additional conditions. However, the similar condition development may follow from the assumption in the case study that all sections exhibit all three types of distress. If some distress types are not present at all, changing the weights might have more serious effect on the OCI at the section level. At the network level, a similar M&R program is obtained even when OCI is based on the maximum formula (Figure 6.A.2(d)).

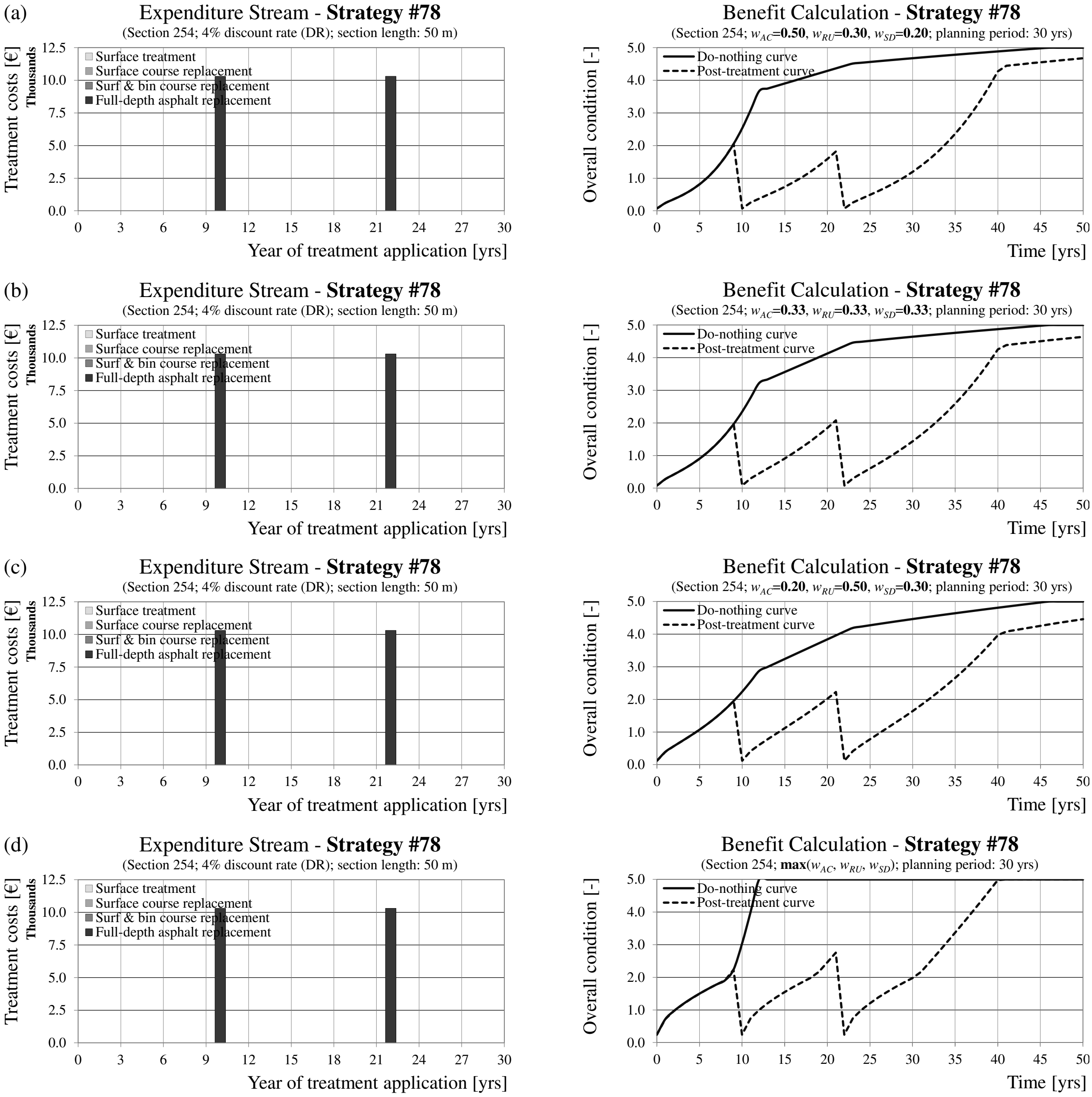


FIGURE 6.A.4 Expenditure stream diagram and benefit calculation for the same road section (Section 254) and the same M&R strategy (Strategy #78) but using different weights for calculation of OCI.

6.A.2 Supplemental figures for BIP, YBY and WF

The annual agency costs and network condition for strategies maximizing benefits without RB (BIP) under different budget scenarios are shown in Figure 6.A.5, whereby Figure 6.A.5(d) is identical to Figure 6.10(b) in Section 6.4.4. The results using BIP are quite similar to the results using IBC (cf. Figure 6.9), with exception of the case of €0.40 million budget. In contrast to IBC, BIP fully utilizes the available budget in years 10-13 and 27-30 (cf. Figure 6.10(a)). Thus, the BIP solution results in significantly less failed sections at the end of the planning period (13% instead of 23%). Regardless of the used solution method, the budget in the last year is never fully utilized, if benefits without RB are being maximized. This is due to the fact that M&R treatments in the last year provide no additional benefit. On the other hand, considering RB leads to the reversed effect with huge concentration of treatments in the last two years (cf. Figure 6.8(d)).

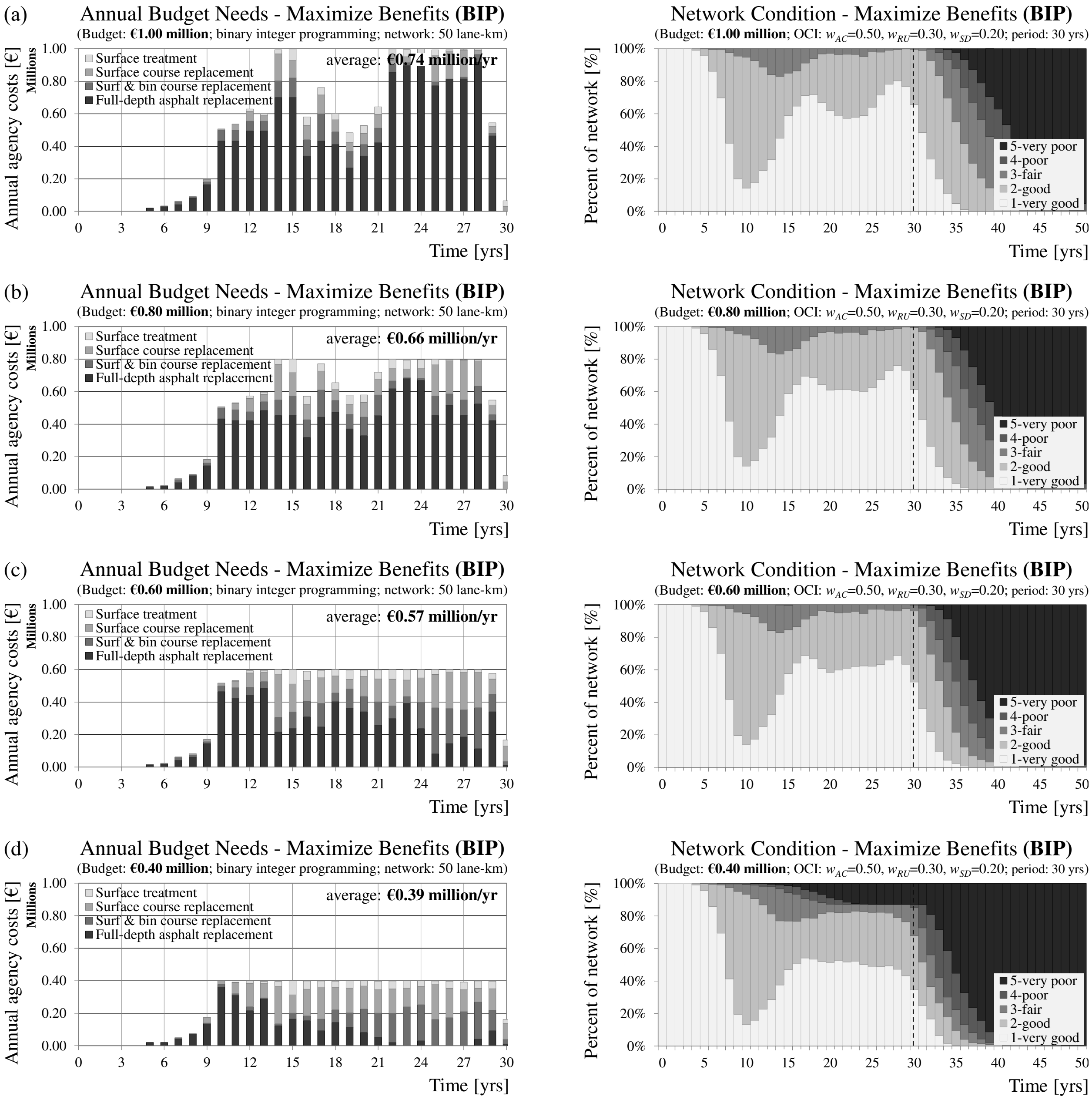


FIGURE 6.A.5 Annual budget needs and network condition for different funding levels, resulting from maximization of the total discounted benefits without RB using BIP.

The annual agency costs and network condition for strategies minimizing cost with RV (BIP) under different budget constraints are depicted in Figure 6.A.6, whereby Figure 6.A.6(d) is identical to Figure 6.10(c) in Section 6.4.4. In contrast to benefit maximization, the average costs are significantly lower for a large budget (€1.00 million), and cutting the budget to €0.40 million per year results in significantly fewer failures (only 8%) in comparison to other approaches. The jumps in the years 28 to 30 result most likely from the definition of RV. Thus, for treatments that are applied in the last years of the planning period, the residual value will be equal (in the last year) or almost equal to the costs of the treatment. This concept of RV is perhaps more suitable for comparison of single-treatment alternatives with different service lives (fixed timing) but less suitable for determination of treatment timing in the life cycle. The solution to this problem is to extend the planning period beyond 30 years. If the planning period is extended long enough in the future, the computation of residual value may be omitted completely. Nevertheless, extending the planning period will drastically increase the number of feasible M&R strategies.

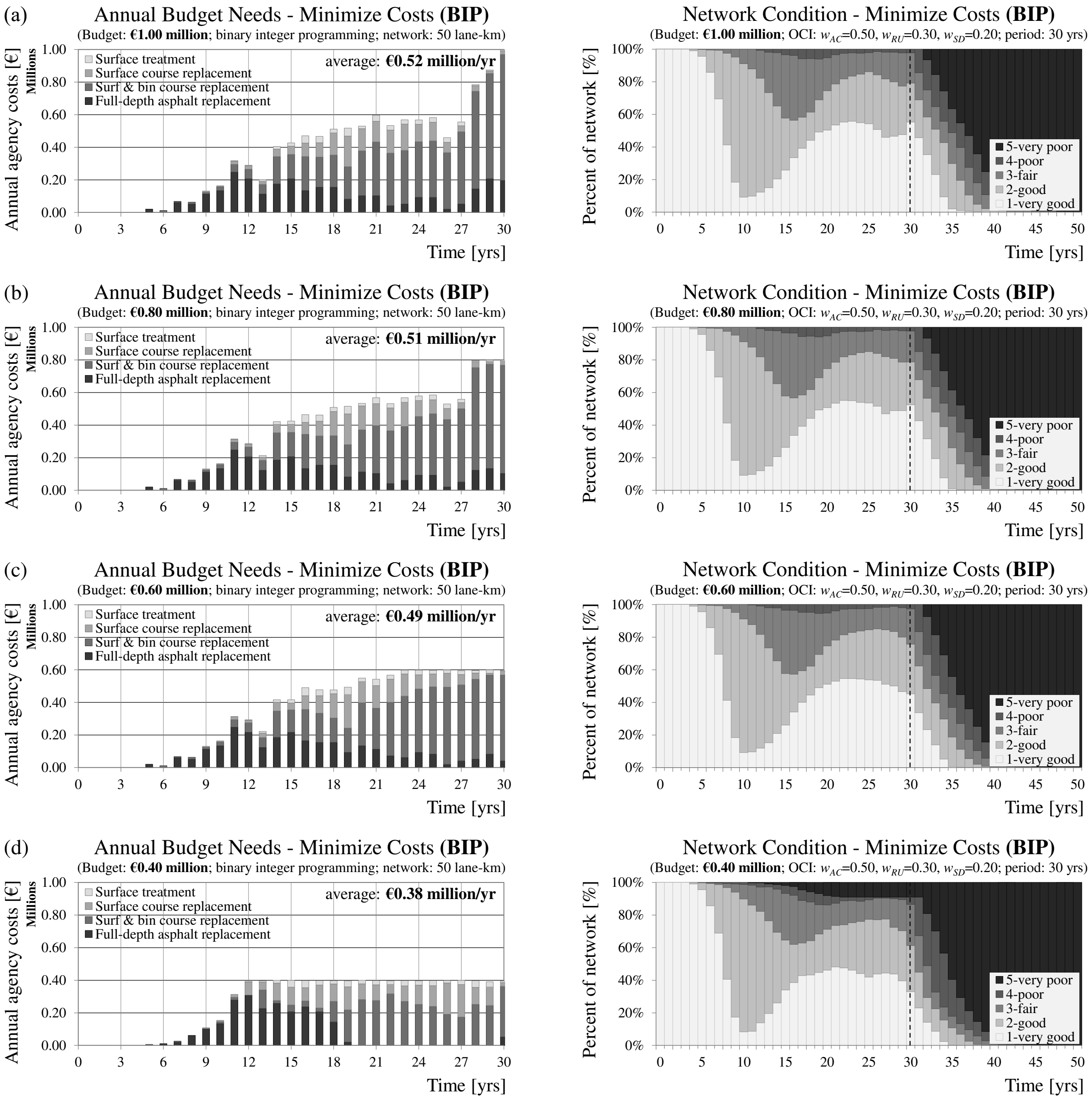


FIGURE 6.A.6 Annual budget needs and network condition for different funding levels, resulting from minimization of the total discounted agency costs with RV using BIP.

The annual agency costs and network condition for the cases of YBY benefit maximization and WF approach are illustrated in Figure 6.A.7, supplementing Figure 6.11 and Figure 6.12. The total discounted costs for the YBY approach and budget of €1.00 million and €0.60 million are €8,967,131 and €6,257,110, respectively. Accordingly, the present value of the total costs based on the WF approach and the same budget constraints amounts to, respectively, €5,650,321 and €5,349,619. Figure 6.A.7(a) clearly shows that the YBY approach is equivalent to a do-maximum strategy. In some years the budget is not fully utilized due to the trigger rules, not allowing treatments when all three distress indices are below 3.0. Although roughly the same share of sections per year are repaired following a WF strategy (see Figure 6.A.7(c)), the annual costs are much lower due to the large proportion of maintenance treatments. It is remarkable how close the WF approach is to the minimum-cost solution for sufficiently large budgets. However, the down-side of the WF approach becomes visible for €0.60 million budget (Figure 6.A.7(d)).

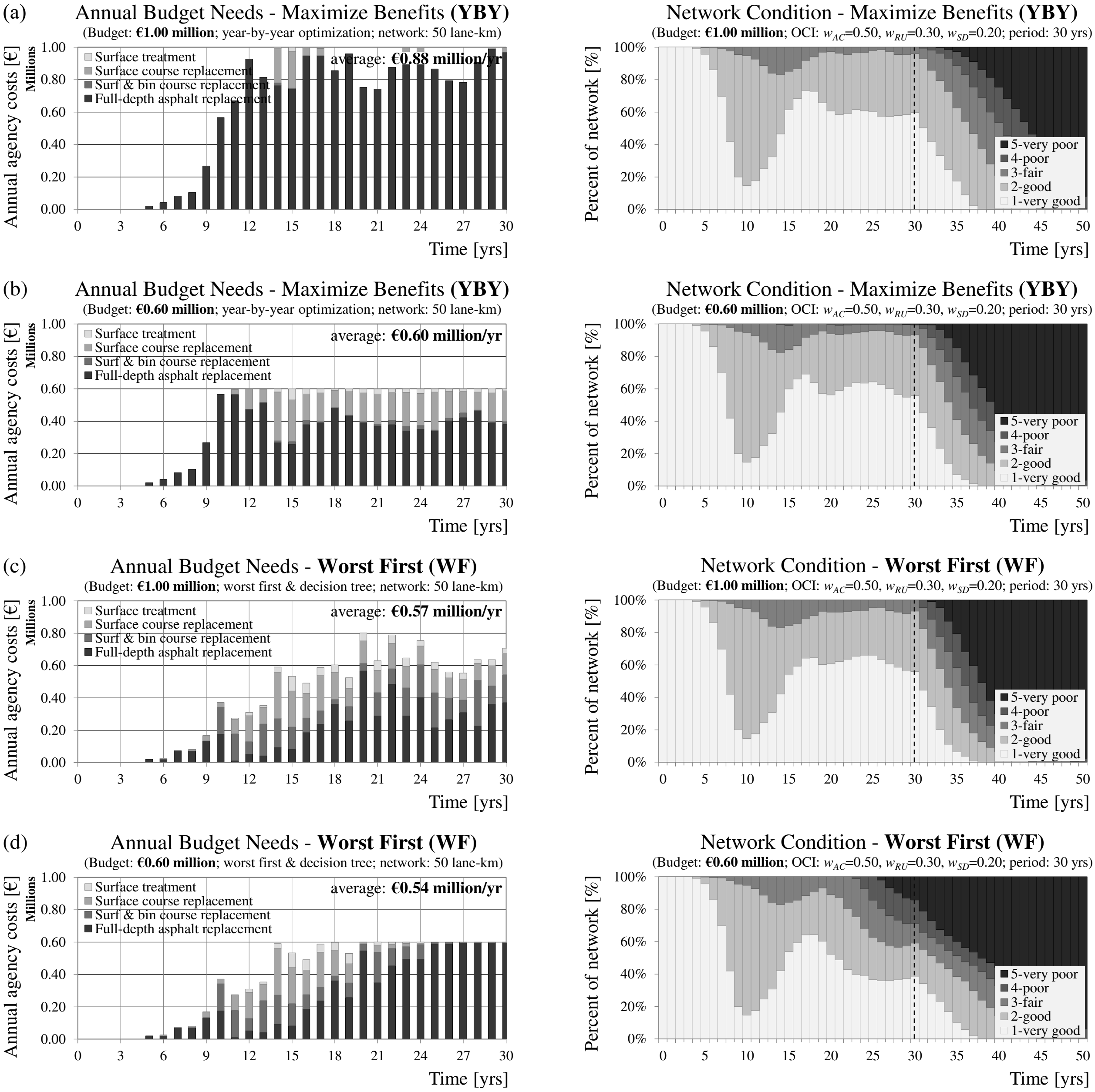


FIGURE 6.A.7 Annual budget needs and network condition for the cases of YBY benefit maximization without RB (a,b) and WF approach (c,d) for budget levels €1.00 million and €0.60 million per year.

CHAPTER 7

Appendix

7.1 LIST OF FIGURES