Design Optimization for Railway Transition Zones

Shang, Y.

DOI
10.4233/uuid:5ecbe3d3-f275-4c17-90fc-c8ec906a1316

Publication date
2024

Document Version
Final published version

Citation (APA)
Shang, Y. (2024). *Design Optimization for Railway Transition Zones*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:5ecbe3d3-f275-4c17-90fc-c8ec906a1316

DESIGN OPTIMIZATION FOR RAILWAY TRANSITION ZONES

DESIGN OPTIMIZATION FOR RAILWAY TRANSITION ZONES

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 5 June 2024 at 12:30 o'clock

by

Yue SHANG

Master of Science in Construction Management and Engineering,
Delft University of Technology, the Netherlands
born in Xi'an, China

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. A.R.M. Wolfert,	Delft University of Technology, promotor
Dr. ir. M. Nogal,	Delft University of Technology, copromotor

Independent members:

Prof. dr. ir. L.A. Tavasszy,	Delft University of Technology
Prof. dr. E. J. OBrien,	University College Dublin, Ireland
Dr. A.A. Nunez Vicencio,	Delft University of Technology
Dr. D. Cantero,	Norwegian University of Science and Technology, Norway
Dr. ir. A.P. de Man,	edilon)(sedra B.V.
Prof. dr. A.V. Metrikine,	Delft University of Technology, reserve member

This research was partially funded by the Chinese Scholarship Council, China, under Grant 201907720116.

Keywords: Railway transition zone, performance-based design, vehicle–track interaction, surrogate modeling, optimization, sensitivity analysis

Printed by: Ipskamp Printing

Cover by: Yue Shang

ISBN 978-94-6384-589-2

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

To my beloved families.

CONTENTS

SUMMARY

Transition zones in railway tracks, which occur at changes in track form and or substructure properties, are known to cause issues with track degradation. Examples can be found where a track section, usually a ballast track, transits to a slab track to cross a roadway or waterway through supporting structures such as bridges, level crossings, and culverts. These changes in track form and properties cause an abrupt change in track support stiffness, which, in turn, generates additional dynamic forces when a vehicle passes by. Over time, the effect of dynamic impact leads to the development of differential settlement between the settlement-free supporting structures and the connecting track. This further increases loads and accelerates the degradation of the track through successive deterioration of track geometry and components.

The underlying degradation mechanism has been extensively studied in previous work, where engineers typically engage in the development of mechanistic models to facilitate the understanding of the physical processes and develop adapted designs (or countermeasures) to mitigate the dynamic impact at transition zones. This typically involves a parametric study to analyze how changes in the properties of track components, or the countermeasure itself, affect the track performance. However, a parametric study may only consider a limited range of plausible designs, representing a manual and intuition-based design approach. This design process can be time-consuming and may only yield partial improvements in system performance.

On the other hand, parametric optimization is a more structured approach to improving the performance of engineering systems, which seeks to automatically navigate the design space through thousands of model runs. This, however, places greater demands on the models and solvers involved. In the context of railway engineering, this can be seen in vehicle-track models, which take into account the dynamic interaction between the track and moving vehicles. These models tend to be black-box (i.e., simulation-based), high-dimensional, stochastic, and expensive to evaluate.

This observation has led to the following research goal:

- To develop dedicated modeling approaches for optimizing the design of transition zones while addressing three main challenges, including (i) improving the computational efficiency of the optimization process, (ii) embedding multiple design aspects, and (iii) dealing with complexities arising from parameter uncertainties and high-dimensional cases.

To achieve this goal, specific research objectives are defined, which result in four main research outcomes as listed in the following.

- A co-simulation approach combining the capabilities of COMSOL for structural dynamic analysis and MATLAB for dedicated mathematical analysis is proposed to

model the vehicle-track coupling dynamics. This approach aims to ease the generation of complex track models while providing flexibility in model adjustment and data postprocessing. This simulation approach is validated against an existing track model to ensure the accuracy of the vehicle-track dynamic simulation in the optimization process.

- A simulation methodology that incorporates the finite element model into an adaptive surrogate modeling scheme is presented, where a set of optimization problems is formulated to minimize dynamic amplification caused by structural discontinuity within a transition zone. The adaptive scheme allows for efficient exploration of the design space while achieving a reasonable balance between solution quality and computational effort.

- Further, a design approach that builds upon the previous one is introduced to integrate engineering and managerial aspects to model design problems for railway tracks. This approach ensures that stakeholder requirements and preferences, such as passenger riding comfort, are integrated into the engineering design process.

- Finally, a novel sensitivity method called the threshold-based sensitivity method is proposed. This method focuses on the part of the output space that yields failure and allows for the evaluation of model sensitivity near system limit states, making it closely aligned with the formulation of reliability-based design optimization (RBDO). This method can be used as a screening tool for factor prioritization and mechanism reduction, thereby informing the selection of the most influential design variables (related to the intended objectives) and guiding the formulation of RBDO problems.

The developed models have been applied to several engineering examples, including a rail level crossing, a train-track-bridge system, and a truss example, which are all formulated based on the finite element method. Further, different types of surrogate models are applied in this context, including radial basis function, Kriging or Gaussian process modeling, polynomial chaos expansion, and polynomial chaos-Kriging, with considerations of two basic surrogate modeling workflows. The performance of these models is compared. Main findings and observations made during the development and implementation process are drawn in Chapter 6, where limitations and proposed avenues for future research are also outlined.

In conclusion, this thesis contributes to the state-of-the-art developments in the context of railway transition zones. It lays out a methodological basis to connect vehicle-track models with surrogate modeling and formulate optimization problems with specific design purposes to mitigate the degradation problems at transition zones. For future extensions and applications, the methodologies can be easily adapted by researchers and practitioners to specific design intentions, constraints, and variables. The integration of more sophisticated vehicle-track models and other types of surrogate modeling can also be considered.

SAMENVATTING

Overgangszones in spoorlijnen, die voorkomen bij veranderingen in de vorm van het spoor en/of de eigenschappen van de onderconstructie, staan er bekend om problemen te veroorzaken met degradatie van het spoor. Voorbeelden zijn te vinden waar een spoorsectie, meestal een spoor met ballast, overgaat in een spoor met platen om een rijweg of waterweg te kruisen via ondersteunende constructies zoals bruggen, overwegen en duikers. Deze veranderingen in spoorvorm en -eigenschappen veroorzaken een abrupte verandering in de stijfheid van de ondersteuning, die extra dynamische krachten genereert wanneer een voertuig passeert. Na verloop van tijd leidt het effect van dynamische impact tot de ontwikkeling van differentiële zetting tussen de civiele constructie en het aansluitende spoor. Dit verhoogt de krachten nog meer en versnelt de degradatie van het spoor door opeenvolgende verslechteringen van de spoorgeometrie en de componenten.

Het onderliggende degradatiemechanisme is uitgebreid bestudeerd in eerder werk, waarbij ingenieurs meestal mechanistische modellen ontwikkelen om de fysieke processen beter te begrijpen en aangepaste ontwerpen (of tegenmaatregelen) te ontwikkelen om de dynamische impact in overgangszones te beperken. Dit houdt meestal een parametrische studie in om te analyseren hoe veranderingen in de spooreigenschappen, of de tegenmaatregel zelf, de spoorprestaties beïnvloeden. Een parametrische studie kan echter slechts een beperkt aantal plausibele ontwerpen in overweging nemen, wat neerkomt op een handmatige en op intuïtie gebaseerde ontwerpbenadering. Dit ontwerpproces kan tijdrovend zijn en slechts gedeeltelijke verbeteringen van de systeemprestaties opleveren.

Aan de andere kant is parametrische optimalisatie een meer gestructureerde aanpak om de prestaties van engineeringsystemen te verbeteren, waarbij geprobeerd wordt om automatisch door de ontwerpruimte te navigeren via duizenden modelruns. Dit stelt echter hogere eisen aan de betrokken modellen en oplossers. In de context van spoorwegtechniek is dit te zien bij voertuig-spoor modellen, die rekening houden met de dynamische interactie tussen het spoor en voertuigen in beweging. Deze modellen zijn meestal black-box (d.w.z. simulatie-gebaseerd), hoog-dimensionaal, stochastisch en duur om te evalueren.

Deze observatie heeft geleid tot het volgende onderzoeksdoel:

- Het ontwikkelen van specifieke modelbenaderingen voor het optimaliseren van het ontwerp van overgangszones, waarbij drie belangrijke uitdagingen worden aangepakt, waaronder (i) het verbeteren van de computerefficiëntie van het optimalisatieproces, (ii) het integreren van meerdere ontwerpaspecten en (iii) het omgaan met complexiteiten die voortkomen uit onzekerheden in parameters en hoog-dimensionale gevallen.

Om dit doel te bereiken, zijn specifieke onderzoeksdoelstellingen gedefinieerd, die resulteren in vier belangrijke onderzoeksresultaten, zoals hieronder opgesomd.

- Een co-simulatiebenadering die de mogelijkheden van COMSOL voor structuurdynamische analyse en MATLAB voor mathematische analyse combineert, wordt voorgesteld om de dynamica van de interactie tussen voertuig en spoor te modelleren. Deze aanpak is bedoeld om het genereren van complexe spoormodellen te vereenvoudigen en tegelijkertijd flexibiliteit te bieden bij het aanpassen van het model en het achteraf verwerken van gegevens. Deze co-simulatiebenadering wordt gevalideerd tegen een bestaand spoormodel om de nauwkeurigheid van de voertuig-spoor dynamische simulatie in het optimalisatieproces te garanderen.

- Een simulatie methodologie die het eindige-elementen model integreert in een adaptief surrogaat modellering schema wordt gepresenteerd, waar een set van optimalisatie problemen wordt geformuleerd om dynamische versterking veroorzaakt door structurele discontinuïteit binnen een overgangszone te minimaliseren. Het adaptieve schema maakt efficiënte exploratie van de te ontwerpen ruimte mogelijk terwijl een redelijke balans tussen oplossingskwaliteit en rekeninspanning wordt bereikt.

- Verder wordt er een ontwerpbenadering geïntroduceerd die voortbouwt op de vorige om engineering- en managementaspecten te integreren om ontwerpproblemen voor het spoor te modelleren. Deze aanpak zorgt ervoor dat de eisen en voorkeuren van belanghebbenden, zoals het rijcomfort van passagiers, worden geïntegreerd in het engineering ontwerp proces.

- Eindelijk wordt er een nieuwe sensitiviteitmethode voorgesteld, de op drempelwaarden gebaseerd sensitiviteitsmethode. Deze methode richt zich op het deel van de uitvoerruimte dat tot falen leidt en maakt het mogelijk om de gevoeligheid van het model te evalueren in de buurt van systeemlimiettoestanden, waardoor het nauw aansluit bij de formulering van op betrouwbaarheid gebaseerde ontwerpoptimalisatie (RBDO). Deze methode kan worden gebruikt als een screeningsinstrument voor het prioriteren van factoren en het reduceren van mechanismen, waardoor de selectie van de meest invloedrijke ontwerpvariabelen (gerelateerd aan de beoogde doelstellingen) en de formulering van RBDO-problemen mogelijk wordt.

De ontwikkelde modellen zijn toegepast op verschillende technische voorbeelden, waaronder een overweg, een trein-spoor-brugsysteem en een trussstructuur, die allemaal geformuleerd zijn op basis van de eindige-elementen methode. Verder worden verschillende soorten surrogaatmodellen toegepast in deze context, waaronder radiale basisfunctie, Kriging of Gaussische procesmodellering, polynomiale chaos-expansie en polynomiale chaos-Kriging. De prestaties van deze modellen worden vergeleken. De resultaten en observaties tijdens het ontwikkelings- en implementatieproces worden beschreven in hoofdstuk 6, waar ook de beperkingen en voorgestelde wegen voor toekomstig onderzoek worden geschetst.

Concluderend draagt dit proefschrift bij aan de state-of-art ontwikkelingen in de context van overgangszones. Het legt een methodologische basis om voertuig-spoor modellen te verbinden met surrogaat modellering en optimalisatieproblemen te formuleren met specifieke ontwerpdoelen om degradatieproblemen in overgangszones te verminderen. Voor toekomstige uitbreidingen en toepassingen kunnen de methodologieën gemakkelijk door onderzoekers en praktijkmensen worden aangepast aan specifieke ontwerpintenties, beperkingen en variabelen. De integratie van meer geavanceerde voertuig-spoor modellen en andere soorten surrogaat modellering kan ook worden overwogen.

ABBREVIATIONS

ASF achievement scalarizing function

CDF cumulative distribution function

CV coefficient of variation

DEM discrete element model

DOF degrees of freedom

ED experimental design

ERS embedded rail system

FE finite element

FST floating slab track

GA genetic algorithm

GPR ground penetrating radar

KV Kelvin-Voight

LAR least angle regression

LHS Latin hypercube sampling

MOO multi-objective optimization

OLS ordinary least squares

PC polynomial chaos

PCA principal component analysis

PCE polynomial chaos expansion

PCK polynomial chaos-Kriging

PDF probability distribution function

PSD power spectral density

RBDO reliability-based design optimization

RBF radial basis function

RDO robust design optimization

ROSA reliability-oriented sensitivity analysis

SA sensitivity analysis

SVM support vector machines

TQI track quality index

TRV track recording vehicle

TTB train-track-bridge

VTI vehicle-track interaction

1

INTRODUCTION

1.1. RAILWAY TRANSITION ZONE. PROBLEM DESCRIPTION

RAIL infrastructure refers to the physical assets used for the functioning of a railway system. These assets include various components such as lines/tracks, signaling systems, power installations, and other supporting facilities and equipment. Specifically, the railway track is linear and non-redundant [1], the safety and availability of a track line depend on the condition of each individual track section. These sections can be either regular parts of the track, known as the *open track*, or supporting structures that are built to cross waterways, roadways, or valleys through bridges, level crossings, or tunnels [2].

Compared to the open track, the supporting structures are geographically distributed and limited in length, but they have significant implications for train safety, riding comfort, and operational expenses due to local intensified degradation. This frequently occurs in the areas connecting the open track, commonly referred to as *transition zones*. Local degradation often manifests itself as irregularities in track geometry, typically longitudinal level (i.e., differential settlement). For example, Figure 1.1 displays longitudinal level measurements obtained from a level crossing on the Dutch railway network. The red dotted lines represent the interface between the crossing and open track. It shows that the track geometry degradation at the level crossing, especially for the locations near the interface, is more pronounced than the open track, and the level of differential settlements on both sides of transitions is largely influenced by the moving direction.

The underlying degradation mechanism has been extensively studied in the literature. The main factors contributing to the degradation are (i) abrupt variations in mechanical properties and (ii) (unloaded) differential settlement (cf. [2]–[5]).

(i) The substantial variations of mechanical properties (e.g., stiffness, damping, and mass) can be attributed to the presence of non-consistent track forms, such as connections between ballast track and slab track, or a sudden change in geotechnical foundations, such as an embankment transiting onto/off a bridge. These track discontinuities are mainly reflected in longitudinal variations in track sup-

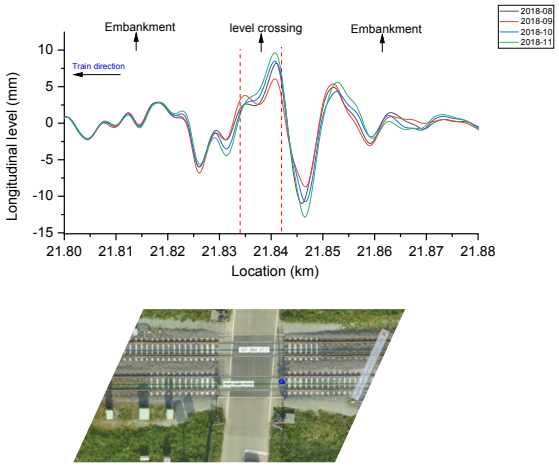

Figure 1.1: A level crossing and track longitudinal level measurement at the location between August and November 2018, wherein the red dotted lines indicate the interface between the level crossing and associated transition zones [3].

port stiffness[1] seen by train, which give rise to differential elastic rail deflections under passing wheels [8]. The uneven deflections disturb the wheel-rail interface and affect how stresses are distributed beneath the track [9], which leads to local but often strong amplification of the response field in the structure [10].

(ii) The open track is often formed by sections of ballast tracks, which over time are susceptible to settlement or plastic deformations. This is particularly true for the ballast layer made up of granular materials, which settles resulting from particle fouling, compaction, and abrasion [3], [8], and for tracks laid on soft soils [11], [12]. In contrast, tracks laid on supporting structures (e.g., bridge decks) or slab tracks (with concrete plates) are designed for minimal settlement. This leads to uneven deformation between sections over a number of loading cycles.

Both factors are sources of disturbance to the train-track interaction. They are also interactive such that the additional dynamic response (e.g., wheel-rail forces) caused by factor (i) is associated with rapid changes in the vertical position of the moving wheels, which combined with factor (ii) will lead to the development of differential permanent settlements under repeated train loading [13]. This further amplifies the response field and exacerbates the local degradation. Without interventions, degradation in track geometry is often associated with *hanging sleepers* [13], i.e., loss of contact between the sleeper and ballast layer when the track is unloaded, and defects in track components, e.g., breakage of ballast particles [14].

In normal conditions, defects in track geometry can be corrected through ballast tamping, which is a regular network-level maintenance activity. Extensive work has fo-

[1] Track support stiffness refers to the equivalent stiffness provided by all the track components located beneath the rail [6], [7].

cused on measuring the track geometry condition and modeling the degradation process, followed by making decisions on optimal scheduling of inspection and maintenance activities [15].[2] However, a further concern is the effectiveness of tamping at transition zones. A trackside measurement was set up on a level crossing in the UK [13]. It showed that hanging sleepers observed before tamping reoccurred very soon after a track renewal, stressing that the local defects in some transition zones may not be effectively rectified by tamping.

MODEL DEVELOPMENT FOR ANALYSIS VERSUS OPTIMIZATION

Studies on transition zones typically involve (i) establishing field measurements to gain insight into transient or long-term track responses at specific locations and (ii) developing mechanistic models to analyze the dynamic behavior of the track under moving loads or vehicles. The mechanistic models can be summarized under two categories.[3]

(i) From the fundamental-theoretical perspective, researchers generally associate the amplification of response field in transition zones to the phenomenon of *transition radiation*, referring to the radiation emitted into the track (in the form of waves) when a train approaches and crosses an inhomogeneity in a railway track (i.e., a transition zone) [10], [16]. The phenomenon is mainly investigated using integral-transform methods. Examples of recent studies are [5], [17], [18]. These studies are able to provide fundamental insight into the underlying mechanism. However, they are often associated with a high level of mathematical abstractness. The models are simplified and may not provide a specific spatial characterization of a system. Therefore, they may be less suitable for making precise predictions of system behavior, especially when dealing with highly localized or detailed features of a system.

(ii) From the engineering application perspective, researchers in this field develop time-domain models with different levels of complexity to study case-specific transition zones. This is often supplemented by field instrumentation for model calibration and validation. The time-domain models are mainly developed using the finite element method and solved by numerical integration, e.g., [4], [19]. This method can be computationally intensive, partly due to the fact that the finite element method deals with finite domains while railway tracks are practically infinite. To eliminate boundary effects, a large volume of the computational domain has to be considered when modeling a railway track. However, the time-domain models are flexible in characterizing structural behavior with regard to various factors, such as the complex geometry of the track [9], [20], nonlinear behavior of the ballast layer [21], [22], and interaction with vehicle motions, soil [23], and supporting structures (e.g., bridges [24]).

The current state-of-the-art in modeling transition zones reflects a learning process where engineers typically engage in the development of increasingly complex models to

[2] This research stream typically adopts a data-driven (or empirical modeling) approach, which is elaborated in Chapter 2.

[3] See an extensive discussion in Chapter 2.

better understand the physical processes leading to the degradation in transition zones. This understanding can then be reflected in adapted designs that incorporate mitigation measures to reduce dynamic impact and improve long-term track performance.

Various mitigation measures have been developed for transition zones in newly built or renewed track sections. A general design guiding principle is to minimize the variations of track support stiffness in the vicinity of the transition zones [25]. Examples of these measures include reducing the sleeper spacing at the interface of a ballast-slab connection [26], varying subgrade filling materials for a tunnel-culvert transition [27], and the use of under sleeper pad for a bridge approach [28]. A comprehensive review can be found in [2].

These adapted designs mostly rely on the analysis and adjustment of design parameters through repetitive model evaluations. The parameters typically include the geometric and mechanical properties of track components, which are tested for an arbitrary range of values. However, determining an optimal combination of the parameters can be challenging, especially when a large set of design variables is involved with high variability. Therefore, adapted designs are often evaluated by perturbing one or two parameters at a time while keeping others constant, a method known as *parametric studies*. This manual, intuition-based iterative design process can be time-consuming and may only yield partial improvements in performance.

A more systematic way of improving the track design is *parametric optimization*, which leverages optimization algorithms to facilitate the exploration of the design space and accelerate the design cycle. The search process allows for a trade-off between a set of design parameters and aims for an optimal combination of the studied parameters that fulfill certain design objective(s) and constraint(s). While a parametric study may only consider a limited range of plausible designs, optimization seeks to explore the design space through thousands of model runs, placing greater demands on the models and solvers involved.

DEVELOPMENT GAP

Previous research in the field of railway engineering has focused on optimizing various design aspects, including rail profiles [29]–[31] and track stiffness [31], [32] in turnouts, railway alignments [33], rail corrugation [34], and vehicle suspension [35]. However, there has been limited research on the design optimization specific to transition zones. The most relevant study in this context [25] employed a multi-objective genetic algorithm (GA) to search for optimal design in a generic ballast-slab connection. GA is an evolutionary algorithm that can handle a range of optimization problems, such as non-convex, mixed-integer, and multi-objective ones, but the method is population-based and operates with many individuals through many generations [36]. This results in thousands of function evaluations to find optimal solutions, and the overall computational cost of optimization can become prohibitively high when the functions involved are computationally expensive.

Optimization in engineering design typically involves the modeling of physical systems to evaluate objective (and constraint, if any) function values for a given design [37]. For railway track design, it is necessary to consider the coupling dynamics between the track and moving vehicles. This is particularly relevant to transition zones where track

discontinuities give rise to unfavorable wheel-rail interaction, which in turn amplifies the response of both systems.

While numerical models have improved in accuracy to simulate the coupling dynamics, the computational time required to run these models has not necessarily decreased. In practice, a single evaluation of these models may take hours. This poses a significant challenge for engineering tasks that rely on them, such as design optimization, reliability analysis, and uncertainty propagation, which require exploration of the design space and involve a large number of model evaluations. Additionally, most of the models used in this context are simulation-based, meaning they are given as black-box models without reliable analytical or numerical derivative information [38]. In such cases, gradient-based algorithms that are efficient in exploring the design space cannot be effectively utilized.

Another obstacle in optimizing vehicle-track systems is the high dimensionality of the models. Improvement in model accuracy, such as capturing the deformation of ballast using the discrete element method[4], has led to models with a growing number of input parameters. This imposes greater computational demands on the optimization process. Specifically, the solver has to examine a larger number of objective function measuring locations, which can quickly render the design process intractable.

Dealing with these high-dimensional models is further complicated by parameter uncertainty. Some vehicle-track parameters are inherently random due to material variability, manufacturing-induced tolerances, and varied operating conditions. This variability may turn a feasible design solution obtained from optimization with deterministic inputs into an unfeasible one [39]. Therefore, it is necessary to account for uncertainties in optimization processes. This can be typically achieved through robust design optimization (RDO) and reliability-based design optimization (RBDO), which address uncertainty in different ways [37].

Both RDO and RBDO represent probabilistic formulations of design optimization problems. Behind the probabilistic optimization is the propagation of input uncertainties through a computational model to output variability as captured by statistical measures such as variance and quantiles in RDO [39], [40], and reliability measures in RBDO [41]. Solving these problems requires the use of forward-propagation methods, e.g., sampling-based techniques. Specifically, RBDO solutions are sought by jointly solving an optimization problem and performing reliability analysis. This can incur significant computational costs, in addition to the optimization process itself, especially when computationally intensive and high-dimensional models are involved.

The development gap identified by this study is the absence of dedicated methods for optimizing the design of railway transition zones, while effectively addressing the challenges posed by the expensive, black-box, and high-dimensional nature of the models used in the search process. Moreover, from an asset management viewpoint, there is a lack of research on integrating the technical and social relevance of the railway track into the engineering design process.

The need to consider long-term track performance in the design process becomes increasingly important due to aging infrastructure and growing traffic demand. This calls for a proper integration of both engineering and managerial aspects into the

[4] See an extensive discussion in Chapter 2.

early design stage. To achieve this, a shift is required from the traditional, single-sided mechanics-based design approach to an integrated, multidisciplinary design optimization approach. This approach accounts for various design aspects such as asset safety, serviceability, and affordability, effectively reflecting the requirements and preferences of multiple stakeholders [42].

1.2. OBJECTIVES AND SCOPE

Following the problem description, the goal of this research is to develop dedicated modeling approaches to optimize the design of transition zones in railway tracks. The methods proposed aim to improve the multifaceted performance of the track structure while targeting three main challenges:

CH-1 Improvement in computational efficiency of the optimization process, specifically concerning the refined use of optimization solvers;

CH-2 Embedding multiple design aspects, transitioning from a strictly mechanical performance focus to include socio-technical considerations;

CH-3 Dealing with complexities arising from parameter uncertainties and high-dimensional cases, specifically concerning the development of a tool for dimension reduction. This tool simplifies the design process and contributes further to improving computational efficiency.

To achieve this goal, the following specific objectives are defined.

O-1 Design optimization of transition zones requires the modeling of the corresponding structure to compute the objective (and constraint) function values for a given design. The first objective is therefore to develop a computational model to characterize the dynamic behavior of transition zones, considering the interaction with the moving vehicle.

O-2 The second objective is to formulate a set of optimization problems with design purposes that cover the technical requirement solely or in combination with the social relevance of railway transition zones. This process translates the design intent into a mathematical statement that can then be linked to the parameterized model developed in objective O-1 and solved by an adequate optimization solver.

O-3 The third objective is to tackle the complexities related to parameter uncertainties and high-dimensional cases, thus addressing challenge CH-3. The aim is to develop a method that can guide or simplify the formulation of RBDO problems for design-under-uncertainty cases. This method serves as a screening tool to identify input factors that are most relevant to the target objective(s), which in essence is a *sensitivity method*. Implementing this sensitivity method can reduce the problem dimensions by screening the most important input factors as design variables while fixing the less important ones as parameters. Consequently, this can eliminate unnecessary model evaluations, thereby reducing the overall computational costs involved in solving RBDO problems.

The methods proposed are grounded in a design principle aiming to smooth variations in track support stiffness along the line. Such variations are a leading cause of degradation in these zones, as discussed under factor (i) in Section 1.1. This focus indicates that the current research primarily concerns the initiation phase of degradation while formulating design optimization problems for transition zones. The idea is to design a transition zone in such a way that the major source of degradation, or the risk of differential settlements, is minimized.

1.3. METHODOLOGY

In line with the defined objectives, this research is conducted in five steps as presented in Figure 1.2.

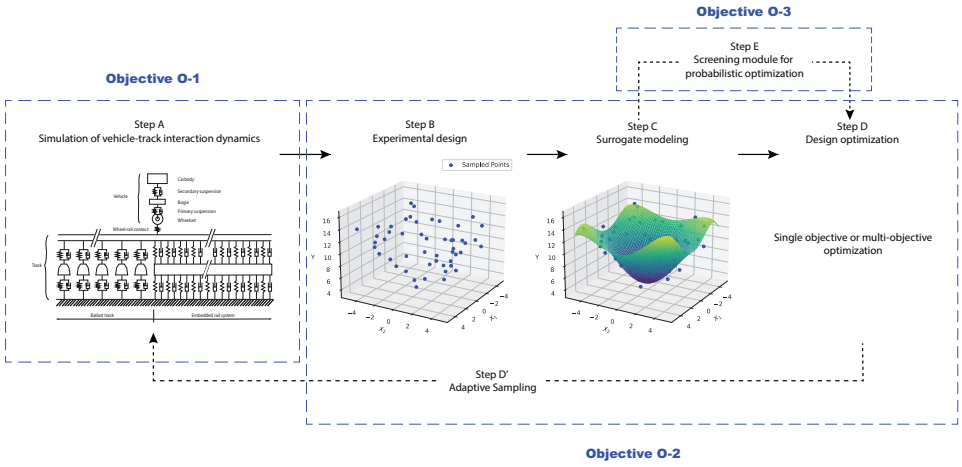

Figure 1.2: Research workflow: Step A provides a schematic representation of a vehicle-track interaction model. Steps B and C are demonstrative figures showing the idea of surrogate modeling based on a two-dimensional function. Step D formulates and solves a set of optimization problems. Step D' is an additional step that can be included or bypassed in surrogate-based optimization. Step E concerns the development of a screening module to assist in the formulation of RBDO problems.

Step A is designed to fulfill objective O-1. It consists in developing a vehicle-track model that is used to analyze the dynamic behavior of a track section (transition zone) under a moving vehicle and compute objective and constraint functions for any given design configurations.

Steps B-D, including Step D', are designed to address objective O-2. These steps form a workflow focused on formulating and solving optimization problems in an efficient manner, which further aids in railway track design. This process leverages *surrogate modeling* techniques and employs the simulation model developed in Step A.

Surrogate modeling, also known as metamodeling or response surface methodology, can be an efficient means to solve engineering problems that involve expensive simulation models. It replaces the original or true function with an approximation that is faster

to evaluate while retaining sufficient accuracy away from the observed points. The decision to implement surrogate modeling techniques in this research was guided by literature research covering domains of operations research, civil engineering (including railway engineering), and machine learning.

As demonstrated in Figure 1.2, surrogate models are constructed by querying the original model (Step A) at given input points. These input points are sampled based on an *experimental design* (Step B). The resulting function values are known as *high-fidelity solutions*, referring to solutions that are evaluated based on the original but computationally expensive functions [43]. Once a surrogate is constructed from the sampled points (Step C), it can be queried to predict output for new input points.

Surrogate-based optimization uses surrogate model(s) to perform optimization tasks. In this context, a surrogate model can be either constructed entirely upfront, thereby bypassing Step D', or it can be sequentially updated through a process known as *adaptive sampling* (Step D'). In the latter case, optimization algorithms interrogate between the surrogate and original models.

Finally, Step E addresses objective O-3 by developing a screening tool or a sensitivity method to reduce the problem dimensions and streamline the probabilistic optimization process, particularly in the context of formulating RBDO problems.

1.4. THESIS CONTRIBUTIONS

A mind map of the thesis is shown in Figure 1.3. This figure highlights the main contributions or outcomes of the thesis in relation to the overarching goal and specific objectives detailed in Section 1.2. It also indicates the corresponding scientific challenges that have been tackled throughout the work.

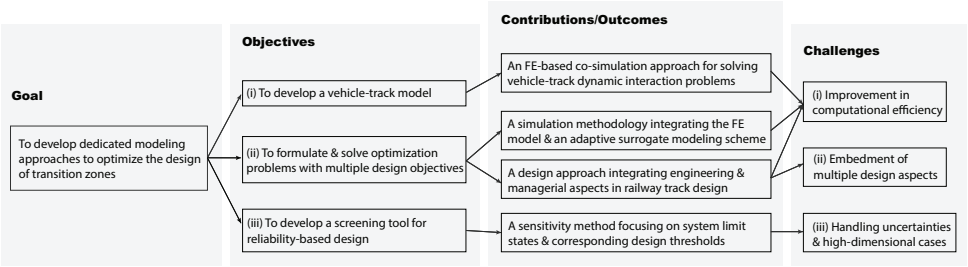

Figure 1.3: Mind map of the thesis.

First, a co-simulation approach combining the capabilities of COMSOL for structural dynamic analysis and MATLAB for dedicated mathematical analysis is proposed to model the vehicle-track coupling dynamics. This approach aims to ease the generation of complex track models while providing flexibility in model adjustment and data postprocessing.

In this approach, the track structure is first generated in COMSOL [44]. The corresponding system matrices are exported to MATLAB [45] via the Livelink interface [46]. The Livelink allows for seamless integration of the two platforms and it is used here for model adjustment and data postprocessing. Simultaneously, the vehicle system ma-

trices are formulated in MATLAB. These two sets of matrices are then coupled to form global system matrices, and the coupled equations of motion that govern the vehicle-track dynamics are solved in the time domain by implementing the Newmark-β integration scheme. Finally, the results generated from this process are post-processed in MATLAB for further analysis. This co-simulation approach is validated against an existing track model to ensure the accuracy of the vehicle-track dynamic simulation in the optimization process.

The second outcome presents a simulation methodology that incorporates the finite element (FE) model into an adaptive surrogate modeling scheme, where a set of optimization problems is formulated to minimize dynamic amplification caused by structural discontinuity within a transition zone. The adaptive scheme allows for efficient exploration of the design space while achieving a reasonable balance between solution quality and computational effort. This methodology is implemented in MATLAB, where the FE model, specifically the track part developed in COMSOL, is parameterized via the Livelink interface. The configuration is fully compatible with the optimization solver and the setup of the defined optimization problems. This includes defining the design variables with their respective bounds, objective functions, and constraints. The objective functions and constraints are defined based on design variables and are implemented as callback functions, which are then passed to the solver. As the solver conducts the search process, it calls back these functions whenever necessary. Additionally, each evaluation of the objective functions requires calling the model in COMSOL.

The third outcome, which builds upon the previous one, introduces a design approach that integrates engineering and managerial aspects to model design problems for railway tracks. This approach ensures that stakeholder requirements and preferences, such as passenger riding comfort, are integrated into the engineering design process. The integration is achieved by incorporating preference modeling into the FE model, thereby optimizing the design of transition zones with stakeholder preferences incorporated.

Finally, a novel sensitivity method called the *threshold-based sensitivity method* is proposed in the thesis. This method allows for the evaluation of model sensitivity near system limit states by generalizing the optimization-based sensitivity method [47] to a wider range of targeted portions of the output space, specifically from extreme values (either maximum or minimum) to thresholds. These quantities are highly relevant to reliability-based design problems, where the focus lies on the regions of the output space that lead to failure. Therefore, the present sensitivity method is targeted to address the specific regions to investigate the relevance of input factors on the intended design objective(s).

This thesis has also led to the development of open-source contributions, including publications and MATLAB-based code files that drive the research outcomes. The author of the thesis is the main author of these publications and is responsible for writing (original & editing) the work presented, as well as the conceptualization, development, implementation, and analysis of the proposed methodology. Table 1.1 provides a summary of the publications included in the thesis, along with their respective code contributions.

Table 1.1: Overview of thesis contributions.

Chapter	Publication[1]	MATLAB code files
2	[3]	Not applicable due to qualitative nature
3	[48]	Available on request due to confidentiality reasons
4	[49]	PreferenceBasedOpt[2]
5	[50]	SA-TTB[3]

[1] References:
[3] Y. Shang, M. Nogal, H. Wang, A.R.M. Wolfert. *Systems thinking approach for improving maintenance management of discrete rail assets: a review and future perspectives*, Structure and Infrastructure Engineering, 19(2), 197-215 (2021).
[48] Y. Shang, M. Nogal, R. Teixeira, A.R.M. Wolfert. *Optimal design of rail level crossings and associated transition zones using adaptive surrogate-assisted optimization*, Engineering Structures, 282, 115740 (2023).
[49] Y. Shang, R. Binnekamp, A.R.M. Wolfert. *Multi-stakeholder service life design for rail level crossings*, in Life-Cycle of Structures and Infrastructure Systems (pp. 949-956). CRC Press (2023).
[50] Y. Shang, M. Nogal, R. Teixeira, A.R.M. Wolfert. *Extreme-oriented sensitivity analysis using sparse polynomial chaos expansion. Application to train-track-bridge systems*, Reliability Engineering & System Safety, 243, 109818 (2024).
[2] Available on 4TU Research Data repository with doi: 10.4121/cc86b3ce-6149-4b77-86ff-fb8f10b48fc7.
[3] Available on 4TU Research Data repository with doi: 10.4121/9a572878-a5b9-4976-a5a4-1de891f55fc8.

1.5. THESIS OUTLINE

The outline of this thesis is presented in Figure 1.4, which is linked to the main contributions as detailed in Section 1.4.

Chapters 2 - 5 constitute the main body of the thesis. Specifically, Chapter 2 is based on [3] and presents a systematic review of studies related to transition zones in railway tracks. The review includes an assessment of track behavior under moving vehicles and modeling of track geometry degradation, which is a primary factor driving frequent maintenance in these areas. The dynamic behavior of the track at transition zones is typically assessed through experimental studies and mechanistic modeling, with the latter approach being the focus of this review. The review proposes a classification of short-term and long-term analyses to structure existing approaches. Meanwhile, the modeling of track geometry degradation largely employs a data-driven approach. The review proposes a taxonomy based on the methods used and modeling purposes to facilitate model comparisons. The interdependencies and synergies between the mechanistic and data-driven approaches are elaborated, which points out limitations in existing studies and possible paths for future research.

Chapter 3 is based on [48], which aligns with objectives O-1 and O-2 (see Figure 1.3) and focuses on the mechanical performance of railway tracks. This chapter addresses one of the gaps identified in Chapter 2 by focusing on optimizing the design of level crossings and associated transition zones. It introduces a surrogate-based simulation methodology including the development of a vehicle-track model to search for an optimal combination of parameters relevant to the geometry and elasticity of track structures.

Chapter 4 is based on [49], which corresponds to objective O-2 by focusing on the

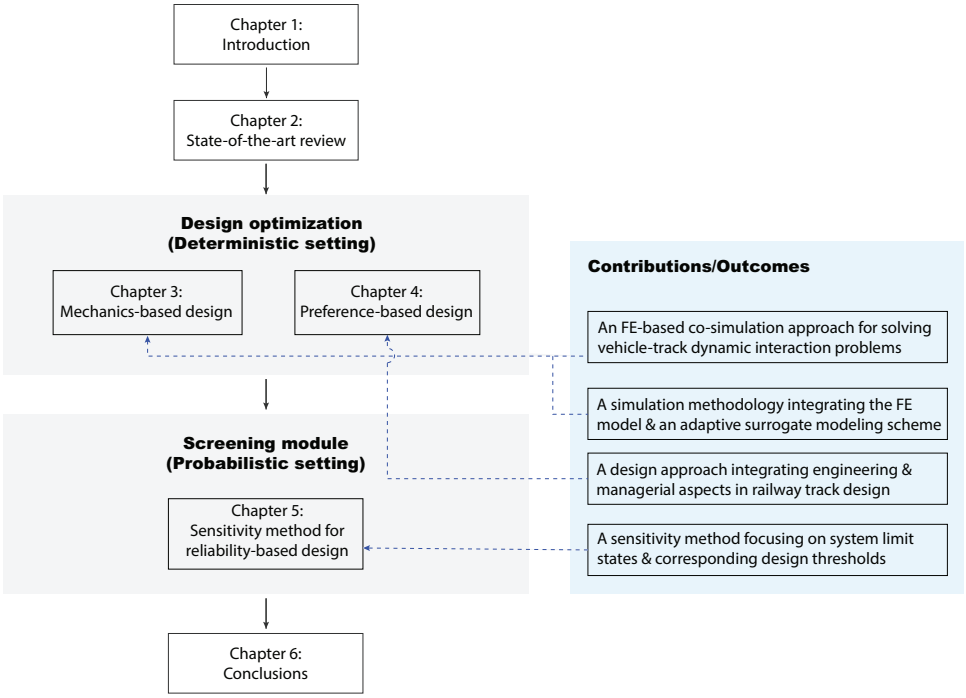

Figure 1.4: Outline of the thesis.

socio-technical relevance of railway track design. This chapter presents a design approach that integrates engineering and managerial aspects to model design problems for railway tracks, where the requirements and preferences of stakeholders are aligned within the engineering design process.

The previous two chapters have addressed the formulation of optimization problems considering deterministic inputs. Chapter 5 moves to the probabilistic formulation, which corresponds to objective O-3. This chapter is based on [50] and introduces a novel sensitivity method that focuses on the part of the output space that yields failure, i.e., for reliability-based design.

Finally, Chapter 6 concludes the thesis with the main findings and limitations of this research, which also outlines potential avenues for future research.

REFERENCES

[1] H. Khajehei, M. Haddadzade, A. Ahmadi, I. Soleimanmeigouni, and A. Nissen, "Optimal opportunistic tamping scheduling for railway track geometry", *Structure and Infrastructure Engineering*, vol. 17, no. 10, pp. 1299–1314, 2020.

[2] B. Indraratna, M. B. Sajjad, T. Ngo, A. G. Correia, and R. Kelly, "Improved performance of ballasted tracks at transition zones: A review of experimental and modelling approaches", *Transportation Geotechnics*, vol. 21, p. 100 260, 2019.

[3] Y. Shang, M. Nogal, H. Wang, and A. R. M. Wolfert, "Systems thinking approach for improving maintenance management of discrete rail assets: A review and future perspectives", *Structure and Infrastructure Engineering*, vol. 19, no. 2, pp. 197–215, 2021.

[4] H. Wang, "Measurement, assessment, analysis and improvement of transition zones in railway track", PhD dissertation, Delft University of Technology, 2018.

[5] A. B. Faragau, "Understanding degradation mechanisms at railway transition zones using phenomenological models", PhD dissertation, Delft University of Technology, 2023.

[6] L. Le Pen, D. Milne, D. Thompson, and W. Powrie, "Evaluating railway track support stiffness from trackside measurements in the absence of wheel load data", *Canadian Geotechnical Journal*, vol. 53, no. 7, pp. 1156–1166, 2016.

[7] C. Shen, R. Dollevoet, and Z. Li, "Fast and robust identification of railway track stiffness from simple field measurement", *Mechanical Systems and Signal Processing*, vol. 152, p. 107 431, 2021.

[8] H. Wang and V. Markine, "Dynamic behaviour of the track in transitions zones considering the differential settlement", *Journal of Sound and Vibration*, vol. 459, p. 114 863, 2019.

[9] D. Milne, J. Harkness, L. Le Pen, and W. Powrie, "The influence of variation in track level and support system stiffness over longer lengths of track for track performance and vehicle track interaction", *Vehicle System Dynamics*, vol. 59, no. 2, pp. 245–268, 2019.

[10] M. J. Steenbergen, "Physics of railroad degradation: The role of a varying dynamic stiffness and transition radiation processes", *Computers & Structures*, vol. 124, pp. 102–111, 2013.

[11] J. N. V. da Silva Ferreira, "Long-term behaviour of railway transitions under dynamic loading application to soft soil sites", PhD dissertation, Universidade NOVA de Lisboa (Portugal), 2013.

[12] B. Z. Coelho, "Dynamics of railway transition zones in soft soils", PhD dissertation, Delft University of Technology, 2011.

[13] L. Le Pen, G. Watson, W. Powrie, G. Yeo, P. Weston, and C. Roberts, "The behaviour of railway level crossings: Insights through field monitoring", *Transportation Geotechnics*, vol. 1, no. 4, pp. 201–213, 2014.

[14] H. Wang and V. L. Markine, "Methodology for the comprehensive analysis of railway transition zones", *Computers and Geotechnics*, vol. 99, pp. 64–79, 2018.

[15] I. Soleimanmeigouni, A. Ahmadi, and U. Kumar, "Track geometry degradation and maintenance modelling: A review", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 232, no. 1, pp. 73–102, 2018.

[16] A. B. Fărăgău, A. V. Metrikine, and K. N. van Dalen, "Transition radiation in a piecewise-linear and infinite one-dimensional structure–a Laplace transform method", *Nonlinear Dynamics*, vol. 98, pp. 2435–2461, 2019.

[17] M. Sadri and M. Steenbergen, "Effects of railway track design on the expected degradation: Parametric study on energy dissipation", *Journal of Sound and Vibration*, vol. 419, pp. 281–301, 2018.

[18] M. Sadri, T. Lu, and M. Steenbergen, "Railway track degradation: The contribution of a spatially variant support stiffness-local variation", *Journal of Sound and Vibration*, vol. 455, pp. 203–220, 2019.

[19] A. L. M. Paixão, "Transition zones in railway tracks: An experimental and numerical study on the structural behaviour", PhD dissertation, Universidade do Porto (Portugal), 2014.

[20] A. Paixao, E. Fortunato, and R. Calcada, "A contribution for integrated analysis of railway track performance at transition zones and other discontinuities", *Construction and Building Materials*, vol. 111, pp. 699–709, 2016.

[21] J. N. Varandas, P. Hölscher, and M. A. Silva, "Settlement of ballasted track under traffic loading: Application to transition zones", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 228, no. 3, pp. 242–259, 2014.

[22] D. Mishra, Y. Qian, H. Huang, and E. Tutumluer, "An integrated approach to dynamic analysis of railroad track transitions behavior", *Transportation Geotechnics*, vol. 1, no. 4, pp. 188–200, 2014.

[23] P. Galvín, A. Romero, and J. Domínguez, "Fully three-dimensional analysis of high-speed train–track–soil-structure dynamic interaction", *Journal of Sound and Vibration*, vol. 329, no. 24, pp. 5147–5163, 2010.

[24] W. Zhai, Z. Han, Z. Chen, L. Ling, and S. Zhu, "Train–track–bridge dynamic interaction: A state-of-the-art review", *Vehicle System Dynamics*, vol. 57, no. 7, pp. 984–1027, 2019.

[25] E. Aggestam and J. C. Nielsen, "Multi-objective optimisation of transition zones between slab track and ballasted track using a genetic algorithm", *Journal of Sound and Vibration*, vol. 446, pp. 91–112, 2019.

[26] R. Sañudo, M. Cerrada, B. Alonso, and L. dell'Olio, "Analysis of the influence of support positions in transition zones. A numerical analysis", *Construction and Building Materials*, vol. 145, pp. 207–217, 2017.

[27] P. Hu, C. Zhang, S. Wen, and Y. Wang, "Dynamic responses of high-speed railway transition zone with various subgrade fillings", *Computers and Geotechnics*, vol. 108, pp. 17–26, 2019.

[28] A. Paixão, J. N. Varandas, E. Fortunato, and R. Calçada, "Numerical simulations to improve the use of under sleeper pads at transition zones to railway bridges", *Engineering structures*, vol. 164, pp. 169–182, 2018.

[29] C. Wan, V. Markine, and I. Shevtsov, "Improvement of vehicle–turnout interaction by optimising the shape of crossing nose", *Vehicle System Dynamics*, vol. 52, no. 11, pp. 1517–1540, 2014.

[30] B. A. Pålsson and J. C. Nielsen, "Track gauge optimisation of railway switches using a genetic algorithm", *Vehicle system dynamics*, vol. 50, no. sup1, pp. 365–387, 2012.

[31] D. Nicklisch, E. Kassa, J. Nielsen, M. Ekh, and S. Iwnicki, "Geometry and stiffness optimization for switches and crossings, and simulation of material degradation", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 224, no. 4, pp. 279–292, 2010.

[32] C. Wan, V. Markine, and I. Shevtsov, "Optimisation of the elastic track properties of turnout crossings", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 2, pp. 360–373, 2016.

[33] W. Li, H. Pu, P. Schonfeld, H. Zhang, and X. Zheng, "Methodology for optimizing constrained 3-dimensional railway alignments in mountainous terrain", *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 549–565, 2016.

[34] O. Oyarzabal, J. Gomez, J. Santamaria, and E. Vadillo, "Dynamic optimization of track components to minimize rail corrugation", *Journal of sound and vibration*, vol. 319, no. 3-5, pp. 904–917, 2009.

[35] R. Alkhatib, G. N. Jazar, and M. F. Golnaraghi, "Optimal design of passive linear suspension using genetic algorithm", *Journal of Sound and vibration*, vol. 275, no. 3-5, pp. 665–691, 2004.

[36] N. Namura, K. Shimoyama, and S. Obayashi, "Expected improvement of penalty-based boundary intersection for expensive multiobjective optimization", *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 6, pp. 898–913, 2017.

[37] J. Martins and A. Ning, *Engineering Design Optimization*. Cambridge University Press, 2021, ISBN: 9781108833417.

[38] Y. Wang and C. A. Shoemaker, "A general stochastic algorithmic framework for minimizing expensive black box objective functions based on surrogate models and sensitivity analysis", *arXiv preprint arXiv:1410.6271*, 2014.

[39] M. Moustapha, B. Sudret, J.-M. Bourinet, and B. Guillaume, "Quantile-based optimization under uncertainties using adaptive Kriging surrogate models", *Structural and multidisciplinary optimization*, vol. 54, pp. 1403–1421, 2016.

[40] M. Moustapha, A. Galimshina, G. Habert, and B. Sudret, "Multi-objective robust optimization using adaptive surrogate models for problems with mixed continuous-categorical parameters", *Structural and Multidisciplinary Optimization*, vol. 65, no. 12, pp. 1–22, 2022.

[41] D. Jerez, H. Jensen, and M. Beer, "Reliability-based design optimization of structural systems under stochastic excitation: An overview", *Mechanical Systems and Signal Processing*, vol. 166, p. 108 397, 2022.

[42] A. R. M. Wolfert, *Open Design Systems*. IOS Press, 2023.

[43] R. Hussein and K. Deb, "A generative kriging surrogate model for constrained and unconstrained multi-objective optimization", in *Proceedings of the Genetic and Evolutionary Computation Conference*, 2016, pp. 573–580.

[44] *COMSOL Multiphysics® version 5.6*. Stockholm, Sweden: COMSOL AB, 2020.

[45] MATLAB, *version 9.11.0 (R2021b)*. Natick, Massachusetts, United States: The MathWorks Inc., 2021.

[46] *LiveLink for Matlab User's Guide*. COMSOL AB, 2020.

[47] M. Nogal and A. Nogal, "Sensitivity method for extreme-based engineering problems", *Reliability Engineering & System Safety*, vol. 216, p. 107 997, 2021.

[48] Y. Shang, M. Nogal, R. Teixeira, and A. R. M. Wolfert, "Optimal design of rail level crossings and associated transition zones using adaptive surrogate-assisted optimization", *Engineering Structures*, vol. 282, p. 115 740, 2023.

[49] Y. Shang, R. Binnekamp, and A. Wolfert, "Multi-stakeholder service life design for rail level crossings", in *Life-Cycle of Structures and Infrastructure Systems*, CRC Press, 2023, pp. 949–956.

[50] Y. Shang, M. Nogal, R. Teixeira, and A. R. M. Wolfert, "Extreme-oriented sensitivity analysis using sparse polynomial chaos expansion. Application to train–track–bridge systems", *Reliability Engineering & System Safety*, vol. 243, p. 109 818, 2024.

2

STATE-OF-THE-ART: TRANSITION ZONES & TRACK GEOMETRY DEGRADATION

This chapter provides a systematic review of ongoing research on the mechanistic modeling of track behavior at transition zones. This is complemented by a thorough synthesis of studies examining the degradation of track geometry, as this degradation is a primary factor driving frequent maintenance in transition zones. In the review, a classification of short-term and long-term analyses is proposed to structure existing approaches for mechanistic modeling of transition zones. Besides, the modeling of track geometry degradation largely employs a data-driven approach, where a taxonomy based on the methods used and modeling purposes is proposed to facilitate model comparisons. These two research streams are interdependent and synergistic, which are reviewed and compared in detail. Limitations in existing studies are identified and directions for future research are outlined.

Parts of this chapter have been published verbatim in Structure and Infrastructure Engineering, 19(2), 197-215 (2021) [1].

2.1. INTRODUCTION

The local degradation of transition zones has been known for a long time, and extensive research has been conducted to understand the underlying mechanisms and evaluate the effectiveness of countermeasures, which can be summarized under two main categories, namely experimental studies and computational modeling.

Experimental studies use laboratory tests or field measurements to investigate the response of different track components at transition zones under various situations. Due to limitations in scale and composition to accommodate relevant features [2], limited work [3], [4] has been carried out in laboratories to model transition zones, whereas more research effort has been made on in-situ measurements to monitor and assess the track response in real-time scenarios. Examples can be found in investigating the dynamic behavior of the ballast-slab connections [4], [5] and transitions to bridges [6]–[11], level crossings [12], [13], and culverts [14]. These investigations are often combined with the assessment of the efficacy of modified design or maintenance solutions to alleviate degradation problems. For example, under sleeper pads [7], wedge-shaped backfills [9], and asphalt underlaid substructure [13] are countermeasures that have been incorporated in modified designs. In addition to conventional tamping, stone blowing is another alternative method used to adjust track geometry conditions [11]. This method involves the use of specialized machines that 'blow' or inject stone ballast underneath the sleepers, raising the track level without disturbing the pre-established packing arrangement of the existing ballast [11].

These studies are very useful in assessing the main features of degradation at transition zones, which highlight the importance of identifying the governing degradation mechanism(s) before implementing a proper mitigation measure. However, they are less viable to evaluate and optimize mitigation measures, as any alteration would require new construction [15]. Besides, the knowledge developed through trackside measurements tends to be site-specific and difficult to generalize [16]. Long-term monitoring of track response is necessary to gain a comprehensive understanding of transition performance, but this is challenging considering many types of transition zones operating under varied conditions.

The use of computational modeling provides an alternative method to investigate the behavior of railway tracks at transition zones. With a single model, multiple design options for a specific transition can be evaluated, and the response of the track structure can be analyzed under various operating scenarios. As such, most of the work has focused on developing mechanistic models to evaluate the track dynamic response at transition zones under train loading.

This chapter presents a systematic review of ongoing research on the mechanistic modeling of track behavior at transition zones, which are classified into short-term and long-term analyses in Section 2.2. The degradation of track geometry, a primary cause for frequent maintenance in transition zones, is explored in Section 2.3. These studies typically follow a data-driven approach that integrates elements of condition measurement, degradation modeling, and maintenance planning. The objective of this approach is to identify the condition of the track geometry and various influencing variables, such as tonnage, train speed, ballast fouling, and moisture content. This information forms the basis for predicting future track conditions and assists in the maintenance decision-

making process. These two research areas are interdependent and synergistic, which are investigated and compared in Section 2.4.

2.2. MECHANISTIC APPROACH

Mechanistic models are developed based on the mechanical properties and layout of all the components that constitute the track structure and vehicles [17]. The track components can be classified based on their principal properties, including those with mass and inertia properties (e.g., rails and sleepers), those with elastic properties (e.g., rail pads), or both (e.g., ballast). Together with the track design, these mechanical properties define the relationship between the forces acting on the track and the corresponding track responses, such as forces, stresses, and displacement [17]. In this context, two paths of modeling solutions are recognized, namely analytical modeling and numerical modeling.

Analytical modeling is suitable for solving problems in a continuous support condition with a limited number of connections and loading positions, which facilitates the retrieval of closed-form solutions to track responses [17]. For railway transition zones, the fundamental insight into the underlying degradation mechanism can be obtained by simplified models, where many researchers have proposed analytical or semi-analytical solutions and they associate the amplification of the response field in transition zones to the phenomenon of transition radiation [18].

The first study on transition radiation of elastic waves was conducted in [18], where an infinite string on a piecewise-homogeneous Winkler foundation is subjected to a constant moving load. Later in [19], both abrupt and smooth variations were considered in the support stiffness to evaluate the corresponding effect on transient vibrations of a string. To account for the flexural rigidity, transition radiation in a beam resting on a Winkler foundation was studied in [20]–[22]. More recent works in this field have been extended to consider a 2D continuum [23], nonlinear elastoplastic foundation [24], vehicle inertia [25], and sleeper periodicity [26].

These studies mainly use integral-transform methods to investigate the transition radiation phenomenon in railway tracks, which facilitates a fundamental understanding of the underlying mechanism. However, the methods often exhibit a high level of mathematical abstraction [27]. The complexity of the equations can increase significantly for nonlinear systems. Therefore, the models in these studies often involve simplifying assumptions and may not provide specific spatial characterization, which renders them less suitable for making accurate predictions [16], especially when dealing with highly localized or detailed features of a system.

Numerical modeling, on the other hand, allows for a more accurate representation of system characteristics, including geometry and (nonlinear) material behavior. Therefore, it has been widely used to simulate the dynamic response with specific features of track inhomogeneity, e.g., stiffness variations, track geometry irregularities, and loss of contact between sleepers and ballast. This represents another research stream that adopts time-domain methods to investigate the dynamic behavior of the railway track at transition zones. Analyses are often case-specific and supplemented with field measurements. These measurements provide detailed site information, which can be used to develop, calibrate, and validate the models.

The time domain models are mainly developed using the FE method, with focus on addressing track geometry irregularities [28], [29], nonlinear behavior of the ballast layer [30], [31], and interaction with vehicle motions, supporting structures (e.g., bridges [32]), and soil [33]. The inclusion of vehicle motions is necessary when examining the dynamic behavior of transition zones, as track discontinuities can result in undesirable wheel-rail interactions that amplify the dynamic response of both systems. Consequently, most time-domain models incorporate VTI dynamics to assess the transition performance. These models typically include three subsystems that represent the behavior of the vehicle, the track, and the interaction between the two.

The vehicle system can be represented by a moving force, moving mass, or moving vehicle-system model. The moving force is the simplest but the dynamic behavior of trains and the corresponding effect on the track vibration are not considered. The moving mass model accounts for the mass and inertia of the running vehicle but neglects the vibration absorbing effect of the suspension system. The moving vehicle-system model can represent the mechanical properties of the vehicles and vary in complexity concerning the vehicle degrees of freedom (DOF). It is established based on the theory of multibody simulation, where the vehicle is represented by an assembly of rigid bodies connected by flexible and massless elements [32]. The bodies typically include a carbody, two bogies, and four wheelsets. Each of them has a maximum of six DOFs and a simplified vehicle model can be achieved by setting physical constraints according to the simulation purpose [34].

The characterization of wheel-rail contact is of significance for analyzing the track performance at transition zones, where the vehicle at some points passes over the unlevelled (track geometry irregularities) and suspended (hanging sleepers) track may cause oscillations and influence the interaction with the track [34]. The behavior of wheel-rail contact is complex and a multitude of contact theories have been proposed to understand, model, and optimize the wheel-rail contact mechanics problem. A comprehensive review of the contact models and related experimental studies can be found in [35]. In general, the problem can be divided into normal and tangential contact [35], where Hertizan and non-Hertzian theories are formulated to solve normal ones, and Kalker's linear theory and its derivatives are the well-known theories for tangential contact.

Despite several restrictive assumptions (see details in [35]), Hertzian contact theory is widely used to analyze wheel-rail contact in studies relevant to transition zones. This means that most time-domain models for analyzing the dynamic performance of transition zones only account for normal contact in wheel-rail interaction. The reason behind this is the necessity for large-scale models to characterize the track performance, considering that one transition zone can extend to tens of meters. Such models can be computationally expensive, especially when complex geometries, nonlinear material behavior, and interactions with other systems are incorporated. Therefore, using simplified wheel-rail interaction models can help reduce overall computational costs.

Modeling of the track structure can be distinguished from the representation of track components. At the superstructure level, the rails are usually modeled using beam elements, such as Euler-Bernoulli or Timoshenko beams. These beam elements can also represent sleepers, which can alternatively be modeled using mass or solid elements. The railpads and ballast are usually modeled by spring elements, and some more ad-

vanced ways to model ballast behavior include solid elements, discrete elements, and lattice models. The track substructure can be represented as rigid (for engineering structures beneath the track), a mass-spring-damper system (Winkler-type), or a 2D/3D continuum.

Numerical models have been proposed extensively to evaluate the track performance at transition zones, which differ in levels of complexity in terms of dimensions, forms of track inhomogeneity, types of mitigation measures, material behavior, and many more. An overview of these studies is provided below, where two aspects are distinguished, namely (i) transient analysis that focuses on the instantaneous or short-term dynamic response during train passage, and (ii) long-term prediction that emphasizes the static change of the track geometry resulting from repeated loading.

2.2.1. SHORT-TERM PERFORMANCE EVALUATION

As discussed in Section 1.1, the main factors contributing to the degradation in transition zones are (i) variations in mechanical properties, typically stiffness variations, and (ii) unloaded differential settlement.

To investigate the effect of factor (i) on the track performance, the track structure was treated as a stochastic system in [36]. An Euler-Bernoulli beam resting on a Kelvin foundation was used to represent a railway track, and the variation in vertical support stiffness was described by a weakly homogeneous random process. Since the randomness in stiffness properties prevents direct analytical solutions, a perturbation approach was proposed to compute the stationary responses of the beam and vehicle, which was further validated against an FE model. Similarly, a stochastic track model was developed in [37] to investigate the influence of randomness in railpad stiffness, ballast stiffness, and dynamic ballast-subgrade mass on track responses, where field and laboratory tests were combined to support the stochastic model. Further in [38], track stiffness data retrieved from a rolling stiffness measurement vehicle was incorporated into the track model to analyze the effect of stiffness variations on the system responses.

The impact study of stiffness variations contributes to the adapted design solutions for transition zones, with the key idea of smoothing the variations along the track line. For instance, a 3D FE model was configured to calculate various geometries of cross sections and geotechnical features of materials [39]. The model was then applied to the key points with abrupt stiffness variations on a Spanish railway line to calculate the corresponding vertical stiffness value, and cross-section designs were proposed for those locations to control the variations. Similarly, various subgrade fillings were evaluated in a 3D FE model for a tunnel-culvert transition to explore economic filling materials [40]. Alternative measures including longer sleepers, auxiliary rails, and improved subgrade were investigated in [41]. For further details on the mitigation measures, the reader is referred to [2].

Factor (ii) is mostly attributed to ballast settlement from fouling and degradation (particle compaction and abrasion), and settlement of fill and subgrade layers. This often appears in bridge approaches and ballast-slab track transitions, where the ballast track at the approaches settles more than the adjacent track on bridge abutments or slab tracks. This issue is particularly severe in soft soil regions [14]. Field monitoring on a culvert transition in the Netherlands revealed that the track settlement consists of two

stages: initial ballast compaction after maintenance, followed by major settlement from embankment and peat layers [14]. By contrast, settlement measurements on a bridge approach in the US showed that the ballast layer is the primary source of differential settlement [8].

Various techniques have been used to incorporate differential settlement in track models, such as reducing local track vertical stiffness [42], describing track irregularities through random processes [43], and imposing assumed uneven settlement profiles [44]–[47], measured variations in track longitudinal level [28], [29], and a transition angle at the rail elevation [48]. Some studies have evaluated the impact of both contributors on transition performance and concluded that differential settlement is more critical than stiffness variations [28], [42], [47], [48].

While the continuum models in FE modeling are widely used, some studies have proposed alternative ways to gain insight into the particle-to-particle nature of load transfer within the ballast layer. This is particularly relevant to the ballast settlement from fouling and degradation. A discrete element model (DEM) that accounts for the particulate nature of varying-sized and -shaped ballast particles was developed to predict the magnitude of field ballast settlement under repeated loading [49]. However, simulating the behavior of ballast particles and transition as a whole using DEM can be time-consuming due to the extended length of transition zones. To address this issue, an integrated approach was proposed in [31], where loading profiles simulated from a validated analytical track model were used as input for a DEM to predict ballast particle accelerations. Further in [15], a lattice model, an alternative to DEM, was proposed to represent ballast particles while assuming that they are of equal size, regularly distributed, and their contacts remain constant. This approach bypasses the need for contact calculations, thereby reducing computational time.

Differential settlement often leads to the issue of hanging sleepers, which has led some researchers to focus on modeling the sleeper-ballast interface in their studies on transition-related problems. Different methods have been utilized to achieve this. For instance, a settlement law was used in [4] to calculate the ballast settlement value at each sleeper. This value was then used as a threshold at the interface to identify gaps between the sleeper and ballast. Further, a piecewise equation was formulated to represent the on/off contact between the sleeper and ballast layer, which was incorporated into a 3D FE model for a culvert transition [50]. A similar approach was reported in [28]. An alternative method involves using solid elements to model the sleeper and the ballast layer, allowing for their interaction to be represented by surface-to-surface contact [14], [42], [47].

The aforementioned studies provide valuable insights by simulating transient dynamic responses under train passages at various levels and locations of track components. They provide a comprehensive understanding of the physical mechanisms driving track degradation. However, the resultant transient analysis is not directly applicable to represent and predict the evolution of track behavior, as this requires an extension of the analysis to include the following long-term analysis.

2.2.2. LONG-TERM PREDICTION OF TRACK DEGRADATION

Numerical modeling can be used to predict long-term track performance, where much attention has been paid to the degradation of track geometry at the longitudinal level, i.e., track settlement. The prediction process typically involves integrating a VTI model for transient dynamic analysis with an empirical equation to account for permanent deformation in the ballast or subgrade. This is then followed by an iterative procedure that consists of two modules in the following.

As shown in Figure 2.1, one is the *dynamics calculation module* with a VTI model to obtain track responses during train passage, including the wheel force [51], [52], sleeper-ballast contact force [53], [54], track-subgrade contact stress [55], sleeper deflection [30], and ballast stress [56]. The other is the *cumulative settlement calculation module* considering the repeated loading, where an empirical settlement equation is coupled with the VTI model by the following calculation procedure: (1) the simulated response is used as input to the settlement equation; (2) the transient dynamic analysis is updated in each iteration to account for the new state of the track response; (3) the accumulated settlement is calculated through repeated procedures until a certain limit value is reached, e.g., the total number of loading cycles.

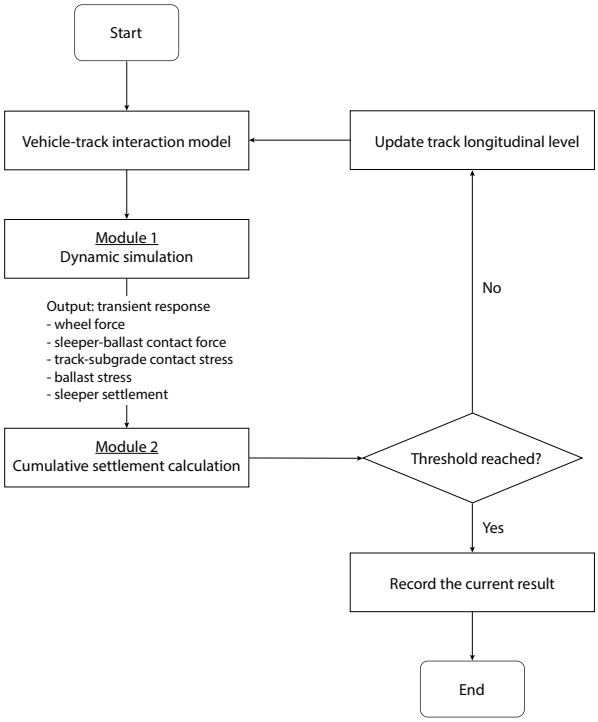

Figure 2.1: Flowchart of the iterative procedure for settlement prediction.

The track models used in Module 1 are mostly developed using FE modeling, with exceptions using the finite difference method [51] and multibody simulation [53]. For the prediction of track settlement, various empirical models have been developed over the

years, and a critical review of the relevant works is presented in [57]. Typically, in transition zones, Sato's settlement law [58] has been coupled to FE models for the prediction of ballast settlement at bridge approaches [54], [56]. This model takes into account various parameters, including annual tonnage, train speed, structure factor, whether the rail is jointed or continuously welded, and the quality of the subgrade. However, considering that the existing settlement models cannot accurately incorporate the loading history (e.g., with continuous changes in loading magnitude), a new settlement model focusing on the magnitude and evolution of ballast forces was proposed in [30]. This model was configured to account for hanging sleepers and the nonlinear constitutive behavior of the ballast. Subsequently, the focus shifted from predicting ballast settlement to considering the subgrade plastic deformation in slab tracks [55]. An FE model was developed to account for the weight of the track and any local contact loss between the track and subgrade.

Those coupled models for simulating long-term track degradation can be seen as a hybrid approach that fuses mechanical responses from physics-based models with empirical relations. Physics-based models are crucial in mapping the relationships between mechanical properties and responses, thereby contributing to the fundamental understanding of the subject. However, the use of empirical relations has some limitations.

First, these empirical equations were extrapolated site specifically, mainly relying on the number of loading cycles and/or load magnitude, but not considering the properties of the ballast and subgrade [57]. This may limit their ability to account for the actual track condition and to be generalized across varied operational and environmental conditions, especially at transition zones. In these track areas, the degradation process is accelerated compared to the open track, and the empirical relations for settlement prediction may become inapplicable.

Second, the applicability of these empirical equations in the track models cannot be validated at the current stage. This restricts the reliability of the prediction results for maintenance planning. Nonetheless, settlement prediction is useful for comparing the performance of different mitigation measures aimed at reducing the need for regular maintenance, such as tamping.

In view of the limitations, a recent update in [59] integrated a geotechnical rheological model into a 3D FE model. This model aims to include the viscoelastic–plastic behavior of the substructure layers in the prediction of settlement at transition zones. To model the plastic response of geomaterials, slider elements were used, which are described by constitutive laws rather than empirical relations used in most of the previous studies. The methodology was then applied to a bridge approach and results showed that increasing the thickness of the granular layer can enhance the performance of transition zones with weak subgrade.

2.3. DATA-DRIVEN APPROACH

Instead of relying on the hybrid approach to predict track settlement, there is another research stream that uses a data-driven approach to predict the degradation process of track geometry. The scope of these studies is not restricted to transition zones characterized by one-dimensional track geometry degradation (settlement), but to quantify the overall track geometry quality by looking at a set of track geometry parameters, including

longitudinal level, gauge, alignment, cross-level, and twist, as shown in Figure 2.2.

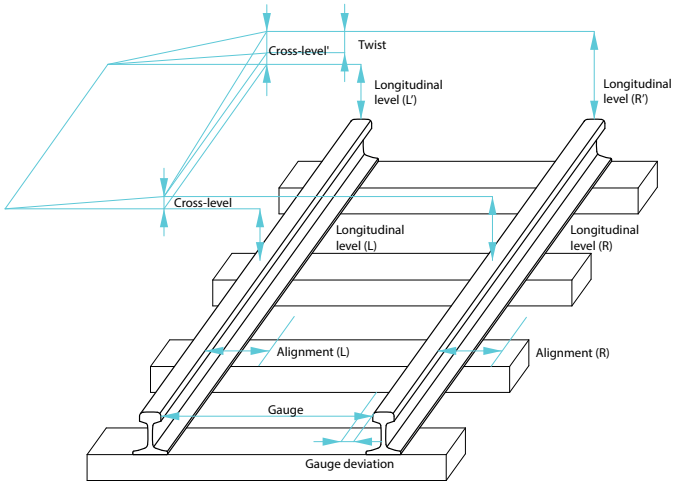

Figure 2.2: Schematic representation of track geometry parameters.

Initially, railway tracks are laid with an ideal geometry condition, which is defined by nominal values. Each geometry parameter is associated with a nominal value. However, as a result of usage, local deformation and wear lead to deviations from these nominal values. Once these deviations reach a certain threshold, they become evident as track defects, alerting the need for maintenance. Among various types of defects, longitudinal level defects are of primary concern. They act as the main source of excitation in the train-track system, making them a critical factor in studies that evaluate the quality of track geometry. This is particularly relevant for studies that predict track settlement at transition zones, which rely on a combination of mechanistic modeling and empirical settlement equations, as discussed in the previous section.

Within condition-based maintenance regimes, the modeling of track geometry degradation relates directly to the process of condition measurement and maintenance planning. Condition measurement involves regular inspection or sensor-based condition monitoring, followed by data processing to extract key features that characterize the track condition. Subsequently, degradation modeling is employed to construct deterioration curves, which analyze patterns of track degradation over time while accounting for variables such as tonnage and train speed. This process provides valuable insights into the mechanisms behind the geometry degradation and can also be used to predict future track conditions. Maintenance optimization strategies are then developed to ensure optimal intervention planning, ensuring track safety while balancing service continuity and cost efficiency. Further elaboration on these aspects is provided in the following sections.

2.3.1. TRACK CONDITION MEASUREMENT AND CHARACTERIZATION

Methods for track geometry measurement generally fall into two categories, namely, track recording vehicle (TRV) and onboard vehicle dynamics measurements. TRV is a

mature way of measuring the track geometry. It is a special self-propelled vehicle dedicated to measuring, processing, assessing, and storing track geometry parameters. Depending on the measurement techniques, two main principles are distinguished, i.e., chord and inertial measuring systems.

(i) The chord method measures the track geometry based on a straight-line chord reference, where the mid-chord amplitude is taken as the measured output. Examples can be found in EM120 of Iran [60], [61].

(ii) The inertial method requires an inertial system as a reference, such as carbody, to measure its relative position with the rail in different dimensions. Examples refer to STRIX and IMV100 in Sweden [62], [63], GJ-4 in China [64], [65], and UFM120 in the Netherlands [66].

More recently, research has increasingly focused on monitoring track geometry through onboard measurements [67]. A comprehensive review of onboard sensors for track geometry measurement can be found in [68]. Compared to TRV, this approach is more cost-efficient and allows for frequent geometry assessments, which facilitates extensive data analytics on track geometry. Relevant studies in this field can be broadly categorized into two approaches: model-based and signal-based [67]. Model-based approaches establish mathematical relationships between input and output signals of a dynamic system (e.g., [69]–[71]), whereas signal-based approaches involve the use of signal processing, statistical analysis, and machine learning techniques on system response signals to infer conclusions about the input data (e.g., [67], [72], [73]). In these studies, track geometry parameters are considered as the input signals, while vehicle dynamics, measured from the axlebox, bogie, or carbody, represent the system responses. These measurement locations are depicted in Figure 2.3.

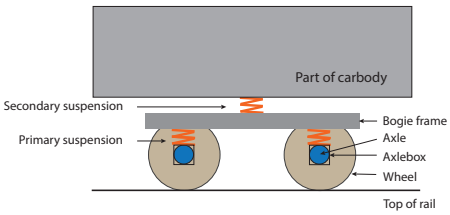

Figure 2.3: Main components in a vehicle system and locations of axlebox, bogie, and carbody.

Once gathering the data from measurements, evaluating and making decisions on each track geometry parameter per unit length is practically difficult as this results in large volumes of data. Often, track quality index (TQI) is utilized to aggregate various track geometry parameters with wavelength variations. The change of the TQI values provides an aggregate-level picture of individual track segments for asset managers to design interventions, where standard deviation and mean over a defined length and power spectral density (PSD) are among the standardized TQIs in EN 13848-5 [74].

Various railroad administrations have tailored their own TQIs to fit the local network characteristics, which can be generally categorized into objective TQI and synthetic/artificial TQI [75]. The former refers to using individual-parameter measurements to formulate an indicator and addresses a specific aspect of the track geometry quality, e.g., track roughness index (the US) [76] and Canadian TQI [77]. The latter develops a mathematical function that describes the track geometry quality by aggregating all the parameters into one equation, e.g., China railway TQI [64], [78], Q index from ProRail of the Netherlands, and Sweden TQI [79].

However, they lack consensus in approach, resulting in conflicting inferences on the track condition [61]. Synthetic TQIs are often dimensionless and lack physical meaning. Some select specific track geometry parameters while dropping others, or assign subjective weights of the parameters to the synthetic indexes; also, aggregating track geometry measurements for an extended length of track may miss exceptions implying safety risk [75].

In order to address the identified limitations, several researchers have developed objective TQIs using unsupervised machine learning. In [75], principal component analysis (PCA) was applied to combine 31 track geometry features into a low-dimensional form that retained much of the data variability. The resulting three principal components were tested better at predicting the defects than the synthetic TQIs. In later work [80], they extended their approach by utilizing both PCA and T-stochastic neighbor embedding as dimension reduction techniques on the track geometry data. Additionally, they integrated safety concerns into the TQI to capture track geometry outliers in the index [81].

Statistical analysis approach has also been used in the development of novel TQIs. In [82], the correlation between existing and previous values of geometry parameters was measured by Pearson correlation, where a strong correlation was identified in gauge and twist, and they were incorporated in TQI development. Similarly, track geometry data was investigated by statistical distributions [76]. A normal distribution pattern was found that best fits frequency curves of the geometry parameters, and new TQIs were developed based on the distribution features. The current deterministic TQIs cannot capture the inherent uncertainty when classifying the track condition against maintenance thresholds. To address this issue, a stochastic TQI based on a Bayesian framework was proposed and applied to a 900-km long track line in Iran [61].

2.3.2. PREDICTION OF TRACK GEOMETRY DEGRADATION

After converting raw data to a quantifiable TQI, a mathematical function can be formulated based on the TQI to model the track degradation process and predict its future state. This information can then inform the maintenance decision-making process.

Initially, settlement is considered to be the controlling degradation factor in the ballast tracks, and many researchers developed degradation models describing the ballast settlement due to its major role in the overall track settlement. The quantitative modeling of degradation for granular and porous materials is extremely complex and sensitive to specific material properties so many settlement relations have been tuned to fit the particular data either from in-situ or laboratory tests [57]. A detailed review of these models is presented in [57]. These mathematical formulations are empirical mod-

els. Compared with the mechanistic models, they are generally easier to handle but lack physical interpretations. Renewed interests in these models refer to their coupling with the vehicle-track interaction models for numerical prediction of track settlement at transition zones, as mentioned in Section 2.2.2.

The significant improvement in track geometry measurement techniques, especially the onboard measurements, enables access to large volumes of data reflecting the real track condition, where statistical models have been widely developed and an emerging research stream applies machine learning tools for predictive analytics of track degradation. Several review articles have been published on statistical modeling of track geometry degradation. The current review built upon these reviews incorporates the recent literature on machine learning applications and provides a taxonomy based on the methods and modeling purposes to facilitate the model comparison, which is presented in Table 2.1.

The field of track geometry degradation modeling includes both statistical and machine learning models. The latter as a sub-field of computer science refers to the ability of a system to learn and improve performance from experience, which is widely understood as methods that analyze data, extract patterns, and make predictive analysis from often rich and unwieldy data. Machine learning has its foundation in statistics, but the primary difference lies in the volume of data involved [101].

The use of machine learning techniques for track geometry data analytics can be broadly classified as unsupervised and supervised learning methods. In unsupervised learning, data is not labeled and response variables are not observed. The motive is to uncover hidden patterns within the input data, which can then be used to extract objective TQIs from the track geometry data, as described in Section 2.3.1. Clustering analysis and dimension reduction are the primary classes of algorithms used in unsupervised learning. On the other hand, supervised learning involves the use of observable response variables to guide the learning process. It deals with predictive analytics based on labeled data for both input and response variables. Classification and regression algorithms are the primary groups of algorithms used in supervised learning.

STATISTICAL APPROACH

The first sub-category of statistical approaches involves using deterministic models to describe track geometry degradation. Regression techniques have been widely used to establish relationships between track geometry degradation and explanatory variables, such as time, accumulated tonnage, speed, and subsoil type. These techniques range from simple linear regression to multivariate regression and exponential regression. Considering the nonlinear degradation process, the process was divided into narrow time slots in [65], [78], and least squares regression was used to approximate the degradation over each time slot. A repeated substitution was made in the process by using updated inspection data, resulting in a family of estimated regression equations that forms a prediction model.

Multiple explanatory variables such as the subsoil type, sleeper type, tonnage, and engineering structures beneath the track were incorporated in a multivariate regression model to analyze their effect on the track geometry degradation [66]. Further, a log-transformed regression model was developed to map the degradation with explanatory

Table 2.1: Summary of data-driven approaches used for track geometry data analytics.

Approach	Subcategory	Example models	Modeling purposes	Reference
Statistical approach	Deterministic models	Linear regression, exponential regression	To simply model track geometry degradation with regard to time/tonnage	[65], [78], [83]–[85]
		Multivariate regression, logistic regression	To establish a relation between degradation and influencing factors	[66], [86]–[88]
		Logistic regression, survival analysis	To estimate the probability of occurrence of isolated geometry defects	[83]–[85], [89]–[91]
	Probabilistic models	(Continuous) Gamma process, Wiener process; (Discrete) Markov chain	- To capture the uncertainty of track geometry degradation over time - To estimate the time period when (1) the degradation path hits maintenance thresholds (continuous) or (2) the condition state is transferred to the next state (discrete)	[64], [92]–[94]
	Hybrid models	Linear regression coupled with autoregressive moving average model; Bayesian framework coupled with regression and conditional autoregressive model	To account for spatial correlation of degradation in adjacent track sections	[84], [85]
Machine learning approach	Classification	Support vector machine, decision tree, ensemble learning (e.g., random forest), linear discriminant analysis, Naïve Bayes	- To predict the (discretized) track state for next intervention - To predict the occurrence of geometry defects	[67], [75], [89], [95]–[97]
	Regression	Decision tree regression, random forest regression, support vector regression, Neural networks	To predict the track condition represented by a continuous value of TQI	[98], [99]
		Neural networks	To predict the track condition considering complex relationships between track degradation and influencing variables	[98], [100]
	Clustering	Hierarchical clustering, k-means clustering	Group geometry data points according to their similarity to evaluate the effect of interventions on track geometry condition	[99]
	Dimension reduction	Principal component analysis, T-stochastic neighbor embedding	- Reduce geometry data from higher-dimensional space to lower dimensions - Produce objective TQIs to characterize the degradation	[75], [80], [81], [89], [99]

variables, where a survival model characterizing the derailment risk and an optimization model for maintenance planning were coupled [86]. Similar approaches have been applied in [87], [88] with different operating contexts and influencing variables. Some other studies also adopted linear regression to model the track geometry degradation path but extended the models to link covariates such as the tamping effect and spatial dependencies between adjacent track sections to make the models more realistic [83], [84].

The current synthetic TQIs provide an overall assessment of the track segment condition, but they may not detect isolated track geometry defects that exceed the set thresholds. Some studies have addressed this limitation by including these defects in the model development to predict corrective maintenance needs. This has been done using methods such as logistic regression, survival analysis, and classification algorithms. Classification algorithms, as a branch of machine learning, will be discussed in the following section. Logistic regression, a form of regression analysis, is specifically designed to address binary classification problems. When the degradation path exceeds the threshold, it is identified as a geometry defect, allowing the track condition to be categorized into either normal or defective states. In such scenarios, logistic regression proves to be an effective technique for addressing this classification problem. Relevant applications of this method are documented in[83], [84], [89], [90]. Additionally, survival analysis is extensively used to model the uncertainty in system lifetimes. Within this context, the Weibull distribution, a prevalent choice in survival analysis, is applied to estimate the probability of track defects [91].

Track degradation is a stochastic process that is influenced by heterogeneous factors along the track. Deterministic models that only consider the nominal degradation behavior may not be sufficient to develop robust maintenance policies in the presence of randomness. The second sub-category of statistical approaches uses probabilistic models to describe the degradation dynamics and incorporate theories from stochastic processes, Bayesian inference, and other related fields to account for the inherent uncertainty in the degradation process.

The gamma process is a stochastic process with independent and non-negative gamma distributed increments [102]. This feature makes it well-suited for characterizing monotonic degradation processes, such as track geometry irregularities that can only grow without intervention. A gamma process model has been proposed in [93] to describe the evolution of longitudinal level defects. Based on this model, a cost model was developed to optimize maintenance. Further, the work was extended by incorporating alignment into the prediction through the use of a bivariate gamma process [94].

The Wiener process relaxes the monotonicity property, thereby allowing for variations in the degradation level caused by interventions and measurement errors. The process starts at zero and it is continuous in time with independent and Gaussian increments. The degradation path of track geometry parameters was formulated through the Wiener process in [92]. The model parameters were estimated using Bayesian inference, and the failure time within a maintenance cycle was estimated from the degradation sampling paths. Rather than treating the degradation process as continuous, the track condition was classified into four ranks based on Chinese TQI [64]. The deterioration process was described as Markov chains, where the transition probabilities between

states incorporated tonnage and line horizontal layout as explanatory variables.

Stochastic process-based models are valuable tools for maintenance decision making, as they can predict the track condition within a maintenance cycle. In continuous processes, the track section is considered defective when the selected TQI exceeds a predefined threshold, calling for intervention. Figure 2.4 shows that the first hitting time when the degradation path surpasses the threshold can be estimated and its inherent uncertainty is quantified by a probability distribution. This information contributes to maintenance knowledge regarding the remaining time before intervention. In discrete processes, track condition is classified into finite states, each associated with a maintenance decision. However, the limitation of these Markov processes is their basic working principle, Markovian property, where the future state is only based on the current state, independent of the past state. Additionally, the complexity of the models may restrict their applicability when analyzing more track geometry data covering multiple lines or an entire network.

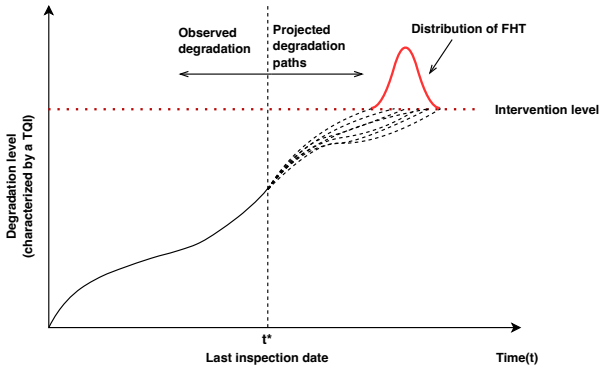

Figure 2.4: Schematic representation of a degradation path modeled by a continuous stochastic process (FHT: first hitting time).

Several researchers have analyzed the temporal variability of the degradation process using stochastic processes, but only a few have considered its spatial dependencies in consecutive track segments. Since neighboring segments with similar structural and operational features tend to exhibit similar degradation patterns, this type of modeling requires a hybrid model that combines a regression model with techniques specifically designed to address spatial variability in regression parameters [84], [85]. Hence, this type of modeling is classified as a hybrid model in Table 2.1.

MACHINE LEARNING APPROACH

Supervised learning algorithms are mostly applied to predict the track geometry quality, which can be further categorized into classification and regression problems. The former is applicable when the output variables are discrete or categorical, while the latter deals with continuous variables.

Classification is the process of assigning a new instance to a specific category based on the knowledge acquired from the training of previously observed instances. Considering the track condition (featured by a TQI) is either within or beyond a threshold, the

problem is a well-posed binary classification problem [67]. Decision trees and support vector machines (SVM) have been used to predict the occurrence of defects in this type of problem. SVM separates data into different classes by transforming them into high-dimensional space with a kernel function and dividing them with decision boundaries [67], [89]. The kernel trick makes SVM a unique classifier that can map non-separable data into high-dimensional space and make it separable [67]. Decision trees, on the other hand, have better interpretability and representation. It is a rooted tree that splits a complex decision into several simpler and more interpretable ones [99]. The TQIs serve as predictors, and the classes to be mapped are target variables, which formulate a top-down approach to constructing a decision tree.

Initially, SVM was proposed to predict changes in defect amplitude while accounting for the effects of track class, traffic volume, and inspection intervals [96]. The model was trained using irregularities in longitudinal level, cross-level, and dip (the maximum value of track longitudinal leveling within a certain length). Later in [97], the same types of defects were considered as inputs, and an ensemble methodology was applied to estimate the occurrence of more severe defects. Further in [89], the performance of three models (random forest, SVM, and logistic regression) was compared for predicting the defect occurrence based on a record of track geometry defects. A total of 30% of recorded defects were used as test data, and random forest demonstrated the best prediction performance.

Instead of binary classification, the track condition was divided into four rank states based on the Chinese TQI and maintenance policy [95]. A tree-augmented naïve Bayes classifier was developed to predict the state of the track for the next inspection. While most studies focus on analyzing longitudinal level irregularities measured from TRV, de Rosa et al. [67] used SVM and decision tree to monitor alignment and cross-level irregularities from lateral and roll bogie accelerations. During the training phase, they only utilized data simulated from a multibody simulation, and the trained models were tested against the onboard measurement.

A few studies have approached the prediction of track geometry quality as a regression problem. Martey et al. [99] studied a mile of a railway track in the US, where a renewal was conducted during the analysis period. They combined unsupervised and supervised learning on TRV data to estimate the effect of geocell installation on the track geometry condition. Considering track geometry degradation is influenced by heterogeneous factors, a neural network model was trained using field data to map the relationships between the degradation and influencing variables [100]. The influencers in the model include eight mechanical inputs and four environmental factors, which were mostly treated as dummy variables. The study showed that neural networks are particularly useful in learning complex relationships between track conditions and multiple interacting factors related to track design, environment, and operational conditions. Another example of this approach is provided in [98].

2.3.3. MAINTENANCE INTERVENTION PLANNING

Based on the prediction from degradation modeling, maintenance planning determines when and where to perform the maintenance over a planning horizon. Tamping is a widely studied method for correcting track geometry defects. Several optimization prob-

lems for tamping scheduling have been formulated, such as integer linear programming [103], [104] and mixed-integer linear/nonlinear programming [105]–[109]. Parameters to optimize include cost, possession time, the total number of tamping operations over a planning horizon, and track conditions captured by TQIs. Another important consideration in optimization is to model the effect of tamping on track geometry quality, which includes changes in degradation level and rate, as tamping is an imperfect maintenance action.

At transition zones, however, local track geometry issues may not be effectively rectified by tamping. This is shown by Le Pen et al. [12], who found that hanging sleepers reoccur soon after tamping is applied at a level crossing approach. Later in [63], [110], the dependency between track geometry quality and longitudinal variations of vertical stiffness was investigated using field data. Results showed that uneven profiles with high degradation rates often occur on track sections where there is a combination of high gradient and low substructure stiffness.

Moreover, the occurrence of track geometry defects was linked to subgrade parameters measured by ground penetrating radar (GPR) [111]. GPR is a continuous non-destructive testing method that measures layer configuration, moisture content, and fouling condition to provide a detailed picture of the ballast and substructure conditions. The results revealed a significant relationship between high rates of track geometry degradation and poor track subsurface conditions. In such cases, tamping may not be a cost-effective long-term solution, and upgrading the ballast or subgrade layer should be considered to tackle the problem.

2.4. PERSPECTIVES AND FUTURE DIRECTION

Research on track geometry degradation, while not directly targeted to transition zones, complements studies in analyzing the track behavior at transition zones. Table 2.2 summarizes a comparison of the two approaches. Mechanistic models focus on micro-level investigations and rely on first principles to inform the choice of variables and the form of the model. On the other hand, the data-driven approach scales up to the macro level and relies on historical data from large-scale network measurements gathered through TRV or onboard measurements.

Mechanistic models may consider short- and long-term analyses. Short-term performance evaluation is useful in investigating the track responses at various levels and locations of track components during train passage. They can adapt to different site situations, and one single model can work out multiple design options through parametric studies, lending itself to testing various track design solutions to improve track performance and reduce maintenance routines. The long-term numerical prediction focuses on track settlement, mostly embedding an empirical settlement law in an iterative computation procedure. However, the mechanical properties of vehicles and tracks may vary in time and space, while the settlement equations were extrapolated site specifically and not based on constitutive laws. They are limited in accounting for the spatial and temporal variability in degradation, hampering the accurate prediction of the track settlement and generalization of the results to other sites of interest.

Data-driven models have better predictive capability. Probabilistic models can account for the uncertainty in the evolution of track degradation; multivariate regression

Table 2.2: Comparison of mechanistic and data-driven approach.

	Mechanistic approach	Data-driven approach
Aim	- To provide physical interpretations of structural degradation - To analyze the impact of system characteristics and operational variables on track degradation - To improve engineering design and maintenance measures at an individual asset level	- To discover complex relationships between track degradation and influencing factors (both endogenous and exogenous variables) - To predict track degradation for maintenance planning
Required data/information	- Site-specific measurement (e.g., accelerometer, deflectometer, geophones, and digital image correlation) - Track & vehicle design parameters - Operational characteristics	- Network measurement: track geometry measurement (TRV or onboard sensors) - Maintenance history - Operational characteristics - Environmental conditions
Approach	- Analytical modeling - Numerical modeling	- Statistical approach - Machine learning approach
Applicable context	- Mostly transient analysis to investigate dynamic track responses under train loading - Long-term analysis where a numerical model is coupled with an empirical model (or a constitutive model) for settlement prediction	- Long-term analysis for track geometry defect prediction and maintenance planning - Synchronize track measurements to investigate potential correlation and identify the root cause of track degradation

and neural network models can reveal complex relationships between track performance and exogenous variables such as the ballast fouling condition and moisture content in the substructure, which cannot be handled by the mechanistic approach. However, the prediction accuracy of the data-driven models depends on the quality and quantity of historical data. These models are blind to physical sources of degradation and unable to account for internal factors related to the structure itself.

Both types of research contribute to a deeper understanding of the degradation mechanisms of the railway track. Mechanistic modeling aims to investigate the root cause within the train-track system, whereas the data-driven approach can reveal the relationships between degradation and exogenous variables related to operational characteristics, maintenance history, and environmental conditions that cannot be addressed by the mechanistic approach. Moreover, it allows for mapping potential correlations between different types of defects, such as the effect of track geometry irregularities on the occurrence of rail defects [112], [113], and the effect of substructure conditions on the development of irregularities [63], [110], [111]. Gaining such insights provides valuable information for the development of innovative maintenance techniques or design solutions.

Through the systematic review provided above, the following gaps are identified in the existing research. There has been extensive research on the degradation mechanisms of transition zones. However, limited attention has been paid to level crossings and associated transition zones, aside from a few experimental studies [12], [13]. There are various ways to carefully design transition zones onto bridges and over culverts to minimize dynamic loading. However, there is a lack of recognized transition designs for level crossings, which is also confirmed in Le Pen et al. [12]. Level crossings are common areas where ballasted tracks meet slab tracks (e.g., the use of embedded rail systems in the crossing). The optimal design for both crossing and transition zones is seldom studied and synchronized. Also, as level crossings represent jointly used areas by rail and road traffic, a balanced crossing design is necessary that provides gradual and smooth transitions to both the roadway and highway approaches. The potential problems caused by the lack of an effective track design at level crossings could be alleviated with an alternative design solution, where more field experiments and numerical studies dedicated to the level crossings are suggested.

From a macro perspective, it was found that the mainstream in existing track geometry data analytics focuses on the degradation of open tracks while neglecting the localized degradation features in the vicinity of transition zones, which is reflected in the selection of TQI and the type of maintenance:

(i) TQI: the intervention planning predominantly relies on TQI-based trend analysis, where the standard deviation of longitudinal level (in wavelength 3–25 m) over a 200-m track segment is a decisive factor. However, the aggregate TQI may not necessarily capture the localized degradation feature or highlight the higher degradation rates at transition zones, as they typically extend only a few meters or tens of meters. In practice, level crossings are often inspected through regular manual checks [114]. Hence, there is a pressing need to translate track geometry data into specific track features in the vicinity of transition zones to guide local attention more effectively.

(ii) Maintenance type: the TQI-based trend analysis is generally used for tamping optimization, mostly coupled with a condition recovery model and replacement concern. Tamping is effective in packing the ballast layer but may not help correct track geometry defects and hanging sleepers at transition zones, where the root cause may lie at the substructure level.

The question of which interventions to undertake in the vicinity of transition zones necessitates the proper evaluation of track sub-structure condition, which requires the synchronized measurement of track layout and stiffness. This brings another important perspective that the data-driven approach can provide to the management of transition zones. The data-driven approach allows for synchronizing results from various track measurements, such as track geometry, GPR, and stiffness measurement, which enables a proper and comprehensive evaluation of track conditions. This, in turn, informs decisions on the most suitable interventions for specific transition zones. Compared to site-specific measurements used in mechanistic modeling (see Table 2.2), these measurements can be conducted on a larger scale, such as an entire track line or a small network, which can therefore account for the spatial and temporal variability in track degradation.

Specifically, stiffness measurements can reveal potential substructure-related problems, and GPR provides valuable insights into the layer configuration, moisture content, and fouling condition of the track. As filtering track geometry data in different wavelengths has specific indications about the types of defects, it can be synchronized with stiffness and GPR measurements to refine the defect diagnosis along the track lines. This approach is particularly useful for capturing localized degradation features at specific locations, and interventions specific to the identified issues can be followed, such as tamping for regular geometry issues, ballast upgrading for ballast fouling, and subgrade reinforcement to strengthen the weak subgrades.

While existing TQI may not capture the defect-proneness of the track section near transition zones, hybrid TQI derived from the combined measurement is suggested as it can ease the track characterization and provide a more precise detection of the poor track condition for proper maintenance treatment. Feature extraction techniques can be applied to define TQI: PCA and T-stochastic neighbor embedding have been tested in mining the track geometry data for the open track, and many other methods are underexplored.

As wavelength contents of track geometry defects are inherently related to the specific issues of vehicle-track interaction, full-spectra track geometry filtering is suggested to be incorporated in the TQI development and degradation modeling in order to avoid omission of potential types of defects, where most studies only considered the wavelength range 3–25 m. Besides, the evolution of track geometry irregularities is mainly investigated in the time domain, application of spectral analysis is not as widely used as the time-domain methods. Since the spectral analysis reveals frequency components in the track geometry defects, it provides a better understanding of the vehicle-track interaction mechanism, showing great potential to link to the track models.

Further, the hybrid TQI that refines the defect diagnosis can be linked to degradation modeling and maintenance decision support and extended from the individual asset level (a transition zone) to the system level (multiple transition zones). Determin-

istic models are not recommended in the degradation modeling as this type of model only captures the nominal degradation dynamics and the resulting maintenance policies may not be robust enough in the presence of randomness, e.g. the heterogeneity in asset features, degradation levels, and operating context, considering the large number of supporting structures (e.g., bridges, level crossings, as introduced in the beginning of Chapter 1) on a network. Group maintenance for multiple supporting structures with similar conditions on the network can be added in this context by using optimization tools. Economic dependence encompassing the set-up cost (sharing) dependence and operational downtime dependence over multiple assets is embedded as a potential benefit in system-level maintenance decision making.

2.5. CONCLUSIONS

This chapter presents a synthesis of studies related to transition zones in railway tracks. This includes an assessment of track behavior under moving vehicles and modeling of track geometry degradation, which is a leading factor contributing to frequent maintenance in these areas. The dynamic behavior of railway tracks at transition zones is typically assessed through experimental studies and mechanistic modeling, with the latter approach being the focus of this review. Besides, modeling of track geometry degradation mainly follows a data-driven approach.

The two categories of studies complement each other in providing insights into different aspects of track degradation. The mechanistic approach involves establishing a relationship between system characteristics, such as geometric and mechanical properties, and system response. It aims to investigate the root cause of degradation within the train-track system and contribute to a fundamental understanding of the underlying mechanisms at transition zones. The use of data-driven models complements the field of knowledge by mapping relationships between track degradation and exogenous factors, such as operational characteristics, maintenance history, and environmental conditions. This approach demonstrates better predictive capability on both spatial and temporal scales and is useful to inform maintenance decisions.

Unlike bridges and culverts, the optimal design for level crossings and associated transition zones is seldom studied and synchronized. Besides, as a regular intervention, tamping may not help solve the local defects at transition zones, where ballast upgrading or other measures are alternatives to testing and comparing. This necessitates the synchronized measurement and refined diagnosis of the track issues, especially at the substructure level. Hybrid TQI derived from the combined measurement is suggested. This can be linked with maintenance decision support to ease and optimize the management process. Relevant techniques that can be used in this process are also elaborated.

REFERENCES

[1] Y. Shang, M. Nogal, H. Wang, and A. R. M. Wolfert, "Systems thinking approach for improving maintenance management of discrete rail assets: A review and future perspectives", *Structure and Infrastructure Engineering*, vol. 19, no. 2, pp. 197–215, 2021.

[2] B. Indraratna, M. B. Sajjad, T. Ngo, A. G. Correia, and R. Kelly, "Improved performance of ballasted tracks at transition zones: A review of experimental and modelling approaches", *Transportation Geotechnics*, vol. 21, p. 100 260, 2019.

[3] Y. Momoya, T. Takahashi, and T. Nakamura, "A study on the deformation characteristics of ballasted track at structural transition zone by multi-actuator moving loading test apparatus", *Transportation Geotechnics*, vol. 6, pp. 123–134, 2016.

[4] A. Namura and T. Suzuki, "Evaluation of countermeasures against differential settlement at track transitions", *Quarterly Report of RTRI*, vol. 48, no. 3, pp. 176–182, 2007.

[5] R. Sañudo, I. Jardí, J.-C. Martínez, *et al.*, "Monitoring track transition zones in railways", *Sensors*, vol. 22, no. 1, p. 76, 2021.

[6] H. Wang, "Measurement, assessment, analysis and improvement of transition zones in railway track", PhD dissertation, Delft University of Technology, 2018.

[7] A. Paixão, J. N. Varandas, E. Fortunato, and R. Calçada, "Numerical simulations to improve the use of under sleeper pads at transition zones to railway bridges", *Engineering structures*, vol. 164, pp. 169–182, 2018.

[8] D. Mishra, H. Boler, E. Tutumluer, W. Hou, and J. P. Hyslip, "Deformation and dynamic load amplification trends at railroad bridge approaches: Effects caused by high-speed passenger trains", *Transportation research record*, vol. 2607, no. 1, pp. 43–53, 2017.

[9] E. Fortunato, A. Paixão, and R. Calçada, "Railway track transition zones: Design, construction, monitoring and numerical modelling", *International Journal of Railway Technology*, vol. 2, no. 4, pp. 33–58, 2013.

[10] S. T. Wilk, T. D. Stark, and J. G. Rose, "Evaluating tie support at railway bridge transitions", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 4, pp. 1336–1350, 2016.

[11] H. Boler, D. Mishra, E. Tutumluer, S. Chrismer, and J. P. Hyslip, "Stone blowing as a remedial measure to mitigate differential movement problems at railroad bridge approaches", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 233, no. 1, pp. 63–72, 2019.

[12] L. Le Pen, G. Watson, W. Powrie, G. Yeo, P. Weston, and C. Roberts, "The behaviour of railway level crossings: Insights through field monitoring", *Transportation Geotechnics*, vol. 1, no. 4, pp. 201–213, 2014.

[13] J. G. Rose, "Rehabilitation techniques to improve long-term performances of highway-railway at-grade crossings", in *ASME/IEEE Joint Rail Conference*, vol. 54594, 2011, pp. 31–43.

[14] B. Z. Coelho, "Dynamics of railway transition zones in soft soils", PhD dissertation, Delft University of Technology, 2011.

[15] J. M. de Oliveira Barbosa, A. B. Fărăgău, and K. N. van Dalen, "A lattice model for transition zones in ballasted railway tracks", *Journal of Sound and Vibration*, vol. 494, p. 115 840, 2021.

[16] A. B. Faragau, "Understanding degradation mechanisms at railway transition zones using phenomenological models", PhD dissertation, Delft University of Technology, 2023.

[17] A. P. de Man, "Dynatrack: A survey of dynamic railway track properties and their quality", PhD dissertation, Delft University of Technology, 2002.

[18] A. Vesnitskii and A. Metrikin, "Transition radiation in one-dimensional elastic systems", *Journal of applied mechanics and technical physics*, vol. 33, no. 2, pp. 202–207, 1992.

[19] A. Metrikine, A. R. M. Wolfert, and H. Dieterman, "Transition radiation in an elastically supported string. Abrupt and smooth variations of the support stiffness", *Wave motion*, vol. 27, no. 4, pp. 291–305, 1998.

[20] A. Vesnitskii and A. Metrikin, "Transition radiation in mechanics", *Physics-Uspekhi*, vol. 39, no. 10, p. 983, 1996.

[21] Z. Dimitrovová and J. Varandas, "Critical velocity of a load moving on a beam with a sudden change of foundation stiffness: Applications to high-speed trains", *Computers & Structures*, vol. 87, no. 19-20, pp. 1224–1232, 2009.

[22] Z. Dimitrovová, "A general procedure for the dynamic analysis of finite and infinite beams on piece-wise homogeneous foundation under moving loads", *Journal of Sound and Vibration*, vol. 329, no. 13, pp. 2635–2653, 2010.

[23] K. N. van Dalen, A. Tsouvalas, A. V. Metrikine, and J. S. Hoving, "Transition radiation excited by a surface load that moves over the interface of two elastic layers", *International Journal of Solids and Structures*, vol. 73, pp. 99–112, 2015.

[24] A. B. Fărăgău, A. V. Metrikine, and K. N. van Dalen, "Transition radiation in a piecewise-linear and infinite one-dimensional structure–a Laplace transform method", *Nonlinear Dynamics*, vol. 98, pp. 2435–2461, 2019.

[25] A. B. Fărăgău, T. Mazilu, A. V. Metrikine, T. Lu, and K. N. van Dalen, "Transition radiation in an infinite one-dimensional structure interacting with a moving oscillator—the Green's function method", *Journal of Sound and Vibration*, vol. 492, p. 115 804, 2021.

[26] M. Sadri, T. Lu, and M. Steenbergen, "Railway track degradation: The contribution of a spatially variant support stiffness-local variation", *Journal of Sound and Vibration*, vol. 455, pp. 203–220, 2019.

[27] M. J. Steenbergen, "Physics of railroad degradation: The role of a varying dynamic stiffness and transition radiation processes", *Computers & Structures*, vol. 124, pp. 102–111, 2013.

[28] D. Milne, J. Harkness, L. Le Pen, and W. Powrie, "The influence of variation in track level and support system stiffness over longer lengths of track for track performance and vehicle track interaction", *Vehicle System Dynamics*, vol. 59, no. 2, pp. 245–268, 2019.

[29] A. Paixao, E. Fortunato, and R. Calcada, "A contribution for integrated analysis of railway track performance at transition zones and other discontinuities", *Construction and Building Materials*, vol. 111, pp. 699–709, 2016.

[30] J. N. Varandas, P. Hölscher, and M. A. Silva, "Settlement of ballasted track under traffic loading: Application to transition zones", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 228, no. 3, pp. 242–259, 2014.

[31] D. Mishra, Y. Qian, H. Huang, and E. Tutumluer, "An integrated approach to dynamic analysis of railroad track transitions behavior", *Transportation Geotechnics*, vol. 1, no. 4, pp. 188–200, 2014.

[32] W. Zhai, Z. Han, Z. Chen, L. Ling, and S. Zhu, "Train–track–bridge dynamic interaction: A state-of-the-art review", *Vehicle System Dynamics*, vol. 57, no. 7, pp. 984–1027, 2019.

[33] P. Galvín, A. Romero, and J. Domínguez, "Fully three-dimensional analysis of high-speed train–track–soil-structure dynamic interaction", *Journal of Sound and Vibration*, vol. 329, no. 24, pp. 5147–5163, 2010.

[34] S. Iwnicki, *Handbook of railway vehicle dynamics*. CRC press, 2006.

[35] S. Z. Meymand, A. Keylin, and M. Ahmadian, "A survey of wheel–rail contact models for rail vehicles", *Vehicle System Dynamics*, vol. 54, no. 3, pp. 386–428, 2016.

[36] L. Andersen and S. R. Nielsen, "Vibrations of a track caused by variation of the foundation stiffness", *Probabilistic Engineering Mechanics*, vol. 18, no. 2, pp. 171–184, 2003.

[37] J. Oscarsson, "Simulation of train-track interaction with stochastic track properties", *Vehicle System Dynamics*, vol. 37, no. 6, pp. 449–469, 2002.

[38] M. Li and E. Berggren, "A study of the effect of global track stiffness and its variations on track performance: Simulation and measurement", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 224, no. 5, pp. 375–382, 2010.

[39] I. Gallego, S. Sánchez-Cambronero, A. Rivas, and E. Laguna, "A mixed slab-ballasted track as a means to improve the behaviour of railway infrastructure", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 7, pp. 1659–1672, 2016.

[40] P. Hu, C. Zhang, S. Wen, and Y. Wang, "Dynamic responses of high-speed railway transition zone with various subgrade fillings", *Computers and Geotechnics*, vol. 108, pp. 17–26, 2019.

[41] M. Shahraki, C. Warnakulasooriya, and K. J. Witt, "Numerical study of transition zone between ballasted and ballastless railway track", *Transportation Geotechnics*, vol. 3, pp. 58–67, 2015.

[42] M. Banimahd, P. K. Woodward, J. Kennedy, and G. M. Medero, "Behaviour of train–track interaction in stiffness transitions", in *Proceedings of the Institution of Civil Engineers-Transport*, Thomas Telford Ltd, vol. 165, 2012, pp. 205–214.

[43] X. Lei and N.-A. Noda, "Analyses of dynamic response of vehicle and track coupling system with random irregularity of track vertical profile", *Journal of sound and vibration*, vol. 258, no. 1, pp. 147–165, 2002.

[44] H. Hunt and Winkler, "Settlement of railway track near bridge abutments", in *Proceedings of the Institution of Civil Engineers-Transport*, Thomas Telford-ICE Virtual Library, vol. 123, 1997, pp. 68–73.

[45] J. Nicks, "The bump at the end of the railway bridge", PhD dissertation, Texas A&M University, 2009.

[46] A. Paixão, E. Fortunato, and R. Calçada, "A numerical study on the influence of backfill settlements in the train/track interaction at transition zones to railway bridges", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 3, pp. 866–878, 2016.

[47] H. Wang and V. Markine, "Dynamic behaviour of the track in transitions zones considering the differential settlement", *Journal of Sound and Vibration*, vol. 459, p. 114 863, 2019.

[48] X. Lei and L. Mao, "Dynamic response analyses of vehicle and track coupled system on track transition of conventional high speed railway", *Journal of sound and vibration*, vol. 271, no. 3-5, pp. 1133–1146, 2004.

[49] E. Tutumluer, Y. Qian, Y. M. Hashash, J. Ghaboussi, and D. D. Davis, "Discrete element modelling of ballasted track deformation behaviour", *International Journal of Rail Transportation*, vol. 1, no. 1-2, pp. 57–73, 2013.

[50] J. Varandas, P. Hölscher, and M. Silva, "Three-dimensional track-ballast interaction model for the study of a culvert transition", *Soil Dynamics and Earthquake Engineering*, vol. 89, pp. 116–127, 2016.

[51] R. D. FRÖHLING, "Low frequency dynamic vehicle/track interaction: Modelling and simulation", *Vehicle System Dynamics*, vol. 29, no. S1, pp. 30–46, 1998.

[52] C. Vale and R. Calçada, "A dynamic vehicle-track interaction model for predicting the track degradation process", *Journal of Infrastructure Systems*, vol. 20, no. 3, p. 04 014 016, 2014.

[53] A. de Miguel, A. Lau, and I. Santos, "Numerical simulation of track settlements based on an iterative holistic approach", *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 40, pp. 1–12, 2018.

[54] J. C. Nielsen and X. Li, "Railway track geometry degradation due to differential settlement of ballast/subgrade–numerical prediction by an iterative procedure", *Journal of Sound and Vibration*, vol. 412, pp. 441–456, 2018.

[55] Y. Guo and W. Zhai, "Long-term prediction of track geometry degradation in high-speed vehicle–ballastless track system due to differential subgrade settlement", *Soil Dynamics and Earthquake Engineering*, vol. 113, pp. 1–11, 2018.

[56] H. Wang and V. Markine, "Modelling of the long-term behaviour of transition zones: Prediction of track settlement", *Engineering structures*, vol. 156, pp. 294–304, 2018.

[57] T. Dahlberg, "Some railroad settlement models—A critical review", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 215, no. 4, pp. 289–300, 2001.

[58] Y. Sato, "Japanese studies on deterioration of ballasted track", *Vehicle System Dynamics*, vol. 24, no. sup1, pp. 197–208, 1995.

[59] P. Punetha and S. Nimbalkar, "An innovative rheological approach for predicting the behaviour of critical zones in a railway track", *Acta Geotechnica*, pp. 1–27, 2023.

[60] M. Mehrali, M. Esmaeili, and S. Mohammadzadeh, "Application of data mining techniques for the investigation of track geometry and stiffness variation", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 234, no. 5, pp. 439–453, 2020.

[61] M. Movaghar and S. Mohammadzadeh, "Intelligent index for railway track quality evaluation based on bayesian approaches", *Structure and Infrastructure Engineering*, vol. 16, no. 7, pp. 968–986, 2020.

[62] M. Germonpré, J. Nielsen, G. Degrande, and G. Lombaert, "Contributions of longitudinal track unevenness and track stiffness variation to railway induced vibration", *Journal of Sound and Vibration*, vol. 437, pp. 292–307, 2018.

[63] J. C. Nielsen, E. G. Berggren, A. Hammar, F. Jansson, and R. Bolmsvik, "Degradation of railway track geometry–Correlation between track stiffness gradient and differential settlement", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 234, no. 1, pp. 108–119, 2020.

[64] L. Bai, R. Liu, Q. Sun, F. Wang, and P. Xu, "Markov-based model for the prediction of railway track irregularities", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit*, vol. 229, no. 2, pp. 150–159, 2015.

[65] R. Liu, P. Xu, and F. Wang, "Research on a short-range prediction model for track irregularity over small track lengths", *Journal of Transportation Engineering*, vol. 136, no. 12, pp. 1085–1091, 2010.

[66] F. Westgeest, R. Dekker, and R. Fischer, "Predicting rail geometry deterioration by regression models", *Adv Saf Reliab Risk Manag ESREL*, vol. 146, pp. 926–933, 2011.

[67] A. De Rosa, R. Kulkarni, A. Qazizadeh, *et al.*, "Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 235, no. 1, pp. 107–120, 2020.

[68] P. Weston, C. Roberts, G. Yeo, and E. Stewart, "Perspectives on railway track geometry condition monitoring from in-service railway vehicles", *Vehicle system dynamics*, vol. 53, no. 7, pp. 1063–1091, 2015.

[69] E. J. OBrien, P. Quirke, C. Bowe, and D. Cantero, "Determination of railway track longitudinal profile using measured inertial response of an in-service railway vehicle", *Structural Health Monitoring*, vol. 17, no. 6, pp. 1425–1440, 2018.

[70] M. Odashima, S. Azami, Y. Naganuma, H. Mori, and H. Tsunashima, "Track geometry estimation of a conventional railway from car-body acceleration measurement", *Mechanical Engineering Journal*, vol. 4, no. 1, pp. 16–00 498, 2017.

[71] S. Strano and M. Terzo, "Review on model-based methods for on-board condition monitoring in railway vehicle dynamics", *Advances in mechanical engineering*, vol. 11, no. 2, p. 1 687 814 019 826 795, 2019.

[72] P. Salvador, V. Naranjo, R. Insa, and P. Teixeira, "Axlebox accelerations: Their acquisition and time–frequency characterisation for railway track monitoring purposes", *Measurement*, vol. 82, pp. 301–312, 2016.

[73] X. Wei, F. Liu, and L. Jia, "Urban rail track condition monitoring based on in-service vehicle acceleration measurements", *Measurement*, vol. 80, pp. 217–228, 2016.

[74] "Railway applications - Track - Track geometry quality Part 5: Geometric quality levels - Plain line, switches and crossings", European Committee for Standardization, Brussels, Standard, Aug. 2017.

[75] A. Lasisi and N. Attoh-Okine, "Principal components analysis and track quality index: A machine learning approach", *Transportation Research Part C: Emerging Technologies*, vol. 91, pp. 230–248, 2018.

[76] J. Sadeghi, "Development of railway track geometry indexes based on statistical distribution of geometry data", *Journal of Transportation Engineering*, vol. 136, no. 8, pp. 693–700, 2010.

[77] A. Roghani, "A quantitative evaluation of the impact of soft subgrades on railway track structure", PhD dissertation, University of Alberta, 2017.

[78] P. Xu, Q. Sun, R. Liu, and F. Wang, "A short-range prediction model for track quality index", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 225, no. 3, pp. 277–285, 2011.

[79] N. O. Attoh-Okine, *Big data and differential privacy: analysis strategies for railway track engineering*. John Wiley & Sons, 2017.

[80] A. Lasisi, A. Merheb, A. Zarembski, and N. Attoh-Okine, "Rail track quality and T-stochastic neighbor embedding for hybrid track index", in *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, 2019, pp. 1470–1477.

[81] A. Lasisi and N. Attoh-Okine, "An unsupervised learning framework for track quality index and safety", *Transportation Infrastructure Geotechnology*, vol. 7, pp. 1–12, 2019.

[82] A. Falamarzi, S. Moridpour, and M. Nazem, "A time-based track quality index: Melbourne tram case study", *International Journal of Rail Transportation*, vol. 9, no. 1, pp. 23–38, 2021.

[83] I. Soleimanmeigouni, A. Ahmadi, H. Khajehei, and A. Nissen, "Investigation of the effect of the inspection intervals on the track geometry condition", *Structure and Infrastructure Engineering*, vol. 16, no. 8, pp. 1138–1146, 2019.

[84] A. R. Andrade and P. F. Teixeira, "Unplanned-maintenance needs related to rail track geometry", in *Proceedings of the Institution of Civil Engineers-Transport*, Thomas Telford Ltd, vol. 167, 2014, pp. 400–410.

[85] I. Soleimanmeigouni, X. Xiao, A. Ahmadi, M. Xie, A. Nissen, and U. Kumar, "Modelling the evolution of ballasted railway track geometry by a two-level piecewise model", *Structure and Infrastructure Engineering*, vol. 14, no. 1, pp. 33–45, 2018.

[86] Q. He, H. Li, D. Bhattacharjya, D. P. Parikh, and A. Hampapur, "Track geometry defect rectification based on track deterioration modelling and derailment risk assessment", *Journal of the Operational Research Society*, vol. 66, no. 3, pp. 392–404, 2015.

[87] H. Guler, S. Jovanovic, and G. Evren, "Modelling railway track geometry deterioration", in *Proceedings of the Institution of Civil Engineers-Transport*, Thomas Telford Ltd, vol. 164, 2011, pp. 65–75.

[88] N. Lyngby, "Railway track degradation: Shape and influencing factors", *International Journal of Performability Engineering*, vol. 5, no. 2, p. 177, 2009.

[89] S. Sharma, Y. Cui, Q. He, R. Mohammadi, and Z. Li, "Data-driven optimization of railway maintenance for track geometry", *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 34–58, 2018.

[90] H. Khajehei, A. Ahmadi, I. Soleimanmeigouni, and A. Nissen, "Allocation of effective maintenance limit for railway track geometry", *Structure and Infrastructure Engineering*, vol. 15, no. 12, pp. 1597–1612, 2019.

[91] N. Alemazkoor, C. J. Ruppert, and H. Meidani, "Survival analysis at multiple scales for the modeling of track geometry deterioration", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 232, no. 3, pp. 842–850, 2018.

[92] S. A. G. Núnez, "Hybrid Bayesian-Wiener process in track geometry degradation analysis", Ph.D. dissertation, University of Delaware, 2017.

[93] C. Meier-Hirmer, G. Riboulet, F. Sourget, and M. Roussignol, "Maintenance optimization for a system with a gamma deterioration process and intervention delay: Application to track maintenance", *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 223, no. 3, pp. 189–198, 2009.

[94] S. Mercier, C. Meier-Hirmer, and M. Roussignol, "Bivariate Gamma wear processes for track geometry modelling, with application to intervention scheduling", *Structure and Infrastructure Engineering*, vol. 8, no. 4, pp. 357–366, 2012.

[95] L. Bai, R. Liu, Q. Sun, F. Wang, and F. Wang, "Classification-learning-based framework for predicting railway track irregularities", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 2, pp. 598–610, 2016.

[96] C. Hu and X. Liu, "Modeling track geometry degradation using support vector machine technique", in *ASME/IEEE Joint Rail Conference*, American Society of Mechanical Engineers, vol. 49675, 2016, V001T01A011.

[97] I. Cárdenas-Gallo, C. A. Sarmiento, G. A. Morales, M. A. Bolivar, and R. Akhavan-Tabatabaei, "An ensemble classifier to predict track geometry degradation", *Reliability Engineering & System Safety*, vol. 161, pp. 53–60, 2017.

[98] J. S. Lee, S. H. Hwang, I. Y. Choi, and I. K. Kim, "Prediction of track deterioration using maintenance data and machine learning schemes", *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 9, p. 04 018 045, 2018.

[99] E. N. Martey, L. Ahmed, and N. Attoh-Okine, "Track geometry big data analysis: A machine learning approach", in *2017 IEEE international conference on big data (big data)*, IEEE, 2017, pp. 3800–3809.

[100] H. Guler, "Prediction of railway track geometry deterioration using artificial neural networks: A case study for Turkish state railways", *Structure and Infrastructure Engineering*, vol. 10, no. 5, pp. 614–626, 2014.

[101] D. Bzdok, N. Altman, and M. Krzywinski, "Statistics versus machine learning", *Nature Methods*, vol. 15, no. 4, pp. 233–234, 2018.

[102] J. M. van Noortwijk, "A survey of the application of gamma processes in maintenance", *Reliability Engineering & System Safety*, vol. 94, no. 1, pp. 2–21, 2009.

[103] C. Dao, R. Basten, and A. Hartmann, "Maintenance scheduling for railway tracks under limited possession time", *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 8, p. 04 018 039, 2018.

[104] C. Vale, I. M. Ribeiro, and R. Calçada, "Integer programming to optimize tamping in railway tracks as preventive maintenance", *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 123–131, 2012.

[105] S. M. Famurewa, T. Xin, M. Rantatalo, and U. Kumar, "Optimisation of maintenance track possession time: A tamping case study", *Proceedings of the institution of mechanical engineers, Part F: journal of rail and rapid transit*, vol. 229, no. 1, pp. 12–22, 2015.

[106] E. Gustavsson, "Scheduling tamping operations on railway tracks using mixed integer linear programming", *EURO Journal on Transportation and Logistics*, vol. 4, no. 1, pp. 97–112, 2015.

[107] H. Khajehei, M. Haddadzade, A. Ahmadi, I. Soleimanmeigouni, and A. Nissen, "Optimal opportunistic tamping scheduling for railway track geometry", *Structure and Infrastructure Engineering*, vol. 17, no. 10, pp. 1299–1314, 2020.

[108] M. Wen, R. Li, and K. B. Salling, "Optimization of preventive condition-based tamping for railway tracks", *European Journal of Operational Research*, vol. 252, no. 2, pp. 455–465, 2016.

[109] C. Vale and I. M. Ribeiro, "Railway condition-based maintenance model with stochastic deterioration", *Journal of Civil Engineering and Management*, vol. 20, no. 5, pp. 686–692, 2014.

[110] A. Roghani and M. T. Hendry, "Quantifying the impact of subgrade stiffness on track quality and the development of geometry defects", *Journal of Transportation Engineering, Part A: Systems*, vol. 143, no. 7, p. 04 017 029, 2017.

[111] D. Yurlov, A. M. Zarembski, N. Attoh-Okine, J. W. Palese, and H. Thompson, "Probabilistic approach for development of track geometry defects as a function of ground penetrating radar measurements", *Transportation Infrastructure Geotechnology*, vol. 6, pp. 1–20, 2019.

[112] R. Mohammadi, Q. He, F. Ghofrani, A. Pathak, and A. Aref, "Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects", *Transportation research part C: emerging technologies*, vol. 102, pp. 153–172, 2019.

[113] A. M. Zarembski, D. Einbinder, and N. Attoh-Okine, "Using multiple adaptive regression to address the impact of track geometry on development of rail defects", *Construction and Building Materials*, vol. 127, pp. 546–555, 2016.

[114] Y. Shang, M. van den Boomen, A. P. de Man, and A. R. M. Wolfert, "Reliability-based life cycle costing analysis for embedded rails in level crossings", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 234, no. 8, pp. 821–833, 2019.

3

MECHANICS-BASED DESIGN OPTIMIZATION FOR LEVEL CROSSINGS

This chapter presents a surrogate-based simulation methodology to search for an optimal combination of parameters relevant to the geometry and elasticity of track structures. Specifically, this methodology integrates FE modeling with surrogate-assisted optimization: (i) the FE model is developed to characterize the dynamic behavior of a level crossing under a moving vehicle; (ii) the optimization problem is formulated by incorporating the expensive FE simulations into an adaptive surrogate modeling scheme. This integration facilitates an efficient exploration of the track design space (thereby reducing the computational cost), and a reasonable balance can be achieved between solution quality and computational effort.

The proposed methodology is applied to a Dutch railway case. The results indicate a significant improvement in performance indicators relevant to wheel-rail contact forces and energy dissipation in the ballast layer when compared to a reference design. The solution has great potential to achieve a more desirable vehicle–track interaction and improve the connecting performance between level crossings and transitions. While the methodology is demonstrated on a level crossing case, it is generic and can be applied to other railway structures, which also contributes to improvements in current track design practices.

Parts of this chapter have been published verbatim in Engineering Structures, 282, 115740 (2023) [1].

3.1. INTRODUCTION

As introduced in Chapter 1, transition zones often exhibit non-consistent track configurations (e.g., connections between ballast track and slab track) and variations in geotechnical foundations (e.g., embankment to a bridge). The first type is typically encountered in many level crossings, where a slab track, e.g., embedded rail system (ERS), is placed in the crossing area, and the ballast track forms the adjacent sections (see Figure 3.1).

Figure 3.1: An example level crossing with embedded rail system [2].

In ERS, the ballast and sleepers are replaced by concrete slabs with channels, where the rails are placed and fixated by an elastic poured compound. The slabs provide an obstacle-free surface with the road pavement for crossing traffic. The elastic compound provides homogeneous continual support to the rail, differing from periodic sleeper support in the traditional ballast track. With design benefits such as noise reduction and savings in construction height and weight [3], ERS also has wide applications in bridges, tunnels, and tramlines [4]–[6].

Despite the design advantages, the use of ERS in these areas presents structural discontinuities with adjacent ballast tracks. The discontinuities lead to the amplification of stresses and strains when trains pass by, resulting in the accumulation of differential settlements over time. In practice, due to the interaction with road traffic, several degradation problems specific to ERS-based level crossings are experienced by asset managers, including debonding of the elastic compound and rail corrosion [2], [7], as shown in Figure 3.2. They suggest that rail corrosion is related to degradation in track geometry. The differential settlement leads to an additional amplification of dynamic responses when trains pass over the level crossings, causing the rail to bend and the elastic compound to debond. The gap created by the debonding can then allow water, road debris, and de-icing salt to penetrate, making the steel rail more vulnerable to corrosion [2], [7].

Another concern is the potential disruption in maintenance operations that may occur at the interface between crossing and transition zones. Due to the minimum continuous depth of ballast required for tamping, it is difficult to apply maintenance right up to and over the interface. As a result, some sleepers and underlying ballast are never mechanically maintained, which can lead to hanging sleepers and further exacerbates the degradation [8].

To alleviate the track issues while considering the maintenance constraint at the interface, an improved design or maintenance solution is needed for level crossings. Al-

Figure 3.2: Reduced profiles in rail foot caused by corrosion [2].

though there has been extensive research on the degradation mechanisms of transition zones in railway tracks, as discussed in Chapter 2, limited attention has been paid to the level crossings and associated transitions, aside from a few experimental studies [8], [9]. Optimal design solutions for crossing and transition areas, especially for the ERS type, have not been explored in the literature. This motivates the aim of the current chapter, which is to develop a design optimization methodology for this type of railway structure in order to minimize the dynamic amplification and improve the connecting performance between crossing and transition zones.

The novelty of the current chapter is twofold. Firstly, an FE model is developed to characterize the dynamic behavior of the ERS-type level crossings. To simulate VTI, a co-simulation approach is proposed by combining the capabilities of COMSOL and MATLAB. Secondly, based on this mechanistic model, an optimization problem is formulated with the aim of minimizing the dynamic amplification caused by the structural discontinuity, where the FE simulations are incorporated into an adaptive surrogate modeling scheme. This integrative methodology allows for efficient exploration of the design space while achieving a reasonable balance between solution quality and computational effort.

Several response quantities that capture the dynamic amplification of the track are proposed and compared as design criteria/objectives, which are transient responses generated from the FE model. Those showing higher solution quality and sensitivity to parametric variation are selected to formulate a multi-objective optimization (MOO) problem, which is solved by embedding an achievement scalarizing function (ASF) in the surrogate modeling scheme. The optimized design solutions are obtained by minimizing the proposed objectives. It is considered that if the transition zones are designed to reduce the transient responses (the amplitude) of the vehicle–track system, the risk of potential track degradation can also be reduced. This may further reduce maintenance needs and the consequent impact on system life-cycle cost and network performance.

This chapter is organized as follows. Section 3.2 presents four building blocks in the proposed methodology, including a method of simulating the vehicle–track coupling dynamics, formulation of a general optimization problem for the level crossings (i.e., definition of design variable, objectives, and mathematical formulation), a surrogate-assisted optimization scheme, and a method of integration. Section 3.3 presents numer-

ical examples to demonstrate the applicability of the proposed procedure. Section 3.4 discusses the quality of obtained solutions and effectiveness of the proposed objectives, using a reference design as a benchmark. Section 3.5 concludes this chapter with some final remarks.

3.2. METHODOLOGY

3.2.1. MODELING OF VEHICLE–TRACK INTERACTION (VTI) DYNAMICS

The interaction between vehicle and track dynamics is critical, particularly in transition zones. Track discontinuities lead to undesirable wheel-rail interactions, which in turn amplify the responses of both the track and the vehicle. By accounting for the coupled dynamics of these systems, it is possible to capture variations in the vertical momentum of the moving vehicle and the consequent impact on track vibrations. To this end, a VTI model is developed to simulate the dynamic response of a level crossing under a passing vehicle. Figure 3.3 shows a schematic representation of the model, which is used as a basis for design alternatives comparison and is represented in a parametric way for optimization purposes. It consists of two subsystems, a vehicle modeled by multibody simulation and a track structure modeled by the FE method. The two subsystems are coupled through wheel-rail contact to formulate an integrated time-dependent system. The model characterizes the vertical dynamics of the vehicle-track system since it is typically pronounced in transition zones [10]. Symmetrical load distribution is assumed between the rails, where half of the track is considered in dynamic simulations.

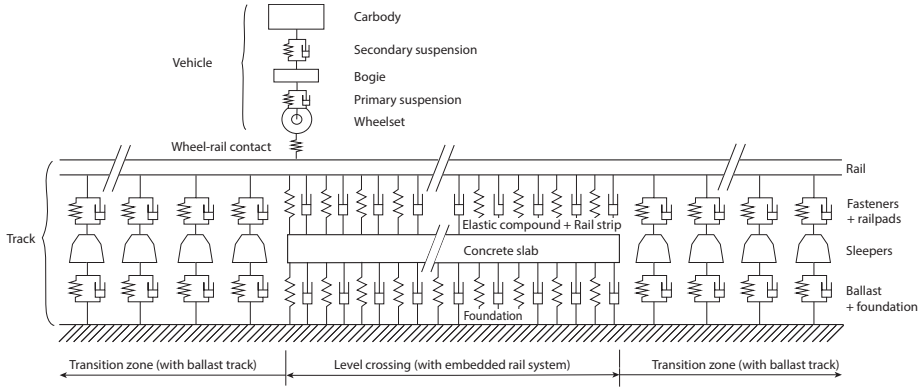

Figure 3.3: Illustration of the vehicle-track interaction model.

VEHICLE MODEL

The vehicle system is represented by a quarter-car model traveling at a constant speed v, as shown in Figure 3.3. It is treated as a multibody system including a carbody, bogie frame, and wheelset. The wheelset is connected to the bogie through the primary suspension, and the carbody and bogie are linked through the secondary suspension. Note that more advanced vehicle systems, i.e., a full-car system or a series of cars, can be implemented. Here, as the focus is to examine the surrogate-assisted optimization in

the level crossing design, a vehicle system with three DOFs is considered in the coupling dynamics with the track.

The equations of motion of the vehicle model can be written as

$$\mathbf{M}_v\ddot{\mathbf{U}}_v + \mathbf{C}_v\dot{\mathbf{U}}_v + \mathbf{K}_v\mathbf{U}_v = \mathbf{F}_v, \tag{3.1}$$

where $\mathbf{M}_v, \mathbf{C}_v$, and $\mathbf{K}_v$ denote, respectively, the mass, damping, and stiffness matrices of the vehicle. $\mathbf{U}_v, \dot{\mathbf{U}}_v$, and $\ddot{\mathbf{U}}_v$ denote, respectively, the displacement, velocity, and acceleration vectors of the vehicle. $\mathbf{F}_v$ is the force vector applied on the vehicle, which contains gravity loads and the wheel-rail contact force. The expression of the system matrices depends on what vehicle DOFs are considered in the simulation. For the current 3-DOF vehicle system, the corresponding system matrices are given by

$$\mathbf{M}_v = \text{diag}\begin{bmatrix} m_c & m_b & m_w \end{bmatrix}, \tag{3.2}$$

$$\mathbf{C}_v = \begin{bmatrix} c_{s2} & -c_{s2} & 0 \\ -c_{s2} & c_{s2}+c_{s1} & -c_{s1} \\ 0 & -c_{s1} & c_{s1} \end{bmatrix}, \tag{3.3}$$

$$\mathbf{K}_v = \begin{bmatrix} k_{s2} & -k_{s2} & 0 \\ -k_{s2} & k_{s2}+k_{s1} & -k_{s1} \\ 0 & -k_{s1} & k_{s1} \end{bmatrix}, \tag{3.4}$$

where m_c, m_b, and m_w denote the mass of the carbody, bogie, and wheelset. c_{s1} and c_{s2} represent damping properties of the primary and secondary suspension; k_{s1} and k_{s2} denote stiffness of the corresponding suspension systems.

TRACK MODEL

The track model consists of an ERS-based level crossing in the middle and transition zones on both sides in the ballast track form. The ballast track is represented by a two-layer discretely supported model; see Figure 3.3. It consists of one rail meshed by Euler-Bernoulli beam elements, railpads as Kelvin-Voight (KV) elements, sleepers as mass elements, and underlying ballast and foundation collectively as the KV elements. Each rail element has two nodes with 2 DOFs, vertical translation and rotation, at each node. The KV element consists of one linear spring and one linear damper placed in parallel, which is commonly used to represent viscoelastic materials in railway structures, such as railpads and ballast layer (e.g., [11]).

The embedded track comprises a rail, fastening, concrete slabs, and a foundation layer. Figure 3.4 presents a simplified ERS cross-section. The fastening is typically an elastic poured compound bonding the rail and a resilient rubber strip under the railbase to provide track elasticity and constrain the vertical rail deflection. Space-saving components can be used, and PVC tubes are for cable installation.

Previous works such as [12] analyzed the dynamic behavior of the embedded track, where a model with two beams (rail and slab) is compared against a more advanced model that accounts for the lateral flexibility of the slab, i.e., two beams (rails) and a flexible plate (slab). The comparison was made in terms of wave propagation and dynamic

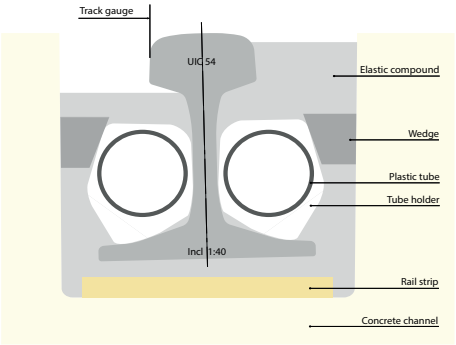

Figure 3.4: Schematic representation of the ERS cross-section.

responses (e.g., vertical displacement of the rail and stresses in the concrete slab). Results showed that the former model can be employed for a quick and sufficiently accurate assessment of the dynamic behavior of the embedded track. The current work therefore adopts the simplified version, where the rail and slab are modeled by the Euler-Bernoulli beam elements, connected by parallel KV elements as the rail fastening and supported by viscoelastic (Winkler-type) foundation.

The equations of motion of the track model can be expressed as

$$\mathbf{M}_t\ddot{\mathbf{U}}_t + \mathbf{C}_t\dot{\mathbf{U}}_t + \mathbf{K}_t\mathbf{U}_t = \mathbf{F}_t, \tag{3.5}$$

where $\mathbf{M}_t, \mathbf{C}_t$, and $\mathbf{K}_t$ denote, respectively, the mass, damping, and stiffness matrices of the track structure. $\mathbf{U}_t$, $\dot{\mathbf{U}}_t$, and $\ddot{\mathbf{U}}_t$ denote, respectively, the displacement, velocity, and acceleration vectors of the track. $\mathbf{F}_t$ is the force vector applied on the track by the running train.

COUPLING OF VEHICLE AND TRACK MODEL

Numerical simulations of the VTI dynamics are generally solved by two types of computational methods. One treats the vehicle and track structure as a unified system and formulates coupled system matrices, e.g., [13], [14]. The other separates the two subsystems and solves individual equations of motion based on an iterative procedure, e.g., [15]. This thesis adopts the former method that couples the vehicle and track to form an integrated time-dependent system. It results in a global system of equations that can be solved in a direct manner without the need for an iterative procedure.

Combining the equations of motion for the vehicle, Eq. (3.1), and track, Eq. (3.5), the unified formulation results in the following global system equations of motion.

$$\mathbf{M}_g\ddot{\mathbf{U}}_g + \mathbf{C}_g\dot{\mathbf{U}}_g + \mathbf{K}_g\mathbf{U}_g = \mathbf{F}_g, \tag{3.6a}$$

where $\mathbf{M}_g$, $\mathbf{C}_g$, and $\mathbf{K}_g$ denote, respectively, the mass, damping, and stiffness matrices of the global integrated system. $\mathbf{U}_g$,$\dot{\mathbf{U}}_g$, and $\ddot{\mathbf{U}}_g$ are the displacement, velocity, and acceleration vectors of the global system, respectively. $\mathbf{F}_g$ is the global force vector. Their expressions are given by

$$\mathbf{M}_g = \begin{bmatrix} \mathbf{M}_v & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_t \end{bmatrix}, \mathbf{C}_g = \begin{bmatrix} \mathbf{C}_v & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_t \end{bmatrix}, \mathbf{K}_g = \begin{bmatrix} \mathbf{K}_{vv} & \mathbf{K}_{vt} \\ \mathbf{K}_{tv} & \mathbf{K}_{tt} \end{bmatrix}, \mathbf{U}_g = \begin{bmatrix} \mathbf{U}_v \\ \mathbf{U}_t \end{bmatrix}, \mathbf{F}_g = \begin{bmatrix} \mathbf{F}_{vt} \\ \mathbf{F}_{tv} \end{bmatrix}. \quad (3.6b)$$

The wheel-rail contact is modeled by Hertz contact theory. The contact stiffness k_w is linearized and calculated by $k_w = \frac{3}{2G} P^{1/3}$, where P is the static wheel axle load and G represents a contact constant [16]. As the wheel-rail contact is modeled by a Hertzian spring k_w, the vehicle and track are coupled through the stiffness matrices only, which is indicated through the entries of $\mathbf{K}_g$ in Eq. (3.6b). $\mathbf{K}_g$ is time-variant since the position of the train changes with time. The detailed formulation is given below.

(a) $\mathbf{K}_{vv}$ is the vehicle stiffness matrix and can be expressed as

$$\mathbf{K}_{vv} = \mathbf{K}_v + \mathbf{K}_v', \quad (3.7)$$

where $\mathbf{K}_v$ is the stiffness matrix of the vehicle itself, as shown in Eq. (3.4). $\mathbf{K}_v'$ is the stiffness matrix of the vehicle induced by the wheel-rail contact. For the current 3-DOF vehicle, it is written as

$$\mathbf{K}_v' = \mathrm{diag}\begin{bmatrix} 0 & 0 & k_w \end{bmatrix}. \quad (3.8)$$

(b) $\mathbf{K}_{tt}$ is the track overall stiffness matrix and is written as

$$\mathbf{K}_{tt} = \mathbf{K}_t + \mathbf{K}_t', \quad (3.9)$$

with

$$\mathbf{K}_t' = k_w \cdot \mathbf{N}^{\mathrm{T}} \cdot \mathbf{N}, \quad (3.10)$$

where $\mathbf{K}_t$ represents the stiffness matrix of the track itself; $\mathbf{K}_t'$ is related to the rail displacement under the wheel, representing a portion of the track stiffness matrix induced by the vehicle. $\mathbf{N}$ is a location vector that defines the correspondence between the vehicle position and rail element in contact. For one wheel-rail contact point, the vector $\mathbf{N}$ is given by

$$\mathbf{N} = \begin{bmatrix} 0 & \dots & 0 & N_{1j} & N_{2j} & N_{3j} & N_{4j} & 0 & \dots & 0 \end{bmatrix}'_{n\times 1}, \quad (3.11a)$$

with n denoting the total number of DOFs of the track structure. The non-zero entries of $\mathbf{N}$ represent the rail element in contact with the wheel, where Hermitian shape functions for the Euler-Bernoulli beam are applied. By letting $\xi_j \in [0, l_j]$ be the local coordinate of rail element j with length l_j, the non-zero entries of $\mathbf{N}$ can

be expressed as

$$N_{1j}(\xi_j) = 1 - 3\left(\frac{\xi_j}{l_j}\right)^2 + 2\left(\frac{\xi_j}{l_j}\right)^3, \tag{3.11b}$$

$$N_{2j}(\xi_j) = \xi_j\left(1 - \frac{\xi_j}{l_j}\right)^3, \tag{3.11c}$$

$$N_{3j}(\xi_j) = 3\left(\frac{\xi_j}{l_j}\right)^2 - 2\left(\frac{\xi_j}{l_j}\right)^3, \tag{3.11d}$$

$$N_{4j}(\xi_j) = \frac{\xi_j^2}{l_j}\left(\frac{\xi_j}{l_j} - 1\right). \tag{3.11e}$$

(c) $\mathbf{K}_{tv}$ and $\mathbf{K}_{vt}$ are coupling matrices induced by the wheel-rail interaction and given by

$$\mathbf{K}_{tv} = \begin{bmatrix} 0 & 0 & -k_w\mathbf{N} \end{bmatrix}_{n\times 3}, \quad \mathbf{K}_{vt} = \mathbf{K}_{vt}{}'. \tag{3.12}$$

At the right-hand side of Eq. (3.6a), the global force vector $\mathbf{F}_g$ is formulated by two load vectors, i.e., $\mathbf{F}_{vt}$ and $\mathbf{F}_{tv}$. $\mathbf{F}_{vt}$ is the load vector of the vehicle and is given by

$$\mathbf{F}_{vt} = \mathbf{F}_{vg} + \mathbf{F}_{vr}, \tag{3.13}$$

with $\mathbf{F}_{vg}$ denoting the load vector induced by the vehicle gravity and $\mathbf{F}_{vr}$ being the load vector induced by the track geometry irregularity. They can be calculated by

$$\mathbf{F}_{vg} = \begin{bmatrix} m_c g & m_b g & m_w g \end{bmatrix}', \quad \mathbf{F}_{vr} = \begin{bmatrix} 0 & 0 & k_w r(x) \end{bmatrix}', \tag{3.14}$$

where g is gravitational acceleration, i.e., 9.8 m/s^2. $r(x)$ denotes the track irregularity in vertical profile, which is location specific.

$\mathbf{F}_{tv}$ is the load vector of the track and is written as

$$\mathbf{F}_{tv} = -k_w r(x) \cdot \mathbf{N}. \tag{3.15}$$

3.2.2. FORMULATION OF AN OPTIMIZATION PROBLEM

A common guideline for improving the transition performance is to smooth the variation in vertical track stiffness, which is also specified in EN 16432-2 [17]. According to this principle, various countermeasures in transition zones have been developed to mitigate the degradation, as discussed in Chapter 2. Likewise, the design principle in this chapter is to obtain a gradual change of stiffness in the track, where relevant response quantities defined to capture the local dynamic amplification are optimized or minimized.

DEFINITION OF DESIGN VARIABLES

A *reference design* that follows a typical Dutch practice in level crossings is defined as a benchmark to assess the optimized design solutions. The parameter values and the justification for the chosen values are given in Section 3.3.1. In the crossing area with ERS, an alternative to the reference design is to install another type of rail strip (herein referred to as *Type II rail strip*). The rail strip is a resilient component fitting between

the underside of the rail and slab channel, as depicted in Figure 3.4. The component provides elasticity to the structure, improves the load distribution, and controls the rail deflection. It has predefined stiffness properties that can be chosen to meet specific system requirements.

For the ballast track in transition zones, the vertical track stiffness is influenced by the flexural stiffness of the rail and supporting stiffness contributed from track components below the rail, such as sleepers and ballast bed. The contribution from the substructure is also significant to the vertical track stiffness, especially in soft soil regions [18], [19]. Modifying the stiffness of rail, sleepers, and ballast can cause problems in track stability and resistance [20]. An adequate solution is to introduce elastic components in the track structure, such as railpads, under sleeper pads, and under ballast mats. The railpad stiffness is the main characteristic parameter of these elements and comes with a wide range of values [20]. Stiff pads such as standard Dutch pads FC9 reduce rail deflection and vibration, while soft pads can mitigate the damage in sleepers and reduce vibration in sleepers and ballast.

Moreover, sleeper parameters such as size, spacing, and material type have a considerable effect on track dynamic behavior, where effort has been made to evaluate the variation of these parameters on transition performance, e.g., [10], [21]–[23]. Referring to the Dutch practice, two types of concrete sleepers are considered in the design space, i.e., a commonly used NS90 sleeper and the other option with larger weight and stability (herein referred to as *strengthened sleepers*). The strengthened type is mainly used in tight curves or in connection with level crossings, switches, and other structures to ensure gradual track stiffness changes. According to Prorail (Dutch railways) system specifications, at least 5 to 8 strengthened sleepers should be placed next to the ERS-type level crossings. And the sleeper bay right next to the concrete slab should be reduced to 0.4 m (for comparison, the standard sleeper spacing is 0.6 m). However, questions remain in approach zones regarding the optimal number of strengthened sleepers being installed next to the junction and the distances between the centers of those neighboring sleepers.

The above design parameters are considered influential to the dynamic performance of the level crossings. In this chapter, they are collected as four major types of design variables in formulating optimization problems, which are displayed in Table 3.1.

Table 3.1: Types of track design variables, $\mathbf{x}$.

Track type	Component	Variables
Ballast	Railpad	Stiffness, $(x_{ri}, i = 1,2,...)$
	Sleeper	Number of strengthened sleeper, (x_n)
		Sleeper spacing, $(x_{sj}, j = 1,2,...)$
ERS	Rail strip	Length of placing Type II strip, (x_l)

ASSESSMENT CRITERIA OF TRACK PERFORMANCE

Track degradation is generally reflected at two levels: the one related to the wheel-rail interface and the other concerning the supporting elements below the rail, such as sleepers and ballast layer.

At the wheel-rail interface, undesirable contact degrades both rail and wheel profiles, and the contact properties influence the rate of degradation [24]. Track degradation at the contact level is often manifested in rail defects, such as rail corrugation and rolling contact fatigue, which represent short-wave components of track irregularities. These defects can significantly affect the magnitude of wheel-rail interaction forces and are therefore relevant to driving safety and vehicle stability [25].

At the lower supporting level, the ballast layer and underlying substructure are important contributors to the deterioration in track geometry. Due to the sliding and breakage of granular particles, the ballast layer presents a progressive deformation with the passage of trains. The layers below the ballast can also experience plastic deformations due to consolidation and cyclic loading, which further contributes to the development of accumulated settlement. However, the slab track in level crossings is designed for minimal settlement, resulting in a differential settlement between the two forms of track.

Based on these typical features, bi-level criteria are defined to evaluate the sensitivity of a track design to expected levels of performance. The first-level criteria are relevant to wheel-rail contact and are assessed based on the magnitude of dynamic contact forces. A larger magnitude of wheel-rail vertical forces F indicates more dynamic amplification in the track structure induced by the passing vehicle, and consequently, reducing the amount of F represents damage mitigation at the wheel-rail contact and reduction in potential rail defects. The second-level measure concerns the damage to the ballast layer, as it is a significant contributor to track geometry degradation. The mechanical energy dissipated in the track ballast is selected as an indicator to assess the sensitivity of a track design to the expected damage in the substructure, which is proposed in [22] and further elaborated in [26], [27]. For one wheel passage, the energy dissipated by the ballast damping in the i^{th} sleeper support is given by [22],

$$E_i = \int_{-\infty}^{\infty} c_{b,i}\, v_{s,i}^2(t)\, dt \tag{3.16}$$

where $c_{b,i}$ is the ballast damping under the i^{th} sleeper; $v_{s,i}(t)$ is the velocity response of the i^{th} sleeper in the time domain. The higher the energy dissipated into the substructure, the stronger the degradation that can be expected. Accordingly, reducing the amount of energy dissipated in the ballast layer represents an important aim for damage reduction in the ballast and therefore in overall track geometry.

The response quantities, i.e., F and E, are generated from VTI dynamic simulations and presented in time series. To better capture the features embedded in the responses, the simulation data are further processed by two statistical metrics, i.e., the root mean square (rms) and maximum-to-minimum (max) value. The former captures the average dynamic amplification over the influenced track section. The latter represents the full range of response variability in the influenced area. The evaluation of the metrics results in four design objectives, i.e., $F_{rms}, F_{max}, E_{rms}$ and E_{max}, which are treated as dynamic benchmarks for track design comparison and parametric optimization.

As previously mentioned, the ballast and substructure are modeled by the KV elements. The damping parameter of the ballast layer is considered constant, referring to characteristics of the typical Dutch track, as elaborated in Section 3.3.1. Although this representation can describe the elastic resistance and damping provided by the layers

supporting the sleepers during train passage, more accurate quantification of energy dissipation in these layers requires a 3D representation that accounts for track-soil interactions and nonlinear material behavior. However, the primary focus of this chapter is on parametric optimization. The energy dissipation (Eq. 3.16) is used to quantify the susceptibility of a track design to expected degradation, where the assessment of the effect of parametric variations on the dissipated energy is relative, aiming to compare different track design solutions. Therefore, it is considered sufficient given the aim of the current work.

THE OPTIMIZATION PROBLEM

The design variables listed in Table 3.1 are collected in a design vector $\mathbf{x} = \left[x_{ri}, x_n, x_{sj}, x_l\right]'$ with $i, j \in \mathbb{N}$, which can be varied according to the problem setting. For instance, in the case of three railpads ($i = 3$) and four sleeper spacings ($j = 4$), it yields a $3+1+4+1 = 9$ dimensional optimization problem. To minimize the dynamic amplification at the ballast-to-ERS transition, a general single-objective optimization problem can be formalized in the following form.

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & x_{ri} \in \mathbb{R} : x_{ri}^l \leq x_{ri} \leq x_{ri}^u,\ i \in \mathbb{N}, \\ & x_n \in \mathbb{Z} : 0 \leq x_n \leq x_n^u, \\ & x_{sj} \in \mathbb{R} : x_{sj}^l \leq x_{sj} \leq x_{sj}^u,\ j \in \mathbb{N}, \\ & x_l \in \mathbb{R} : 0 \leq x_l \leq x_l^u, \end{aligned} \tag{3.17}$$

where the superscripts l and u represent the lower and upper bound of each design variable, respectively; $f(\mathbf{x})$ is an objective function defined by each of the performance criteria described in the previous section. It represents a true function that maps a given input design vector $\mathbf{x}$ and a function value $f(\mathbf{x})$ calculated from the vehicle-track dynamic simulation.

The bi-level objectives, i.e., F and E, are expected to be conflicting since certain design variables inherently have opposite effects on the objectives. For instance, stiff railpads can reduce noise and vibration from wheel-rail contact, but soft pads can lower the effect of loads transmitted to underlayers, thereby reducing vibration in sleepers and ballast particles [20]. For this reason, it is necessary to simultaneously minimize the objectives, which can be achieved by formulating an MOO problem to search for the optimal compromise solution.

Statistical metrics (i.e., rms and max) used to evaluate the dynamic responses are compared based on the results obtained from single objective problems. Those showing higher solution quality and sensitivity to parametric variation are chosen for the MOO problem formulation.

Without loss of generality, a multi-objective optimization problem with $k\,(\geq 2)$ objective functions can be expressed as

$$\begin{aligned} &\min_{\mathbf{x}} \ \mathbf{F}(\mathbf{x}) = \left\{f_1(\mathbf{x}), ..., f_k(\mathbf{x})\right\}, \\ &\text{s.t.} \quad \mathbf{x} \in \Omega, \end{aligned} \tag{3.18}$$

where $f_i(\mathbf{x})$ is the i^{th} objective and $\mathbf{F}: \Omega \rightarrow \Lambda$ maps the design variables ($\mathbf{x}$) to the vector ($\mathbf{y} = [y_1, ..., y_k]$) in the objective function space Λ.

As introduced in Chapter 1, surrogate-based optimization uses surrogate model(s) to perform optimization tasks. When dealing with multiple objectives in surrogate-based optimization, the most common approach is to build a separate surrogate for each objective function. However, this method is time-consuming as it requires training multiple surrogates [28]. An alternative method is to use an ASF to convert the multi-objective problem into a single-objective problem. A single surrogate can then be built on the scalarizing function, which is applied to search for optimal solutions. The ASFs, introduced by [29], map k objective functions to a scalar, which represents the a priori preference articulation in multi-objective optimization problems. Certain properties of the ASFs guarantee Pareto optimality of the solutions obtained from a scalarizing problem (See [30] for a detailed description). This approach reduces the computational complexity as only one surrogate is built and one infilling criterion (see Section 3.2.3) is used in the optimization workflow.

Given the simplicity and computational efficiency, the ASF-based mono surrogate approach is adopted to deal with the expensive MOO problem. Specifically, the (conflicting) objectives selected from the single-objective simulation round are scalarized into one global function by an ASF of augmented Chebyshev type in a form [31],

$$g\left(\mathbf{x}, \mathbf{f}^T\right) = \max_{i \in \{1,...,k\}} \left\{\kappa_i \left(f_i(x) - f_i^r\right)\right\} + \rho \sum_{i \in \{1,...,k\}} \kappa_i \left(f_i(x) - f_i^r\right), \tag{3.19}$$

where f_i ($i = 1, ..., k$) are performance objectives selected from the single-objective optimization problems; $\rho > 0$ is an arbitrary small parameter and $\mathbf{f}^T = [f_1^r, ..., f_k^r]$ is a vector that defines a reference point. κ_i are non-negative normalization coefficients. The idea of this function is to minimize the deviations from the reference objective, and any reasonable or desired point in the objective space specified by the decision maker can be considered as a reference objective [32]. The second term of the function guarantees that all the objectives play a role, not only the one more deviated from the reference value. Through the scalarization, a multi-objective problem aimed at minimizing track dynamic amplification is formulated. This involves minimizing the objective function defined in Eq. (3.19), subject to the constraints defined in Eq. (3.17).

3.2.3. SURROGATE-ASSISTED OPTIMIZATION

A SURROGATE MODEL: RADIAL BASIS FUNCTION INTERPOLATION

Various types of surrogate models have been applied to support engineering tasks, such as Kriging [11], [33], radial basis function [34], and neural networks [35], [36]. In the railway field, Kriging models have been trained to approximate the relationship between track parameters and frequency response function features [11], and neural networks have been used to predict responses of a vehicle-bridge system [35], [36], which are all developed based on FE numerical simulations.

Surrogate models can be either interpolating or non-interpolating, and parametric or nonparametric [37]. In the current VTI model, each simulation run is deterministic. While training a surrogate for such mechanical problems, it is often reasonable to assume the true objective function can be evaluated precisely or with a minor approxi-

mation error at sampled input points [38]. For this purpose, radial basis function (RBF) interpolation is employed to approximate the solutions of the VTI simulation, given its powerful convergence properties and easily adjustable smoothness [39]. Specifically, a cubic RBF interpolation is used as it has shown a competitive performance profile compared to other surrogates in [37].

An RBF interpolant is defined as

$$\hat{f}(\mathbf{x}) = \sum_{k=1}^{n} \lambda_k \phi(\|\mathbf{x}-\mathbf{x}_k\|) + p(\mathbf{x}), \tag{3.20}$$

where in the first term, $\mathbf{x}_k, k = 1,...,n$, denotes the points that have been evaluated by the true objective function; $\|\cdot\|$ is the Euclidean norm; $\lambda_k \in \mathbb{R}, k = 1,...,n$ are coefficients; $\phi(\cdot)$ is a radial basis function and many function forms are available as described in [40]. The cubic RBF interpolant uses the cubic function $\phi(r) = r^3$ (where $r = \|\mathbf{x}-\mathbf{x}_k\|$).

The second term in Eq. (3.20) represents a polynomial tail whose order depends on the chosen RBF [41]. The general form is defined as $p(\mathbf{x}) = \sum_{l=1}^{m} \beta_l p_l(\mathbf{x})$, with m denoting the order of the basis $p_l(\cdot)$ and β_l as the coefficients. For the cubic RBF, it should be at least a linear polynomial, and it becomes $p(\mathbf{x}) = a + \mathbf{b}^T\mathbf{x}$ with coefficients $a \in \mathbb{R}$ and $\mathbf{b} = \left[b_1,...,b_d\right]^T \in \mathbb{R}$ [42]. The coefficients λ_k, a and $\mathbf{b}$ are determined by solving the following linear system of equations

$$\begin{bmatrix} \Phi & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{f}} \\ 0 \end{bmatrix}, \tag{3.21}$$

where $\Phi_{kv} = \phi(\|\mathbf{x}_k - \mathbf{x}_v\|)$, $k, v = 1,...,n$, and

$$\mathbf{P} = \begin{bmatrix} \mathbf{x}_1^T & 1 \\ \mathbf{x}_2^T & 1 \\ \vdots & \vdots \\ \mathbf{x}_n^T & 1 \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \\ a \end{bmatrix}, \quad \hat{\mathbf{f}} = \begin{bmatrix} \hat{f}(\mathbf{x}_1) \\ \hat{f}(\mathbf{x}_2) \\ \vdots \\ \hat{f}(\mathbf{x}_n) \end{bmatrix}. \tag{3.22}$$

The matrix in Eq. (3.21) is invertible if and only if rank$(\mathbf{P}) = d+1$ [39], [41].

AN EFFICIENT INFILLING SCHEME

Once a surrogate model is built, optimal solutions can be searched using surrogate function values. The original function $f(\mathbf{x})$ in Eq. (3.17) is replaced by a surrogate model $\hat{f}(\mathbf{x})$, which shifts the problem to minimize the function $\hat{f}(\mathbf{x})$. However, the estimated function values (i.e., $\hat{f}(\mathbf{x})$) obtained from exploring the surrogate are subject to the model accuracy. Strategies are required that improve the surrogate model accuracy while guiding the search to promising areas of the design space [34]. This type of strategy typically balances exploration and exploitation. Exploration samples the regions far from any optimum searched before, thereby having high uncertainty. Exploitation concerns the search in local (promising) areas with the hope of improving the current optimum.

The key idea of such strategies is to pick the next function evaluation point based on the surrogate predictions and a measure of the error in this model [40]. It is essentially

a way of refining the surrogate model while guiding the search towards the optimum within a relatively limited number of function evaluations (i.e., to keep the total computational time manageable). The process of adding the next evaluation point based on this 'strategic' sampling is the so-called infilling scheme. The chosen point for the next function evaluation generally refers to the *infilling* or *adaptive point*, and the measures to determine the point are known as *infilling criteria* (also *acquisition* or *merit functions*).

The adaptive surrogate modeling technique has been applied to many mechanical systems for various engineering purposes. For example, it has been used for design optimization of vehicle crashworthiness [43], aerodynamic shape [34], and wing typology [44], as well as for reliability analysis of a hydrokinetic turbine blade and a hysteretic oscillator [45] and for material parameter identification of a specimen bending system [46]. However, its application in railway structural design requires further exploration. The technique is well-suited to expensive simulations with a limited budget of function evaluations, typically ranging from 20 to 200 [47]. Therefore, it is a promising tool to address optimization problems in railway track design that involve intensive calculations of vehicle-track dynamic simulations.

An efficient balancing strategy is developed by Regis and Shoemaker [48], where an algorithmic framework called Metric Stochastic Response Surface is introduced for global optimization of expensive functions. The framework executes an adaptive learning process consisting of the following steps: 1) choosing the adaptive point based on a merit function from a sequence of random candidate points, and 2) evaluating the true objective function value of the adaptive point and using it to update the surrogate model. This process continues iteratively until a stopping criterion is satisfied.

The merit function proposed in [48] is a weighted combination of function values from the current surrogate (*response surface criterion*) and distances to previously evaluated points (*distance criterion*). At each iteration, a candidate point set Ω_c is generated randomly by adding perturbations to the best point found so far x_{best} (see [48] for a detailed discussion on the random perturbations). Each candidate point in Ω_c is given a score by evaluating the corresponding merit function and the one with the lowest score is selected as the adaptive point.

Define $x_k (k = 1, ..., n)$ as n previously evaluated points in Ω_k and $x_c^v (v = 1, ..., t)$ as t candidate points in Ω_c. For each candidate point x_c^v, the merit function is expressed as

$$u\left(x_c^v\right) = \omega \frac{\hat{f}\left(x_c^v\right) - \hat{f}_{\min}}{\hat{f}_{\max} - \hat{f}_{\min}} + (1-\omega) \frac{d_{\max} - d\left(x_c^v\right)}{d_{\max} - d_{\min}}, \tag{3.23}$$

where on the right-hand side, the first term corresponds to the response surface criterion and the second refers to the distance criterion. The parameter ω is a weight with $0 < \omega < 1$ that balances minimizing the surrogate value (*response surface criterion*) while exploring the space to improve the model accuracy (*distance criterion*). $\hat{f}\left(x_c^v\right)$ is the current surrogate model evaluated at point x_c^v, $\hat{f}_{\min} = \min\left(\hat{f}\left(x_c^v\right)\right)$ and $\hat{f}_{\max} = \max\left(\hat{f}\left(x_c^v\right)\right)$; $d\left(x_c^v\right) = \min\left(d_{kv}\right)$ with d_{kv} being the distance from an evaluated point x_k to a candidate point x_c^v, $d_{\min} = \min\left(d\left(x_c^v\right)\right)$ and $d_{\max} = \max\left(d\left(x_c^v\right)\right)$.

The Metric Stochastic Response Surface framework in [48] was originally developed for continuous optimization problems. When dealing with integrality constraints, which are often encountered in engineering problems, it is necessary to ensure that the points

generated from random perturbations satisfy these constraints. This is achieved by relaxing the integer condition while sampling and then rounding the values of the integer variables in the obtained points. A similar approach can be found in Müller [41].

To supplement the random perturbations and search for the minimum of the merit function, a Branch-and-Bound approach can be used in parallel. This approach has been embedded in other surrogate-based optimization frameworks to deal with mixed-integer problems [44], [49]. The idea is to enhance exploration of the space. Since the RBF is a continuous fit, evaluating only the points from the random perturbations (by rounding the numbers) may miss regions of the space that contain relevant information. Therefore, the sampling is supplemented by branching and searching locally in the RBF function. In this case, the candidate point set Ω_c is enhanced by the Branch-and-Bound approach (in addition to the random perturbations), where the point with the lowest merit function score is chosen for the next iteration in the sequential enrichment of the surrogate.

3.2.4. INTEGRATIVE SIMULATION METHODOLOGY

The methodology for optimizing geometric and elastic properties of level crossings involves an integration of the VTI dynamic simulations (Section 3.2.1) and an adaptive surrogate-based technique (Section 3.2.3). The proposed integrated simulation methodology has an iterative procedure, which is illustrated in Figure 3.5 and synthesized as follows.

First, n distinct sample points of the design vector $\mathbf{x}$ are generated from a space-filling design, such as a Latin hypercube sampling (LHS) (Step A). Subsequently, FE simulations of vehicle-track dynamics are evaluated at the sampled points, and the corresponding function values are obtained (Step B). Further details regarding Step B are provided in the following paragraphs. In Part II, the surrogate-based optimization workflow consists of two phases, namely (i) the surrogate construction and (ii) adaptive learning phase:

(i) Initially, a surrogate of the objective function is constructed by interpolating a cubic RBF interpolant, given by Eq. (3.20), through the n evaluated points (Step C - at the first iteration).

(ii) In the adaptive learning phase, the surrogate function is updated (Step C - at successive iterations), where the next point of evaluation is selected according to the merit function, given by Eq. (3.23), from a sequence of random candidate points (Step D and E).

The solution of this new point is evaluated with the expensive/true function, i.e., FE simulations in Step B. This process continues until a stopping criterion is reached (Step G). Here, the maximum number of function evaluations is selected as the stopping condition, as emphasized in [41], [50], [51]. In many surrogate-assisted optimization problems, termination depends more on the computational budget than on a measure of convergence due to the high computational cost of the function evaluation.

Numerical solutions for VTI models are often developed using FE software or self-programmed codes. While FE software provides greater flexibility in generating track

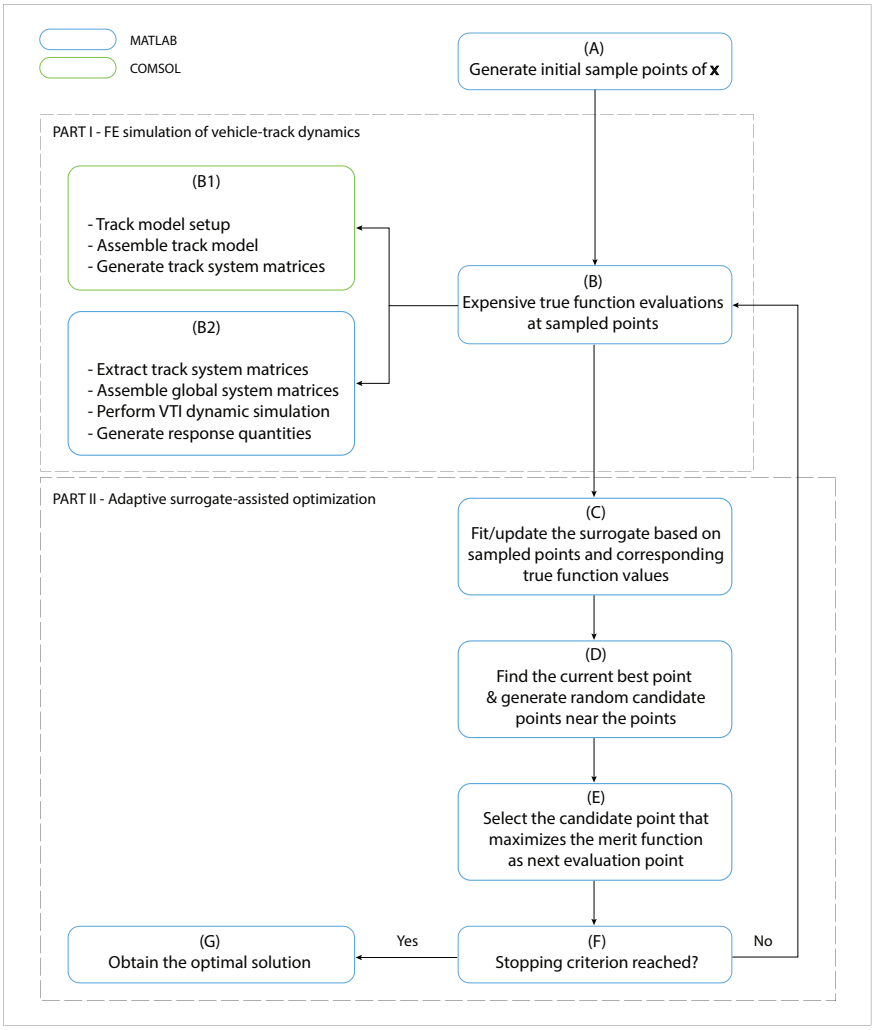

Figure 3.5: Illustrative flowchart of the simulation methodology integrating (a) PART I: FE simulation of vehicle-track dynamics; and (b) PART II: adaptive surrogate-assisted optimization.

models with complex structural configurations, it lacks the necessary adaptability for model adjustment and data postprocessing. To address these limitations and ease the generation of the track model, a combined simulation approach is proposed to model the vehicle-track coupling dynamics. The approach of using a general FE software COMSOL and MATLAB interface has been demonstrated in a baseline case by Shang et al. [52], where a general beam model subject to a moving vehicle is simulated and verified by a benchmark case coded in MATLAB. Here the methodology is extended to the case of level crossings.

Specifically, in Step B (Part I of the integrated methodology), COMSOL is used to

establish a ballast-ERS transition, which characterizes the track used in a typical heavy-duty level crossing. The track system matrices $\mathbf{M}_t, \mathbf{C}_t$, and $\mathbf{K}_t$ in Eq. (3.5) are generated in COMSOL and exported to MATLAB through the Livelink interface. The software used in these steps is highlighted in Figure 3.5. The vehicle system matrices $\mathbf{M}_v, \mathbf{C}_v$, and $\mathbf{K}_v$ in Eq. (3.1) are then formulated and coupled with the extracted track matrices in MATLAB to form global system matrices (Eq. 3.6b). The resulting coupled equations of motion (Eq. 3.6a) that govern the vehicle-track dynamics are solved in the time domain using the Newmark-β integration scheme, where the commonly used combination of $\beta = 1/4$ and $\gamma = 1/2$ was adopted, providing unconditional stability for linear elastic systems [53]. A fixed time step of 0.001s was chosen for the simulations used in the optimization process. Post-processing of the numerical results is carried out in MATLAB and the response quantities given a parameter set are also generated accordingly.

The simulation approach that couples COMSOL and MATLAB to model the vehicle-track dynamics is validated against the result in [54], which considers a similar track form, i.e., a connection between a floating slab track (FST) and a ballast track. In [54], the floating slabs are modeled in discrete precast sections, and the rails are periodically supported. The simulation setting for the track and vehicle system is modified according to the case in [54], including two axles in the simulation for comparison purposes. Figure 3.6 presents the wheel and rail displacement at the contact point when a vehicle moves from the FST to ballast track, showing that the patterns agree well with those generated from the reference case (see Figure 3 in [54]). The current simulation results indicate that the averaged rail and wheel displacement (for both axles) decreases by about 2.8 and 2.7 mm, respectively, when moving from the FST to the ballast track. In the reference case, 2.8 mm was reported for the change in the rail displacement, and 2.6 mm was for the wheel displacement. Therefore, it is considered that the current methodology to simulate the vehicle-track dynamics can be properly used as a basis for track parametric optimization.

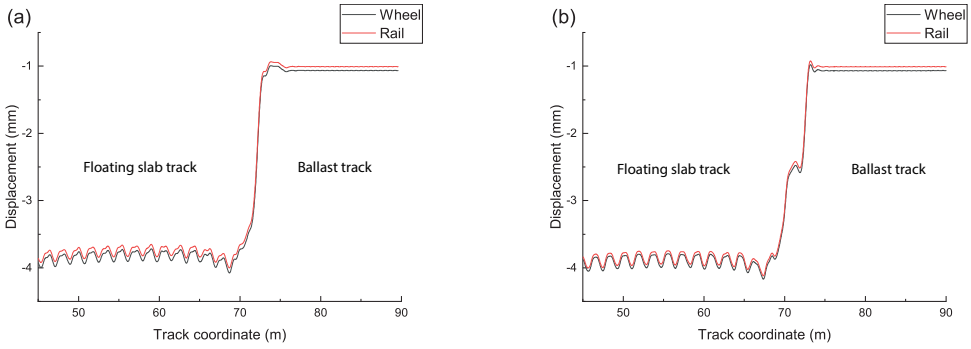

Figure 3.6: Wheel and rail displacement of the front axle (a) and rear axle (b) when a vehicle moves from the FST to ballast track.

3.3. NUMERICAL STUDY

3.3.1. CHARACTERIZATION OF THE VTI MODEL

Numerical examples are presented in this section to demonstrate the applicability of the proposed procedure. Vehicle parameters used in the model refer to VIRM trains, which are double-deck trains operated by NS (Dutch Railways). Table 3.2 lists parameter values of a full-car system according to VIRM technical parameters. A conversion is made in the computation to reduce a full car to the current 3-DOF model. Additionally, the vehicle parameters are further reduced to half following the half track.

Table 3.2: Vehicle parameters.

Parameter	Notation	Value
Carbody mass	m_c	56378 kg
Bogie mass	m_b	3772 kg
Wheelset mass	m_w	1819 kg
Primary suspension stiffness	k_{s1}	2445 kN/m
Secondary suspension stiffness	k_{s2}	2227 kN/m
Primary suspension damping	c_{s1}	2 kN·s/m
Secondary suspension damping	c_{s2}	50.1 kN·s/m
Static wheel load	P	100 kN
Contact constant	G	$5.13 \times 10^{-8}\text{m/N}^{\frac{2}{3}}$
Velocity	v	140 km/h

Track parameters correspond to characteristics of a typical Dutch track, which are based on a full-track scale and presented in Table 3.3. The rail properties conform to the nominal values of UIC54 rail, which are consistent in both track forms. The rail is pinned at both ends, which implies reflections of the vibration will occur at the ends. However, the boundary effect can be limited or negligible when the track length is sufficient [14]. To limit the boundary effect, the total track length in the demonstration cases has been extended to 134 m. This includes a ballast track section (100 m, on the left), an ERS-based level crossing (12 m, middle), and another ballast track section (22 m, right), as shown in Figure 3.3.

Sleepers provide periodic rail support in the ballast track. The commonly accepted spacing between centers of adjacent sleepers is 0.6 m [55], which is considered here with six rail elements per sleeper bay. Railpads and sleepers refer to Dutch standard components in the ballast track, i.e., FC9 4.5-mm cork-rubber pad and NS90 concrete sleeper. The properties are determined referring to Prorail system specifications and relevant works [56]–[58].

The embedded track in crossings corresponds to the standard solution for heavy-duty (ERS-type) level crossings. It consists of several concrete slabs, each with a length of 6 or 9 m. By combining different slabs, level crossings can be installed with variations in length. The slab is lowered onto a ditch filled with a mixed granulate and stabilized sand layer. As the ability to adjust the track geometry after the construction is limited,

Table 3.3: Track parameters.

Track Type	Component	Parameter	Value
Ballast/ERS	Rail (54E1)	Young's modulus	210 GPa
		Poisson's ratio	0.3
		Density	7800 kg/m^3
		Moment of inertia	2.337×10^{-5} m^4
		Area of cross section	6.977×10^{-3} m^2
Ballast	Railpad (FC9)	Stiffness	1120 MN/m
		Damping	120 kN·s/m
	Sleeper (NS90)	Mass	251 kg
		Support spacing	0.6 m
	Ballast/subgrade	Stiffness	45 MN/m
		Damping	96 kN·s/m
ERS	Elastic compound incl. rail strip	Stiffness	54 MN/m/m
		Damping ratio	0.2
	Slab	Young's modulus	31 GPa
		Poisson's ratio	0.3
		Density	2500 kg/m^3
		Width	2.37 m (bottom); 2.23 m (top)
		Length	6 m × 2
		Height	0.58 m
	Subgrade	Stiffness	500 MN/m/m
		Damping	20 kN·s/m

the requirement for the substructure is generally very high for slab track [57]. To reinforce the substructure layers, measures such as geogrids and injection mortar should be applied. Accordingly, in this chapter, a stiff substructure underneath the slabs is applied; the parameter values are chosen according to [56], which are collected in Table 3.3. The geometric and mechanical properties of the concrete slab correspond to the actual design of the ERS-type level crossings. Stiffness and damping properties of elastic fastening are gathered from product specifications calibrated based on laboratory experiments.

Note that the material properties of the track are different when they are measured by either quasi-static or dynamic loading tests. Material properties measured by quasi-static loading tests are commonly referred to as static properties, while those measured by dynamic loading tests are referred to as dynamic properties [59]. Although most previous works use static material properties to simulate vehicle-track dynamics [59], which is also the case in some works of railway design optimization (e.g., [60]), a proper dynamic simulation requires dynamic properties as input.

This work considers the effect of dynamic material properties on the vehicle-track dynamic simulation and chooses specific track component values based on the suggestions concluded from the comparative study in [59]. The most relevant material proper-

ties to dynamic excitations in the track are the stiffness of elastic elements and the modulus of elasticity of concrete. The dynamic stiffness of railpads significantly increases the dynamic impact factor based on the wheel-rail contact force, while the dynamic modulus of elasticity has a less effect compared with the former [59]. Therefore, dynamic stiffness values of railpads (FC9) and elastic compounds in ERS are adopted in the simulation, which are also the main design variables and, therefore, significant in the optimization problems.

It is also worth mentioning that the properties of railpads are temperature and frequency dependent and are also affected by preload and aging [56], [58], [59]. In railway practice, railpads are commonly simplified as spring and dashpot elements, and constant values are used to describe their viscoelasticity characteristics. This representation is adopted here, but more advanced models that take into account the sensitivity of railpads to these factors (such as [58]) can be incorporated for a more accurate representation. However, this would require more computational effort, especially when combined with simulations that involve vehicle-track interaction dynamics.

3.3.2. CASE STUDY: THE DUTCH LEVEL CROSSING DESIGN

Single-objective optimization problems are formulated for demonstration, where performance criteria, i.e., $F_{rms}, F_{max}, E_{rms}$ and E_{max}, are compared as dynamic benchmarks for track parametric optimization. Each measure is tested against the reference design (as defined in Table 3.3) to evaluate its effectiveness concerning solution quality and sensitivity to parametric variation. It is expected that the vehicle velocity highly influences optimization results as studies have shown that track response and expected degradation are sensitive to the train speed (e.g., [22], [59]). Here as the main purpose is to demonstrate the applicability of the integrative simulation approach, the velocity considered in the examples is defined as 140 km/h, referring to the standard speed of conventional passenger trains (e.g., VIRM trains) in the Dutch railway lines.

Table 3.4 showcases the variables and corresponding range of definitions used in the numerical examples. The first type of variable considers the stiffness distribution of railpads placed right next to the level crossing. The case study includes three pads, resulting in three design variables, $x_{ri}, i = 1,2,3$. The upper and lower bounds of the variables refer to the scope outlined in [61], where railpads are classified according to their vertical stiffness k_r (in unit: MN/m): very soft ($k_r = 100$), soft ($k_r = 200$), stiff ($k_r = 400$), and very stiff ($k_r = 800$). The current range of definitions covers the suggested values and also includes the consideration of the FC9 railpad (used in the reference design; see Table 3.4) to provide a reasonable design space. The stiffness change affects the railpad damping, which is scaled linearly with the stiffness values in the search process.

The second type of variable is related to sleeper parameters. The number of strengthened sleepers (x_n) applied in the transition zone is limited to 20, which forms an approximately 12m-long track section. This length is considered sufficient for the typical length of a level crossing approach, as it is prescribed to use $5 \sim 8$ strengthened sleepers in the approach to ERS-type level crossings. Additionally, based on experimental analysis of bridge approaches, Wang et al. [62] suggested 4.5 m as the upper limit for the length of the studied transition.

Another variable related to sleepers is spacing ($x_{sj}, j = 1,2,3$), which concerns the

Table 3.4: Track design variables and corresponding range of definition.

Track type	Component	Variables	Unit	Range of definition
Ballast	Railpad	Stiffness, $(x_{ri}, i = 1,2,3)$	MN/m	$x_{ri} \in \mathbb{R} : x_{ri} \in [50,1200], i = 1,2,3$
	Sleeper	Number of strengthened sleeper, (x_n)	-	$x_n \in \mathbb{Z} : x_n \in [0,20]$
		Sleeper spacing, $(x_{sj}, j = 1,2,3)$	-	$x_{sj} \in \mathbb{Z} : x_{sj} \in [3,7], j = 1,2,3$
ERS	Rail strip	Length of Type II strip, (x_l)	-	$x_l \in \mathbb{Z} : x_l \in [0,120]$

distances between the centers of three adjacent sleepers placed next to the junction. As previously mentioned, due to insufficient depth of ballast, there is an inherent operational discontinuity of tamping works at the ballast-to-slab connections, which may lead to a group of sleepers never being mechanically maintained. Apart from the ballast requirement, tamping machine operability also specifies the distances between sleepers. To avoid further disturbance to the regular tamping works, the number of sleepers is limited to three in the studied case. As shown in Table 3.4, the range of variables (x_{sj}) is adapted to the FE-based simulation environment, where the variables are discrete values rather than continuous (see Eq. 3.17) to align the optimization with the discretization of the FE method. The value of x_{sj} implies the number of finite elements, and each represents a 0.1 m-long rail element. The lower bound $(x_{sj} = 3)$ refers to the smallest sleeper spacing (0.3 m), and the upper bound $(x_{sj} = 7)$ defines the largest span as 0.7 m.

The last type of variable concerns the length of Type II rail strip (with a stiffness of 42 MN/m/m) in the embedded rail channel. Again the range of definition is adapted to the FE simulation environment. The lower bound $(x_l = 0)$ implies that Type II strip is not applied in the channel. The upper bound ($x_l = 120$, with each element in a length of 0.1 m) corresponds to the full length (12 m) of the example level crossing, meaning Type II strip is applied in full.

The general design vector $\mathbf{x}$ is reduced to $\mathbf{x} = \left[x_{r1}, x_{r2}, x_{r3}, x_n, x_{s1}, x_{s2}, x_{s3}, x_l\right]'$ in the studied case. In single-objective problems, the general objective function $f(\mathbf{x})$ is defined by each of the performance measures, i.e., $F_{rms}(\mathbf{x}), F_{max}(\mathbf{x}), E_{rms}(\mathbf{x})$, and $E_{max}(\mathbf{x})$.

As previously mentioned, termination depends more upon the computational budget in the surrogate-assisted optimization problems [50], [51]. The case study specifies a fixed budget of 200 function evaluations in the single objective problems to assess the effectiveness of the proposed objectives and identify appropriate ones for the follow-up search process. The statistical metrics, namely, rms and max values, are compared, and those showing higher solution quality and sensitivity to the parametric variation are chosen to formulate the MOO problem. A fixed budget of 400 function evaluations is assigned to the MOO problem to guarantee that the desired improvement level can be achieved.

In the multi-objective case, the scalarized global function, Eq. (3.19), contains two objectives, i.e., f_i $(i = 1,2)$. The parameter $\rho = 0.05$. The normalization vector is calculated by $\kappa_i = \frac{1}{f_i^{max} - f_i^r}$ $(f_i^r \leq f_i \leq f_i^{max}, i = 1,2)$, with f_i^r and f_i^{max} being the reference

(ideal) point and worst value obtainable for objective i, respectively. The ideal point defines the desired improvement level for each objective. Objectives are set of equal importance and the desired level is defined as a 20% improvement from the current reference design. The worst value, f_i^{max}, is determined according to the results simulated from the single-objective problems.

3.4. RESULTS AND DISCUSSION

3.4.1. SINGLE-OBJECTIVE OPTIMIZATION

The detailed processes of minimizing force (F)-related and energy (E)-related measures are presented in Figures 3.7 and 3.8, respectively, where the role of surrogate modeling and adaptive learning can be observed. All black triangles and dots represent sample points. Each point corresponds to a specific design solution $\mathbf{x}$ with eight variables ($\mathbf{x} = \left[x_{r1}, x_{r2}, x_{r3}, x_n, x_{s1}, x_{s2}, x_{s3}, x_l\right]^T$), which is evaluated by the objective or true function (i.e., the FE model). The objective function values shown in Figures 3.7 and 3.8 (the vertical axes) represent response quantities simulated from the FE model, which is dependent on $\mathbf{x}$ (the design alternatives).

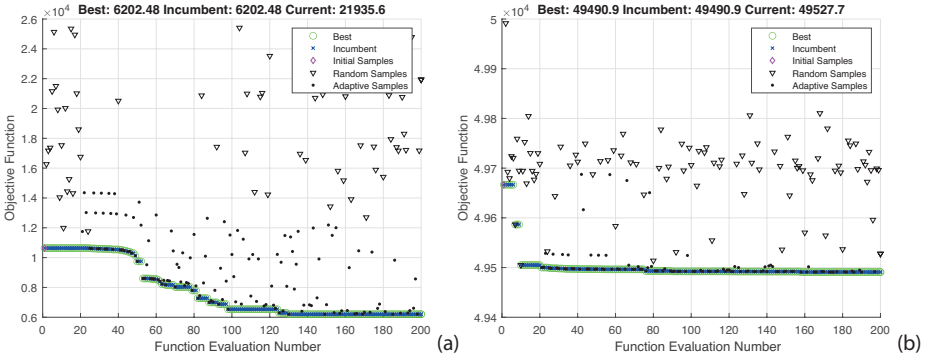

Figure 3.7: Search process of optimizing objective F_{max} (a) and F_{rms} (b).

In the surrogate construction phase, the first sequence of 20 points is used to build a cubic RBF, which are indicated by a pink diamond (as the initial point) and 19 black triangles (the random samples generated from the experimental design (ED)). The pink diamond, as noted by the legend 'Initial Samples' in the figures, represents the point(s) that is(are) specified beforehand, which in the studied case is the reference design. It means that all the optimization problems share the same starting point for comparison purposes. Besides, the size of the random samples generated in the vicinity of the starting point is determined referring to Regis and Shoemaker [48]. It highlights the measure of min $(2d, 10)$ with d being the problem dimensions to start a simulated annealing algorithm. The studied case has dimensions $d = 8$, and the sample size is considered sufficient for the initial surrogate construction.

In the adaptive learning phase, i.e., after evaluation number 20, samples are generated for surrogate updating and optimum search. Both black dots ('Adaptive Sam-

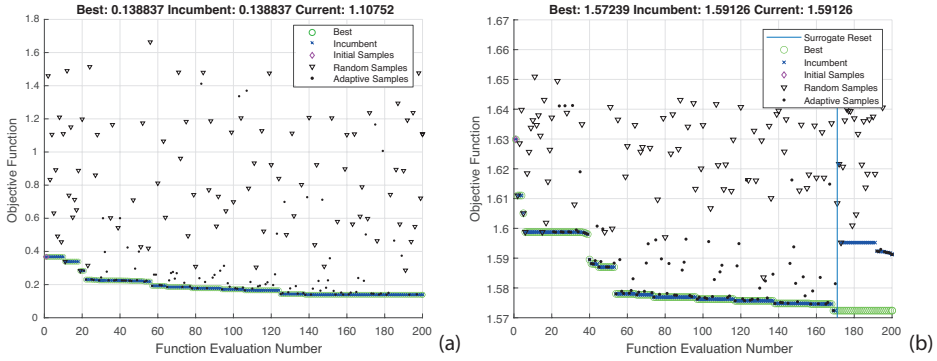

Figure 3.8: Search process of optimizing objective E_{max} (a) and E_{rms} (b).

ples') and triangles ('Random Samples') are shown in this phase. The adaptive samples are those found through random perturbations, and the random samples are searched by the Branch-and-Bound approach, as discussed in Section 3.2.3. All these points are evaluated by the FE model, and the thick green line keeps track of the best value found among all evaluated points. The current problems are formulated as minimization so that the best value implies the lowest objective function value (i.e., the minimum of the F- and E -related response quantities).

Specifically, Figure 3.7 (a) shows a noticeable drop in the best value near evaluation number 50, indicating the search process encounters a markedly better design option. Then the best point slowly drops in value and becomes stable after evaluation number 130. In Figure 3.7 (b), a rapid change in value is captured at the initial surrogate construction phase, whereas it shows little improvement afterward. Figure 3.8 (a) shows a trend similar to Figure 3.7(a), where the search process becomes stable around the evaluation number 130 ~ 140.

The solver may get stuck in a local minimum, which is measured by a distance-based tolerance parameter. In this case, the surrogate will be reset and returned to a new construction phase. While it may seem better to retain the past optimization trajectory in the next run, Regis and Shoemaker [48] suggested otherwise, based on the computational experience. This is because the past trajectory may bias the selection of candidate points towards the previously found optimum. The incumbent value, represented by the blue cross in the figures, is for tracking the lowest objective function value since the recent surrogate reset. In the process of optimizing the objective E_{rms}, a surrogate reset has occurred, which is indicated in Figure 3.8 (b) by a vertical straight line. It means that the evaluated points are tightly clustered around the best point, and a reset is needed to avoid getting trapped in the local optimum. For comparison purposes, further function evaluations are not assigned, where all the single-objective problems presented share the same starting point and computational budget. Besides, by comparing the scales of objective values, it can be observed from both figures that the rms-based values are less sensitive to changes in parameters, implying the statistical metric rms may not be a suitable indicator for optimization purposes.

Apart from investigating the search process, the solutions of the single-objective optimization and reference case are collected in Tables 3.5 and 3.6. Table 3.5 indicates the force (F)-related measures in general suggest higher railpad stiffness than the energy (E)-related measures, especially for F_{rms}, where the optimized values of variable $x_{ri}, (i = 1,2,3)$ almost reach the predefined upper bound. This can be explained by the fact that stiffer railpads contribute to reducing vibration from the wheel-rail contact (quantified by the F-related measures); however, this may lead to a higher effect of loads transmitted to underlayers, thereby causing vibration in sleepers and ballast (quantified by the E-related measures).

Table 3.5: Design solutions from single-objective optimization.

Design variables	Reference design	Design that optimizes			
		F_{rms}	F_{max}	E_{rms}	E_{max}
x_{r1} [MN/m]	1120	1199	103	59	50
x_{r2} [MN/m]	1120	1199	50	62	77
x_{r3} [MN/m]	1120	1199	72	50	50
x_n	8	0	4	0	1
x_{s1} [# of 0.1-m FE]	6	5	6	6	7
x_{s2} [# of 0.1-m FE]	6	7	6	3	5
x_{s3} [# of 0.1-m FE]	4	4	4	7	4
x_l [# of 0.1-m FE]	0	120	0	0	0

Moreover, all solutions indicate a limited number of strengthened sleepers (variable x_n) in the studied transition compared to the reference design. Regarding the spacing (variable $x_{si}, i = 1,2,3$), with the exception of E_{rms}, all objectives recommend the shortest distance (x_{s3}) between the structural interface and the adjacent sleeper.

Table 3.6: Objective values in single-objective optimization. Minimum (optimal) values are highlighted for each optimization problem.

	F_{rms} [N]	F_{max} [N]	E_{rms} [N·m]	E_{max} [N·m]
Design that optimizes F_{rms}	**49491**	14683	1.6367	0.7222
Design that optimizes F_{max}	49903	**6263**	1.6153	0.5620
Design that optimizes E_{rms}	49939	21508	**1.5724**	0.5575
Design that optimizes E_{max}	49961	12814	1.6049	**0.1388**
Reference design	49667	10636	1.6299	0.3676

The optimization process indicates that the full-length application of a Type II rail strip is the solution to minimizing F_{rms} with respect to x_l. However, this finding may be questionable, as the ballast track with FC9 pads was found to be stiffer than the ERS in the reference case. Applying the softer Type II strip in the ERS may lead to greater dynamic amplification in the structure. This is supported by the consistent solutions obtained from optimizing the other objectives, which resulted in $x_l = 0$.

Table 3.6 presents a comparison of the performance of the optimized designs to the reference design. As shown in column F_{rms} and E_{rms}, the rms-based values are less sensitive to parametric variation, which is consistent with the findings presented in Figures 3.7 and 3.8. It is worth noting that no single solution exists that can simultaneously optimize each objective. The conflicting nature of the objectives F_{max} and E_{max} can be captured from this table. The design that optimizes F_{max} (second row) has associated very large values of E_{max}. On the contrary, the design that optimizes E_{max} (fourth row) has associated very large values of F_{max}.

To visually evaluate the performance of rms and max metrics, Figure 3.9 compares the performance of designs obtained from optimizing F-related measures (a) and E-related measures (b), where the data are normalized based on Table 3.6 to account for different scales used in the objectives. Figure 3.9 (a) shows that optimizing F_{max} results in a reduction of two objective values, i.e., F_{max} and E_{rms}, whereas optimizing F_{rms} only improves itself. Similarly, Figure 3.9 (b) indicates that optimizing E_{max} can effectively improve E_{max} and E_{rms} without significantly degrading other objectives, as observed in the case of optimizing E_{rms}. Based on these results, it can be concluded that the max metric generally performs better than the other in terms of solution quality and sensitivity to parameter changes, making it more suitable for parametric optimization purposes.

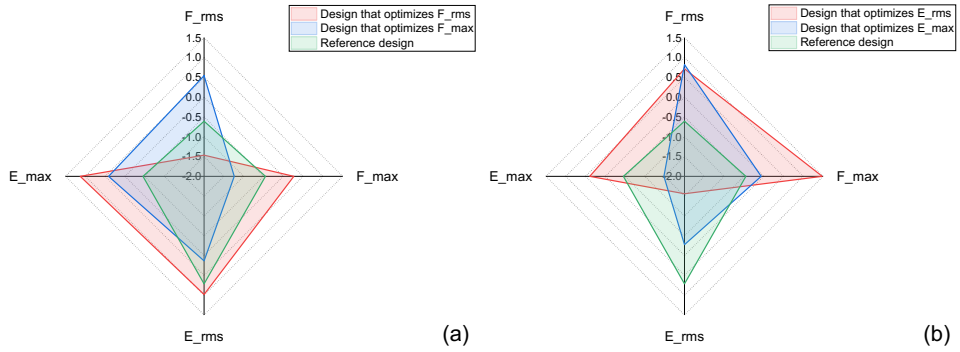

Figure 3.9: Comparison of metric rms and max in terms of solution quality. For representative purposes, the values are normalized with respect to objective values obtained from single-objective problems.

3.4.2. MULTI-OBJECTIVE OPTIMIZATION

The evaluation of single-objective problems in the previous section suggests that the metric max outperforms the others in terms of both solution quality and sensitivity to parametric variation. Figure 3.10 compares the normalized objective values obtained by minimizing F_{max} and E_{max}. The reference design serves as a benchmark, revealing the conflicting nature between the two objectives and suggesting the need for simultaneous optimization.

The scalarized global function (Eq. 3.19) is formulated by two objectives (F_{max} and E_{max}). Figure 3.11 shows the process of minimizing the global function. It can be observed that the search process becomes stable after evaluation number 280. In com-

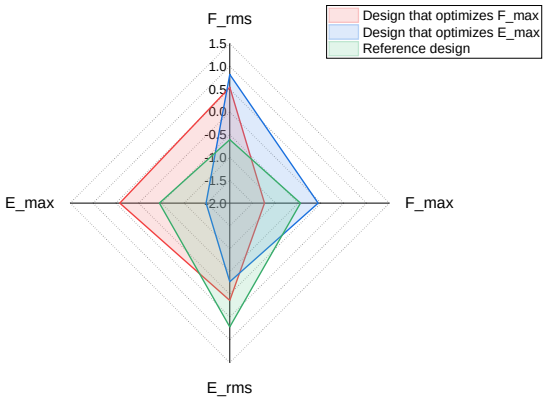

Figure 3.10: Comparison of objective F_{max} and E_{max} in terms of solution quality (normalized values; see also Figure 3.9).

parison to single-objective cases, the multi-objective problem is allocated an additional computational budget of 400 function evaluations. This is because the main purpose of the single-objective problems is to assess the effectiveness of candidate objectives and identify appropriate ones for the follow-up search. In contrast, the multi-objective problem is formulated to search for the optimum while guaranteeing that the desired improvement level can be achieved.

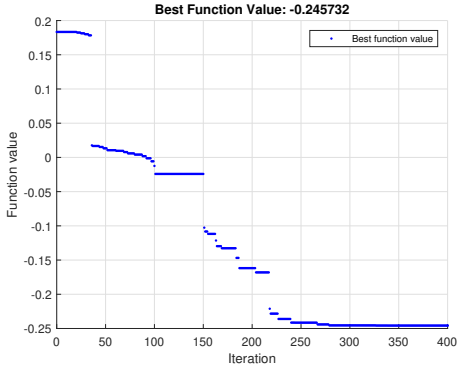

Figure 3.11: Process of simultaneous optimization of F_{max} and E_{max}.

As mentioned in Eq. (3.19), the ideal point $f_i^r (i = 1,2)$ represents the desired improvement level, which is defined as a 20% improvement from the reference design. This corresponds to target values of F_{max} and E_{max} of 8508.5 N and 0.2941 N·m, respectively. Table 3.7 lists the design solutions and corresponding objective values from single- and multi-objective problems. The optimized function values from the multi-objective problem (i.e., 5662.6 N for F_{max} and 0.1919 N·m for E_{max}) are below the targets, indicating the optimization has reached the desired improvement level. The results show a 46.8%

(F_{max}) and 47.8% (E_{max}) improvement from the reference design, which exceeds the target improvement. Notably, the value of F_{max} is even better than that obtained from the single-objective problem, demonstrating the complexity and nonlinearity of the objective function (simulated from the FE model). The solver may get trapped in regions containing a local optimum, and some searches may provide better solutions, as in the current multi-objective case.

Table 3.7: Design solutions and objective values for different optimization problems.

Objectives	Design vector $\mathbf{x}$ $\left[x_{r1}, x_{r2}, x_{r3}, x_n, x_{s1}, x_{s2}, x_{s3}, x_l\right]'$	F_{max} [N]	E_{max} [N·m]
Min F_{max}	$[103, 50, 72, 4, 6, 6, 4, 0]'$	6202.5	0.5620
Min E_{max}	$[50, 77, 50, 1, 7, 5, 4, 0]'$	12814	**0.1388**
Reference design	$[1120, 1120, 1120, 8, 6, 6, 4, 0]'$	10636	0.3676
MOO (F_{max},E_{max})	$[214, 155, 50, 5, 6, 7, 3, 0]'$	**5662.6**	0.1919

Moreover, using fewer strengthened sleepers in the transition, specifically reducing their number from 8 to 5, is preferable when compared to the reference. This is because an increased number of strengthened ones can make the ballast track stiffer. However, the solution still suggests a few numbers, which might be helpful in vibration isolation for the underlayers (typically the ballast) due to the larger size and improved stability. A comparison between the multi-objective and reference scenarios can be seen in Figure 3.12, showing a profile of energy dissipation under each sleeper. A total of 20 sleepers is presented, which corresponds to the upper bound of variable X_n. Sleeper number 1 is the one closest to the structural interface. It can be seen that near the interface, the energy dissipated in the optimized design is lower than that in the reference case, indicating an expected reduction in ballast degradation and the consequent impact on local track geometry.

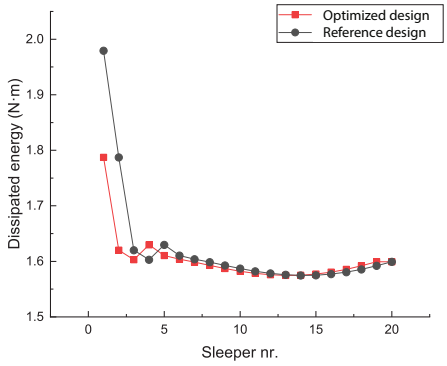

Figure 3.12: Comparison of optimized and reference design in terms of energy dissipation in track substructure.

The optimized sleeper spacing does not change much with the reference design. A

shorter distance is generally recommended between the structural interface and the first sleeper to ensure a smooth transition. The last variable indicates that the current Type I rail strip is preferable to the alternative as the former is stiffer and can balance the arrangement of track support stiffness between different track forms.

Note that the optimized results are influenced by the default parameter setting (Table 3.4). For instance, the railpad stiffness comes in a wide range of values. The default setting considers the stiff FC9 type, which makes the ballast track stiffer than the ERS structure. In this case, the solver guides the search toward the default (i.e., Type I) rail strip used in the ERS to guarantee a smooth stiffness transition. However, using other pads in the ballast track may make the ERS structure stiffer than the other, resulting in different design solutions. For instance, the Type II rail strip may become the preferred option. The case discussed is for demonstration purposes, and the default parameter setting refers to the Dutch standard practice.

3.5. CONCLUSIONS

This chapter focuses on the design optimization of level crossings and associated transition zones. A VTI model is developed to simulate the dynamic behavior of a level crossing, where an embedded rail structure is installed in the crossing with connections to the conventional ballast track. Based on this, an integrative simulation methodology is proposed that embeds the VTI model in the adaptive modeling scheme to facilitate the efficient exploration of the design space. The goal is to optimize the local track performance by varying geometric and elastic properties in track structures. The mitigation of track degradation may further reduce maintenance needs and the consequent impact on, e.g., system life-cycle cost and network performance.

Four different objectives are proposed to capture the performance of the track structure, taking into account the wheel-rail contact and lower supporting level. Several optimization problems are formulated accordingly, including single-objective problems for comparing the effectiveness of the candidate objectives, and the multi-objective problem to search for the optimal compromise solution. Compared to the reference design, the optimized solution shows a significant improvement in the most relevant objectives (46.8% in F_{max} and 47.8% in E_{max}). Therefore, it can be considered the best-known solution to support relevant decision-making in track preliminary design. While 400 function evaluations are assigned to obtain the solution, more computational resources can be allocated to the proposed methodology to search for (possibly) better solutions. Besides the final solution from the MOO formulation, evaluating the other solutions is meant to inform and improve the current design practice for transition zones, especially in areas using embedded rail structures.

REFERENCES

[1] Y. Shang, M. Nogal, R. Teixeira, and A. R. M. Wolfert, “Optimal design of rail level crossings and associated transition zones using adaptive surrogate-assisted optimization”, *Engineering Structures*, vol. 282, p. 115 740, 2023.

[2] Y. Shang, M. van den Boomen, A. P. de Man, and A. R. M. Wolfert, “Reliability-based life cycle costing analysis for embedded rails in level crossings”, *Proceed-*

ings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, vol. 234, no. 8, pp. 821–833, 2019.

[3] Z. Yang, P. Zhang, and L. Wang, "Wheel-rail impact at an insulated rail joint in an embedded rail system", *Engineering Structures*, vol. 246, p. 113 026, 2021.

[4] H. Wang and V. Markine, "Dynamic behaviour of the track in transitions zones considering the differential settlement", *Journal of Sound and Vibration*, vol. 459, p. 114 863, 2019.

[5] L. Ling, J. Han, X. Xiao, and X. Jin, "Dynamic behavior of an embedded rail track coupled with a tram vehicle", *Journal of Vibration and Control*, vol. 23, no. 14, pp. 2355–2372, 2017.

[6] V. Stančík, P. Ryjáček, and M. Vokáč, "Thermal and load rate-dependent interaction between embedded rail system and bridge", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 233, no. 3, pp. 326–336, 2019.

[7] C. Valkenburg, "Technisch onderzoek en LCM-analyse conserveringsmethoden (Technical research and LCM analysis of preservation methods)", in *Civiele Techniek*, 7, Routledge, 2015, pp. 2–3.

[8] L. Le Pen, G. Watson, W. Powrie, G. Yeo, P. Weston, and C. Roberts, "The behaviour of railway level crossings: Insights through field monitoring", *Transportation Geotechnics*, vol. 1, no. 4, pp. 201–213, 2014.

[9] J. G. Rose, "Rehabilitation techniques to improve long-term performances of highway-railway at-grade crossings", in *ASME/IEEE Joint Rail Conference*, vol. 54594, 2011, pp. 31–43.

[10] B. Indraratna, M. B. Sajjad, T. Ngo, A. G. Correia, and R. Kelly, "Improved performance of ballasted tracks at transition zones: A review of experimental and modelling approaches", *Transportation Geotechnics*, vol. 21, p. 100 260, 2019.

[11] C. Shen, R. Dollevoet, and Z. Li, "Fast and robust identification of railway track stiffness from simple field measurement", *Mechanical Systems and Signal Processing*, vol. 152, p. 107 431, 2021.

[12] M. Shamalta and A. Metrikine, "Comparison of the dynamic response of one- and two-dimensional models for an embedded railway track to a moving load", *Heron*, vol. 47, no. 4, pp. 243–262, 2002.

[13] H. Xia and N. Zhang, "Dynamic analysis of railway bridge under high-speed trains", *Computers & Structures*, vol. 83, no. 23-24, pp. 1891–1901, 2005.

[14] E. Aggestam, J. C. Nielsen, and R. Bolmsvik, "Simulation of vertical dynamic vehicle–track interaction using a two-dimensional slab track model", *Vehicle system dynamics*, vol. 56, no. 11, pp. 1633–1657, 2018.

[15] X. Lei, S. Wu, and B. Zhang, "Dynamic analysis of the high speed train and slab track nonlinear coupling system with the cross iteration algorithm", *Journal of Nonlinear Dynamics*, vol. 2016, 2016.

[16] X. Lei, *Model for Vertical Dynamic Analysis of the Vehicle-Track Coupling System*. Springer, 2017, pp. 161–199.

[17] "Railway applications - Ballastless track systems - Part 2: System design, subsystems and components.", European Committee for Standardization, Standard, 2017.

[18] J. N. V. da Silva Ferreira, "Long-term behaviour of railway transitions under dynamic loading application to soft soil sites", PhD dissertation, Universidade NOVA de Lisboa (Portugal), 2013.

[19] B. Z. Coelho, "Dynamics of railway transition zones in soft soils", PhD dissertation, Delft University of Technology, 2011.

[20] M. Sol-Sánchez, F. Moreno-Navarro, and M. C. Rubio-Gámez, "The use of elastic elements in railway tracks: A state of the art review", *Construction and building materials*, vol. 75, pp. 293–305, 2015.

[21] R. Sañudo, M. Cerrada, B. Alonso, and L. dell'Olio, "Analysis of the influence of support positions in transition zones. A numerical analysis", *Construction and Building Materials*, vol. 145, pp. 207–217, 2017.

[22] M. Sadri and M. Steenbergen, "Effects of railway track design on the expected degradation: Parametric study on energy dissipation", *Journal of Sound and Vibration*, vol. 419, pp. 281–301, 2018.

[23] M. Shahraki, C. Warnakulasooriya, and K. J. Witt, "Numerical study of transition zone between ballasted and ballastless railway track", *Transportation Geotechnics*, vol. 3, pp. 58–67, 2015.

[24] M. Sedghi, O. Kauppila, B. Bergquist, E. Vanhatalo, and M. Kulahci, "A taxonomy of railway track maintenance planning and scheduling: A review and research trends", *Reliability Engineering & System Safety*, vol. 215, p. 107 827, 2021.

[25] M. Steenbergen, A. Metrikine, and C. Esveld, "Assessment of design parameters of a slab track railway system from a dynamic viewpoint", *Journal of Sound and Vibration*, vol. 306, no. 1-2, pp. 361–371, 2007.

[26] M. Sadri, T. Lu, and M. Steenbergen, "Railway track degradation: The contribution of a spatially variant support stiffness-global variation", *Journal of Sound and Vibration*, vol. 464, p. 114 992, 2020.

[27] M. Sadri, T. Lu, and M. Steenbergen, "Railway track degradation: The contribution of a spatially variant support stiffness-local variation", *Journal of Sound and Vibration*, vol. 455, pp. 203–220, 2019.

[28] T. Chugh, "Scalarizing functions in Bayesian multiobjective optimization", in *2020 IEEE Congress on Evolutionary Computation (CEC)*, IEEE, 2020, pp. 1–8.

[29] A. P. Wierzbicki, "The use of reference objectives in multiobjective optimization", in *Multiple criteria decision making theory and application*, Springer, 1980, pp. 468–486.

[30] K. Miettinen, *Nonlinear multiobjective optimization*. Springer Science & Business Media, 1999, vol. 12.

[31] X.-D. Zhang, *A matrix algebra approach to artificial intelligence*. Springer, 2020.

[32] G. Chiandussi, M. Codegone, S. Ferrero, and F. E. Varesio, "Comparison of multi-objective optimization methodologies for engineering applications", *Computers & Mathematics with Applications*, vol. 63, no. 5, pp. 912–942, 2012.

[33] Q. Cheng, S. Wang, Z. Liu, and Y. Yuan, "Surrogate-based simulation optimization approach for day-to-day dynamics model calibration with real data", *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 422–438, 2019.

[34] M. Urquhart, E. Ljungskog, and S. Sebben, "Surrogate-based optimisation using adaptively scaled radial basis functions", *Applied Soft Computing*, vol. 88, p. 106 050, 2020.

[35] X. Han, H. Xiang, Y. Li, and Y. Wang, "Predictions of vertical train-bridge response using artificial neural network-based surrogate model", *Advances in Structural Engineering*, vol. 22, no. 12, pp. 2712–2723, 2019.

[36] H. Li, T. Wang, and G. Wu, "Probabilistic safety analysis of coupled train-bridge system using deep learning based surrogate model", *Structure and Infrastructure Engineering*, pp. 1–20, 2021.

[37] J. Müller and C. A. Shoemaker, "Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems", *Journal of Global Optimization*, vol. 60, no. 2, pp. 123–144, 2014.

[38] A. Bacigalupo, G. Gnecco, M. Lepidi, and L. Gambarotta, "Computational design of innovative mechanical metafilters via adaptive surrogate-based optimization", *Computer Methods in Applied Mechanics and Engineering*, vol. 375, p. 113 623, 2021.

[39] M. D. Buhmann, *Radial basis functions: theory and implementations*. Cambridge university press, 2003, vol. 12.

[40] H.-M. Gutmann, "A radial basis function method for global optimization", *Journal of global optimization*, vol. 19, no. 3, pp. 201–227, 2001.

[41] J. Müller, "MISO: Mixed-integer surrogate optimization framework", *Optimization and Engineering*, vol. 17, no. 1, pp. 177–203, 2016.

[42] Y. Duan, "A note on the meshless method using radial basis functions", *Computers & Mathematics with Applications*, vol. 55, no. 1, pp. 66–75, 2008.

[43] L. Shi, R.-J. Yang, and P. Zhu, "An adaptive response surface method for crashworthiness optimization", *Engineering Optimization*, vol. 45, no. 11, pp. 1365–1377, 2013.

[44] S. Roy, W. A. Crossley, B. Stanford, K. T. Moore, and J. S. Gray, "A mixed integer efficient global optimization algorithm with multiple infill strategy-applied to a wing topology optimization problem", in *AIAA Scitech 2019 Forum*, 2019, p. 2356.

[45] Y. Liu, L. Li, S. Zhao, and S. Song, "A global surrogate model technique based on principal component analysis and Kriging for uncertainty propagation of dynamic systems", *Reliability Engineering & System Safety*, vol. 207, p. 107 365, 2021.

[46] H. Wang, F. Ye, E. Li, and G. Li, "A comparative study of expected improvement-assisted global optimization with different surrogates", *Engineering optimization*, vol. 48, no. 8, pp. 1432–1458, 2016.

[47] E. C. Garrido-Merchán and D. Hernández-Lobato, "Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes", *Neurocomputing*, vol. 380, pp. 20–35, 2020.

[48] R. G. Regis and C. A. Shoemaker, "A stochastic radial basis function method for the global optimization of expensive functions", *INFORMS Journal on Computing*, vol. 19, no. 4, pp. 497–509, 2007.

[49] E. Davis and M. Ierapetritou, "A kriging based method for the solution of mixed-integer nonlinear programs containing black-box functions", *Journal of Global Optimization*, vol. 43, no. 2, pp. 191–205, 2009.

[50] V. Picheny, R. B. Gramacy, S. Wild, and S. Le Digabel, "Bayesian optimization under mixed constraints with a slack-variable augmented Lagrangian", *Advances in neural information processing systems*, vol. 29, 2016.

[51] Y. Wang and C. A. Shoemaker, "A general stochastic algorithmic framework for minimizing expensive black box objective functions based on surrogate models and sensitivity analysis", *arXiv preprint arXiv:1410.6271*, 2014.

[52] Y. Shang, M. Nogal, and A. R. M. Wolfert, "A co-simulation solution for vehicle-track interaction dynamics problems", in *The fifth international conference on railway technology: Research, development and maintenance*, Civil-Comp Press, 2022, pp. 31–15.

[53] Y.-B. Yang, J. Yau, Z. Yao, and Y. Wu, *Vehicle-bridge interaction dynamics: with applications to high-speed railways*. World Scientific, 2004.

[54] Z. Li and T. Wu, "On vehicle/track impact at connection between a floating slab and ballasted track and floating slab track's effectiveness of force isolation", *Vehicle System Dynamics*, vol. 47, no. 5, pp. 513–531, 2009.

[55] R. S. Ortega, J. Pombo, S. Ricci, and M. Miranda, "The importance of sleepers spacing in railways", *Construction and Building Materials*, vol. 300, p. 124 326, 2021.

[56] A. P. de Man, "Dynatrack: A survey of dynamic railway track properties and their quality", PhD dissertation, Delft University of Technology, 2002.

[57] C. Esveld, *Modern railway track*. MRT-productions Zaltbommel, 2001, vol. 385.

[58] M. Oregui, A. Núnez, R. Dollevoet, and Z. Li, "Sensitivity analysis of railpad parameters on vertical railway track dynamics", *Journal of Engineering Mechanics*, vol. 143, no. 5, p. 04 017 011, 2017.

[59] T. Li, Q. Su, and S. Kaewunruen, "Influences of dynamic material properties of slab track components on the train-track vibration interactions", *Engineering Failure Analysis*, vol. 115, p. 104 633, 2020.

[60] E. Aggestam and J. C. Nielsen, "Multi-objective optimisation of transition zones between slab track and ballasted track using a genetic algorithm", *Journal of Sound and Vibration*, vol. 446, pp. 91–112, 2019.

[61] B. Ripke and K. Knothe, "Simulation of high frequency vehicle-track interactions", *Vehicle System Dynamics*, vol. 24, no. sup1, pp. 72–85, 1995.

[62] H. Wang, V. Markine, and X. Liu, "Experimental analysis of railway track settlement in transition zones", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit*, vol. 232, no. 6, pp. 1774–1789, 2018.

4

PREFERENCE-BASED DESIGN OPTIMIZATION FOR LEVEL CROSSINGS

With the previous chapter focusing on the technical perspective, this chapter moves to a wider problem-solving context and introduces a design approach to incorporate social relevance (reflected in multi-stakeholder preferences) into railway track design. This approach combines FE modeling with preference modeling to optimize the design of transition zones while considering the objectives of relevant stakeholders. The FE model characterizes the dynamic behavior of the track, providing insights into the expected level of track degradation and serving as performance measures for design optimization. The preference modeling is utilized to incorporate stakeholder preferences into optimization problems. A case study focused on the optimization of track support stiffness for level crossings is presented. The results highlight the significant influence of integrating stakeholder preferences on the optimal track design configuration. This allows the level crossing design to be managed focusing on the best fit for common purpose rather than solely on mechanical behavior.

Parts of this chapter have been published verbatim in Life-Cycle of Structures and Infrastructure Systems, 949-956 (2023) [1].

4.1. INTRODUCTION

Railway infrastructure consists of tracks, signaling systems, power installations, and other supporting facilities. Beyond these physical assets, it constitutes a technically and organizationally complex system where the physical components are interconnected and interact with human entities throughout their operational life [2]. Within such systems, the overall performance at the system level, including technical, socio-economic, and environmental aspects, is influenced by the behavior and decisions of various stakeholders [3]. A good synergy of all system components is therefore necessary to ensure the proper functioning of the system.

Given the nature of the complex systems, the decisions of railway activities should be examined in relation to its internal and external environment [2]. In the context of railway design, the internal factor is related to the mechanical behavior of the track structure, specifically addressing the technical needs. This is typically manifested in the traditional approach to designing a classic railway track, which emphasizes load-bearing considerations and track maintainability [4]. The external environment can be synthesized by considering the perspectives of different stakeholders, such as the riding comfort of train users and the maintenance effort expected from service providers. Above all, ensuring the safe operation of rolling stock (i.e., fulfilling the technical requirements) is the primary objective of rail infrastructure and should be prioritized in all track design activities. However, while adhering to these technical constraints, rational and optimal system-level decisions would require the alignment of management objectives from the stakeholders with engineering systems design.

Let us consider the design of transition zones to highlight the need for an integrated design approach. As elaborated in Chapter 3, a common strategy to reduce dynamic amplification at transition zones is to ensure a gradual change in vertical stiffness along the track. This can be achieved through various means, for instance, by incorporating elastic elements with varying properties such as the railpads and under sleeper pads. These elements can effectively modify the track support stiffness and mitigate vibration and noise [5]. Additionally, the design parameters of sleepers, which distribute vehicle loads onto the track, have a significant impact on the dynamic behavior of the railway track [6]. However, these design adjustments may require additional use of track components, which drives up construction costs. A study by [7] highlights the economic implications of sleeper spacing. It suggests that a substantial reduction in construction per kilometer, around 40%, can be achieved by increasing the sleeper spacing to 1 m as compared to the standard 0.6 m spacing.

Beyond their economic consequences, variations in the parameters of track components also influence maintenance routines and the quality of service. A more uniform distribution of track support stiffness can reduce dynamic impacts, thereby alleviating track degradation. This further lessens the need for maintenance and reduces the associated costs, aligning with the management goals of maintenance service providers. Moreover, the interaction between the track and vehicles means that the level of dynamic amplification in the railway track also affects the vehicle responses, such as carbody accelerations. These responses are directly linked to passenger comfort, thus influencing the overall quality of service.

It shows from the above that the variations in track design hold relevance to differ-

ent parties. This stresses the need for an integrative approach that accounts for both technical feasibility and stakeholder preferences at the track design stage. However, the coordination between engineering and management practices is often insufficient due to the limited involvement of stakeholders in the design process, which is also emphasized in [8]. Traditional design methods for transition zones mainly focus on meeting technical requirements, where the mitigation measures typically consider mechanical responses such as wheel-rail contact forces, without adequately addressing their social implications.

To address this gap, this chapter presents a design approach that integrates engineering and managerial aspects to model design problems for railway tracks. Since stakeholder preferences may often conflict (where no single design solution exists that simultaneously satisfies all), it becomes necessary for engineers to navigate a space of trade-offs, where optimization problems are solved to strike a balance between the stakeholder preferences and the mechanical performance of the track. For this purpose, a preference-based optimization tool is combined with the FE model developed in Chapter 3. This combination allows for the consideration of a wider variety of design aspects: the preference-based optimization tool translates vague societal needs into crisp engineering design variable values, while the FE model characterizes the dynamic behavior of the track under moving vehicles and serves as a basis for parametric optimization.

4.2. PROBLEM DESCRIPTION

Among the stakeholders involved in the lifecycle management of railways, three groups are particularly relevant: infrastructure managers, train users, and maintenance service providers. This consideration stems from the organizational transit in European railways, i.e., a separate administration between infrastructure and operation [2], [4]. Specifically, in the Netherlands, the primary tasks of infrastructure management include infrastructure planning, infrastructure maintenance, and capacity management & traffic control [2]. The maintenance part is outsourced through a performance-oriented maintenance contract (PGO). Maintenance of the entire railway network is divided over 21 geographically defined contract areas, and several recognized maintenance parties carry out a large part of the small-scale maintenance in these contract areas. For this reason, infrastructure managers and maintenance service providers are separated in the following analysis to reflect this situation.

OBJECTIVE 1: CONSTRUCTION COSTS

One of the direct concerns of infrastructure managers is the costs associated with constructing new or renovated track lines. The construction cost, as defined by [2], refers to the amount of resources spent on the construction of a railway activity, which is influenced by various factors, including layout characteristics (mainly the number of civil engineering structures, switches, and crossings), labor cost, expropriation cost, and the number of electrical substations. This study specifically addresses transition zones where the construction costs are linked to the implementation of mitigation measures. These measures may involve variations in sleeper spacings and the use of alternative types of railpads and sleepers.

OBJECTIVE 2: RIDING COMFORT

While operating in a complex track condition (for example, with the curve radius, line ramp, and track irregularities), trains are exposed to excitation through contact with the rail. These vibrations are then transmitted from the floor and seats, impacting the comfort of passengers. Passenger ride comfort remains a pressing topic as the level of service quality can influence user preferences for a particular means of transport. Ensuring ride comfort relies on various factors, such as vehicle operations and conditions of the track and vehicle. This study focuses on the evaluation of riding comfort influenced by design variations in transition zones, thus emphasizing the track condition.

The riding comfort is quantified by *max-to-min difference of carbody accelerations* induced when a vehicle passes through a transition zone. Lower carbody accelerations are indicative of a higher expected level of service. The selection of carbody accelerations as the index is based on the vibration transmission path of a vehicle [9]. A running vehicle is assembled with two levels of suspension systems. The primary suspension system mainly influences the running stability of the vehicle [10], and the secondary level of suspension, which forms a direct connection to the carbody, plays an essential role in ensuring the ride comfort of the vehicle [9], [11].

Further, as concluded in the preceding chapter, the statistic metric of the max-to-min difference proves effective in terms of solution quality and sensitivity to parameter changes. This metric is used for the carbody acceleration response in this chapter. Note that it is also possible to associate the vibration data (i.e., carbody accelerations) with human subjective feelings. This can be achieved using methods such as Sperling's index [12]. In this chapter, the direct usage of vibration data is considered as an indicator of comfort levels. The assessment of Sperling's index will be introduced in the subsequent chapter.

OBJECTIVE 3: EXPECTED LEVEL OF TRACK DEGRADATION

For maintenance contractors, track degradation is of particular concern because it directly affects the amount of required maintenance work, the associated costs, and potential penalties due to excessive track possession time. Consequently, their primary objective is to minimize track degradation issues and maintain track conditions at an acceptable level. While infrastructure managers share this objective, it is not part of their direct management goals since maintenance work is outsourced through PGO contracts. Under these contracts, contractors have the flexibility to plan and carry out maintenance work within specified timeframes and geographical regions [13].

Track degradation refers to the deviation of the track condition from the ideal state, occurring at various levels of track structures with different scales. It can range from metallurgical defects within the microstructure of the rail to large-scale track settlements [4]. Typically, as discussed in Chapter 1, the degradation in transition zones is manifested as irregularities in track geometry. Therefore, as the main driver causing the geometry degradation, this chapter mainly concerns the expected damage to the ballast and underlying layers, which can be linked to the mechanical energy dissipated in these layers, as elaborated in the previous chapter.

The *maximum differential energy dissipation* between adjacent sleepers is selected as an indicator to assess the sensitivity of a track design to the expected damage [14].

This is computed through Eq. (3.16). The higher the energy dissipated into the ballast layer, the stronger the expected degradation. Consequently, one of the main objectives in reducing damage to the overall track geometry condition and saving on maintenance efforts (and costs) is to minimize the amount of energy dissipated.

OVERVIEW OF DESIGN OBJECTIVES

Table 4.1 provides a summary of the objectives and corresponding performance measures outlined in the current design problem. The aim is to provide a methodological basis that integrates socio-technical perspectives with stakeholder preferences into the decision-making process for railway track design. Starting from this basis, future work may consist in incorporating additional design perspectives, performance measures, or specific models capable of addressing more complex track degradation phenomena in the proposed design method.

Table 4.1: Objectives and performance measures.

Objectives	Performance measures	Unit
Minimize construction cost, $g_1(\cdot)$	C_{cap} - construction cost	€
Maximize riding comfort, $g_2(\cdot)$	A_{max} - max-to-min of carbody acceleration	m/s^2
Minimize expected long-term degradation, $g_3(\cdot)$	E_{max} - max. differential energy dissipation	$N \cdot m$

4.3. SOLUTION METHOD

Multi-objective optimization is applicable when decisions need to be taken in the presence of trade-offs between several objectives. The previous chapter has formulated an MOO problem given the conflicting nature of the force-related and energy-related objectives. This chapter dives into this topic and focuses on how stakeholder preferences can be modeled and incorporated into the design process. Details are presented in the following sections.

4.3.1. OVERVIEW AND RELATED CONCEPTS

The process of solving multi-objective optimization problems is typically divided into two stages: search and decide [15]. The *search* stage involves the optimization of objective functions. The *decide* stage refers to making decisions about what kind of tradeoffs are appropriate, which requires the designer to articulate the preferences of stakeholders. These preferences represent the stakeholders' opinions about points in the objective space [16].

The classification of multi-objective optimization methods is fundamentally based on the different strategies used to articulate preferences. In *a priori* preference articulation, stakeholder preferences are quantified before the search process begins, meaning the preferences are determined prior to examining the points in the objective space [16]. On the other hand, *a posteriori* methods impose stakeholder preferences directly onto a set of points in the objective space, leading to final solutions that align with the stakeholders' regions of interest. This set of points in the objective space is commonly referred

to as the *Pareto front*. When these points are projected back into the design space, the corresponding solutions are known as the *Pareto optimal solutions*, which collectively form a *Pareto optimal set*. A more extensive explanation of these terms can be found in [15]–[17].

Building on this concept, Figure 4.1 presents a mind map that illustrates the preference-based design process for railway tracks. The diagram shows potential pathways for design optimization, which is divided into three main blocks: definition, optimization, and selection. The *definition* block formulates a design problem that defines both the design and objective space. The *optimization* block refers to the search process, with various types of optimization techniques available to find optimal solutions. The *selection* block focuses on the decision-making process, which links stakeholder preferences to final solutions.

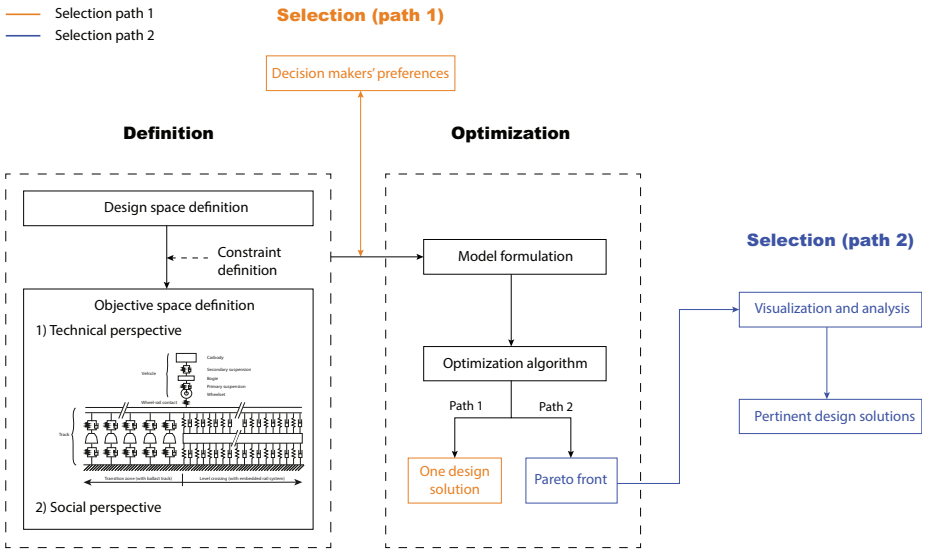

Figure 4.1: Overview of the preference-based design process for railway tracks.

Specifically, in the definition block, the objective space showcases two design perspectives: technical and social. These are determined based on the problem definition in Section 4.2. The social perspective is linked to the objective of minimizing the construction costs, $g_1(\cdot)$ in Table 4.1, which is defined by an explicit function and will be elaborated in Section 4.4. The technical perspective concerns the mechanical performance of the railway track, with the aim of maximizing riding comfort (i.e., minimizing A_{max}) and minimizing the expected long-term degradation (i.e., E_{max}). These objectives are represented by $g_2(\cdot)$ and $g_3(\cdot)$, respectively, in Table 4.1. While improving riding comfort and service quality has social relevance, the corresponding performance measure is evaluated based on the mechanical responses of the system. Therefore, it falls within the scope of the technical perspective when formulating optimization problems.

A computational model is required to evaluate the objective functions $g_2(\cdot)$ and $g_3(\cdot)$. This model is designed to characterize the relationship between the design parameters and dynamic responses of a vehicle-track system. In this context, the FE model developed in Chapter 3 is applied, and the current design problem can be formulated as

$$\begin{aligned} &\min_{\mathbf{x}} \mathbf{G}(\mathbf{x}) = \min_{\mathbf{x}} \left\{ g_1(\mathbf{x}), g_2(\mathbf{x}), g_3(\mathbf{x}) \right\}, \\ &\text{s.t.} \quad \mathbf{x} \in \Omega, \end{aligned} \tag{4.1}$$

where $\mathbf{x} = [x_1, x_2, ..., x_d]$ is a vector containing a list of design parameters defined in d-dimensional space, e.g., railpad stiffness, sleeper spacing. $g_i(\mathbf{x})$ is the i^{th} objective and $\mathbf{G} : \Omega \rightarrow \Lambda$ maps the design variables, $\mathbf{x}$, to the vector, $\mathbf{y} = \left[y_1, ..., y_k \right]$, in the objective function space Λ. Here, $k = 3$ objectives are considered.

The selection block shown in Figure 4.1 presents two general pathways: Path 1 and Path 2, which respectively represent a priori and a posteriori methods of articulating preferences. These strategies chosen to handle preferences significantly influence the design outcomes. Essentially, Path 2 leads to a set of solutions that corresponds to the idea of Pareto optimality, whereas Path 1 results in a single solution point. This is because Path 1 incorporates preferences before the optimization process, directing the search towards the preferred part of the Pareto front. It is worth noting that in this case, the parameters that define stakeholder preferences can be systematically altered over repeated runs (i.e., resulting in different problem settings) to explore multiple solutions. These solutions collectively provide an approximate representation of the Pareto set [18]. However, to highlight the fundamental difference between the a priori and a posteriori methods (i.e., the number of solutions derived from one problem formulation), this is not visualized in Figure 4.1.

Various methods can be used to articulate preferences in an a priori manner. For instance, a single solution can be seen as a compromise solution. This refers to an optimal point where the distance between the solution point and the ideal point (from the stakeholder's perspective) is minimized in the objective space [16]. Alternatively, it can be a synthesized solution sought by aggregating and maximizing the preferences of all stakeholders. This strategy is demonstrated in the IMAP (Integrative Maximized Aggregated Preference) method, as introduced in [8]. In this chapter, the compromise solution approach, a standard form of a priori methods, is employed to demonstrate how engineering and managerial aspects are integrated into railway design problems. For other a priori methods, readers can refer to [8].

The following sections present a solution method for the current multi-objective optimization problem. This method involves three key steps that are designed to incorporate preference modeling into the search process and to achieve a balanced design solution for all stakeholders. First, a global criterion method is employed to search for a compromise solution. Second, a preference handling scheme is integrated into each objective function, which is elaborated in Section 4.3.2. This allows stakeholders to specify their preferences in a flexible and intuitive manner, aligning the decision-making process more effectively with the desired goals. Finally, in Section 4.3.3, surrogate modeling is combined with the preference-based design process to reduce the computational costs associated with evaluating the objective functions.

4.3.2. PREFERENCE-BASED DESIGN OPTIMIZATION

The global criterion method aims to minimize a function (i.e., a global criterion) that measures how close the designer can approach the ideal vector. The ideal vector is defined by $\mathbf{G}^0 = \left[g_1\left(\mathbf{x}^0\right), g_2\left(\mathbf{x}^0\right), ..., g_k\left(\mathbf{x}^0\right)\right]'$, where $g_i\left(\mathbf{x}^0\right)$ represents the optimal value (by default, the minimum value) of the i^{th} function achieved at $\mathbf{x}^0$. In simpler terms, the ideal vector gathers the optimum of each objective achieved at the same point, $\mathbf{x}^0$ [15]. This forms an ideal or utopia point in the design space. However, achieving this ideal point is typically unfeasible with conflicting objectives. Therefore, the practical alternative is to aim for a solution that is as close as possible to this point, i.e., a compromise solution, as previously discussed.

The closeness to the ideal point can be measured by a family of L_p metrics [15], [19], which is defined as

$$L_p(\mathbf{G}) = \left[\sum_{i=1}^{k} \left|g_i\left(\mathbf{x}^0\right) - g_i\left(\mathbf{x}\right)\right|^p\right]^{1/p}, \quad 1 \le p \le \infty, \tag{4.2}$$

where the value of p determines the type of distance. When $p = 1$, the L_p metric emphasizes 'group utility', where all deviations from $g_i\left(\mathbf{x}^0\right)$ are equally considered. When $2 \le p \le \infty$, larger deviations are given greater weights in the L_p metric. In particular, $p = 2$ indicates the use of the Euclidean norm [16]. When $p \to \infty$, the L_p metric only considers the largest deviation, which results in 'individual utility', i.e., the *min-max* method.

There are various ways to incorporate preference information in engineering optimization. Many of them draw from the classical multi-criteria decision analysis literature [20]. The common methods include using coefficients (or parameters), rankings, constraints, and preference functions. The weighting scheme and reference points (such as the ideal point as mentioned earlier) belong to the category of using coefficients. Rankings can be classified into solution rankings and objective rankings: the former allows the designer to directly express preferences over a set of solutions (e.g., [21], [22]), whereas the latter focuses on the order of objectives (e.g., [23]). Some also combine the methods of handling preferences, such as the lexicographic approach that uses objective rankings and constraints to define preference information in optimization problems [19]. The reader is referred to [19] for fundamental concepts of preference modeling and [20] for a systematic review of applications of preference incorporation in evolutionary optimization.

The focus of this chapter is the method based on preference functions, which involves associating each objective with a preference function. This preference function can be seen as an individual utility function that provides stakeholders with the flexibility to express their preferences for each objective. The preferences can be related to concerns or constraints about certain values of an objective. For each objective, stakeholders are invited to provide a numerical range (or ranges) that represents degrees of satisfaction. These ranges are determined by the limits (also called preference thresholds [24]) of the corresponding objective values.

Essentially, a preference function transforms objective values into preference scores, which operate on a dimensionless scale, such as 0 to 100. This transformation effectively shifts the problem into a preference-based decision-making domain, thereby creating a

different objective space. From this perspective, the integration of preference functions also serves as function transformations in Eq. (4.2) to allow for a proper comparison among multiple objectives (with different measurement units and scales).

The idea of preference functions has evolved into many variant forms over recent decades. The relevant concepts/methods are *desirability functions* [25], [26] and *physical programming* [27]–[29]. Specifically, physical programming has been applied to many engineering problems, ranging from structural design (such as aircraft wings, beam and truss structures [27]) to production planning [30], [31]. In its original form [28], the construction of the preference functions (which are called the class functions in physical programming literature) is conditioned to the use of gradient-based methods for solving optimization problems. For this reason, the preference functions are defined as splines to preserve specific mathematical properties such as derivativity and convexity. Further in [27], [32], the method is extended to *global physical programming* to relax the constraints on the curvature of preference functions, where piecewise exponential or linear functions can be used.

In most cases, individual preference functions are aggregated into a measure of 'group utility' using various forms, such as summation, exponential, and logarithmic functions. These aggregated functions are then merged with multi-objective evolutionary algorithms to guide the search process to a specific region of the Pareto front to improve the pertinency of the design solutions (e.g., [29]).

In this chapter, preference functions are introduced to the objective functions within the framework of the min-max approach (see Eq. (4.2), when $p \to \infty$). By implementing the min-max approach, which focuses on the 'individual utility' or the (single) objective that causes the largest deviation with respect to the reference point, the preference information (preference scores) of different objectives encoded in this problem formulation is essentially not aggregated. This relaxes the required constraint when constructing preference functions in global physical programming, where the limits of each preference range must present the same image in the preference function for all objectives involved. Further details of this constraint can be found in [27].

As previously discussed, the min-max approach aims to find a solution that minimizes the maximum deviation between the reference values for the objectives and values of a potential solution. When preference functions are incorporated into the min-max approach, all solutions are mapped to preference scores on a specified scale. Considering a 0-100 scale, the reference values correspond to target values of 100. This mapping allows stakeholder preferences to be translated into relative deviations from the goal. Thus, the minimization of the largest deviation can be formalized as

$$\min_{\mathbf{x}} \max_{i} \left\{ w_i \left[100 - P_i \left(g_i \left(\mathbf{x} \right) \right) \right] \right\}, \tag{4.3}$$

where $P_i(\cdot)$ is the preference function corresponding to i^{th} objective. It translates $g_i(\mathbf{x})$ to a preference measurement, where preference information is embedded in the function and used to rank design solutions. w_i is the weighting factor associated to i^{th} objective.

4.3.3. KRIGING OR GAUSSIAN PROCESS MODELING

As mentioned in Chapter 1, the use of optimization algorithms in railway track design tends to be computationally demanding due to the complexity of the computational models. Here, surrogate modeling techniques are integrated into the present design method to reduce the computational cost. In particular, the responses being approximated are E_{max} and A_{max}, which correspond to the functions $g_2(\cdot)$ and $g_3(\cdot)$, respectively, in Table 4.1.

The effectiveness of using surrogate models is influenced by various factors, including problem types and modeling conditions (such as dimensionality, (non)linearity, and sample size). Kianifar & Campean [33] systematically compared the performance of several surrogate modeling options, including polynomials, radial basis function, and Kriging, across well-defined problem categories. The experiment used two sample sizes: a small ED size of $10d$ and a large size of $30d$, with d denoting the problem dimension. The results indicate that the Kriging model with Matérn 5/2 correlation function shows competing performance among the candidates in terms of accuracy and robustness for both sample sizes. Kriging is especially useful in capturing the local variability of the model output [34]. Since the current design approach adopts a fixed ED size, and optimal values are usually local phenomena, Kriging is chosen here to better capture the local characteristics of the computational model. A brief review of Kriging basics is presented in the following.

Kriging, also known as Gaussian process modeling, is one particular surrogate model that considers the function to approximate as a realization of a stochastic process [35]. It can be expressed as

$$\mathscr{M}(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\,\boldsymbol{\beta} + Z(\mathbf{x}), \tag{4.4}$$

where the first term, $\mathbf{f}^T(\mathbf{x})\,\boldsymbol{\beta}$, is the mean value of $\mathscr{M}(\mathbf{x})$, including q arbitrary functions $\{f_j; j = 1,...,q\}$ and the corresponding coefficients $\{\beta_j; j = 1,...,q\}$. It represents the global characteristics (also called the trend) of the model. The second term, $Z(\mathbf{x})$, captures the local deviations by a Gaussian process with expectation being zero and variance being σ^2.

In the case of scalar output, once training samples $\boldsymbol{X} = \{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, ..., \mathbf{x}_{(N)}\}^T$ are determined, the corresponding output $\boldsymbol{Y} = \{y_{(1)}, y_{(2)}, ..., y_{(N)}\}^T$ can be obtained by querying the FE model at the sampled points. This process creates a training set, denoted as $\Phi = \{(\mathbf{x}_{(\varrho)}, y_{(\varrho)}) \,|\, \varrho = 1,...,N\}$, and with this training set, the covariance of $Z(\mathbf{x})$ can be expressed by

$$\operatorname{cov}\left[Z(\mathbf{x}), Z(\mathbf{x}')\right] = \sigma^2 R(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}), \tag{4.5}$$

where $R(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})$ is the correlation function for any pair of input samples $(\mathbf{x}, \mathbf{x}')$, whose hyperparaters are gathered in the vector $\boldsymbol{\theta}$. The choice of the correlation function is based on assumptions regarding the level of smoothness and regularity of the underlying model [36]. Here, the Matérn 5/2 correlation function is considered given its accuracy and robustness as evaluated in [33], which can be defined in the one-dimensional case as

$$R(x, x'; \theta) = \left(1 + \sqrt{5}\frac{|x - x'|}{\theta} + \frac{5}{3}\left(\frac{|x - x'|}{\theta}\right)^2\right) \exp\left[-\sqrt{5}\frac{|x - x'|}{\theta}\right], \tag{4.6}$$

where θ represents a scale parameter. When multidimensional problems are considered ($d > 1$), the term $\frac{|x-x'|}{\theta}$ in Eq. (4.6) is replaced by $\left[\sum_{r=1}^{d}\left(\frac{x_r-x'_r}{\theta_r}\right)^2\right]^{0.5}$ and θ becomes a vector $\boldsymbol{\theta}$ with corresponding parameters θ_r ($r = 1,...,d$). This follows an ellipsoidal construction of the multi-dimensional correlation function for Gaussian process models [37].

The training of a Kriging model is based on the set $\Phi = \left\{\left(\mathbf{x}_{(\varrho)}, y_{(\varrho)}\right) | \varrho = 1,...,N\right\}$, from which the generalized least-square estimates of the coefficients $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{F}^T\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{R}^{-1}\boldsymbol{Y}, \tag{4.7}$$

and the estimate for the variance σ^2 is obtained by

$$\hat{\sigma}^2 = \frac{1}{N}\left(\boldsymbol{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right)^T\mathbf{R}^{-1}\left(\boldsymbol{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right), \tag{4.8}$$

where $\mathbf{F} = [\mathbf{f}^T\left(\mathbf{x}_{(1)}\right),...,\mathbf{f}^T\left(\mathbf{x}_{(N)}\right)]^T$ is the $N \times q$ matrix gathering the regression functions evaluated on the training points $\boldsymbol{X}$.

Then, the scale parameters in the correlation function, $\boldsymbol{\theta}$, need to be estimated to fully define a Kriging model. This can be achieved through leave-one-out cross-validation by solving the following optimization problem [34], [38],

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\left[\boldsymbol{Y}^T\mathbf{R}^{-1}\,\mathrm{diag}\left(\mathbf{R}^{-1}\right)^{-2}\mathbf{R}^{-1}\boldsymbol{Y}\right]. \tag{4.9}$$

Once the model is determined, the Kriging predictor for any given new point $\boldsymbol{x}$ is assumed to follow a Gaussian distribution $\hat{\mathscr{M}}(\boldsymbol{x}) \sim N\left(\mu_{\hat{\mathscr{M}}}(\boldsymbol{x}), \sigma^2_{\hat{\mathscr{M}}}(\boldsymbol{x})\right)$, which is defined as

$$\begin{aligned}\mu_{\hat{\mathscr{M}}}(\boldsymbol{x}) &= \mathbf{f}^T(\boldsymbol{x})\hat{\boldsymbol{\beta}} + \mathbf{r}^T(\boldsymbol{x})\mathbf{R}^{-1}\left(\boldsymbol{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right),\\ \sigma^2_{\hat{\mathscr{M}}}(\boldsymbol{x}) &= \sigma^2\left(1-\mathbf{r}^T(\boldsymbol{x})\mathbf{R}^{-1}\mathbf{r}(\boldsymbol{x}) + \mathbf{u}^T(\boldsymbol{x})\left(\mathbf{F}^T\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{u}^T(\boldsymbol{x})\right),\end{aligned} \tag{4.10}$$

where $\mathbf{u}(\boldsymbol{x}) = \mathbf{F}^T\mathbf{R}^{-1}\mathbf{r}(\boldsymbol{x}) - \mathbf{f}(\boldsymbol{x})$, and $\mathbf{r}(\boldsymbol{x}) = \left[R\left(\boldsymbol{x},\mathbf{x}_{(1)};\boldsymbol{\theta}\right),...,R\left(\boldsymbol{x},\mathbf{x}_{(N)};\boldsymbol{\theta}\right)\right]^T$ represents the correlation vector between the point $\boldsymbol{x}$ and the observed points $\boldsymbol{X} = \left\{\mathbf{x}_{(1)},\mathbf{x}_{(2)},...,\mathbf{x}_{(N)}\right\}^T$ [35], [36], [39].

The accuracy and predictive quality of the models can be evaluated by the *relative generalization error* [39], ε_{gen}, which is given by

$$\varepsilon_{gen} = \frac{\mathbb{E}[(g(\mathbf{x}) - \hat{g}(\mathbf{x}))^2]}{\mathrm{Var}_Y}. \tag{4.11}$$

4.4. NUMERICAL EXAMPLE

A level crossing design case is selected to demonstrate the application of the proposed method. This optimization problem is formulated in line with the design principle discussed in Chapter 3, which aims for a smooth stiffness transition between the connecting ballast track and ERS-based level crossing.

The design variables are listed in Table 4.2, which are collected in a design vector $\mathbf{x}$ ($\mathbf{x} = [x_s, x_n, x_{r1}, x_{r2}, x_{r3}, x_l]$). The variable x_l is specific to the ERS design, while the others

are related to the ballast track. The rail strips in ERS are elastic components underneath the rails, similar to the railpads in the ballast track. As mentioned in Chapter 3, two types of rail strips with predefined stiffness properties have been developed for ERS. The current practice in the Netherlands utilizes Type I strip. It is also worth mentioning that the variables x_s and x_l are treated as discrete values to align the optimization setting with the FE discretization. The value implies the number of 0.05m-long finite elements. The lower bound of x_s means 0.5 m, and the upper bound is 0.7 m. This range is considered reasonable and can maintain structural integrity, according to Ortega et al. [7]. x_l has a lower bound of 0 and upper bound of 6 m, implying that a 6-m level crossing is considered in the example.

Table 4.2: Definition of design variables.

Variables	Unit	Range of definition	Related measures
Sleeper spacing (x_s)	-	$x_s \in \mathbb{Z} : x_s \in [10, 14]$	$A_{max}, C_{cap}, E_{max}$
Number of strengthened sleepers (x_n)	-	$x_n \in \mathbb{Z} : x_s \in [0, 15]$	$A_{max}, C_{cap}, E_{max}$
Railpad stiffness ($x_{ri}, i = 1, 2, 3$)	MN/m	$x_{ri} \in \mathbb{R} : x_{ri} \in [50, 1000]\,, i = 1...3$	A_{max}, E_{max}
Length of Type II rail strip (x_l)	-	$x_l \in \mathbb{Z} : x_l \in [0, 120]$	A_{max}, E_{max}

The track parameters in the FE model are defined based on a typical Dutch level crossing design. The vehicle parameters specifically correspond to VIRM trains, which are double-deck trains operated by Dutch Railways. The default parameter setting for both the track and vehicle aligns with the ones defined in Chapter 3. However, it should be noted that the design variables considered in this chapter are not consistent with the variables defined in Chapter 3, as this chapter introduces additional dimensions, i.e., social relevance, to the design problem. Therefore, the corresponding FE model needs to be adapted to accommodate the current problem setting.

Three objectives (see Table 4.1) are defined in the optimization problems. Two Kriging models are developed to approximate E_{max} and A_{max}, respectively. Initially, 2000 points of $\mathbf{x}$ are generated based on LHS. A relatively large sample size is considered for two reasons. First, this sample size is further divided into training, validation, and test sets to evaluate the performance of the model, which requires a sufficient dataset. Second, according to the benchmark study in [33], increasing the number of sampling points often improves the accuracy and robustness of modeling techniques, which include Kriging.

Then, the FE model is queried at these input locations to generate the quantities of E_{max} and A_{max}. The input-output formulates a dataset, which is split randomly into training, validation, and test sets with the respective percentages of 64%, 16%, and 20%. The Kriging parameters are tuned based on the leave-one-out cross-validation approach using the training and validation data. It results in two hyperparameter sets that define the corresponding Kriging models. For each quantity, ten iterations were performed and the resulting candidate surrogate models were evaluated against the test set using the relative generalization error ε_{gen} as defined in Eq. 4.11. This evaluation indicates that the Kriging models predicting E_{max} and A_{max} have a relative generalization error of 2% and 1%, respectively.

The calculation of the objective C_{cap} depends on two variables, namely x_s and x_n. The variable x_s is varied within a 5-m section adjacent to the level crossing. This constraint in length is intentionally set to prevent potential interference with the tamping operations, which is elaborated in Chapter 3. Ortega et al. [7] analyzed the impact of sleeper spacing (x_s) in terms of construction cost savings. A summary of the cost reduction with various spacing alternatives was reported compared with the standard spacing of 0.6 m. This cost relation is considered in this chapter, with the total cost of placing sleepers at the standard spacing assumed to be €3000. The costs for other spacing alternatives are calculated based on the cost ratio provided by Ortega et al. [7]. Besides, the variable x_n refers to the number of strengthened sleepers installed in a transition zone. It is assumed that the cost ratio between the strengthened and standard type is 1.5, and the unit cost of using the standard type of sleeper in the transition zone is €400.

4.5. RESULTS AND DISCUSSION

Single-objective optimization problems are first solved. The results are presented in Table 4.3. The optimum produced from Alt. 1-3 represents the preferred track design solution for maintenance service providers, train users, and infrastructure managers, respectively. The maximization problems in Alt. 4-6 are solved to gather extremes for each objective and associate a preference function to each objective in the following multi-objective problem formulation.

Table 4.3: Design solutions and corresponding objective function values from single-objective optimization. Optimal values are highlighted for each problem.

Design alternatives	Design solutions	Objective function values		
	$\mathbf{x} = [x_s, x_n, x_{r1}, x_{r2}, x_{r3}, x_l]$	E_{max} $(N \cdot m)$	A_{max} (m/s^2)	C_{cap} (€)
Alt. 1: E_{max} minimization	$\mathbf{x} = [13, 6, 139, 179, 50, 30]$	**0.1107**	0.2300	9967.7
Alt. 2: A_{max} minimization	$\mathbf{x} = [10, 3, 74, 50, 50, 0]$	0.1814	**0.2148**	10013
Alt. 3: C_{cap} minimization	$\mathbf{x} = [\mathbf{14}, \mathbf{0}, 78, 78, 78, 0]$	0.2836	0.2529	**8562.2**
Alt. 4: E_{max} maximization	$\mathbf{x} = [14, 0, 885, 50, 1000, 41]$	**1.3880**	0.2332	8562.2
Alt. 5: A_{max} maximization	$\mathbf{x} = [14, 4, 50, 593, 792, 84]$	0.6909	**0.2570**	9362.2
Alt. 6: C_{cap} maximization	$\mathbf{x} = [\mathbf{10}, \mathbf{15}, 78, 78, 78, 0]$	0.2160	0.2136	**12413**

The optimization of C_{cap} depends on variables x_s and x_n, as emphasized in the Design Solutions for Alt. 3 & 6. All other variables remain at their default values. The objective values in Alt. 3 & 6 indicate that there is a conflict between C_{cap} and the other objectives. Specifically, a design that incorporates larger sleeper spacing and avoids the use of strengthened sleepers is preferable from a cost perspective. However, such a design does not contribute to reducing the expected damage in the ballast (E_{max}) and maintaining the level of train service (A_{max}).

When comparing solutions from Alt. 1-2 and Alt. 4-5, the first observation is that

softer railpads ($x_{ri}, i = 1,2,3$) are generally preferred to reduce dynamic impact in the vehicle-track system. Particularly in Alt. 1, the optimal value of x_{r3} is significantly lower than those of x_{r1} and x_{r2}. Here, x_{r3} denotes the stiffness of the railpad adjacent to the crossing, the optimal value of which is linked to x_l. A value of $x_l = 30$ suggests a 1.5 m-long Type II (softer) strip in conjunction with a soft railpad (x_{r3}) in the ballast track. This combination allows for a more uniform distribution of track support stiffness, thereby reducing the load effect transmitted to the ballast layer (i.e., reduced E_{max}). This rationale also accounts for the optimal values of the corresponding variables in Alt. 4, which yields a contradictory outcome. In contrast to Alt. 1, Alt. 2 recommends against the use of the softer strip in ERS to minimize A_{max} and improve ride comfort. The reason could be that E_{max} focuses on the dynamics in the track underlayers, while A_{max} pertains to the upper vehicle dynamics.

In Alt. 3, minimizing the cost (C_{cap}) leads to a significant deviation of E_{max} and A_{max} from their minimum values. In contrast, Alt. 1 allows for the minimization of E_{max} without a considerable impact on A_{max}. This observation can be attributed to the interaction between the vehicle and track structure, where E_{max} and A_{max} are correlated. It is also worth noting that the objectives under consideration are influenced by different design variables: C_{cap} depends on x_s and x_n, while the other objectives are influenced by the entire set of design variables. As a result, while these objectives compete, optimizing one does not necessarily lead to the extreme values of the others.

In the multi-objective formulation, a linear preference function is assigned for each objective due to its simplicity and interpretability. To construct a reasonable range for associating a preference function with an objective, the maximum and minimum of each objective are considered. For instance, as depicted in Figure 4.2 (a), for maintenance contractors, a preference score of 0.1 for E_{max} corresponds to 100, indicating the desired level, whereas a preference score of 1.4 for E_{max} corresponds to 0, representing the worst scenario that should be avoided from the contractor's perspective. The same applies to other objectives as presented in Figure 4.2 (b)-(c).

The preference curves in Figure 4.2 (a)-(c) also display the optimal objective values obtained using the minmax method, with equal weights assigned. The preference scores are nearly equal, around 82, with $E_{max} = 0.3398$, $A_{max} = 0.2192$, and $C_{cap} = 9207$. This illustrates the rationale behind the min-max paradigm: the method seeks to find a balanced solution that satisfies all stakeholders.

The design solutions and their corresponding objective function values from single- and multi-objective optimization are summarized in Table 4.4. The design solutions indicate that the soft railpads and strips ($x_{ri}, x_l; i = 1,2,3$) are recommended at the junction between the level crossing and transition zone, according to both the single-objective and multi-objective problems. However, compared to the single-objective problems, the multi-objective formulation integrating stakeholder preferences affects the solution for sleeper parameters (x_s and x_n), as highlighted in the Design Solutions section. This is because these variables directly influence the objective C_{cap}, which is in conflict with the other objectives.

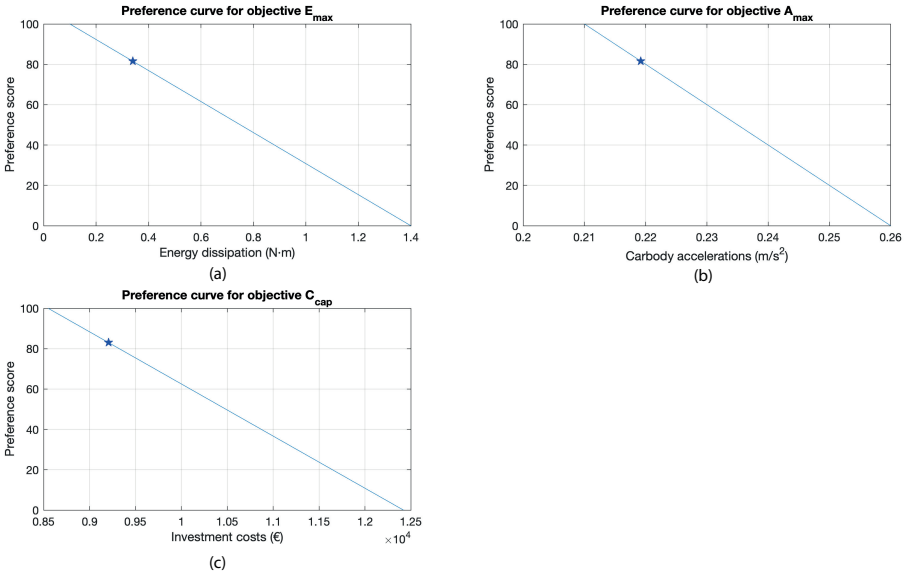

Figure 4.2: Design solutions from multi-objective design optimization: preference curve for objective (a) E_{max}, (b) A_{max}, and (c) C_{cap} including the optimum obtained from the preference-based min-max method.

Table 4.4: Design solutions and objective function values for different optimization problems. Optimal objective values are highlighted.

Optimization method	Design solutions	Objective function values		
	$\mathbf{x} = [x_s, x_n, x_{r1}, x_{r2}, x_{r3}, x_l]$	E_{max} $(N \cdot m)$	A_{max} (m/s^2)	C_{cap} (€)
E_{max} minimization	$\mathbf{x} = [13, 6, 139, 179, 50, 30]$	**0.1107**	0.2300	9967.7
A_{max} minimization	$\mathbf{x} = [10, 3, 74, 50, 50, 0]$	0.1814	**0.2148**	10013
C_{cap} minimization	$\mathbf{x} = [14, 0, 78, 78, 78, 0]$	0.2836	0.2529	**8562.2**
MOO (E_{max}, A_{max}, C_{cap})	$\mathbf{x} = [\mathbf{11,0}, 89, 50, 50, 53]$	0.3398	0.2192	9207

4.6. CONCLUSIONS

Effective service life management of railway assets requires aligning the preferences and requirements of various stakeholders into the engineering design process. For this purpose, this chapter presents a design approach that integrates engineering and managerial aspects to model design problems for railway tracks. This approach enables a more comprehensive evaluation of alternatives for railway track design, considering both the technical and social relevance.

Three representative stakeholder groups and their corresponding interests are defined, which are relevant to rail asset feasibility (technics), affordability, and serviceabil-

ity. These perspectives are translated to performance measures, which are then used to formulate design optimization problems. The solution method integrates FE modeling, Kriging, and preference functions to find an optimal design configuration that balances stakeholder preferences and actual track performance within a reasonable amount of time. To demonstrate the application of the design approach, a case study on level crossing design is provided. The design solutions obtained demonstrate relevance with respect to stakeholder preferences and expected long-term track performance. It shows that the current design approach allows the track design to be managed focusing on best fit for common purpose rather than on mechanical behavior only.

REFERENCES

[1] Y. Shang, R. Binnekamp, and A. Wolfert, "Multi-stakeholder service life design for rail level crossings", in *Life-Cycle of Structures and Infrastructure Systems*, CRC Press, 2023, pp. 949–956.

[2] V. Profillidis, *Railway planning, management, and engineering*. Taylor & Francis, 2022.

[3] W. Chen, F. Ahmed, Y. Cui, Z. Sha, and N. Contractor, "Data-driven preference modelling in engineering systems design", in *Handbook of Engineering Systems Design*, Springer, 2022, pp. 1–34.

[4] M. J. Steenbergen, "Physics of railroad degradation: The role of a varying dynamic stiffness and transition radiation processes", *Computers & Structures*, vol. 124, pp. 102–111, 2013.

[5] M. Sol-Sánchez, F. Moreno-Navarro, and M. C. Rubio-Gámez, "The use of elastic elements in railway tracks: A state of the art review", *Construction and building materials*, vol. 75, pp. 293–305, 2015.

[6] R. Sañudo, M. Cerrada, B. Alonso, and L. dell'Olio, "Analysis of the influence of support positions in transition zones. A numerical analysis", *Construction and Building Materials*, vol. 145, pp. 207–217, 2017.

[7] R. S. Ortega, J. Pombo, S. Ricci, and M. Miranda, "The importance of sleepers spacing in railways", *Construction and Building Materials*, vol. 300, p. 124 326, 2021.

[8] A. R. M. Wolfert, *Open Design Systems*. IOS Press, 2023.

[9] Y. Peng, J. Zhou, C. Fan, *et al.*, "A review of passenger ride comfort in railway: Assessment and improvement method", *Transportation Safety and Environment*, vol. 4, no. 2, tdac016, 2022.

[10] F. Ripamonti and A. Chiarabaglio, "A smart solution for improving ride comfort in high-speed railway vehicles", *Journal of Vibration and Control*, vol. 25, no. 13, pp. 1958–1973, 2019.

[11] N. Wu, J. Zeng, and Y. Wang, "Effect of wheel-rail wear and suspension system failure on vehicle dynamic performance", *Journal of Vibration and Shock*, vol. 34, no. 5, pp. 82–87, 2015.

[12] E. Sperling and C. Betzhold, "Beitrag zur beurteilung des fahrkomforts in schienenfahrzeugen (contribution to the assessment of ride comfort in rail vehicles)", *Glaser's Annals*, vol. 80, pp. 314–320, 1956.

[13] B. Buurman, K. Gkiotsalitis, and E. Van Berkum, "Railway maintenance reservation scheduling considering detouring delays and maintenance demand", *Journal of Rail Transport Planning & Management*, vol. 25, p. 100 359, 2023.

[14] M. Sadri, T. Lu, and M. Steenbergen, "Railway track degradation: The contribution of a spatially variant support stiffness-local variation", *Journal of Sound and Vibration*, vol. 455, pp. 203–220, 2019.

[15] G. Chiandussi, M. Codegone, S. Ferrero, and F. E. Varesio, "Comparison of multi-objective optimization methodologies for engineering applications", *Computers & Mathematics with Applications*, vol. 63, no. 5, pp. 912–942, 2012.

[16] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering", *Structural and multidisciplinary optimization*, vol. 26, pp. 369–395, 2004.

[17] R. H. Stewart, T. S. Palmer, and B. DuPont, "A survey of multi-objective optimization methods and their applications for nuclear scientists and engineers", *Progress in Nuclear Energy*, vol. 138, p. 103 830, 2021.

[18] K.-H. Chang, *Design theory and methods using CAD/CAE: The computer aided engineering design series*. Academic Press, 2014.

[19] S. Greco, J. Figueira, and M. Ehrgott, *Multiple criteria decision analysis*. Springer, 2016, vol. 37.

[20] S. Bechikh, M. Kessentini, L. B. Said, and K. Ghédira, "Preference incorporation in evolutionary multiobjective optimization: A survey of the state-of-the-art", in *Advances in Computers*, vol. 98, Elsevier, 2015, pp. 141–207.

[21] R. Battiti and A. Passerini, "Brain–computer evolutionary multiobjective optimization: A genetic algorithm adapting to the decision maker", *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 671–687, 2010.

[22] J. W. Fowler, E. S. Gel, M. M. Köksalan, P. Korhonen, J. L. Marquis, and J. Wallenius, "Interactive evolutionary multi-objective optimization for quasi-concave preference functions", *European Journal of Operational Research*, vol. 206, no. 2, pp. 417–425, 2010.

[23] L. Rachmawati and D. Srinivasan, "Incorporating the notion of relative importance of objectives in evolutionary multiobjective optimization", *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 4, pp. 530–546, 2010.

[24] P. N. Kodikara, B. Perera, and M. Kularathna, "Stakeholder preference elicitation and modelling in multi-criteria decision analysis–A case study on urban water supply", *European Journal of Operational Research*, vol. 206, no. 1, pp. 209–220, 2010.

[25] T. Wagner and H. Trautmann, "Integration of preferences in hypervolume-based multiobjective evolutionary algorithms by means of desirability functions", *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 5, pp. 688–701, 2010.

[26] N. R. Costa, J. Lourenço, and Z. L. Pereira, "Desirability function approach: A review and performance evaluation in adverse conditions", *Chemometrics and Intelligent Laboratory Systems*, vol. 107, no. 2, pp. 234–244, 2011.

[27] J. Sanchis, M. A. Martínez, X. Blasco, and G. Reynoso-Meza, "Modelling preferences in multi-objective engineering design", *Engineering Applications of Artificial Intelligence*, vol. 23, no. 8, pp. 1255–1264, 2010.

[28] A. Messac, C. Sukam, and E. Melachrinoudis, "Mathematical and pragmatic perspectives of physical programming", *AIAA journal*, vol. 39, no. 5, pp. 885–893, 2001.

[29] G. Reynoso-Meza, J. Sanchis, X. Blasco, and S. Garcia-Nieto, "Physical programming for preference driven evolutionary multi-objective optimization", *Applied Soft Computing*, vol. 24, pp. 341–362, 2014.

[30] X. Lai, M. Xie, and K.-C. Tan, "QFD optimization using linear physical programming", *Engineering optimization*, vol. 38, no. 5, pp. 593–607, 2006.

[31] A. Messac, M. P. Martinez, and T. W. Simpson, "Effective product family design using physical programming", *Engineering Optimization*, vol. 34, no. 3, pp. 245–261, 2002.

[32] J. Sanchis, M. Martinez, and X. Blasco, "Multi-objective engineering design using preferences", *Engineering Optimization*, vol. 40, no. 3, pp. 253–269, 2008.

[33] M. R. Kianifar and F. Campean, "Performance evaluation of metamodelling methods for engineering problems: Towards a practitioner guide", *Structural and Multidisciplinary Optimization*, vol. 61, no. 1, pp. 159–186, 2020.

[34] R. Schobi, B. Sudret, and J. Wiart, "Polynomial-chaos-based Kriging", *International Journal for Uncertainty Quantification*, vol. 5, no. 2, 2015.

[35] M. Moustapha, B. Sudret, J.-M. Bourinet, and B. Guillaume, "Quantile-based optimization under uncertainties using adaptive Kriging surrogate models", *Structural and multidisciplinary optimization*, vol. 54, pp. 1403–1421, 2016.

[36] H. M. Kroetz, M. Moustapha, A. T. Beck, and B. Sudret, "A two-level kriging-based approach with active learning for solving time-variant risk optimization problems", *Reliability Engineering & System Safety*, vol. 203, p. 107 033, 2020.

[37] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2.

[38] F. Bachoc, "Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification", *Computational Statistics & Data Analysis*, vol. 66, pp. 55–69, 2013.

[39] C. Lataniotis, D. Wicaksono, S. Marelli, and B. Sudret, "UQLab user manual – Kriging (Gaussian process modeling)", Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, Tech. Rep., 2021, Report # UQLab-V1.4-105.

5

EXTREME-ORIENTED SENSITIVITY METHOD FOR RELIABILITY-BASED DESIGN

This chapter introduces a novel sensitivity method that focuses on the part of the output space that yields failure, that is, extreme values (either maximum or minimum) and values near design thresholds. These quantities are highly relevant to reliability-based problems such as reliability-based design optimization. The present method allows engineers to perform sensitivity analysis near system limit states and visualize the relevance of input factors to different design criteria (or objectives) and corresponding thresholds. The polynomial chaos expansion is used to approximate the model output and alleviate the computational cost in sensitivity analysis, which features sparsity and adaptivity to enhance efficiency.

To validate the accuracy and efficiency of the method, an analytical example and a truss structure are first examined. Then, the method is applied to a dynamic train-track-bridge system. Results show that the relevance of input factors varies significantly across different design criteria. This highlights the importance of exploring the design space near limit states before formulating reliability-based design optimization problems for high-dimensional models. By providing valuable insights into this particular space, the method assists engineers in identifying the most influential design variables (that align with the intended objectives), thereby guiding the formulation of reliability-based optimization problems.

Parts of this chapter have been published verbatim in Reliability Engineering & System Safety, 243, 109818 (2024) [1].

5.1. INTRODUCTION

Sensitivity analysis refers to the investigation of how changes in input factors affect the model output. It is primarily used to i) identify factors whose accuracy is critical to the quality of the output (*factor prioritization* [2]), ii) fix unimportant factors to reduce the model complexity (*factor fixing* [2]), and iii) identify regions of the input space that lead to undesirable outcomes, such as system failure. This is particularly important for engineering structures that demand a high level of reliability. Based on the factor space of interest, sensitivity methods can be divided into local and global approaches [3], [4]. The former measures the impact of variations in input factors near their nominal values, whereas the latter examines the sensitivity of the model output to input factors over their entire domain of variation.

Various global sensitivity methods have been proposed, including variance-based methods (e.g., the Sobol method [5] and the Fourier amplitude sensitivity (FAST) method [6]), regression-based methods (e.g., standardized regression coefficient [2]), moment-independent methods (e.g., Borgonovo indices [7]), and many others. Of interest herein is the variance-based approach, which leverages the functional decomposition of the variance measuring the contribution of each factor or their combinations to the output variance. Sobol's indices [5] and related total-effect indices [8] are commonly used variance-based measures. These measures can be computed through sampling-based methods [9]–[11] or surrogate models [12]–[15]. The sampling-based methods can come at a significant computational cost due to the large number of function evaluations required, which may render the methods infeasible for computationally demanding models.

Surrogate models have received much attention for the computation of variance-based measures due to their potential to alleviate computational costs. Specifically, PCE and its variants have been used to derive Sobol's indices through post-processing of model coefficients. This approach is originally shown in [12], using the coefficients of generalized PCE [16], which is extended using sparse PCE [13], [17]–[19] and partial least squares-driven PCE (PLS-PCE) [15] to deal with high-dimensional problems. Furthermore, advances have been made to account for the dependence in input factors [20], generalized modeling of both aleatory and epistemic uncertainty in input factors [21], and derivative-based sensitivity measures for efficient screening of unimportant factors [22]. These studies have demonstrated PCE to be a versatile and efficient tool for sensitivity analysis (SA).

Variance-based sensitivity measures seek to evaluate the influence of input factors on the variance of a quantity of interest (QoI). Most often, the QoI refers to the average value of the model output, as used in the Sobol indices and their extensions (e.g., [20], [21]). To clarify, let $g(\cdot)$ be a generic black box representation of a model, which has a set of input factors $\boldsymbol{X} = X_1, X_2, ..., X_d \in \mathbb{R}^d$ and a scalar output Y such that $Y = g(\boldsymbol{X})$. The first-order Sobol index with respect to X_i is obtained by

$$I_i^{Sobol} = \frac{\mathrm{Var}_{X_i}\left(\mathrm{E}_{X_{\sim i}}(Y|X_i)\right)}{\mathrm{Var}(Y)}, \tag{5.1}$$

where $\mathrm{Var}(Y)$ is the unconditional variance of Y, and $\mathrm{E}_{X_{\sim i}}(Y|X_i)$ is the mean of Y when X_i is fixed. By construction, the Sobol-related indices aim to capture the mean behavior

of the systems, which may not be sufficient for sensitivity evaluation of safety-critical structures such as railway tracks.

To address this issue, this chapter focuses on the extreme values of the structural response. *Extreme values* are the maximum and minimum levels the model output can achieve. *Response* refers to the output of a model that characterizes the structural behavior, which can be either static or dynamic. The condition of a structure can be described by a group of limit states, each associated with design criteria (or indicators) and thresholds. *Failure* occurs if the response quantity exceeds a given threshold.

This concept is closely linked to reliability analysis, making reliability-oriented sensitivity analysis (ROSA) the natural way to perform SA for safety-critical structures [23]. ROSA is concerned with evaluating the sensitivity of the output of reliability analysis to the model inputs, where the part of the output space that yields failure is of interest. Approaches dedicated to reliability-oriented sensitivities can also be categorized according to the scope, i.e., local [24], [25] versus global sensitivity measures [26]–[28]. These studies aim to identify critical input factors with more potential to reduce the probability of failure, instead of quantifying the influence of input factors on output extreme values. In parallel, some work directs sensitivity evaluation to quantile-based [29] and high-order measures (e.g., skewness and kurtosis decomposition [30], [31]), which may provide an indication about input factors that drive the output towards its extreme values. However, they do not directly address SA near output extreme values.

Limited research has focused on sensitivity methods for extreme values. Wong et al. [32] compute extremum sensitivities by constraining the input distribution to regions leading to the output extremum. They use Monte Carlo filtering to model the conditional distribution, which is a sampling-based approach that partitions model realizations within or outside targeted regions in the input space. Another approach is optimization-based [3]. It involves discretizing an input factor within its range and searching for the output extremum at each discrete point by fixing the target factor while perturbing the rest (the so-called *A-1AT - all minus one at a time*), which poses an optimization problem at each point. Then, a curve can be defined by the extreme values computed from all discrete points, and the variation of this curve is evaluated to quantify the main effect of the target factor on the output extreme values.

This chapter presents a new sensitivity method called the *threshold-based sensitivity method*. This method allows for evaluating the model sensitivity near system limit states by generalizing the optimization-based sensitivity method [3] to a wider range of targeted portions of the output space, specifically from extreme values to thresholds. The threshold-based method offers flexibility by allowing the evaluation of multiple decision thresholds in SA. It also enables visualizing the structural performance (failure or non-failure) around limit states in the input space, highlighting critical regions of the input space that may lead to system failure under a given threshold. In addition, the performance of PCE is investigated in the context of extreme-oriented SA. While the use of PCE for Sobol's indices has been relatively well-studied, research on its effectiveness in extreme-based cases remains scarce. Therefore, the performance of PCE is compared with other commonly used surrogate models, such as Kriging and polynomial chaos-Kriging (PCK), to provide insights into its suitability for these types of analyses.

The applicability of the present method is examined in vehicle-track systems, which

are safety-critical and of high relevance to reliability-based design (cf. [33]–[35]). Sensitivity evaluation centered on the extremes can provide valuable insights for decision-making in the design and management of such systems. In most cases, the sensitivity of vehicle/structure responses to input factors is assessed using the sampling method, known as One At a Time (OAT) analysis. OAT implies SA is approached by perturbing one single factor at each time, while the rest are fixed in a given value. This method is computationally efficient and well-suited for SA of vehicle-track interaction problems since the models involved can easily become computationally intensive. However, it may not provide reliable results when nonlinear terms are present in the model [4], [36], [37], which in the vehicle-track models can be connected to the nonlinearity in wheel-rail contact [38], [39], pad [40], [41], and ballast behavior [41], [42].

A few studies have applied global sensitivity methods to evaluate vehicle-track systems. Xu et al. [43] evaluate a nonlinear vehicle-track model to identify the factors dominating the system dynamics, with special attention on track irregularities. Later, Wan et al. [36] performed a dynamic global SA in a time-varying train-track-bridge system, comparing a Kriging-based approach against a Monte Carlo simulation scheme. Recently, sensitivity and uncertainty analyses were performed in tandem [44], [45] to analyze the uncertainty propagation in a train-track-bridge system and identify factors that are most responsible for the response uncertainty. However, these studies only directed sensitivity analyses to the average system response using the Sobol method, while overlooking the part of the output space that yields failure.

This chapter is organized as follows. Section 5.2 presents the PCE-based sensitivity method for extreme problems, including the A-1AT sensitivity method with two formulations (Section 5.2.1), the basics of PCE (Section 5.2.2), and the extension to the threshold-based sensitivity method (Section 5.2.3). The methodology is verified in an analytical function and a truss structure (Section 5.3) and further applied to a dynamic train-track-bridge system to demonstrate its applicability (Section 5.4). Section 5.5 discusses the sensitivity results along with evaluating the predictive performance of PCE and the impact of design thresholds on the sensitivity results. Final remarks and future research lines are drawn in Section 5.6.

5.2. METHODOLOGY

5.2.1. EXTREME-ORIENTED SENSITIVITY METHOD

In this section, the A-1AT sensitivity method with two strategies to formulate the extreme problems are presented: the original problem formulation in [3] is first provided for readability, followed by its extension to threshold-based SA.

A-1AT UPON THE OUTPUT EXTREMA

Let us consider a system whose behavior is described by $g(\cdot)$, where a set of input factors $\boldsymbol{X} = \{X_1, X_2, ..., X_d\} \in \mathbb{R}^d$ yields a scalar output Y such that

$$Y = g(\boldsymbol{X}). \tag{5.2}$$

Each input factor is defined within an interval $X_i \in [X_i^l, X_i^u]$. First, when a factor X_i is fixed at a specific point x_i^0, the model defined in Eq. (5.2) can reach an extreme

value (either maximum or minimum) by varying the remaining factors within their corresponding intervals. This extreme value represents an output extremum of the reduced (i.e., $d-1$) dimensional space, which can be denoted as $\bar{Y}_i$. Then, when considering the entire range of the factor X_i, one can determine a curve formed by the output extrema of the $d-1$ dimensional space. The sensitivity of this curve at point x_i^0 can be defined in terms of finite differences, i.e.,

$$S_i^{Ext}\left(x_i^0; X_{\sim i}\right) \simeq \frac{\Delta \bar{Y}_i\left(x_i^0, X_{\sim i}\right)}{\Delta x_i}, \qquad x_i^0 \in \left[X_i^l, X_i^u\right], \forall X_i. \tag{5.3}$$

The procedure of evaluating the sensitivity of $\bar{Y}_i$ to X_i (Eq. 5.3) is provided in the following steps.

Step a. Discretize the factor X_i in n_i points within its range $\left[X_i^l, X_i^u\right]$; denote each discrete point of X_i by $x_{i,j}, j = 1, ..., n_i$.

Step b. For a discrete point $x_{i,j}$, the extreme value of Eq. (5.2) is obtained by fixing X_i at $x_{i,j}$ while varying the remaining factors $X_{\sim i}$ in their corresponding intervals. This poses an optimization problem for each discrete value of X_i, which is defined by

$$\begin{aligned} \min_{\mathbf{X}\in[\mathbf{X^l},\mathbf{X^u}]} \quad & g(\mathbf{X}) \\ \text{s.t.} \quad & X_i = \{x_{i,j}\}, \quad j = 1, ..., n_i, \\ & \mathbf{X}_{\sim i} \in \left[\mathbf{X}_{\sim i}^l, \mathbf{X}_{\sim i}^u\right]. \end{aligned} \tag{5.4}$$

Step c. Save the optimal (i.e., extreme) value at each discrete value of X_i.

Step d. Once the optimal value is determined for all discrete values of X_i, the curve formed by the output extrema of the $d-1$ dimensional space, $\bar{Y}_i$, can be determined and the sensitivity of the curve is calculated according to Eq. (5.3).

Specifically, in Step *a*, different types of discretization can be selected for the input factors, such as equal-, log-, or randomly spaced discretization. The degrees of discretization, meaning the number of discrete points, can also vary for each input factor. Since the optimal search is defined for a specific factor (X_i) at a time, the search process is independent of the discretization strategy for the remaining non-fixed factors ($\mathbf{X}_{\sim i}$). The reader is referred to [3] for further details about the discretization process.

The optimization problem in Eq. (5.4) can be solved using either gradient-based methods or (meta)heuristics. As the method is optimization-based and the interest here is either the maximum or minimum, the term *optimal value* also refers to *extreme value* in Steps *c* and *d*. It is also worth noting that the minimum value is considered by default to conform to the standard form of defining an optimization problem, as shown in Eq. (5.4). A maximization problem can be treated by negating the objective function $g(\mathbf{X})$. This is frequently encountered in structural engineering problems where the maximum value holds significance for structural reliability.

Further, the following importance measure [3] is defined to rank the input factors according to their contribution to the total variance introduced by the individual factors

to the output extreme values,

$$I_i^{Ext} = \frac{\mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)}{\sum_i \mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)}, \tag{5.5}$$

where $\mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)$ represents the variance of the output extreme values associated with factor X_i. $\sum_i \mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)$ represents the total variance introduced by the individual factors to the output extreme values.

As mentioned earlier, the input factors can be discretized either uniformly or non-uniformly. If non-uniform discretization is used, the variance in Eq. (5.5) must be replaced by the weighted variance, i.e.,

$$\mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right) = \sum_{j=1}^{n_i} \frac{w_j}{\sum_{j=1}^{n_i} w_j}\left(\bar{Y}_{i,j} - \bar{\mu}\right)^2, \quad \bar{\mu} = \sum_{j=1}^{n_i} \frac{w_j}{\sum_{j=1}^{n_i} w_j}\bar{Y}_{i,j}, \tag{5.6}$$

where $\bar{Y}_{i,j}$ is the output extremum at the discrete point $x_{i,j}$, and the weight w_j is given by

$$w_j = \frac{1}{2}\left(x_{i,j+1} - x_{i,j-1}\right), \quad \forall\, j \in [2, n_i - 1], \tag{5.7}$$

with $w_1 = \frac{1}{2}\left(x_{i,2} - x_{i,1}\right)$ and $w_{n_i} = \frac{1}{2}\left(x_{i,n_i} - x_{i,n_i-1}\right)$ for the first and last points of the discretized factor.

5

A-1AT UPON THE DEVIATION BETWEEN THE OUTPUT AND GIVEN THRESHOLD

A limit state refers to a state of impending failure, beyond which a structure can no longer perform its intended function satisfactorily, in terms of either safety or serviceability. The basic idea of the limit-state design approach is to identify all possible modes of failure (i.e., limit states) and determine acceptable levels of safety against the occurrence of each limit state [46]. From this, one limit state function can be evaluated by

$$G_k(\mathbf{X}) = g(\mathbf{X}) - t_k \tag{5.8}$$

where $G_k(\mathbf{X})$ denotes the limit state function of a structure; $g(\mathbf{X})$ describes the actual performance of the structure measured by a design criterion, and $t_k (k = 1, ..., \kappa)$ represents the k^{th} allowable level (i.e., the threshold) of the corresponding criterion, e.g., maximum allowable displacement of a beam. κ denotes the number of decision thresholds considered for a single criterion.

The sign of $G_k(\mathbf{X})$ determines the structural reliability state, which is defined as

- failing state if $G_k(\mathbf{X}) > 0$;
- reliable state if $G_k(\mathbf{X}) < 0$;
- limit state if $G_k(\mathbf{X}) = 0$.

The structure needs to be checked for all groups of limit states to ensure sufficient margins between the actual structural behavior and given threshold. However, the determination of the structural state can be significantly affected by the attitudes of decision-makers towards the threshold levels. For example, ride comfort is relevant to structural

serviceability, which is associated with a group of evaluation indices and rating scales (i.e., thresholds) [47]. The demand for traveling quality is however very subjective, and the decision on which rating scale to be included in the design process can influence the outcomes significantly, which may also affect the SA result.

To address this issue, one can frame the preferences of the decision-makers into multiple decision thresholds. This allows for evaluating the sensitivity of the model output to each input factor at different levels. To achieve this, the extreme problem is reformulated by modifying Step *b* in this section as follows.

$$\begin{aligned} \min_{\mathbf{X}\in[\mathbf{X^l},\mathbf{X^u}]} \Delta(\mathbf{X}) \quad &= \min_{\mathbf{X}\in[\mathbf{X^l},\mathbf{X^u}]} \left(g(\mathbf{X}) - t_k\right)^2 \\ \text{s.t.} \quad X_i &= \{x_{i,j}\}, \quad j = 1,...,n_i, \\ \mathbf{X}_{\sim i} &\in \left[\mathbf{X}^l_{\sim i}, \mathbf{X}^u_{\sim i}\right], \end{aligned} \tag{5.9}$$

where $\Delta(\mathbf{X})$ represents the squared deviation of the model output from a given threshold, that is, $\Delta(\mathbf{X}) = [G_k(\mathbf{X})]^2$.

Recall that for a single design criterion, the output space is defined by $Y = g(\mathbf{X})$, the present sensitivity method essentially requires two basic steps: (1) identify the surface of interest in Y, and (2) analyze the sensitivity of that surface. In the first step, the surface can be characterized either by the output extrema through optimization (Eq. 5.4) or by being related to a given threshold (Eq. 5.9). If the surface related to a threshold is of interest, the idea is to search for the model output that approaches the reference threshold, and the goal is to identify which input factors are critical in reaching that threshold, i.e., the limit state of the structure. This process, which allows for the analysis of the sensitivity of the surface around the given threshold, is referred to as *threshold-based sensitivity*.

An illustrative example of threshold-based sensitivity is shown in Figure 5.1. The figure depicts how changes in a single input factor $X \in \left[X^l, X^u\right]$ affect the *threshold-based response* (calculated from Eq. 5.9) near three different design thresholds ($t_k, k = 1,2,3$). Specifically, t_1 is easy to reach, t_2 is a middle level, and t_3 is extremely hard to achieve. These thresholds correspond to Scenario 1, 2, and 3, respectively. The vertical axis ($\tilde{Y}^k$) represents the threshold-based response, which is denoted by a solid line for each threshold level (the dotted line with the same color). The value of $\tilde{Y}^k$ varies depending on whether the maximum output exceeds the corresponding threshold. In other words, it poses a binary classification problem, where (1) for solid lines that fall below the corresponding threshold, the value of $\tilde{Y}^k$ represents the actual maximum output of a system, indicating a reliable state; (2) When a solid line remains stable at a given threshold, it suggests that the maximum output has either reached or exceeded the threshold, indicating the limit or failing state of the system. In such cases, the value of $\tilde{Y}^k$ is determined by the threshold value.

Note that the term *response* is used in the context of threshold-based SA, specifically referring to the threshold-based response, instead of the *output* used in the general context of SA. This distinction is made because the threshold-based response is not exactly equivalent to the model output. Its value depends on whether the system reaches the failure threshold. Therefore, in this chapter, the term *response* is used in two specific contexts: (1) in structural engineering problems where it refers to the reaction or be-

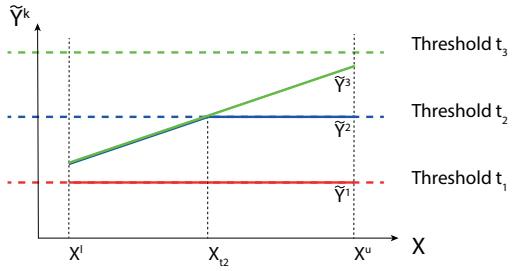

Figure 5.1: Illustrative example of threshold-based sensitivity. The figure shows three scenarios: Scenario 1 (red lines) with threshold t_1 (dotted) and threshold-based response $\tilde{Y}^1$ (solid); Scenario 2 (blue lines) with threshold t_2 (dotted) and threshold-based response $\tilde{Y}^2$ (solid); and Scenario 3 (green lines) with threshold t_3 (dotted) and threshold-based response $\tilde{Y}^3$ (solid). In each scenario, the region above or below the corresponding threshold indicates the failure or non-failure domain, respectively. $\tilde{Y}^2$ is equal to $\tilde{Y}^3$ when $X \in [X^l, X_{t2})$, as they both represent the maximum output within this range, which is independent of the threshold values. Note: the figure is only illustrative. See Section 5.5.3 for a detailed analysis of a large-scale numerical example.

havior of a structure to the input it receives, either static or dynamic, as mentioned in Section 5.1, and (2) in threshold-based SA where the term threshold-based response is introduced to visualize the impact of input factors on the model output near the thresholds, which is elaborated in the following.

As shown in Figure 5.1, in Scenario 1, the value of $\tilde{Y}^1$ (the red solid line) is constant at the level of t_1, indicating that the current threshold t_1 can always be reached for the factor X over its range of definition, $[X^l, X^u]$. Therefore, X is not a critical factor that would impede reaching the threshold at t_1. In Scenario 2, the response surface lies below t_2 when $X \in [X^l, X_{t2})$. This suggests that ensuring $X \in [X^l, X_{t2})$ can create a safe margin between the maximum output (the blue solid line) and the threshold (the blue dotted line). However, when $X \in [X_{t2}, X^u]$, the limit state is (reached or) exceeded, resulting in (impending) system failure. Therefore, X is critical in reaching the threshold t_2 over its entire range of definition $[X^l, X^u]$. Herein, one can see that the threshold-based SA allows engineers to visualize how the input factors affect the model output near the limit states, indicating areas of the input space that may lead to undesirable outcomes, such as system failure. This knowledge can be useful in decision-making and risk management for the design and maintenance of engineering structures.

If the threshold is sufficiently large at t_3, the value of $\tilde{Y}^3$ is determined by the maximum output of the system (the green solid line). Varying $X \in [X^l, X^u]$ will not alter the system reliability state, as there will be sufficient margins between the maximum and t_3. Note that in this case, the problem is equivalent to maximizing $g(\mathbf{X})$, and vice versa. This corresponds exactly to the original formulation in Eq. (5.4). The threshold-based sensitivity calculated through Eq. (5.9) essentially represents an alternative way of describing the extreme problems, while it extends the original setting by taking the limit state concept into account and is flexible to allow for assessing multiple decision-making scenarios, as measured by multi-thresholds, in SA.

An overview of the extreme-based and threshold-based formulation is provided in Figure 5.2 to compare the calculation procedures and application contexts. To summarize, both formulations require discretizing the input factors as the first step, and the

discretization strategy remains consistent. However, they differ in terms of the target portion of the output space (Step 2), formulation of the optimization problem (Step 3), and the way the calculation results are post-processed (Step 4). The extreme-based formulation is mainly used to quantify and visualize the main effects of the input factors on the extreme values of the output. On the other hand, threshold-based sensitivity focuses on visualizing the impact of altering an individual input factor over a specified threshold by introducing the threshold-based response. This formulation clearly indicates the system reliability state around limit states in the input space, highlighting critical regions that may lead to system failure under a given threshold. This knowledge can be leveraged to support analyses centered around system reliability, such as RBDO. Also, the evaluation results obtained from both formulations complement each other, providing valuable insight into the contribution of the input factors across different regions of the output space.

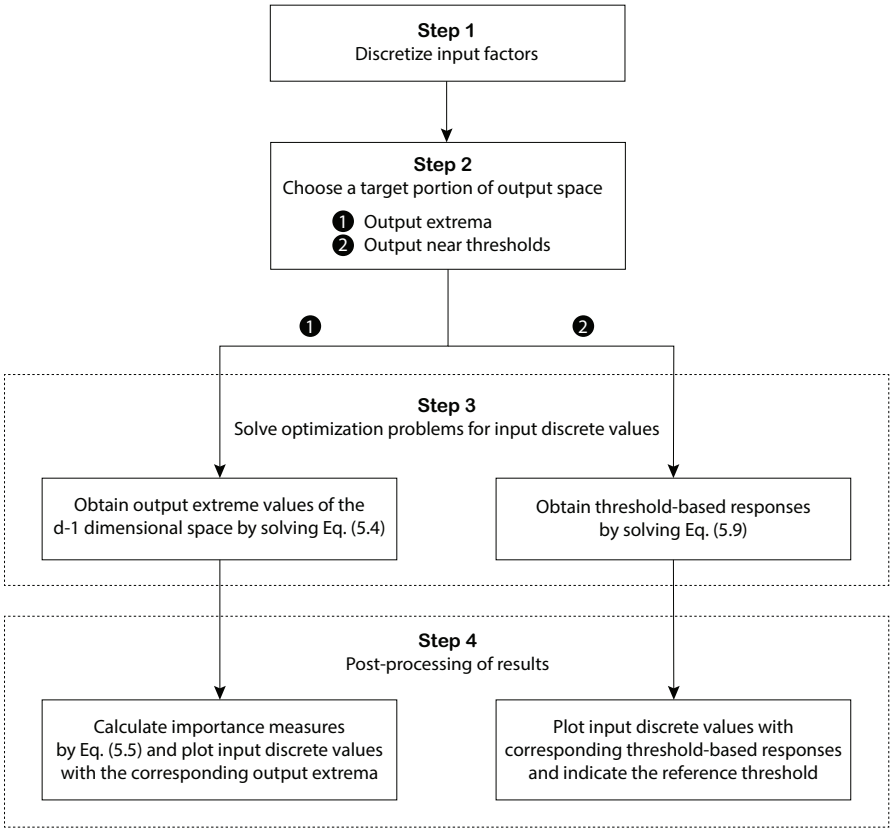

Figure 5.2: Overview of the extreme-based (Path 1) and threshold-based (Path 2) formulation.

5.2.2. POLYNOMIAL CHAOS EXPANSIONS (PCE)

Let us recall the mapping $g(\cdot)$ in Eq. (5.2) to describe the behavior of a system. Assume that there is uncertainty associated with the input vector $\mathbf{X}$, which can be described by a random vector with joint probability distribution function (PDF) $\boldsymbol{f}_{\mathbf{X}}(\mathbf{x})$ and marginal PDFs $f_{X_i}(x_i), i = 1, ..., d$. The components of $\mathbf{X}$ are considered independent, which is the case for the models in the present study. Consequently, the model output Y is also a random variable that can be approximated using PCE [17], [48] such that

$$g(\mathbf{X}) \approx \hat{g}(\mathbf{X}) = \sum_{\boldsymbol{\alpha} \in \mathscr{A}} \mathbf{c}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{X}), \tag{5.10}$$

where $\hat{g}(\mathbf{X})$ denotes the PCE approximation; $\mathscr{A}$ is a set of multi-indices $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_d\}$, $\mathbf{c} = \{\mathbf{c}_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathscr{A}\}$ are polynomial coefficients to be computed, and $\Psi = \{\Psi_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathscr{A}\}$ are multivariate polynomials that are orthogonal with respect to $\boldsymbol{f}_{\mathbf{X}}(\mathbf{x})$. In other words, PCE aims to provide an approximation of the model output by expressing it as a sum of orthogonal polynomials. These polynomials are systematically chosen based on the probability distributions of the input variables of the model. The result is a series where each term is a product of a polynomial function and a coefficient, which together represent how changes in the inputs affect the output. This approach is particularly useful in SA for quantifying how uncertainty in the inputs influences the model output, facilitating a deeper understanding of the model behavior under uncertainty.

The independence of input variables allows for constructing the multivariate polynomials as the tensor product of univariate orthonormal polynomials with respect to $f_{X_i}(x_i)$, i.e.,

$$\Psi_{\boldsymbol{\alpha}}(\mathbf{X}) = \prod_{i=1}^{d} \Psi_{\alpha_i}(X_i), \tag{5.11}$$

where $\Psi_{\alpha_i}(X_i)$ is a polynomial of degree α_i in the i^{th} input variable. The expansion is originally formulated with standard Gaussian random variables with Hermite polynomials [49]. It was later extended into a broader framework known as the *generalized* PCE [16] to employ basis functions from the Askey scheme of orthogonal polynomials with their underlying random variables. For example, Legendre polynomials can be associated with uniform random variables, and Laguerre polynomials correspond to the Gamma distribution. If other types of random variables appear in the input vector $\mathbf{X}$, it is possible to perform an isoprobabilistic transform such that the generalized PCE can be applied to this variable [48], [50].

For computational purposes, the PCE in Eq. (5.10) has to be truncated, where the polynomials $\{\Psi_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathscr{A}\}$ are generally retained with total degree up to p such that

$$\mathscr{A}^{d,p} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^d : \quad \|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^{d} \alpha_i \leq p \right\}, \qquad \mathbf{card}\, \mathscr{A}^{d,p} \equiv P = \binom{d+p}{p}, \tag{5.12}$$

where $\|\boldsymbol{\alpha}\|_1$ denotes the degree of the multi-indices $\boldsymbol{\alpha}$; $\mathbf{card}\, \mathscr{A}^{d,p}$ represents the number of multi-indices (i.e., the number of coefficients in the PCE). The coefficient vector $\mathbf{c}_{\boldsymbol{\alpha}}$ can be determined through regression approaches [13], [48]. However, considering the size of the basis ($\mathbf{card}\, \mathscr{A}^{d,p}$) in Eq. (5.12), the computational effort in the regression method grows dramatically with the size of d or p, which makes the full PCE intractable

in the high-dimensional problems (e.g., $d \geq 10$ or $p \geq 10$) [48]. Ref. [51] also highlighted this property limits the application of PCE in the field of vehicle-track interaction problems.

This limitation was addressed by a hyperbolic truncation scheme [48], which defines a new set $\mathscr{A}_q^{d,p}$ of multi-indices as

$$\mathscr{A}_q^{d,p} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^d : \quad \|\boldsymbol{\alpha}\|_q = \left(\sum_{i=1}^{d} \alpha_i^q \right)^{1/q} \leq p \right\}. \tag{5.13}$$

The multi-indices are determined based on q $(0 < q < 1)$ [48]. For $q = 1$, the hyperbolic truncation corresponds to the standard total-degree truncation degree in Eq. (5.12), where the polynomials of maximum total degree of p are retained. When $q < 1$, the truncation penalizes high-degree terms with many interacting variables, while favoring the main effects and low-order interactions. The reasoning behind this truncation scheme is that when there are multi-variables, the system is likely to depend primarily on the main effects and low-order interactions, which is known as the *sparsity-of-effects principle* [52]. The reader is referred to [48] for further details.

Apart from the hyperbolic truncation scheme, an *adaptive* algorithm [48] contributes further to an efficient procedure for the selection of polynomials. The algorithm is based on least angle regression (LAR) that iteratively enhances the polynomials under construction. In brief, it provides a collection of polynomial chaos (PC) representations (surrogates) in such a way that terms in $\mathscr{A}_q^{d,p}$ are added one by one, and at each iteration, the surrogate under construction is given an error estimate ε. By defining $\varepsilon^* = \min(\boldsymbol{\varepsilon})$, one stops the algorithm if ε^* is less than the preset target error ε_{tgt}. The PC representation with ε^* is eventually retained, associated with the optimal subset $\mathscr{A}^*$. It is said to be *sparse* since it contains a reduced number of terms in $\mathscr{A}^*$ compared to a full representation in Eq. (5.10).

Once LAR provides a selected set of terms at each iteration, the coefficients of the related PC representation and the corresponding ε can be computed by ordinary least squares (OLS) regression, which follows the so-called *hybrid* LAR proposed in [53].

Let $\{\mathbf{x}_{(1)}, ..., \mathbf{x}_{(N)}\}$ denote a set of N input realizations from an ED, and $\{y_{(1)}, ..., y_{(N)}\}$ be the corresponding model evaluations, i.e., $\{y_{(\varrho)} = g(\mathbf{x}_{(\varrho)}), \varrho = 1, ..., N\}$. The PC approximation is calculated by Eq. (5.10), where the coefficients $\mathbf{c}$ are chosen by minimizing the mean-square error between the exact value and its PC approximation, i.e.,

$$\mathbf{c} = \arg \min_{\mathbf{c} \in \mathbb{R}^{\mathrm{card}\,\mathscr{A}}} \mathbb{E}\left[\left(g(\mathbf{X}) - \sum_{\boldsymbol{\alpha} \in \mathscr{A}} \mathbf{c}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{X}) \right)^2 \right], \tag{5.14}$$

where $\mathbb{E}$ is the mathematical expectation. The solution of Eq. (5.14) can be obtained based on the OLS estimates [54],

$$\hat{\mathbf{c}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}\mathbf{Y}, \tag{5.15}$$

where $\mathbf{A}$ is a data matrix of size $N \times P$ and its general entry is defined by

$$A_{\varrho j} = \Psi_{\alpha_j}(\mathbf{x}_{(\varrho)}), \quad \varrho = 1, ..., N;\ j = 0, ..., P-1. \tag{5.16}$$

The accuracy of each PC representation in the LAR is evaluated by the *relative leave-one-out error* estimate, denoted by ε as mentioned above. Let $\hat{g}_{(-\varrho)}$ be the surrogate model that is constructed from ED while removing the ϱ^{th} observation. The leave-one-out error is defined as

$$\varepsilon_{LOO} = \frac{1}{N}\sum_{\varrho=1}^{N}\left[g(\mathbf{x}_{(\varrho)}) - \hat{g}_{(-\varrho)}(\mathbf{x}_{(\varrho)})\right]^2, \tag{5.17}$$

where $g(\mathbf{x}_{(\varrho)}$ and $\hat{g}_{(-\varrho)}(\mathbf{x}_{(\varrho)})$ represents the model evaluation at $\mathbf{x}_{(\varrho)}$, and its prediction from $\hat{g}_{(-\varrho)}$, respectively.

Then the relative leave-one-out error can be given by

$$\varepsilon = \frac{\varepsilon_{LOO}}{\text{Var}_Y}, \tag{5.18}$$

where Var_Y denotes the empirical variance of the output Y, calculated by

$$\text{Var}_Y = \frac{1}{N-1}\sum_{\varrho=1}^{N}(g(\mathbf{x}_{(\varrho)}) - \mu_Y)^2, \qquad \mu_Y = \frac{1}{N}\sum_{\varrho=1}^{N} g(\mathbf{x}_{(\varrho)}). \tag{5.19}$$

In case an independent dataset is available next to the training and validation set (used to construct surrogate models), the *relative generalization error*, ε_{gen}, is a measure commonly used to quantify the accuracy and predictive quality of the surrogate models, which is defined in the previous chapter, Eq. (4.11).

5.2.3. THRESHOLD-BASED SENSITIVITY ANALYSIS USING PCE

The proposed method seeks to efficiently evaluate the model sensitivity to the input factors under different design thresholds. The computation process is presented in Algorithm 1, which consists of two main procedures: construction of PCE (Procedure 1) and threshold-based sensitivity evaluation (Procedure 2).

Procedure 1 mainly follows the techniques proposed in [48], as elaborated in Section 5.2.2. To enhance readability, the algorithmic framework for constructing the PCE is presented here. Procedure 1 starts with selecting values of the algorithm parameters (*line 1*). It chooses an ED (e.g., LHS) and evaluates the model output at the corresponding sampled points (*line 2-3*). The least angle regression is applied to select the optimal set of the polynomial basis $\mathscr{A}^*$, which requires an iterative procedure (*line 4-12*) in the following:

Initially, a candidate set of p-order ($p = 1$) polynomials is determined by the hyperbolic truncation scheme, i.e., the set $\mathscr{A}_q^{d,p}$ defined by q-norm according to Eq. (5.13) (*line 5*). The LAR is applied to $\mathscr{A}_q^{d,p}$ to select the optimal set $\mathscr{A}^{(p)}$ with the lowest error $\varepsilon^{(p)}$ according to Eq. (5.17) (*line 6*). Then, $\varepsilon^{(p)}$ is checked against the target error ε_{tgt}. If $\varepsilon^{(p)} < \varepsilon_{tgt}$, it stops the iterative process (*line 9*). Otherwise $p = p+1$ (*line 10*), and repeat the process (*line 4-12*). The optimal set $\mathscr{A}^*$ is eventually retained, associated with the lowest error ε^* (*line 8*). The corresponding coefficients are determined by OLS estimators (*line 13*). The final PCE ($\hat{g}$) can be determined (*line 14*). Note that it is possible for the error $\varepsilon^{(p)}$ to increase from a certain order p, which can be attributed to overfitting. To

Algorithm 1: PCE-based simulation scheme for threshold-based sensitivity analysis

Input: PDFs of input factors $(f_{X_i}(x_i), i = 1, ..., d)$
Output: Threshold-based response $(\tilde{Y}^k_{i,j})$
Procedure 1 Construct a PC approximation of the response for a design criterion
1 $\varepsilon_{tgt}, p_{max}, q, N \leftarrow$ Preset the algorithm parameters
2 $\left\{\mathbf{x}_{(\varrho)}, \varrho = 1, ..., N\right\} \leftarrow$ Read statistic inputs $\mathbf{X}$ with specified $f_{X_i}(x_i)$; select an ED with sample size N
3 $\left\{y_{(\varrho)}, \varrho = 1, ..., N\right\} \leftarrow$ Collect the model evaluations
4 **for** $p = 1 \rightarrow p_{max}$ **do**
5 $\quad \mathscr{A}^{d,p}_q \leftarrow$ Gather a candidate set of p-order polynomials
6 $\quad \mathscr{A}^{(p)}, \varepsilon^{(p)} \leftarrow$ Apply LAR to the candidate set $\mathscr{A}^{d,p}_q$
7 $\quad$ **if** $\varepsilon^{(p)} < \varepsilon_{tgt}$ **then**
8 $\quad\quad \mathscr{A}^* = \mathscr{A}^{(p)}$, $\varepsilon^* = \varepsilon^{(p)}$
9 $\quad\quad$ **break**
10 $\quad$ **else** $p = p + 1$
11 $\quad$ **end if**
12 **end for**
13 $\mathbf{c} \leftarrow$ Compute the coefficients associated with $\mathscr{A}^*$ according to Eq. (5.14)
14 $\hat{g} \leftarrow$ Retain the final PC approximation with $\mathscr{A}^*, \mathbf{c}, \varepsilon^*$ according to Eq. (5.10)
End procedure 1
Procedure 2 Evaluate the threshold-based sensitivity on $\hat{g}$
15 $t_k, n_i, \kappa \leftarrow$ Preset the algorithm parameters
16 $X^l_i, X^u_i \leftarrow$ Define bounds for each input factor X_i according to $f_{X_i}(x_i)$
17 $x_{i,j} \leftarrow$ Discretize X_i in n_i points within $\left[X^l_i, X^u_i\right]$
18 **for** $j = 1 \rightarrow n_i$, $k = 1 \rightarrow \kappa$ **do**
19 $\quad \bar{\chi}^k \leftarrow$ Minimize $\left(\left(\hat{g} - t_k\right)^2, X_i = \{x_{i,j}\}, X_{\sim i} \in \left[X^l_{\sim i}, X^u_{\sim i}\right]\right)$ according to Eq. (5.9)
20 $\quad \tilde{Y}^k_{i,j} \leftarrow$ Collect $\bar{\chi}^k$
21 **end for**
End procedure 2

ensure proper convergence of the algorithm, an early stop criterion is introduced. This criterion terminates the process if the error $\varepsilon^{(p)}$ increases for at least two subsequent iterations ($\varepsilon^{(p)} \geq \varepsilon^{(p-1)} \geq \varepsilon^{(p-2)}$). For further details, the reader is referred to [48]. For the implementation of this algorithm, the target error ε_{tgt} is set equal to 0 to minimize the error in the attained PCE models.

Procedure 2 evaluates the threshold-based sensitivity (Section 5.2.1) on $\hat{g}$. It also begins by presetting the parameter values (*line 15*). Then, for non-uniformly distributed input factors, the lower (respectively upper) bound of X_i is determined by the 1$^{\text{st}}$ per-

centile (respectively 99$^{\text{th}}$ percentile) of the corresponding $f_{X_i}(x_i)$ (*line 16*). X_i is then discretized within these bounds into n_i points (*line 17*), as detailed in Section 5.2.1. For each discrete value of X_i and threshold level t_k, the squared deviation is minimized by fixing X_i at $x_{i,j}$ while varying the remaining factors $X_{\sim i}$ in their corresponding interval (*line 18&19*). Finally, save the threshold-based response, $\tilde{Y}^k_{i,j}$, for all the points n_i of X_i and thresholds $t_k, k = 1, ..., \kappa$ (*line 20*).

It is worth noting that the sensitivity method, whether in the extreme-based or the threshold-based formulation, is independent of the probability distribution of input factors. This is because the sensitivity is evaluated based on the discretization of the input factors, as emphasized in [3]. However, in the proposed simulation scheme, the probability distribution is required as input because the PCE models are developed based on this information.

In terms of computational efficiency, the original A-A1T method requires solving optimization problems in the order of $\mathcal{O}\{n \times d\}$, where n is the average number of discrete points in input factors and d is the total number of input factors. When multiple design criteria are present (denote the number of criteria by N_c), the number of optimizations increases to $\mathcal{O}\{n \times d \times N_c\}$. Let N_a be the average function calls required by solving one optimization problem. Then, the total function evaluations required by the original A-A1T will be $\mathcal{O}\{n \times d \times N_c \times N_a\}$.

On the other hand, the present surrogate-based method requires N evaluations of the original model to construct the surrogate models, which is independent of N_c since responses are generally returned from a single call. For structural models built from the FE analysis (which is the common case in dynamic simulations of vehicle-structure interaction), the computational time of a single run ranges from a few minutes to hours, depending on the complexity of the model. By comparison, the sparse adaptive PCE is faster to build and evaluate, and as the original model becomes more complex, the use of PCE has the potential for more computational gains in this context.

5.3. VERIFICATION

5.3.1. EXAMPLE 1: TWO-DIMENSIONAL ANALYTICAL FUNCTION

This section demonstrates the concept of output extreme surfaces of the $d-1$ dimensional space (Section 5.2.1) and validates the use of PCE for the extreme-based sensitivity method through an analytical function. The six-hump camel function is chosen because it involves only two input factors, allowing for visualization. Figure 5.3 illustrates the global nonlinear behavior of this function, which is defined as

$$f(X_1, X_2) = (4 - 2.1X_1^2 + \frac{X_1^4}{3})X_1^2 + X_1X_2 + (-4 + 4X_2^2)X_2^2, \tag{5.20}$$

with X_1 defined in the interval $[-2, 2]$ and X_2 defined in $[-1, 1]$.

Figure 5.4 depicts the output extreme surfaces ($\bar{f}_{\min}$ and $\bar{f}_{\max}$) of the six-hump camel function when fixing each factor, one at a time. The estimated importance measures are also displayed for each factor. A sample size $N = 100$ is considered given the limited dimension of the problem. The extreme surfaces evaluated from the true function are shown to compare the accuracy of PCE, Kriging, and PCK. It can be observed that both

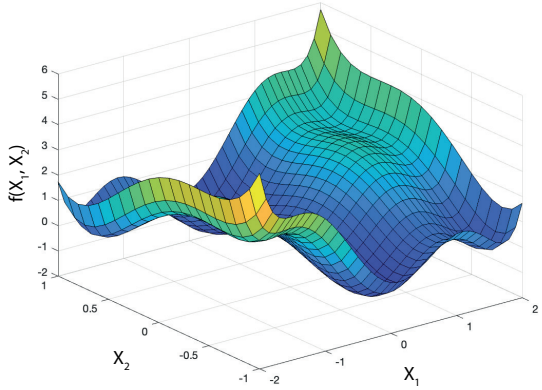

Figure 5.3: Visualization of the six-hump camel function.

PCE and PCK effectively capture the global shape of the function and accurately represent the minimal surfaces ($\bar{f}_{\min}$). However, they fail to identify the minimum point in the maximum surface ($\bar{f}_{\max}$) for X_2. Despite this, their estimated importance measures remain close to the true measures. In contrast, Kriging has limitations in approximating the global behavior of the function, resulting in a large deviation between the estimated and true importance measures, especially for the maximum surfaces ($\bar{f}_{\max}$). Note that the extreme surfaces are obtained by discretizing each factor at a time, with $n_i = 50$. Therefore, the surfaces may not appear smooth.

5.3.2. EXAMPLE 2: TRUSS STRUCTURE

The second example concerns a truss structure shown in Figure 5.5. It consists of 23 bars, and the upper section is subjected to six vertical loads. This particular structure has been studied in many works (e.g., [50], [53], [55]) for different purposes, where SA has been performed using the Sobol method and not in the extreme case. This example is investigated here to further verify the PCE-based approach against the original method [3]. Note that the threshold-based SA is not examined here as the aim is to test whether the PCE models are feasible to approximate the output of interest and compute sensitivity measures (Eq. 5.5), in terms of both efficiency and accuracy.

Ten random variables ($d = 10$) are considered in this example, including the applied loads (P_m, $m = 1,...,6$); Young's modulus and cross sections of the horizontal and diagonal elements (respectively denoted by E_1 and A_1 for the horizontal; E_2 and A_2 for the diagonal bars). Accordingly, the input vector is defined as $\mathbf{X} = [E_1, E_2, A_1, A_2, P_1, ..., P_6]'$. Table 5.1 provides the range of definition and the number of discrete points n_i for each factor X_i. For input factors with large ranges, the discrete points are distributed using a log scale. This choice is based on the suggestion in [3] where, compared to the linear and random discretization methods, using the log-spaced discretization for factors with large ranges ensures the robustness of the importance measures with a relatively small number of n_i.

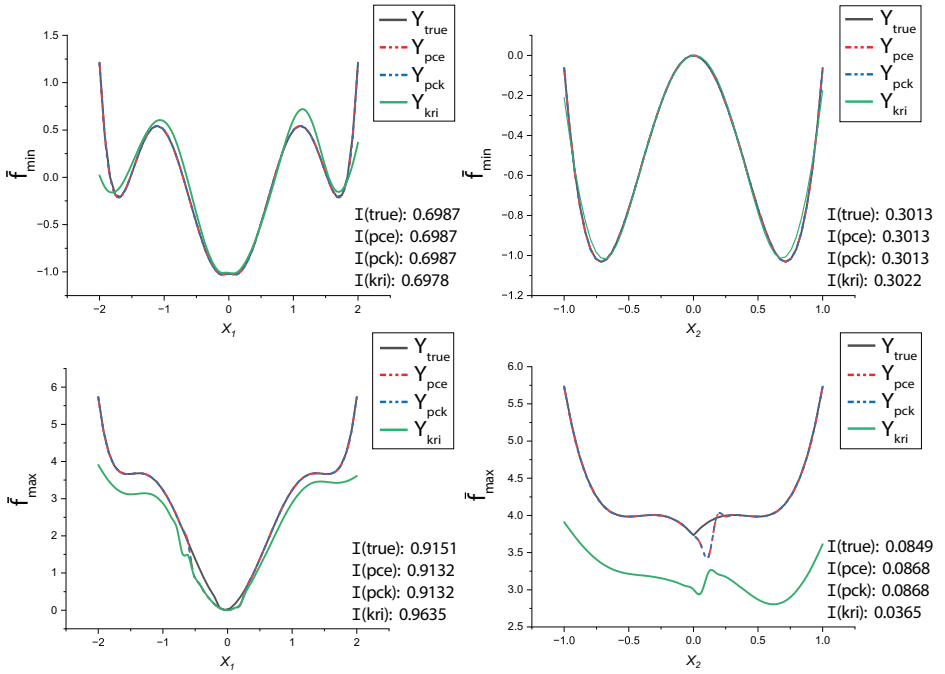

Figure 5.4: Output extreme (minimum $\tilde{f}_{\min}$ and maximum $\tilde{f}_{\max}$) surfaces of the six-hump camel function for each factor.

Table 5.1: Truss structure: definition range and discretization of the input factors.

Variable	Data proposed by [55]			Range	Discretization
	Distribution	Mean	Stdv.	$[P_{0.01}, P_{0.99}]$	
E_1, E_2 (Pa)	Lognormal	2.10×10^{11}	2.10×10^{10}	$[1.54 \times 10^{11}, 2.84 \times 10^{11}]$	50 points (log spaced)
A_1 (m^2)	Lognormal	2.0×10^{-3}	2.0×10^{-4}	[0.0015, 0.0027]	50 points (equally spaced)
A_2 (m^2)	Lognormal	1.0×10^{-3}	1.0×10^{-4}	[0.0007, 0.0014]	50 points (equally spaced)
P_m ($m = 1...6$) (N)	Gumbel	5.0×10^4	7.5×10^3	[26475, 64677]	50 points (log spaced)

The quantity of interest is the midspan deflection ω (counted positively downwards). It is obtained by evaluating an FE model (denoted by g^{truss}) using the Matlab code [56], i.e., $\omega = g^{truss}(\mathbf{X})$.

PCE models ($\hat{g}^{truss}$) are constructed to approximate the value of ω based on $\mathbf{X}$. The random vector $\mathbf{X}$ is transformed into a standard Gaussian distributed vector (i.e., a change of scale in terms of the units or magnitude of the input) to use Hermite polynomials [49] in constructing the PCE model in the form [48], [50],

$$\xi_i = \Phi^{-1}\left(F_{X_i}(x_i)\right), \quad i = 1, ..., 10, \tag{5.21}$$

where Φ is the standard Gaussian cumulative distribution function (CDF) and $F_{X_i}(x_i)$ is

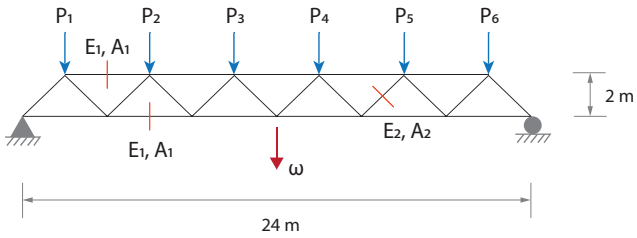

Figure 5.5: Truss structure with 23 members [55].

the CDF of X_i. This results in the following PCE model according to Eq. (5.10),

$$\omega \approx \hat{g}^{truss}(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha}\in\mathscr{A}} \mathbf{c}_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{5.22}$$

Latin hypercube sampling is used to generate the experimental design. A sample size of $N = 1000$ is considered as the base scenario, and the influence of different sample sizes is evaluated subsequently. The FE model is queried at the sampled points to generate the corresponding response ω. The input-output pairs formulate a dataset that is randomly divided into training, validation, and test sets with the respective percentage of 64%, 16%, and 20%, where five iterations were performed, resulting in five candidate PCE models.

All PCE models are constructed by varying the q-norm from 0.5 to 1 and the maximum degree p from 3 to 15. The model performance is evaluated based on the test set using ε_{gen} (Eq. 4.11), where the model with the smallest ε_{gen} yields 2.54×10^{-7} for predicting ω. The optimal degree p is 9 and q-norm is 0.5, associated with $\varepsilon = 1.15 \times 10^{-7}$. This model includes 181 polynomial basis elements, while the size of basis elements for $q = 0.5$ and $p = 9$ is 571, i.e., without applying the LAR algorithm, and the size of full basis elements can be 92378 for $q = 1$ and the same p, i.e., the maximum possible number of polynomial basis elements without any sparsity constraints for a polynomial degree $p = 9$.

Two functions, g^{truss} and $\hat{g}^{truss}$, are applied to the A-1AT method (Section 5.2.1). The importance measures I_i^{Ext} to each factor X_i are reported in Table 5.2. It can be observed that I_i^{Ext} returned from the PCE model $\hat{g}^{truss}$ agree well with those calculated from directly calling the FE model g^{truss} (the reference). The reference case also shows differences with the result in [50]. This is because [50] employs the Sobol method, which measures the impact of input factors based on the average value of the midspan deflection.

The number of function calls to the FE model is also tracked for both methods. As mentioned earlier, an optimization problem is present (Eq. 5.4) for each discrete point of X_i to search for the output extremum. In this example, the Matlab algorithm 'patternsearch' [57] has been used. The average number of model runs per optimization N_a was about 165, and the average number of discrete points n for input factors was 50. Therefore, the reference method (direct calls of g^{truss}) requires 82242 ($= n \times d \times N_a$) times of function calls, which took 42.5 min on a desktop with an 8-core CPU and 16 GB of RAM. In contrast, the PCE-based method needs $N = 1000$ function evaluations to

Table 5.2: Truss structure: comparison of importance measures calculated from the PCE ($\hat{g}^{truss}$) and actual FE model evaluation (g^{truss}) when $N = 1000$.

Variable	Importance measure I_i^{Ext}	
	Reference	PCE-based
E_1	0.3952	0.3965
A_1	0.3703	0.3706
P_3	0.0652	0.0652
P_4	0.0652	0.0649
P_5	0.0314	0.0315
P_2	0.0314	0.0310
A_2	0.0180	0.0177
E_2	0.0154	0.0149
P_6	0.0040	0.0038
P_1	0.0040	0.0038
Function calls	82242	1000

construct and validate the surrogate model. The evaluation of $\hat{g}^{truss}$ in SA took 5.5 min with the same computing condition.

The importance measures are also evaluated using different sample sizes (N) and surrogate models (PCE, Kriging, and PCK), as shown in Figures 5.6 and 5.7. In the alternative surrogate models, an ordinary Kriging model with a Matérn 5/2 correlation function is applied. PCK considers a sequential formulation [58] where the set of polynomials and Kriging are determined sequentially. The LAR algorithm (Section 5.2.2) is applied to select the optimal set of polynomials, varying the maximum degree p from 3 to 15. This set of polynomials is then used as the trend of a Kriging model with a Gaussian correlation function, which is further calibrated using maximum likelihood. For sensitivity evaluation, the model training is repeated five times for each approach, and the one with the lowest ε_{gen} is chosen.

The comparison of the sample size and surrogate models is presented in Figure 5.6. It can be observed that Kriging shows the largest deviation from the true values across all the sample cases. PCE and PCK demonstrate comparable performance, with PCE showing better results especially when the sample size is small.

The analysis is replicated to account for the uncertainties in experimental designs. Figure 5.7 shows the relative errors of the sensitivity measures over 20 trials. For small sample sizes ($N = 100$ or 400), PCE demonstrates the best overall performance in terms of both mean and variation of the errors, while PCK shows the largest variation of the errors when $N = 100$, especially for factors P_1 and P_6 that have a negligible effect on the model output. However, as the sample size N increases ($N = 700$ or 1000), PCK shows a slight improvement over PCE, as indicated by lower mean and variation of the errors. This difference can be attributed to the fact that PCK is more prone to overfitting when dealing with small ED sizes [58]. On the other hand, Kriging shows considerable bias in the mean of the errors across all the sample sizes, and no clear improvement in its performance is observed as the sample size increases.

Further, Figure 5.8 compares the importance measures using different levels of dis-

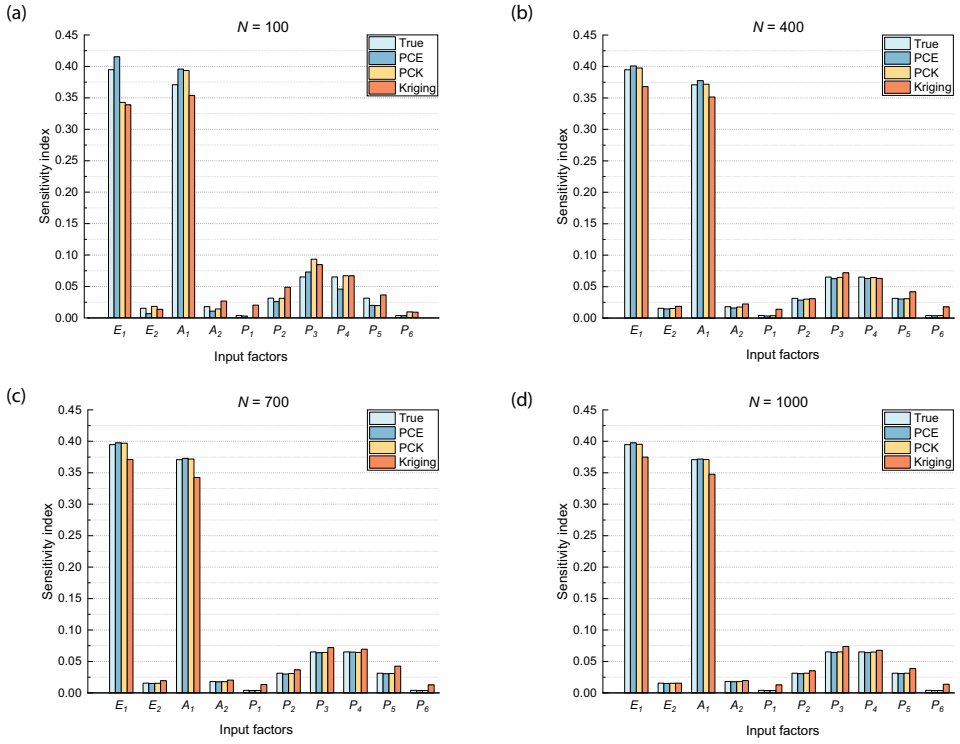

Figure 5.6: Importance measure of each input factor using true function evaluations (True), PCE, PCK, and Kriging model.

cretization. The results demonstrate the robustness of the importance measures across different cases and indicate that the default size of $n_i = 50$ is sufficient for the current analysis. For a more detailed discussion of the discretization aspect, the reader is referred to [3].

5.4. APPLICATION TO A TRAIN-TRACK-BRIDGE SYSTEM

5.4.1. MODELING OF TRAIN-TRACK-BRIDGE DYNAMICS

As a typical example of vehicle-structure interaction problems, the vibration of the train-track-bridge (TTB) system is a fundamental concern in railway engineering, frequently used to evaluate running safety, riding comfort, and performance of railway tracks and bridges. Here, the applicability of the proposed method is demonstrated using a coupled TTB model [59]. The model is FE-based and is implemented in Matlab. It is capable of simulating the vertical dynamic interaction between the subsystems, namely the train, track, and bridge. A schematic representation of the model is depicted in Figure 5.9, where an articulated train is shown traveling over a ballast bridge at a specific speed. Equations of motion are defined for each subsystem, and their features are summarized

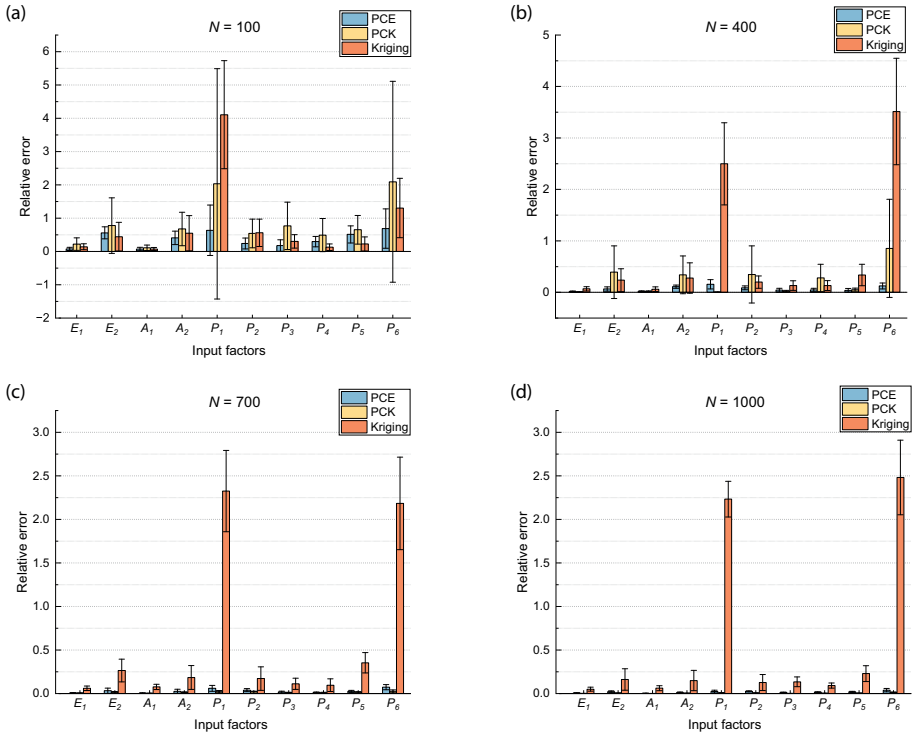

Figure 5.7: Accuracy of the importance measure of each input factor using PCE, PCK, and Kriging model. The bar heights show the mean and error bars represent ±1 standard deviation over 20 trials.

below.

The train is represented by a succession of individual vehicles, with each vehicle consisting of one carbody, two bogie frames, and four wheelsets, treated as rigid bogies. Each bogie is connected to two wheelsets through the primary suspension, and the main body rests on two bogies via the secondary suspension. Both vertical and rotational movements of the carbody (respectively, z_c and θ_c) and bogies (respectively, z_{bi} and $\theta_{bi}, i = 1,2$) are considered. A rigid contact is assumed between the rail and wheelsets. This results in 6 DOFs for each vehicle of the train and the displacement vector of a vehicle can be denoted as $\mathbf{u}_v = [z_c, \theta_c, z_{b1}, \theta_{b1}, z_{b2}, \theta_{b2}]'$.

The railway track concerns a ballast track structure that includes a combination of components, i.e., rail, pads, sleepers, ballast, and subgrade. The rail is modeled as an Euler-Bernoulli beam, with each element having 4 DOFs in terms of vertical and rotational motions. The remaining components are represented as layers of masses and viscoelastic supports, which is a conventional simplification for modeling the dynamic behavior of railway tracks [60]. The model used here is a three-layer track model. This is recommended by UIC (International Union of Railways) [61] that specifies design requirements for railway bridges with regard to train-track-bridge interaction phenomena,

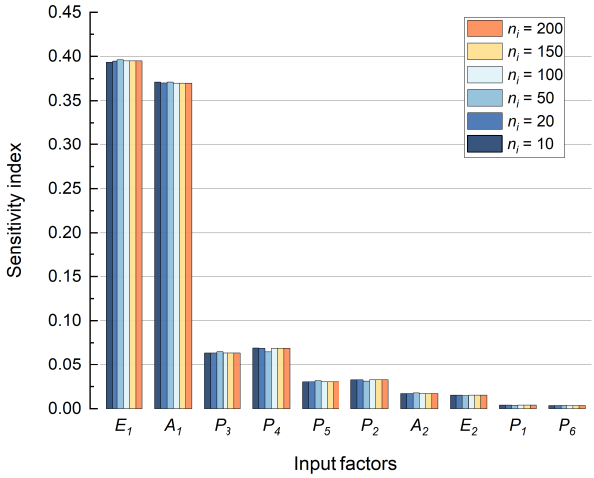

Figure 5.8: Importance measure of the truss structure for different degrees of discretization.

as stressed in [62]. Depending on the particular bridge configuration, a bridge structure can be accurately modeled by finite elements of solid, shell, and beam. For simplicity and without loss of generality, this demonstration considers a simply supported bridge meshed with Euler-Bernoulli beam elements.

Each subsystem is defined by a set of equations of motion, and their coupling can be formulated in the following general matrix form [63],

$$\mathbf{M}_g\ddot{\mathbf{U}}_g + \mathbf{C}_g\dot{\mathbf{U}}_g + \mathbf{K}_g\mathbf{U}_g = \mathbf{F}_g \tag{5.23a}$$

where $\mathbf{M}_g$, $\mathbf{C}_g$, and $\mathbf{K}_g$ denote, respectively, the mass, damping, and stiffness matrices of the global system. $\mathbf{U}_g$,$\dot{\mathbf{U}}_g$, and $\ddot{\mathbf{U}}_g$ are the displacement, velocity, and acceleration vectors of the global system, respectively. $\mathbf{F}_g$ is the global force vector. Their expressions are given by

$$\begin{aligned}
\mathbf{M}_g &= \begin{bmatrix} \mathbf{M}_v & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_b \end{bmatrix}, \ \mathbf{C}_g = \begin{bmatrix} \mathbf{C}_v & \mathbf{C}_{vt} & \mathbf{0} \\ \mathbf{C}_{tv} & \mathbf{C}_t & \mathbf{C}_{tb} \\ \mathbf{0} & \mathbf{C}_{bt} & \mathbf{C}_b \end{bmatrix}, \\
\mathbf{K}_g &= \begin{bmatrix} \mathbf{K}_v & \mathbf{K}_{vt} & \mathbf{0} \\ \mathbf{K}_{tv} & \mathbf{K}_t & \mathbf{K}_{tb} \\ \mathbf{0} & \mathbf{K}_{bt} & \mathbf{K}_b \end{bmatrix}, \ \mathbf{U}_g = \begin{bmatrix} \mathbf{U}_v \\ \mathbf{U}_t \\ \mathbf{U}_b \end{bmatrix}, \ \mathbf{F}_g = \begin{bmatrix} \mathbf{F}_v \\ \mathbf{F}_t \\ \mathbf{F}_b \end{bmatrix}.
\end{aligned} \tag{5.23b}$$

where the subscripts v, t, and b denote respectively the train, track, and bridge subsystems.

The coupled equations of motion, Eq. (5.23a), is solved by direct numerical integration with the Newmark-β method to obtain the dynamic responses of the subsystems [59], corresponding to the design criteria considered in SA.

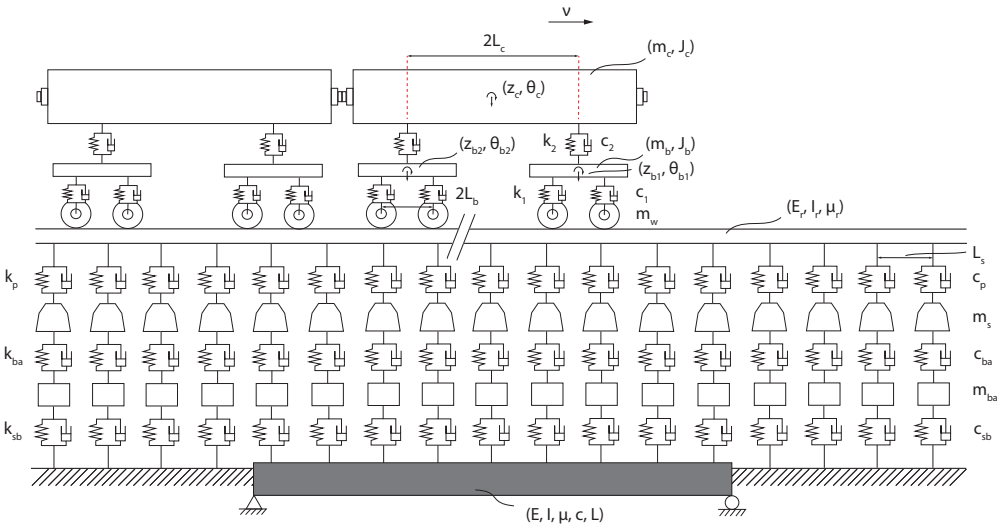

Figure 5.9: Schematic representation of a train-track-bridge model.

5

5.4.2. CHARACTERISTICS OF INPUT FACTORS AND DESIGN CRITERIA

The variability space of input factors is defined to perform SA. This includes determining parameters (that are subject to small deviations and thereby being treated as deterministic) and variables (that may have relevant non-deterministic properties that can lead to a significant change in model response) in the TTB system. This case study considers the variability in factors related to the structural resistance and excitation source. The former is relevant to the train, track, and bridge, while the excitation source focuses on the effect of track geometry irregularities on the TTB dynamics. Table 5.3 provides an overview of the random variables defined in the TTB system, along with their corresponding probability distributions.

Trains exhibit significant variability in terms of their dimensions and properties, e.g., Eurostar [67], Manchester benchmark [68], VIRM trains [69], Koploper ICM and ICR carriage [70]. The ranges of variation for the properties of train catalogs are location specific and have to be estimated according to the operating context in which the analysis is taking place. For demonstration purposes, the Manchester benchmark model is considered here with variations in factors including the carbody mass (m_c), primary suspension stiffness (k_1), and secondary suspension stiffness (k_2) (see Table 5.3). These are selected according to [36], given the fact that the mass of vehicles varies depending on the occupancy rate of passengers and that the suspension stiffness shows variability during its service life. The rest of the properties are treated as constants according to the benchmark model [68], as presented in Table 5.4. In the simulation, three successive vehicles are considered.

Regarding the track structure, the variability of supporting components, namely, the railpad, sleeper, and ballast, is high. The railpad and sleeper are often associated with a wide range of design alternatives, while the ballast properties are very likely to change

Table 5.3: Input random variables. Note: the fourth and fifth columns depend on the distribution type. For Gaussian or lognormal distributed random variables, mean values and coefficient of variation (CV) are used. For uniformly distributed variables, minimum and maximum values are used.

Variable	Unit	Distribution	Mean or Min.	CV or Max.	Reference
Train random variables					
Carbody mass factor	-	Lognormal	1	0.15	[36]
Primary suspension stiffness factor	-	Lognormal	1	0.1	[36]
Secondary suspension stiffness factor	-	Lognormal	1	0.1	[36]
Track random variables					
Railpad stiffness (k_p)	N/m	Uniform	1e8	1.5e9	[64]
Railpad damping (c_p)	N·s/m	Uniform	1e4	7e4	[64]
Sleeper mass (m_s)	kg	Uniform	220	325	[35]
Ballast density (ρ_{ba})	kg/m^3	Uniform	1500	2100	[35]
Ballast elastic modulus (E_{ba})	N/m^2	Uniform	8e7	1.6e8	[35]
Ballast depth (h_{ba})	m	Uniform	0.3	0.6	[35]
Ballast load distribution angle (α_{ba})	°	Uniform	15	35	[35]
Ballast damping (c_{ba})	N·s/m	Uniform	4e4	2.8e5	[64]
Irregularity amplitude (I_{rr})	10^{-7} rad·m	Uniform	4.032	10.80	[65]
Bridge random variables					
Concrete density weight (μ)	kg/m	Gaussian	69000	4%	[35], [66]
Concrete elastic modulus (E)	N/m^2	Gaussian	35e9	8%	[35], [66]
Structural damping ratio (c)	%	Gaussian	2	0.3 stdv.	[35], [44]
Second moment of area (I)	m^4	Gaussian	51.3	5%	[45]

Table 5.4: Vehicle parameters.

Parameter	Notation	Value
Carbody mass	m_c	32000 kg
Bogie mass	m_b	2615 kg
Wheelset mass	m_w	1813 kg
Primary suspension stiffness	k_1	1220 kN/m
Secondary suspension stiffness	k_2	430 kN/m
Primary suspension damping	c_1	4 kN·s/m
Secondary suspension damping	c_2	20 kN·s/m
Distance between centers of car and bogie	L_c	9.5 m
Distance between centers of bogie and wheel	L_b	1.28 m
Full length of vehicle	L_v	25 m
Velocity	v	140 km/h

due to geometry issues and maintenance works. Therefore, they are more appropriately described by bounding limits [71], i.e., uniform distributions shown in Table 5.3.

It is worth mentioning that for the ballast, it is not straightforward to find equivalent properties to capture its behavior using the multi-layer track model (e.g., the present

three-layer model). To this end, Zhai et al. [65], [72] proposed a ballast model to analyze its vibration based on a hypothesis that the load transmission from the sleepers to the ballast follows approximately a cone distribution. This model defines a function mapping between the input factors relevant to the ballast material properties and the output regarding the equivalent parameter values for the ballast block used in the multi-layer track model (see Figure 5.9). It has been validated against a field measurement [72] and applied to relevant works (e.g., [35], [73]). For details of the mathematical formulation, the reader is referred to [65], [72].

As shown in Table 5.3, the variability in ballast properties is considered, including the density (ρ_{ba}), elastic modulus (E_{ba}), depth (h_{ba}), and load distribution angle (α_{ba}), which are specified according to [35]. Then, the above ballast model is applied to determine the equivalent parameter values for the ballast, namely, the vibrating mass of ballast under a sleeper support (m_{ba}) and the corresponding support stiffness k_{ba} (see Figure 5.9). The viscous damping coefficient c_{ba} is considered as a constant in [72]. Here, it is treated as a random variable referring to [35], [44] to cover the variability in energy dissipation mechanisms. Besides, the ballast model also includes dimensions of the sleeper support such as the sleeper spacing, L_s, to determine m_{ba} and k_{ba}. Since the dimensions are normally better defined in design specifications, they are treated as constants according to [72].

The rail is modeled as Euler-Bernoulli beam elements, whose behavior is defined by the elastic modulus (E_r), moment of inertia (I_r), and mass per unit length (μ_r), as presented in Figure 5.9. Compared with the ballast, the properties of the steel rail are relatively easy to determine from nominal design values, and the subgrade properties (k_{sb} and c_{sb}) are also better defined according to the required bearing capacity. These properties are therefore considered deterministic and the corresponding values are referred to [72].

The properties of railway bridges are case-specific, and it is challenging to find a set of properties that can be applied to describe large catalogs of bridge structures in general. Herein, a 50 m long concrete bridge is considered, which refers to a specific case reported in [74]. The reference work [74] also suggested the default parameter setting for the bridge in the adopted TTB model [59]. Consequently, the total track length modeled in the simulation is 110 m, including a 30-m approach, the 50-m bridge itself, and an additional 30-m track section following the bridge. The mean values of bridge properties are first determined according to the selected case. Then, Gaussian distributions are assigned to describe the variability of these properties, as reported in Table 5.3.

Track irregularities are deviations from the ideal track geometry that can significantly affect the dynamic behavior of trains and structures. It is therefore essential to account for their effect in the dynamic analysis of such systems. Random track irregularities are often characterized by PSD functions, which describe the severity of track irregularities as a function of the spatial frequency $\Omega = 1/\lambda$, where λ represents the wavelength in meters. Various PSD functions have been developed by different railway authorities. Given its extensive use in the field of railway engineering, the German track spectrum for the vertical track profile is adopted with the PSD function defined as (e.g., [65]),

$$S_v(\Omega) = \frac{A_v \Omega_c^2}{\left(\Omega^2 + \Omega_r^2\right)\left(\Omega^2 + \Omega_c^2\right)}, \tag{5.24}$$

where the unit of $S_v(\Omega)$ is $m^2/rad/m$; Ω is the spatial frequency; Ω_r and Ω_c are cut-off frequencies, set to 0.0206 and 0.8246 rad/m, respectively. The magnitude of A_v (unit: $m^2 \cdot rad/m$) quantifies the track quality and varies between 4.032×10^{-7} and 1.08×10^{-6} to represents respectively the limit for low disturbance and high disturbance of track irregularities. Here, the track quality is represented by a uniformly distributed random variable within the range. The function considers wavelengths ranging from 3 to 150 m (referring to EN 13848-5 [75] ranges D1-D3). To perform time-domain analyses, track irregularity profiles are generated from Eq. (5.24) using inverse fast Fourier transforms [76]. The random seed is fixed for each run to maintain consistency across simulations and to specifically analyze the effect of track geometry quality on the TTB system.

The performance of the TTB system is controlled by four design criteria, namely Sperling index (W_z), vertical rail displacement (U_r), vertical sleeper acceleration (A_s), and vertical deck deflection (U_d), where the dynamic response of each subsystem is accounted for to capture the overall level of vibration in the system. Of interest for the present SA are the maximum values of the responses, which are relevant to the exceedance of a limit state, i.e., the occurrence of 'system failure', and therefore reflect the most unfavorable condition of the system. While the remaining response quantities can be obtained directly from the dynamic simulation, W_z is a synthetic index calculated based on the carbody acceleration. This index measures the level of riding comfort and details are provided below.

Trains are exposed to excitation from contact with the rail while running in the complex operating environment, which is transmitted from the floor and seats to affect the riding comfort of passengers. Comfort is a subjective feeling, and, because of this, substantial efforts have been made to transform vibration data (either from simulations or measurements) into objective indices and associate the indices with subjective feelings [47]. The severity of human exposure to the vibration environment is generally quantified by the weighted acceleration [77], in which the weighting function is incorporated into the acceleration amplitude to reproduce the sensitive frequency range and resonance strength of the human body. Sperling's method (W_z index) [78] is one of the most common methods for assessing passenger comfort on trains (e.g., [77], [79]).

In its original form, the Sperling index (for a certain frequency) is calculated as [78],

$$W_z = 0.896 \left(\frac{A^3}{f} F(f)\right)^{1/10}, \tag{5.25a}$$

where A is the acceleration amplitude (unit: m/s^2); f denotes the corresponding frequency (unit: Hz). $F(f)$ is a frequency weighting function to represent the human perception of the frequency range. Usually, a weighting function $B(f)$ is introduced, and Eq. (5.25a) can be rewritten as

$$W_z = \left(A^3 B(f)^3\right)^{1/10} = \left(A_B(f)^3\right)^{1/10}, \tag{5.25b}$$

where $A_B(f)$ denotes the weighted acceleration at a certain frequency f.

Considering the discrete signals in a whole spectrum, the total Sperling index can be derived as [77],

$$W_z = \left(\sum_{k=0}^{N-1} (A_B(k))^3 \right)^{1/10}, \tag{5.26}$$

where N denotes the number of sample points in the considered spectrum.

Since the present model is based on time-domain analyses, the method in [77] is applied to convert the calculation process of W_z from the frequency domain to the time domain. Note that there are several different methods to determine W_z, and the index values are inconsistent due to differences in the algorithms used, sampling time, and frequency sample interval. To circumvent this issue, Deng et al. [77] proposed the following generalized form of Eq. (5.26),

$$W_{z,m} = \left(\sum_{k=0}^{N-1} (A_B(k))^m \right)^{0.3/m}, \tag{5.27}$$

where $m = 2$ or 3, i.e., the second or third powers of the acceleration. A consistency check was made between time-domain and frequency-domain analyses over different sample times. It was found the results are stable only when the calculation of $W_{z,m}$ is carried out at $m = 2$ [77]. This leads to the following equivalent calculation of W_z in the time domain [77],

$$W_z = \left(\sqrt{\sum_{k=0}^{N-1} (A_B(k))^2} \right)^{0.3} = \left(\frac{1}{N} \sqrt{\sum_{n=0}^{N-1} (a_B(n))^2} \right)^{0.3}, \tag{5.28}$$

where $a_B(n)$ denotes the weighted time-domain acceleration sequence. It is determined by a time-domain acceleration sequence and the corresponding weighting function. The former is generated from the current model, and the weighting function $H(s)$ for the time-domain response at the time instance s is derived according to Parseval's theorem, which is given by [77],

$$H(s) = \frac{0.59(0.0063s^2 + 0.2200s)}{1.4836 \times 10^{-4} s^3 + 0.0070s^2 + 0.2488s + 1}. \tag{5.29}$$

5.5. RESULTS AND DISCUSSION

5.5.1. PCE REPRESENTATIONS OF THE MODEL OUTPUT

As the first part of the methodology, PCE models are constructed to approximate the model output. This is implemented using the uncertainty quantification toolbox UQLab [80]. Following the workflow in Section 5.3, an experimental design $\left\{ \boldsymbol{x}_{(\varrho)}, \varrho = 1, ..., N \right\}$ with a fixed size of $N = 3000$ is generated using LHS. The choice of a relatively large sample size aims to ensure the accuracy of the surrogate models, as SA is performed around extreme values that may be omitted by global sampling approaches conducted in a single step. When the ED is fixed (or not adaptive), the LHS technique has been preferred in reliability applications due to its global representation of the input space [81]. This aligns with the goal of surrogate modeling in this chapter, which is to develop

a globally accurate surrogate model to approximate the original function and enable the measurement of each input factor's effect on the output.

The TTB model is queried at the sampled points to obtain the response quantities $\left\{\boldsymbol{y}_{(\varrho)}, \varrho = 1,...,3000\right\}$. For a specific sample ϱ, the response vector is defined by $\boldsymbol{y} = \{W_z, U_r, A_s, U_d\}'$. The sample data is split randomly into training, validation, and test sets with the respective percentage of 64%, 16%, and 20%, where for each design criterion, ten iterations were performed and the resulting candidate surrogate models are evaluated based on the test set using ε_{gen} (Eq. 4.11). Figure 5.10 compares the PCE predictions with the actual model evaluations at the test set, and ε_{gen} obtained from the optimal PC approximation is indicated for each criterion.

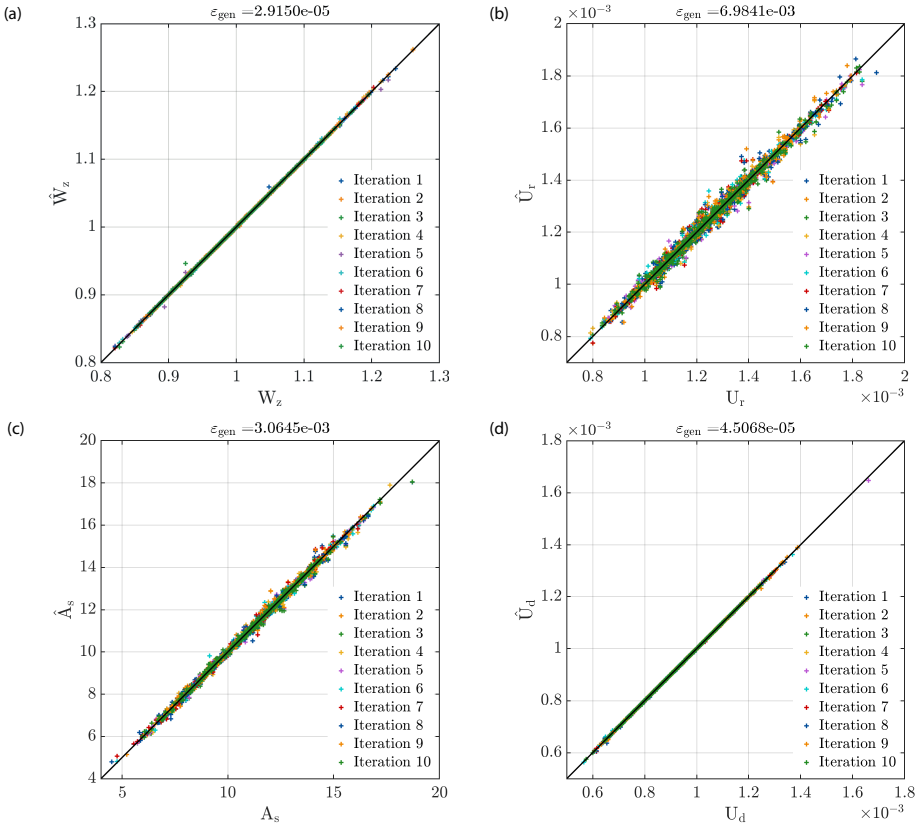

Figure 5.10: Comparison of PCE with the actual model evaluation at the test set: (a) Sperling index W_z; (b) vertical rail displacement U_r (unit: m); (c) vertical sleeper acceleration A_s (unit: m/s^2); and (d) vertical deck deflection U_d (unit: m).

The PCE models are developed by varying the q-norm from 0.5 to 1 and the maximum degree p from 3 to 15. The one yielding the smallest ε_{gen} to predict W_z comprises 371 polynomial basis elements, with $p = 15$ and $q = 0.5$. By comparison, for the same p, the size of full basis elements is 9081 when $q = 0.5$ and around 1.55×10^8 when $q = 1$.

The index of sparsity, defined in [48] as the ratio of the number of elements in the sparse representation to the full size of elements for the same p and q, is $371/9081 \approx 0.041$. The small ratio indicates the computational gain that the sparse PCE can bring to the sensitivity analysis, compared with the full PC representation.

The optimal sparse PCE for predicting U_r consists of 136 basis elements, with $p = 15$ and $q = 0.5$. It has an index of sparsity of $136/9081 \approx 0.015$. As for the response A_s and U_d, the optimal PCE includes 223 and 442 elements, respectively, with the same p and q as in the model for approximating U_r. The corresponding index of sparsity is $223/9081 \approx 0.025$ for A_s and $442/9081 \approx 0.049$ for U_d.

Additionally, the accuracy and efficiency of PCE are compared with the Kriging and PCK models, which are presented in Table 5.5. The training of Kriging and PCK follows the same methodology as described in Section 5.3. Ten independent runs are carried out for each surrogate, and the one with the lowest relative generalization error ε_{gen} is chosen for further comparison. Results from Table 5.5 indicate that PCE outperforms the other approaches in terms of ε_{LOO} and function evaluation time. PCK performs slightly better than PCE (with the same order of accuracy) for the relative generalization error ε_{gen} of two response indicators (W_z and U_d). Furthermore, PCE demonstrates significant computational efficiency compared to these approaches, which is advantageous for conducting sensitivity analyses of large-scale engineering systems.

Table 5.5: Train-track-bridge system: comparing the performance of surrogate models to approximate Sperling index W_z, vertical rail displacement U_r (unit: m), vertical sleeper acceleration A_s (unit: m/s^2), and vertical deck deflection U_d (unit: m). Minimum values are highlighted for each index (ε_{LOO}, ε_{gen} and average function evaluation time). The time refers to a single evaluation of the function.

Response	Surrogate	Minimum ε_{LOO}	Minimum ε_{gen}	Average function evaluation time (s)
W_z	PCE	$\mathbf{7.13 \times 10^{-6}}$	2.92×10^{-5}	**0.0044**
	Kriging	5.00×10^{-4}	5.09×10^{-4}	0.1904
	PCK	9.17×10^{-6}	$\mathbf{2.73 \times 10^{-5}}$	0.2168
U_r	PCE	$\mathbf{5.93 \times 10^{-3}}$	$\mathbf{6.98 \times 10^{-3}}$	**0.0039**
	Kriging	1.77×10^{-2}	1.42×10^{-2}	0.2037
	PCK	7.18×10^{-3}	9.17×10^{-3}	0.2141
A_s	PCE	$\mathbf{1.74 \times 10^{-3}}$	$\mathbf{3.06 \times 10^{-3}}$	**0.0044**
	Kriging	5.62×10^{-3}	5.92×10^{-3}	0.1804
	PCK	2.18×10^{-3}	3.38×10^{-3}	0.2003
U_d	PCE	$\mathbf{1.44 \times 10^{-5}}$	4.51×10^{-5}	**0.0074**
	Kriging	1.55×10^{-3}	4.22×10^{-4}	0.1967
	PCK	1.59×10^{-5}	$\mathbf{3.77 \times 10^{-5}}$	0.2513

Note that the performance of surrogate models depends on various factors, such as the shape and complexity of the function being approximated, the ED size, and the sampling strategies employed (e.g., [82]–[84]). This chapter employs the PCE approach in the

proposed sensitivity method, which allows us to efficiently capture the global stochastic behavior of the system [58], [81] and measure the effect of input factors on the output. The results from the truss example demonstrate that PCE performs particularly well with limited ED sizes, which is advantageous when dealing with expensive models. However, it should be noted that PCE is most effective for functions that can be well-approximated by global smooth polynomials. In two engineering problems, it was observed that PCE outperformed Kriging. Kriging is generally suitable for managing local variability of the output [58]. This suggests that the functions involved show global smoothness rather than (highly) local nonlinearity, thereby contributing to the better performance of PCE.

As mentioned earlier, this chapter uses LHS with a fixed ED to provide a global description of the input space. Instead of sampling the ED at once, it is possible to apply adaptive or sequential sampling techniques to refine sampling in specific regions of interest. This approach allows for a more efficient allocation of computational resources, balancing the exploration and exploitation of the input space while constructing the surrogate model. Note that in this case, the sampling is more focused on achieving accuracy in the proximity of specific regions, depending on the degree of exploitation, rather than aiming for a globally accurate surrogate model throughout the entire domain, which is more exploration-based. The choice of sampling methods can depend on function complexities. For functions with highly local nonlinearity, where the extremes are usually local phenomena, a balance of exploration and exploitation of the input space would be necessary. Considering that the functions in the current engineering problems demonstrate global smoothness, LHS, which captures the global description of the ED, is considered suitable here.

5.5.2. RANKING OF INPUT FACTORS

The PCE models developed in Section 5.5.1 are employed to assess and rank the impact of input factors on the variability of extreme responses. This evaluation is performed on the original SA formulation presented in Section 5.2.1. The input factors are discretized using an equal number of points. Similar to the discretization strategy discussed in Section 5.3, a log-spaced discretization method is applied for factors with ranges exceeding 10^3. After evaluating various discretization options, $n_i = 100$ is chosen for each factor to balance the computational cost and the accuracy of sensitivity index values, which is applied to all design criteria to ensure consistency in the comparison. Given that the optimization of the function is performed at each point, the number of required optimization problems is 1600 for a single criterion and 6400 if all criteria are considered.

For additional information, a comparison of the importance measures for criterion U_r using different discrete points is provided in Figure 5.11. For smaller discrete points ($n_i = 10$ or 20), the importance measures provide a consistent ranking for factors that collectively account for more than 99.58% of the variability of extreme U_r. Some deviations in rankings are observed for factors c_{ba} and k_1, I_{rr} and k_2, with their individual contributions to the response variability being at most 0.23% (factor c_{ba} when $n_i = 10$). From $n_i = 50$ onwards, the rankings become stable for all the factors considered, indicating that the default size of $n_i = 100$ is sufficient for the current analysis of the TTB system.

Further, the importance measures (see Eq. 5.5) of the input factors are compared

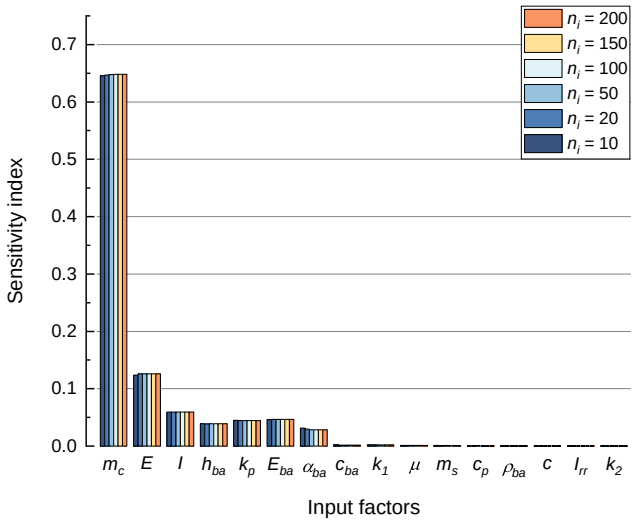

Figure 5.11: Importance measure of U_r for different degrees of discretization.

across responses W_z, U_r, A_s, and U_d, which are presented in Figure 5.12. The indices are first presented in descending order for W_z. Then, different rankings regarding the relevance of the factors are observed for the remaining criteria. The variability in carbody mass (m_c) shows high relevance to all the criteria considered, indicating that the loading magnitude plays a significant role in the vibration of the TTB system.

Specifically, the factors relevant to the vehicle (m_c and k_2) and track geometry quality (I_{rr}) contribute to most of the variability (about 99.47%) in the maximal W_z. Employing the condition $I_i^{Ext} < 0.01$ to sort out unimportant factors allows one to consider 13 out of 16 input factors as unimportant when the design focus lies on the critical riding condition that passengers may experience. This implies that those factors could be given a deterministic value to reduce the model complexity without essentially affecting the extreme response W_z. On the other hand, the above three factors are deserving of further analysis or measurement to assign the appropriate values in the model for a more accurate representation of the riding quality condition.

At the wheel-rail interface, more input factors are involved that have non-negligible effects on the variability of maximal vertical deflection of the rail (U_r). These are m_c, E, I, h_{ba}, k_p, E_{ba}, and α_{ba}, with the first two being dominant and accounting for about 76.60% of variability in maximal U_r. The accuracy of these factors is of high relevance to the quality of the extreme response U_r, which should be carefully defined if the rail deflection is of concern in the design process.

For underlayers of the track structure, the variability of the maximal vertical sleeper acceleration (A_s) is significantly influenced by m_c and k_p (about 87.55%). Apart from m_c, the effect of the railpad stiffness k_p can be explained by the fact that the railpads

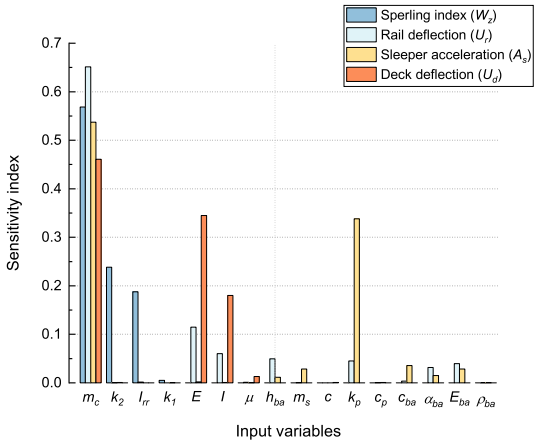

Figure 5.12: Comparison of sensitivity index of each input factor in terms of W_z, U_r, A_s, and U_d.

are elastic components introduced to the track structure. Stiff pads contribute to a reduction in noise and vibration from wheel-rail contact, while soft pads allow for a lower effect of loads transmitted to underlayers and therefore mitigate damage and vibrations to the sleepers and ballast [69], [85]. This indicates if the designer focuses on the critical condition of the supporting layers (for example, in the case of transition zones to the railway bridges [86] where issues with settlement often exist), the values of m_c and k_p should be chosen with careful consideration in the modeling process, especially for k_p since it often comes with a wide range of values and is also influenced by the frequency and environmental factors such as temperature, preload, and aging [40]. If possible, a more advanced model that captures the effect of those factors on the railpad properties should be employed for a more accurate representation of the railpad behavior.

It can also be observed from Figure 5.12 that m_s, as the only property relevant to the sleeper in the modeling, plays a minor role (about 2.83%) in the variability of maximum sleeper accelerations. Despite the common use of mass elements to represent sleepers in railway structures (e.g. [60]), it may be inferred that, in the adopted TTB model, the sleeper modeling might be over-simplified if the condition of track supporting layers is of particular concern. Instead, beams or solids are alternatives to improve the modeling of sleepers. This highlights the value of using SA based on extremes to validate models.

Track ballast provides a supporting layer to the sleepers, and it becomes reasonable that the ballast properties ($\rho_{ba}, E_{ba}, h_{ba}, \alpha_{ba}$, and c_{ba}) play a relatively important role in the sleeper response A_s. However, in Figure 5.12, the sensitivity indices of those factors with regard to A_s are more spread out with the individual importance measure no greater than 5%. This distributed effect may be attributed to the ballast model [65], [72] adopted in the dynamic analysis (see Section 5.4.2), where the five ballast properties were aggregated into two equivalent parameters (m_{ba} and k_{ba}) that were actually used in the dynamic simulation. It may become clear when measuring the total effect of the ballast properties, which account for 9.10% of the variability in maximal A_s.

The maximum deck deflection shows the highest sensitivity to the carbody mass (m_c) and bridge properties (E, I, μ and c), which jointly accounts for almost all of the response variability (about 99.99%). Specifically, factors E and I contribute about 52.51% of the variability, which is physically reasonable, given the importance of EI in beam deflection. Therefore, it is important to carefully define these factors before performing dynamic analysis if the goal is to evaluate the bridge response.

Note that train speed is not considered a variable in the current SA. It is fixed in its given value since the present work aims at evaluating the impact of input factors with high variability on the extreme dynamic response, while the operational speed of the trains is normally predefined with less uncertainty. The Kelvin foundation adopted in the TTB model implies a limitation of the train speed to subcritical velocities. For speeds that are above the critical velocity [87], dynamic amplification effects of the response can be observed. This can cause significant changes in the extreme response surface and therefore affects the result of SA.

5.5.3. IMPACT OF DESIGN THRESHOLDS

The threshold-based sensitivity analysis (Section 5.2.3) is applied in the TTB case. Figure 5.13 shows the threshold-based sleeper acceleration ($\tilde{A}_s^k$) for each fixed factor, considering all input factors for demonstration purposes. Figure 5.14 presents the threshold-based response for the remaining criteria, specifically focusing on the most significant input factors to maintain brevity.

In Figure 5.13, the thresholds are defined referring to [88], where sleeper accelerations for the plain line are observed in the range of $\pm 2g$ ($g = 9.8\text{m/s}^2$, gravitational acceleration). Accordingly, the thresholds of $1.5g, 2g$, and $2.5g$ m/s^2 are determined for Scenarios 1, 2, and 3, respectively.

When $A_s = 1.5g$ m/s^2, it can be seen that only m_c and k_p cause changes in the response curves (the solid red lines in the corresponding subplots), implying their critical contributions to the exceedance of the current limit state. Guaranteeing either $m_c < 27500$ kg or $k_p < 140$ MN/m, the unfavorable system condition can be avoided. Note that a combination of other insignificant factors may also influence the system reliability state. The current SA aims at quantifying the univariate effect of input factors on the model response near the limit states, while the joint effect will require further analysis.

In Scenario 2 ($A_s = 2g$ m/s^2), more input factors influence whether the reliable state of the system can be attained, i.e., m_c, k_p, m_s, E_{ba}, h_{ba}, α_{ba}, and c_{ba}. Compared with Scenario 1, m_c or k_p can take the value in a larger range without taking the risk of 'failure', since the design threshold is less restrictive. However, when $A_s = 2.5g$ m/s^2, the maximal response is always below the threshold. In this case, none of the input factors are considered important, as none of the combinations of the factors taking values in their definition range will cause 'failure'. From this, it is highlighted that the chosen threshold has a substantial impact on the contribution of input factors (whether critical or not) to the variability of the extreme model response.

Figure 5.14 shows the threshold-based response of the remaining criteria, focusing on the factors that contribute to substantial changes in these curves. In Figure 5.14 (a), the limits for W_z are defined according to rating scales of the Sperling index and the response variability space from the ED (see Figure 5.10 (a)). $W_z < 1$ classifies the riding

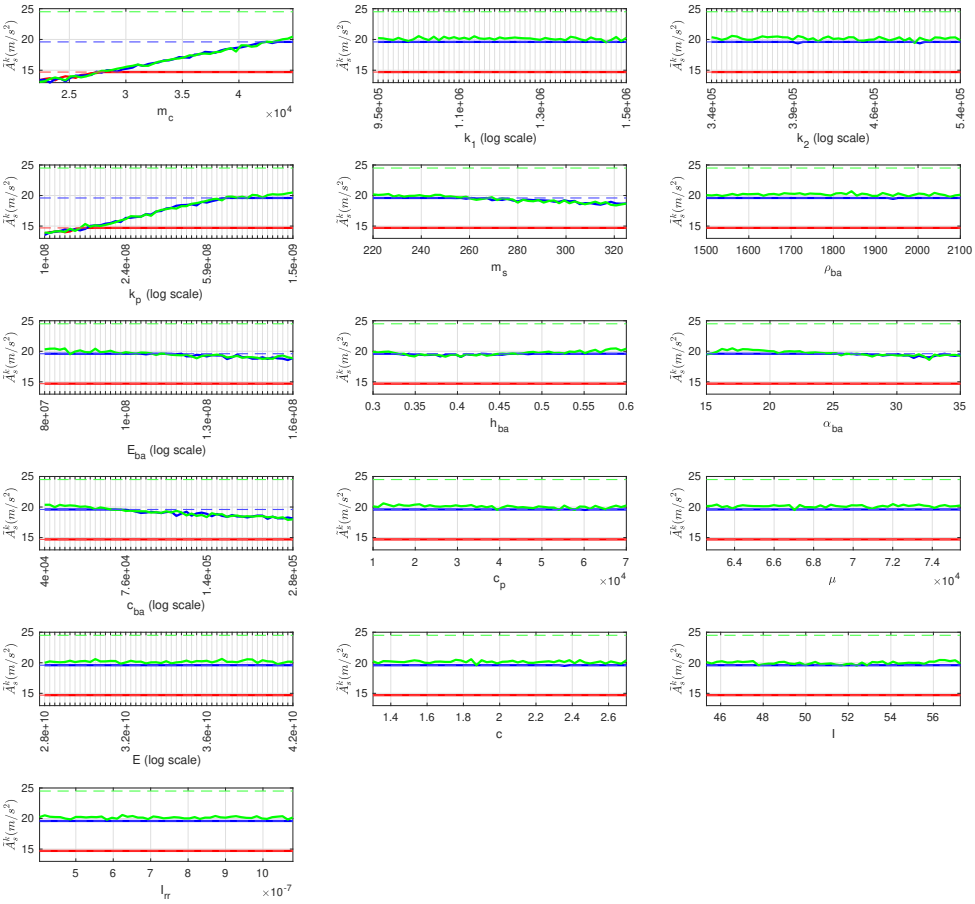

Figure 5.13: Threshold-based response surface of A_s for each fixed factor. Scenario 1 (red lines) with threshold $A_s = 1.5g$ m/s^2 (dotted) and threshold-based response $\tilde{A}_s{}^1$ (solid); Scenario 2 (blue lines) with threshold $A_s = 2g$ m/s^2 (dotted) and threshold-based response $\tilde{A}_s{}^2$ (solid); and Scenario 3 (green lines) with threshold $A_s = 2.5g$ m/s^2 (dotted) and threshold-based response $\tilde{A}_s{}^3$ (solid). In each scenario, the region above or below the corresponding threshold indicates the failure or non-failure domain, respectively. This also applies to Figure 5.14.

condition as 'Just noticeable'; $1 < W_z < 2$ means 'Clearly noticeable'. $W_z = 1.56$ defined for Scenario 2 refers to the nominal value of the limit in [89]. In Figure 5.14 (b), the limits for U_r are defined referring to [89], where a range between 0.00102 and 0.0025 m is considered for the vertical rail deflection. Accordingly, the thresholds of 0.0015, 0.0020, and 0.0025 m are determined for Scenarios 1, 2, and 3, respectively.

In Figure 5.14 (c), the thresholds for Scenarios 1 and 2 are determined based on the response variability space obtained from the ED (see Figure 5.10 (d)). The threshold defined for Scenario 3 refers to EN 1990 [90], which specifies the maximum permissible vertical deflection for railway bridges based on factors such as the number of spans, span length, train speed, and bridge configuration. For the current bridge case, $U_d = 0.058$ m

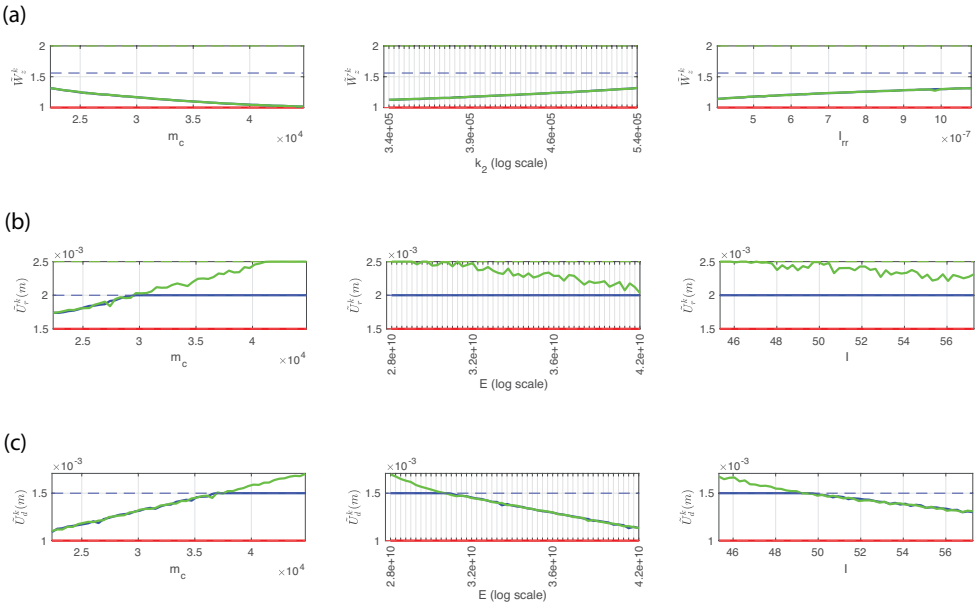

Figure 5.14: Threshold-based response surface of (a) W_z, (b) U_r, and (c) U_d with regard to the most significant factors.

is derived and defined as the decision threshold for Scenario 3.

Firstly, it is observed from Figure 5.14 that in Scenario 1 (the red lines), the response surfaces consistently align with the corresponding thresholds for all the criteria. This indicates that the design thresholds of these criteria can always be attained by a critical combination of the variables, where none of the variables play an impeding role.

In Scenarios 2 and 3, deviations occur in the considered criteria. As shown in Figure 5.14 (a), the response surfaces in Scenario 2 (the blue solid lines) coincide with those simulated in Scenario 3 (the green solid lines), and the maximum response values consistently fall below the threshold for Scenario 2 ($W_z = 1.56$). This implies that varying any factor within its range will not alter the system reliability state, as there are sufficient margins between the actual extreme model response and the given threshold.

However, Figure 5.14 (b) demonstrates the significant influence of m_c in determining the system reliability states ($U_r = 0.0020\,\text{m}$ and $U_r = 0.0025\,\text{m}$). This observation aligns with the sensitivity index shown in Figure 5.12, where m_c alone accounts for about 65.13% variability of the maximum U_r. Besides, in Scenario 3, despite slight fluctuations caused by numerical simulations, it is observed that the response curves of factors E and I intersect with the threshold $U_r = 0.0025\,\text{m}$, indicating their relevance in reaching the corresponding reliability state.

In terms of the criterion U_d, the dominant factors indicated in Figure 5.14 (c) align with those shown in Figure 5.14 (b). Specifically, in Scenario 2, varying the values in the range $m_c < 37000\,\text{kg}$, $E > 3.12 \times 10^{10}\,\text{N/m}^2$, or $I > 50\,\text{m}^4$ can guarantee the maximum U_d below this threshold. In Scenario 3, there are always sufficient margins between the

maximum U_d and given threshold $U_d = 0.058$ m, suggesting that all the factors can be considered as non-critical.

5.6. CONCLUSIONS

Sensitivity analysis provides an understanding of how a given model responds to changes in its input factors, which allows for prioritizing the factors, reducing the model dimensionality, calibrating the model, and evaluating the consistency between the model input and output. Different types of methods have been developed for SA, where the Sobol method is possibly the most prevalent form of global sensitivity method in engineering applications. However, this method focuses on the average behavior of the systems, which may not be sufficient for safety-critical structures where often limit states and the corresponding extreme values of the response are of particular concern.

This chapter focuses on the extreme response that a structure can potentially experience. The 'extreme' can be interpreted as either the maximum (or minimum) [3] or the response near a limit state (i.e., the threshold-based response). When a threshold is large enough, the threshold-based response implies the maximum value of the response, and vice versa. Therefore, the threshold-based response can be considered a generalized extreme response. For this, an efficient method is proposed to evaluate the sensitivity of the extreme model response to input factors, which extends the work [3] by incorporating the limit state design considerations in the formulation of extreme problems, i.e., the so-called threshold-based sensitivity method.

Further, since the sensitivity method is optimization-based, which requires iteratively maximizing (or minimizing) the model to search for the extreme model response, the computational cost involved in SA may become unaffordable in dealing with computationally intensive models. To address this issue, the random model output is represented by PCE, in which the original expensive model is replaced by an approximation that is faster to evaluate.

The method is applied to a dynamic train-track-bridge system. The sensitivity of the maximum dynamic response is assessed while accounting for the existence of track irregularities and uncertainty in the factors of the train, track, and bridge. Four design criteria are defined to capture the overall level of vibration of the system. The result suggests the high relevance of loading magnitude to all the criteria considered. However, it also highlights that the relevance of input parameters can vary significantly across different design criteria. This observation underscores the importance of exploring the design space near limit states before formulating RBDO problems for high-dimensional models. By providing valuable insights into this particular space, the present method allows for selecting appropriate design variables that align with the intended objectives, where the dimensionality of the space is significantly reduced, thereby informing the definition of reliability-based optimization problems.

REFERENCES

[1] Y. Shang, M. Nogal, R. Teixeira, and A. R. M. Wolfert, "Extreme-oriented sensitivity analysis using sparse polynomial chaos expansion. Application to train–track–bridge systems", *Reliability Engineering & System Safety*, vol. 243, p. 109 818, 2024.

[2] A. Saltelli and K. Chan, *Sensitivity analysis*. Wiley Series in Probability and Mathematical Statistics. Chichester etc.: Wiley., 2000.

[3] M. Nogal and A. Nogal, "Sensitivity method for extreme-based engineering problems", *Reliability Engineering & System Safety*, vol. 216, p. 107 997, 2021.

[4] J. Yang, "Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis", *Environmental Modelling & Software*, vol. 26, no. 4, pp. 444–457, 2011.

[5] I. Soboĺ, "Sensitivity estimates for nonlinear mathematical models", *Math. Model. Comput. Exp.*, vol. 1, p. 407, 1993.

[6] A. Saltelli, S. Tarantola, and K.-S. Chan, "A quantitative model-independent method for global sensitivity analysis of model output", *Technometrics*, vol. 41, no. 1, pp. 39–56, 1999.

[7] E. Borgonovo, "A new uncertainty importance measure", *Reliability Engineering & System Safety*, vol. 92, no. 6, pp. 771–784, 2007.

[8] T. Homma and A. Saltelli, "Importance measures in global sensitivity analysis of nonlinear models", *Reliability Engineering & System Safety*, vol. 52, no. 1, pp. 1–17, 1996.

[9] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates", *Mathematics and computers in simulation*, vol. 55, no. 1-3, pp. 271–280, 2001.

[10] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index", *Computer physics communications*, vol. 181, no. 2, pp. 259–270, 2010.

[11] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur, "Statistical inference for Sobol pick-freeze Monte Carlo method", *Statistics*, vol. 50, no. 4, pp. 881–902, 2016.

[12] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions", *Reliability engineering & system safety*, vol. 93, no. 7, pp. 964–979, 2008.

[13] G. Blatman and B. Sudret, "Efficient computation of global sensitivity indices using sparse polynomial chaos expansions", *Reliability Engineering & System Safety*, vol. 95, no. 11, pp. 1216–1229, 2010.

[14] K. Konakli and B. Sudret, "Global sensitivity analysis using low-rank tensor approximations", *Reliability Engineering & System Safety*, vol. 156, pp. 64–83, 2016.

[15] M. Ehre, I. Papaioannou, and D. Straub, "Global sensitivity analysis in high dimensions with PLS-PCE", *Reliability Engineering & System Safety*, vol. 198, p. 106 861, 2020.

[16] D. Xiu and G. E. Karniadakis, "The Wiener–Askey polynomial chaos for stochastic differential equations", *SIAM journal on scientific computing*, vol. 24, no. 2, pp. 619–644, 2002.

[17] G. Deman, K. Konakli, B. Sudret, J. Kerrou, P. Perrochet, and H. Benabderrahmane, "Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model", *Reliability Engineering & System Safety*, vol. 147, pp. 156–169, 2016.

[18] M. Thapa and S. Missoum, "Uncertainty quantification and global sensitivity analysis of composite wind turbine blades", *Reliability Engineering & System Safety*, vol. 222, p. 108 354, 2022.

[19] K. Cheng and Z. Lu, "Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression", *Computers & Structures*, vol. 194, pp. 86–96, 2018.

[20] T. A. Mara and W. E. Becker, "Polynomial chaos expansion for sensitivity analysis of model output with dependent inputs", *Reliability Engineering & System Safety*, vol. 214, p. 107 795, 2021.

[21] R. Schöbi and B. Sudret, "Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions", *Reliability Engineering & System Safety*, vol. 187, pp. 129–141, 2019.

[22] B. Sudret and C. V. Mai, "Computing derivative-based global sensitivity measures using polynomial chaos expansions", *Reliability Engineering & System Safety*, vol. 134, pp. 241–250, 2015.

[23] V. Chabridon, "Reliability-oriented sensitivity analysis under probabilistic model uncertainty–Application to aerospace systems", Ph.D. dissertation, Université Clermont Auvergne [2017-2020], 2018.

[24] I. Papaioannou, K. Breitung, and D. Straub, "Reliability sensitivity estimation with sequential importance sampling", *Structural Safety*, vol. 75, pp. 24–34, 2018.

[25] H. Jensen, F. Mayorga, and M. Valdebenito, "Reliability sensitivity estimation of nonlinear structural systems under stochastic excitation: A simulation-based approach", *Computer Methods in Applied Mechanics and Engineering*, vol. 289, pp. 1–23, 2015.

[26] G. Sarazin, J. Morio, A. Lagnoux, M. Balesdent, and L. Brevault, "Reliability-oriented sensitivity analysis in presence of data-driven epistemic uncertainty", *Reliability Engineering & System Safety*, vol. 215, p. 107 733, 2021.

[27] I. Papaioannou and D. Straub, "Variance-based reliability sensitivity analysis and the FORM α-factors", *Reliability Engineering & System Safety*, vol. 210, p. 107 496, 2021.

[28] M. Ehre, I. Papaioannou, and D. Straub, "A framework for global reliability sensitivity analysis in the presence of multi-uncertainty", *Reliability Engineering & System Safety*, vol. 195, p. 106 726, 2020.

[29] V. Maume-Deschamps and I. Niang, "Estimation of quantile oriented sensitivity indices", *Statistics & Probability Letters*, vol. 134, pp. 122–127, 2018.

[30] A. B. Owen, J. Dick, and S. Chen, "Higher order Sobol' indices", *Information and Inference: A Journal of the IMA*, vol. 3, no. 1, pp. 59–81, 2014.

[31] G. Geraci, P. M. Congedo, R. Abgrall, and G. Iaccarino, "High-order statistics in global sensitivity analysis: Decomposition and model reduction", *Computer Methods in Applied Mechanics and Engineering*, vol. 301, pp. 80–115, 2016.

[32] C. Y. Wong, P. Seshadri, and G. Parks, "Extremum sensitivity analysis with polynomial Monte Carlo filtering", *Reliability Engineering & System Safety*, vol. 212, p. 107 609, 2021.

[33] Z.-W. Li, Y.-L. Zhou, X.-Z. Liu, and M. A. Wahab, "Service reliability assessment of ballastless track in high speed railway via improved response surface method", *Reliability Engineering & System Safety*, vol. 234, p. 109 180, 2023.

[34] P. Ni, J. Li, H. Hao, and H. Zhou, "Reliability based design optimization of bridges considering bridge-vehicle interaction by Kriging surrogate model", *Engineering Structures*, vol. 246, p. 112 989, 2021.

[35] J. Rocha, A. Henriques, and R. Calçada, "Probabilistic safety assessment of a short span high-speed railway bridge", *Engineering Structures*, vol. 71, pp. 99–111, 2014.

[36] H.-P. Wan and Y.-Q. Ni, "An efficient approach for dynamic global sensitivity analysis of stochastic train-track-bridge system", *Mechanical Systems and Signal Processing*, vol. 117, pp. 843–861, 2019.

[37] A. Saltelli, K. Aleksankina, W. Becker, *et al.*, "Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices", *Environmental modelling & software*, vol. 114, pp. 29–39, 2019.

[38] W. Wang, Y. Zhang, and H. Ouyang, "Modeling uncertainties of vehicle-track coupled dynamic systems", *Mechanics Based Design of Structures and Machines*, vol. 49, no. 7, pp. 947–968, 2021.

[39] X. Chen, X. Deng, and L. Xu, "A three-dimensional dynamic model for railway vehicle–track interactions", *Journal of Computational and Nonlinear Dynamics*, vol. 13, no. 7, 2018.

[40] M. Oregui, A. Núnez, R. Dollevoet, and Z. Li, "Sensitivity analysis of railpad parameters on vertical railway track dynamics", *Journal of Engineering Mechanics*, vol. 143, no. 5, p. 04 017 011, 2017.

[41] V. Kumar, V. Rastogi, and P. M. Pathak, "Dynamic analysis of vehicle–track interaction due to wheel flat using bond graph", *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, vol. 232, no. 3, pp. 398–412, 2018.

[42] J. Varandas, A. Paixão, E. Fortunato, B. Z. Coelho, and P. Hölscher, "Long-term deformation of railway tracks considering train-track interaction and non-linear resilient behaviour of aggregates–a 3D FEM implementation", *Computers and Geotechnics*, vol. 126, p. 103 712, 2020.

[43] L. Xu, W. Zhai, and J. Gao, "Global sensitivity analysis for vehicle–track interactions: Special attention on track irregularities", *Journal of Computational and Nonlinear Dynamics*, vol. 13, no. 3, 2018.

[44] X. Liu, Z. Lai, P. Xiang, and Y. Chen, "Sensitivity and dynamic analysis of train-bridge coupled system with multiple random factors", *Engineering Structures*, vol. 221, p. 111 083, 2020.

[45] L. Xin, X. Li, Y. Zhu, and M. Liu, "Uncertainty and sensitivity analysis for train-ballasted track–bridge system", *Vehicle System Dynamics*, vol. 58, no. 3, pp. 453–471, 2020.

[46] N. Anwar and F. A. Najam, *Structural cross sections: analysis and design*. Butterworth-Heinemann, 2016.

[47] Y. Peng, J. Zhou, C. Fan, *et al.*, "A review of passenger ride comfort in railway: Assessment and improvement method", *Transportation Safety and Environment*, vol. 4, no. 2, tdac016, 2022.

[48] G. Blatman and B. Sudret, "Adaptive sparse polynomial chaos expansion based on least angle regression", *Journal of computational Physics*, vol. 230, no. 6, pp. 2345–2367, 2011.

[49] R. G. Ghanem and P. D. Spanos, *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.

[50] B. Sudret, "Uncertainty propagation and sensitivity analysis in mechanical models–Contributions to structural reliability and stochastic spectral methods", *Habilitationa diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, France*, vol. 147, p. 53, 2007.

[51] F. Liu, Y. Zhao, L. Li, and J. Xiao, "Random vibration analysis of an uncertain vehicle-track coupled system based on a polynomial dimensional decomposition", *International Journal of Rail Transportation*, pp. 1–20, 2023.

[52] D. C. Montgomery, *Design and analysis of experiments*. John wiley & sons, 2017.

[53] G. Blatman, "Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis", Ph.D. dissertation, Clermont-Ferrand 2, 2009.

[54] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone, "Approximation of quantities of interest in stochastic PDEs by the random discrete L^2 projection on polynomial spaces", *SIAM Journal on Scientific Computing*, vol. 35, no. 3, A1440–A1460, 2013.

[55] S. H. Lee and B. M. Kwak, "Response surface augmented moment method for efficient reliability analysis", *Structural safety*, vol. 28, no. 3, pp. 261–272, 2006.

[56] A. Ferreira and N. Fantuzzi, *MATLAB Codes for Finite Element Analysis: Solids and Structures* (Solid Mechanics and Its Applications). Springer International Publishing, 2020, ISBN: 9783030479527.

[57] MATLAB, *version 9.11.0 (R2021b)*. Natick, Massachusetts, United States: The MathWorks Inc., 2021.

[58] R. Schobi, B. Sudret, and J. Wiart, "Polynomial-chaos-based Kriging", *International Journal for Uncertainty Quantification*, vol. 5, no. 2, 2015.

[59] D. Cantero, "TTB-2D: Train–Track–Bridge interaction simulation tool for Matlab", *SoftwareX*, vol. 20, p. 101 253, 2022.

[60] A. C. Lamprea-Pineda, D. P. Connolly, and M. F. Hussein, "Beams on elastic foundations–A review of railway applications and solutions", *Transportation Geotechnics*, vol. 33, p. 100 696, 2022.

[61] "Design requirements for rail-bridges based on interaction phenomena between train, track and bridge", UIC International Union of Railways, Paris, leaflet, Jun. 2009.

[62] D. Cantero, T. Arvidsson, E. J. OBrien, and R. Karoumi, "Train–track–bridge modelling and review of parameters", *Structure and Infrastructure Engineering*, vol. 12, no. 9, pp. 1051–1064, 2016.

[63] P. Lou, "Finite element analysis for train–track–bridge interaction system", *Archive of Applied Mechanics*, vol. 77, no. 10, pp. 707–728, 2007.

[64] C. Shen, R. Dollevoet, and Z. Li, "Fast and robust identification of railway track stiffness from simple field measurement", *Mechanical Systems and Signal Processing*, vol. 152, p. 107 431, 2021.

[65] W. Zhai, *Vehicle–Track Coupled Dynamics: Theory and Applications*. Springer Nature Singapore, 2019, ISBN: 9789813292833.

[66] D. Wisniewski, "Safety formats for the assessment of concrete bridges", *Guimarães: University of Minho*, 2007.

[67] A. Doménech, P. Museros, and M. Martínez-Rodrigo, "Influence of the vehicle model on the prediction of the maximum bending response of simply-supported bridges under high-speed railway traffic", *Engineering Structures*, vol. 72, pp. 123–139, 2014.

[68] S. Iwnick, "Manchester benchmarks for rail vehicle simulation", *Vehicle System Dynamics*, vol. 30, no. 3-4, pp. 295–313, 1998.

[69] Y. Shang, M. Nogal, R. Teixeira, and A. R. M. Wolfert, "Optimal design of rail level crossings and associated transition zones using adaptive surrogate-assisted optimization", *Engineering Structures*, vol. 282, p. 115 740, 2023.

[70] J. N. Varandas, P. Hölscher, and M. A. Silva, "Dynamic behaviour of railway tracks on transitions zones", *Computers & structures*, vol. 89, no. 13-14, pp. 1468–1479, 2011.

[71] P. Salcher, H. Pradlwarter, and C. Adam, "Reliability assessment of railway bridges subjected to high-speed trains considering the effects of seasonal temperature changes", *Engineering Structures*, vol. 126, pp. 712–724, 2016.

[72] W. Zhai, K. Wang, and J. Lin, "Modelling and experiment of railway ballast vibrations", *Journal of sound and vibration*, vol. 270, no. 4-5, pp. 673–683, 2004.

[73] A. F. Rodrigues and Z. Dimitrovová, "Optimization of the behaviour of high-speed railway tracks using a genetic algorithm approach", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 229, no. 4, pp. 345–363, 2015.

[74] H. Xia, N. Zhang, and G. De Roeck, "Dynamic analysis of high speed railway bridge under articulated trains", *Computers & structures*, vol. 81, no. 26-27, pp. 2467–2478, 2003.

[75] "Railway applications - Track - Track geometry quality Part 5: Geometric quality levels - Plain line, switches and crossings", European Committee for Standardization, Brussels, Standard, Aug. 2017.

[76] G. Chen and W. Zhai, "Numerical simulation of the stochastic process of railway track irregularities", *Journal of Southwest Jiaotong University*, vol. 34, no. 2, pp. 138–142, 1999.

[77] C. Deng, J. Zhou, D. Thompson, D. Gong, W. Sun, and Y. Sun, "Analysis of the consistency of the sperling index for rail vehicles based on different algorithms", *Vehicle System Dynamics*, vol. 59, no. 2, pp. 313–330, 2021.

[78] E. Sperling and C. Betzhold, "Beitrag zur beurteilung des fahrkomforts in schienenfahrzeugen (contribution to the assessment of ride comfort in rail vehicles)", *Glaser's Annals*, vol. 80, pp. 314–320, 1956.

[79] Y. Jiang, B. K. Chen, and C. Thompson, "A comparison study of ride comfort indices between Sperling's method and EN 12299", *International journal of rail transportation*, vol. 7, no. 4, pp. 279–296, 2019.

[80] S. Marelli and B. Sudret, "UQLab: A framework for uncertainty quantification in Matlab", in *Vulnerability, uncertainty, and risk: quantification, mitigation, and management*, American Society of Civil Engineers, 2014, pp. 2554–2563.

[81] R. Teixeira, M. Nogal, and A. O'Connor, "Adaptive approaches in metamodel-based reliability analysis: A review", *Structural Safety*, vol. 89, p. 102 019, 2021.

[82] M. Hadigol and A. Doostan, "Least squares polynomial chaos expansion: A review of sampling strategies", *Computer Methods in Applied Mechanics and Engineering*, vol. 332, pp. 382–407, 2018.

[83] J. N. Fuhg, A. Fau, and U. Nackenhorst, "State-of-the-art and comparative review of adaptive sampling methods for kriging", *Archives of Computational Methods in Engineering*, vol. 28, pp. 2689–2747, 2021.

[84] N. Lüthen, S. Marelli, and B. Sudret, "Sparse polynomial chaos expansions: Literature survey and benchmark", *SIAM/ASA Journal on Uncertainty Quantification*, vol. 9, no. 2, pp. 593–649, 2021.

[85] M. Sol-Sánchez, F. Moreno-Navarro, and M. C. Rubio-Gámez, "The use of elastic elements in railway tracks: A state of the art review", *Construction and building materials*, vol. 75, pp. 293–305, 2015.

[86] B. Indraratna, M. B. Sajjad, T. Ngo, A. G. Correia, and R. Kelly, "Improved performance of ballasted tracks at transition zones: A review of experimental and modelling approaches", *Transportation Geotechnics*, vol. 21, p. 100 260, 2019.

[87] P. A. Costa, A. Colaço, R. Calçada, and A. S. Cardoso, "Critical speed of railway tracks. Detailed and simplified approaches", *Transportation Geotechnics*, vol. 2, pp. 30–46, 2015.

[88] M. D. Milošević, B. A. Pålsson, A. Nissen, J. C. Nielsen, and H. Johansson, "Reconstruction of sleeper displacements from measured accelerations for model-based condition monitoring of railway crossing panels", *Mechanical Systems and Signal Processing*, vol. 192, p. 110 225, 2023.

[89] L. Xu and W. Zhai, "A new model for temporal–spatial stochastic analysis of vehicle–track coupled systems", *Vehicle System Dynamics*, vol. 55, no. 3, pp. 427–448, 2017.

[90] "Eurocode - Basis of structural and geotechnical design", European Committee for Standardization, Brussels, Standard, Sep. 2021.

6

CONCLUSIONS AND FUTURE DIRECTIONS

All models are wrong,
but some are useful.

George E. P. Box

This research is concerned with the degradation problems that are frequently encountered at transition zones in railway tracks. This issue has been known for a long time, and the underlying degradation mechanism has been extensively studied in previous work. Typically, the investigations into these specific track areas involve (i) establishing field experiments to gain insight into the local track behavior, both transient and long-term, and (ii) developing mechanistic models to analyze the dynamic behavior of the track under moving vehicles.

The model development is often followed by assessing the efficacy of design countermeasures on the studied sites. This typically involves a parametric study to analyze how changes in the properties of track components, or the countermeasure itself, affect the track performance. The performance is assessed through vehicle-track responses, such as rail displacement, wheel-rail force, and energy dissipation within the substructure. In most cases, parameters of interest are tested for an arbitrary range of values, with a single or two variables studied at a time while the remaining variables are fixed. While this approach facilitates understanding of the track behavior and the effects of parameter variations, it may overlook the potential interactions between variables and their consequential impact on overall track responses. As a result, this straightforward approach may not necessarily lead to an optimal design, implying that the intended track performance could only see partial improvements.

Parametric optimization provides a structured approach to improving the performance of engineering systems, which is the focus of this research. Considering the typ-

ical model characteristics used for analyzing the dynamic behavior of the track in transition zones, the goal of this research lies in the development of dedicated modeling approaches to optimize the design of transition zones while addressing three main challenges. These include CH-1. improving the computational efficiency of the optimization process, CH-2. embedding multiple design aspects, and CH-3. dealing with complexities arising from uncertainties and high-dimensional cases.

This thesis delivers four main research outcomes, which are illustrated in Figure 1.3 (revisited below for readability) and detailed in Section 1.4. The subsequent section provides key findings and observations that emerged during the development process, which are organized according to their corresponding chapters for coherence and readability.

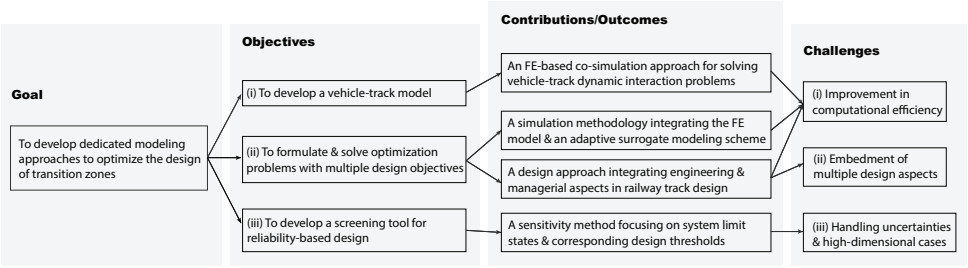

Figure 6.1: Mind map of the thesis.

6.1. MAIN FINDINGS

CONCLUSIONS FROM LITERATURE RESEARCH (CHAPTER 2)

The thesis commences by providing a comprehensive review of studies related to transition zones in the railway track. This review bridges two distinct fields of research, namely, dynamic analysis of track behavior under moving vehicles and modeling of track geometry degradation. The main findings are summarized in the following.

- These two research streams essentially adopt different modeling approaches and analyze at different scales. The dynamic behavior of the track at transition zones is mainly evaluated using mechanistic modeling, which focuses on micro-level investigations and relies on first principles to build a model. On the other hand, modeling of track geometry degradation is generally led by a data-driven approach. This approach operates on a macro scale, following a sequential process that starts with condition measurement to degradation modeling, and finally to maintenance decision support.

- Two aspects can be distinguished in the mechanistic models, including short-term performance evaluation and long-term settlement prediction. The short-term analysis aims for a deeper understanding of the physical mechanisms driving the degradation of the track in the vicinity of transition zones. The contributors that are frequently reported and analyzed in existing studies are variations in

mechanical properties (typically stiffness variations) and unloaded differential settlement.

A key observation to note is that both factors contribute to degradation at transition zones, but the primary cause is not consistently clear and can be case-specific. Some studies have evaluated the impact of both contributors and concluded that differential settlement tends to be more critical than stiffness variations, where an uneven settlement profile (either assumed or measured) is often the way of incorporating differential settlement into the track models. However, a comprehensive assessment of settlement should ideally include a representation of the viscoelastic-plastic behavior of the substructure layers. Without this, comparisons made by simply imposing uneven profiles may not yield valid conclusions.

- The long-term analysis requires an iterative procedure that couples the transient dynamic analysis with an empirical settlement equation. It is worth noting that this approach has certain limitations, where many of the empirical equations are extrapolated from specific sites and lack a foundation in constitutive properties. This may hamper the ability to generalize the results to other sites of interest and thus the accurate prediction of track settlement across different sites. Despite this, the method is advantageous for railway track design. It allows for the comparison of the effectiveness of various countermeasures aimed at reducing differential settlement at transition zones. This is because the empirical equations inherently account for long-term effects, which mainly depend on the number of loading cycles and or load magnitude.

- Data-driven models demonstrate better predictive capability on both spatial and temporal scales, but the accuracy depends on the quality and quantity of historical data. Unlike the mechanistic approach, which seeks to investigate the root cause of degradation within the train-track system and contribute to a fundamental understanding of the underlying mechanisms at transition zones, data-driven models extend the field of knowledge by mapping relationships between track degradation and exogenous factors. These factors include operational characteristics, maintenance history, and environmental conditions. Additionally, the data-driven approach can also establish statistical correlations with other types of track defects.

  From this, it can be concluded that the two categories of studies, while different in their methodologies, complement each other in providing insights into different aspects of track degradation. This points to the need for synchronized modeling approaches and measurements for refined diagnosis of track conditions and proper selection (including development) of maintenance techniques and design solutions, thereby effectively mitigating degradation issues at transition zones.

METHODOLOGY FOR IMPROVING MECHANICAL PERFORMANCE (CHAPTER 3)

Significant attention is given to addressing challenge CH-1 during the model development process. The implementation of the proposed methodology in level crossing design results in the following key findings.

- Design solutions obtained from the single-objective optimization problems show that force (F)-related design criteria in general suggest higher railpad stiffness than the energy (E)-related criteria. This can be explained by the fact that stiffer railpads contribute to reducing vibration from the wheel-rail contact, which is quantified by the F-related measures. However, this may lead to a higher effect of loads transmitted to underlayers, thereby causing vibration in sleepers and ballast, as quantified by the E-related measures. Further, the minimal distance (reaching the lower bound of the corresponding design variable) is generally preferred between the structural interface and the adjacent sleeper.

- The choice of design criteria (as objective functions) is not always obvious. The performance of two statistic metrics (i.e., the root mean square (rms) and maximum-to-minimum (max) value) is compared when solving single-objective optimization problems. Results show that the max metric generally performs better in terms of solution quality and sensitivity to parameter changes, making it more suitable for parametric optimization purposes. From this, it can be concluded that it is necessary to experiment with different objectives, which should be part of the design exploration process.

- Observations from solving single-objective problems reveal the conflicting nature of the objectives F_{max} and E_{max}. This suggests the need for a multi-objective formulation.

  In addition, a comparison of the optimized objective function values in single-objective and multi-objective cases was carried out. It was found that the F_{max} value in MOO surpassed the value obtained from the single-objective problem. This result indicates the complexity and nonlinearity of the current objective function, which is simulated from the FE model. The solver can potentially get trapped in regions with a local optimum, and in some instances, certain formulations of objective functions, like the one used in the current multi-objective case (Eq. 3.19), can yield better results.

- The previous remark sheds light on the selection of surrogate models in applications that involve vehicle-track dynamic simulations. This choice varies depending on how surrogate models participate in the optimization process. Essentially, there are two ways to leverage the computational efficiency of surrogate models in optimization. The first method involves substituting the original model with the surrogate throughout the entire optimization process, as shown in Steps A-D in Figure 1.2. The second method incorporates the surrogate as just one component of the overall optimization process, where the solver interrogates between the surrogate and the original model. To achieve this, an additional sub-step D' is introduced in the workflow, as seen in Figure 1.2.

  For the first method to be employed effectively, an accurate surrogate in the entire design space would be required. This becomes particularly important when dealing with functions that exhibit nonlinear behavior, such as the one found in current vehicle-track dynamic simulations. In these cases, surrogates that can better

capture nonlinearities are more suitable, such as Kriging for handling local nonlinearities, and PCE for global nonlinearities.

The proposed methodology in this chapter adopts RBF as the surrogate model. RBF has been chosen due to its relative simplicity and computational efficiency. There might be concerns regarding the accuracy of RBF within this particular context. However, it is worth noting that the second method (with step D') is considered here, and the primary role of RBF is to guide the search towards promising areas in the design space that may contain optimal values of the objective function. It is not intended to replace the original model. Therefore, this surrogate is not necessarily to be accurate in the entire design space.

- The solution obtained from the multi-objective case shows a notable improvement in the most relevant objectives (46.8% in F_{max} and 47.8% in E_{max}), compared to the reference design. This is achieved with a predefined maximum number of function evaluations, set at 400.

  The duration of an optimization process is based on the number of function evaluations. In surrogate-assisted optimization problems, termination depends more upon the computational budget. Strictly speaking, this stopping criterion does not guarantee convergence to an optimum, either local or global. However, this is not to suggest that the surrogate-assisted methods are not useful. In practical engineering problems, finding improved solutions is often desirable regardless of whether or not they are strictly optimal. Here, the design solution from the multi-objective case can be considered the best-known solution to support decision-making in track preliminary design. While 400 function evaluations have been assigned to obtain the solution, additional computational resources can be allocated to the proposed methodology to search for potentially better solutions.

- The proposed methodology is focused on the formulation of surrogate-based adaptive modeling in railway track design. It is demonstrated through a simplified vehicle-track model, with the potential to incorporate more complex models in terms of track configurations, structural elements, and track-soil coupling effects. While a single simulation run of these models may require more computational resources, the total computational costs of the entire optimization process, used for repetitive calls of function evaluations, are expected to result in significant savings. The reason is that the current methodology requires fewer function evaluations than the common optimization methods used in the field of railway engineering, such as genetic algorithms.

METHODOLOGY COVERING SOCIO-TECHNICAL RELEVANCE (CHAPTER 4)

This chapter focuses on addressing challenge CH-2, emphasizing the importance of integrating engineering and management practices in the design process of railway tracks. The following points highlight the main observations from the development and implementation of the proposed design approach in the level crossing case.

- Design solutions obtained from both single- and multi-objective problems indicate that softer railpads (in the ballast track) and strips (in the embedded rail sys-

tem) are recommended at the junction between the level crossing and transition zone.

However, the single- and multi-objective cases show a large deviation in terms of optimal solutions for sleeper parameters (x_s and x_n). This is because these variables directly impact the objective of construction costs (C_{cap}), which conflicts with the other objectives (E_{max} and A_{max}).

- Comparative analysis of optimized objective values across different single- and multi-objective cases reveal that the objective A_{max} shows limited sensitivity to parametric variations. This could potentially be attributed to the omission of track geometry irregularities within the simulation, despite these irregularities being a significant source of disturbance to the carbody accelerations (A_{max}). The consideration here is that if a prescribed track geometry were incorporated as input in the vehicle-track simulation, the design variables would be optimized to that specific irregularity profile. However, the goal here is to design a transition zone that mitigates the impact of stiffness variations, i.e., focusing on the initiation phase of degradation.

- Energy dissipation in the substructure (E_{max}) is used to quantify the susceptibility of a track design to the expected degradation. This assessment is relative, serving as a tool to compare various track design solutions under the effect of parametric variations. However, it is worth noting that for absolute quantification of long-term degradation, a 3D representation is required, one that accounts for the complex behavior of ballast and interaction between the track and soil. The proposed design method is generic and can be adapted to incorporate these specifics. However, modeling such details would significantly increase computational costs. Therefore, a tradeoff between solution quality and computational affordability would be necessary in the design exploration process.

- This chapter lays out a methodological basis to incorporate socio-technical perspectives with stakeholder preferences into the decision-making process of railway track design. It formulates optimization problems by synthesizing three design objectives that are relevant to infrastructure managers, train users, and maintenance service providers. The objective functions embody design intents that provide a basis for design improvement. As more data becomes available, design perspectives can be extended to include more specific economic implications. It also allows for the consideration of environmental effects, such as carbon emissions in production. Based on the design specifications and physical limitations, these aspects can be addressed either as individual objectives and constraints, or integrated into both functions.

EXTREME-ORIENTED SENSITIVITY ANALYSIS (CHAPTER 5)

This chapter focuses on addressing challenge CH-3 through the development of a sensitivity method. This method is designed to focus on the part of the output space that yields failure, making it closely aligned with the formulation of reliability-based design optimization. Some insights gained from the development and application of the method are discussed in the following.

- Literature review has pointed out an important aspect of sensitivity analysis in the context of vehicle-track models. The most commonly used approach for studying these models has been the local sensitivity analysis, often referred to as 'One At a Time' analysis, as mentioned in Chapter 5. However, this approach may not provide reliable results when the model contains nonlinear terms. These nonlinearities can be associated with factors such as wheel-rail contact, and the behavior of railpads and ballast. It is therefore necessary to carefully consider these factors while conducting sensitivity analyses in vehicle-track models.

- The threshold-based sensitivity method is applied to a train-track-bridge system. This is because the ultimate and serviceability limit states of railway bridges are clearly defined for design purposes in EN 1990 [1] and supporting literature. Four design criteria are defined to capture the overall level of vibration of the system. The sensitivity of the maximum dynamic response (of each design criterion) is assessed while accounting for the existence of track irregularities and uncertainty in the factors of the train, track, and bridge. The result suggests the high relevance of loading magnitude to all the criteria considered.

  However, the relevance of input parameters can vary significantly across different design criteria. This highlights the importance of exploring the design space near limit states before formulating reliability-based optimization problems for high-dimensional models.

- From the case study, it is highlighted that the chosen threshold has a substantial impact on the contribution of input factors to response variability. For this reason, the thresholds that determine the system limit states should be carefully defined in both sensitivity analysis and engineering optimization.

- The current sensitivity method can be used either as a standalone process or can be integrated with surrogate modeling to alleviate computational costs in sensitivity evaluation. Here, PCE is incorporated into the methodology. The random model output is represented through a PC approximation which substitutes the original model with an approximation that is faster to evaluate.

  While implementing the method into several examples, it was found that the performance of surrogate models depends on various factors, such as the shape and complexity of the function being approximated, the ED size, and the sampling strategies employed. The PCE approach allows us to efficiently capture the global stochastic behavior of the system and measure the effect of input factors on the output. The results from the truss example demonstrate that PCE performs particularly well with limited ED sizes, which is advantageous when dealing with expensive models.

  However, it should be noted that PCE is most effective for functions that can be well-approximated by global smooth polynomials. In two engineering problems, it was observed that PCE outperformed Kriging. Kriging is generally suitable for managing the local variability of the output. This suggests that the functions involved show global smoothness rather than (highly) local nonlinearity, thereby contributing to the better performance of PCE.

6.2. LIMITATIONS AND FUTURE DIRECTIONS

Based on the conclusions of this research, the following section outlines the limitations and provides recommendations for future work.

STIFFNESS VARIATIONS AND DIFFERENTIAL SETTLEMENTS

The stiffness variation and differential settlement are the main causes that lead to the degradation in transition zones. The methodologies proposed in this research are primarily designed to address the first cause - variations in track support stiffness. The goal is to smooth these variations along the railway line, thereby mitigating the impact they have on track degradation.

The differential settlement can be viewed from two perspectives: (i) the static change or (ii) the development process.

(i) The static change represents the resulting irregularities in track geometry from the settlement mechanism in the ballast and underlying layers. As presented in Eqs. 3.14 and 3.15, the current vehicle-track model can incorporate these irregularities in the simulation. Such irregularities can be inferred from track geometry measurements or simulated from predefined power spectral densities provided by relevant railway authorities.

However, it should be noted that the condition of track geometry is location specific. Given this concern, the transition zone has been designed without irregularities in track geometry, approximately representing the situation when a line is open to traffic. As previously mentioned, the underlying idea is that if a prescribed track geometry were incorporated as input in the vehicle-track simulation, the design variables would be optimized to that specific irregularity profile.

While the present focus lies on the development of the design methodology, future applications can adapt the track geometry condition to match the specific application context.

(ii) The development process introduces temporal aspects that include dynamic and static forces, as well as cyclic loading and their interactions. A comprehensive understanding of these factors is critical for accurate modeling of settlement mechanisms within the track structure, particularly in the ballast and soil layers. Furthermore, integrating models that account for track-soil interactions and the nonlinear behavior of the ballast represents a promising direction for future research.

On the other hand, incorporating such detailed models entails additional computational costs. This consideration is particularly relevant for the proposed surrogate-based methodology, which is expected to achieve more significant computational savings. This is because the duration of optimization is determined by the number of function evaluations, and the current methodology requires fewer function evaluations compared to the optimization methods commonly employed in railway engineering (e.g., GA).

An additional concern when introducing these detailed models is to clarify the design principle before implementing the proposed methodology. This principle

should specify whether the aim is to mitigate the impact caused by stiffness variations, differential settlements, or both. For instance, when tracks are laid on soft soils, it becomes necessary to address the effects of differential settlements within the design process. This requires a comprehensive analysis of both static and dynamic aspects. Such an approach not only aids in mitigating transient dynamic impacts but also ensures that the track design is robust against the long-term settlement effects.

SIMPLIFICATION OF THE VEHICLE-TRACK MODEL

The current vehicle-track model has certain limitations, the first of which ties back to the previous remark. The ballast and underlying substructure are modeled by the Kelvin-Voight elements, assuming a constant damping parameter. Although this representation can describe the elastic resistance and damping provided by the layers supporting the sleepers during train passage, more accurate quantification of energy dissipation in these layers requires a 3D representation that accounts for track-soil interactions and nonlinear material behavior.

It might also be advantageous to use a 3D model for the embedded rail system. This system provides homogeneous and continual rail support, which can be distinguished from conventional ballast tracks. This configuration allows for a more even distribution of wheel loads, which further influences the vehicle-track dynamics in all three dimensions. Moreover, implementing trackside measurements is considered necessary to gain more insight into the dynamic amplification phenomenon that arises from the structural discontinuity between the ballast track and the embedded rail system. This also holds relevance for field validation of the co-simulation approach and for assessing the design outcomes derived from case studies.

In the demonstration cases, it is assumed that the vehicle remains in contact with the rail, following the VTI dynamics simulation approach in [2]. For future research lines, it is advised to incorporate more refined representations of the wheel-rail contact mechanisms and to address potential instances of the pitch motion of bogies (and the upper carbody) in the dynamic simulations. These refinements are particularly relevant to the consequences of differential settlements at transition zones, as visualized in [3].

While implementing the Newmark-β integration scheme, a fixed time step of 0.001s was chosen for the simulations used in the optimization process. It can be beneficial to implement adaptive time-stepping techniques, where the time step is adjusted dynamically based on the system behavior or design alternatives during the simulation (and optimization). This allows for finer resolution in regions where more detail is needed (e.g., high-frequency content) while optimizing both accuracy and computational efficiency.

The development of the vehicle-track model raises a concern regarding the benchmark studies to evaluate the models that simulate dynamic vehicle-track interactions at transition zones. Although many researchers are dedicated to developing increasingly complex numerical models in this field, there seems to be a lack of benchmark work that coordinates efforts and allows for comparison between different model formulations.

These models primarily focus on understanding the underlying mechanisms. However, as demonstrated by this research, they are also fundamental for other engineering

tasks, such as design optimization, sensitivity analysis, reliability analysis, and uncertainty quantification. It is therefore recommended to conduct relevant benchmark studies. Previous research on switches and crossings [4], [5] highlights the potential benefits of such analyses. The benchmark studies would not only further the development of vehicle-track models for transition zones but also benefit the broader research community and practical engineering applications.

VALIDATION OF DESIGN SOLUTIONS

A post-optimality study is necessary to validate and interpret the design outcomes from case studies. This can be done by evaluating the final design solution(s) in higher fidelity models or by field experiments. At the same time, the effect of the stopping criterion, the maximum number of function evaluations, on the quality of the solutions can be further investigated.

The optimization problems formulated in this research are based on information gathered from existing literature and system specifications (e.g., the embedded rail system). For future applications or extensions of the proposed methods to specific transition zones, data collection efforts are directed to geotechnical information (e.g., the presence of soft soils under the tracks), specifications of track components, and cost information (both for construction and maintenance; if such data is not available, references from similar projects can also be useful). This data collection process would require collaboration among researchers, infrastructure managers, system providers, and maintenance contractors.

In addition, this research applies the min-max approach to accommodate stakeholder preferences in railway design problems. This approach, a classical form of a priori methods, is chosen to lay a methodological basis for integrating socio-technical perspectives (reflected in multi-stakeholder preferences) into railway track design. This approach focuses on the 'individual utility' or the single objective causing the largest deviation from the ideal point (in this context, the preference score of 100). Each stakeholder is treated individually, resulting in a design solution where all stakeholders are equally distant from the score of 100, thus achieving a compromise solution.

As a way forward, the current solution method can be extended with other preference handling methods, such as the IMAP method in [6] that focuses on the 'group utility'. This method aggregates preferences, interlinking all stakeholders and maximizing their group-wise preferences in the decision-making process. It is worth noting that the design outcomes may vary when using different preference handling schemes. This necessitates presenting those outcomes (i.e., the design alternatives) back to the stakeholders for validation and selection of the final design that best meets their requirements and is also technically feasible.

SENSITIVITY ANALYSIS AND SURROGATE MODELING

The case studies of sensitivity analysis use Latin Hypercube sampling with a fixed ED to provide a global description of the input space. Instead of sampling the ED at once, it is possible to apply adaptive or sequential sampling techniques to refine sampling in specific regions of interest. This approach allows for a more efficient allocation of computational resources, balancing the exploration and exploitation of the input space while constructing the surrogate model. Note that in this case, the sampling is more focused

on achieving accuracy in the proximity of specific regions, depending on the degree of exploitation, rather than aiming for a globally accurate surrogate model throughout the entire domain, which is more exploration-based. This choice of sampling methods can depend on function complexities. For functions with highly local nonlinearity, where the extremes are usually local phenomena, a balance of exploration and exploitation of the input space would be necessary.

The current sensitivity method is generic and can be extended to accommodate different surrogate models. By using PCE, the focus is limited to functions that can be well-approximated by global smooth polynomials. Future work would be required to explore alternative approaches to handle functions with highly local non-smoothness or nonlinearity. For example, more involved versions of PCE, such as Stochastic Spectral Embedding, can be explored to fit the PCE in subdomains and capture local nonlinearities. Another extension of the method involves conducting a benchmark study to investigate the connection among function properties (dimensionality and complexity), types of surrogate models (and solvers), the ED size, and sampling schemes.

In addition to this, the present method examines the univariate effect of input factors on the extreme or threshold-based response, where the components of the input vector are independent. A potential avenue for future work could be to incorporate the dependence between input factors into the sensitivity analysis.

REFERENCES

[1] "Eurocode - Basis of structural and geotechnical design", European Committee for Standardization, Brussels, Standard, Sep. 2021.

[2] P. Lou, "Finite element analysis for train–track–bridge interaction system", *Archive of Applied Mechanics*, vol. 77, no. 10, pp. 707–728, 2007.

[3] H. Wang and V. Markine, "Dynamic behaviour of the track in transitions zones considering the differential settlement", *Journal of Sound and Vibration*, vol. 459, p. 114 863, 2019.

[4] B. A. Pålsson, R. Ambur, M. Sebès, *et al.*, "A comparison of track model formulations for simulation of dynamic vehicle–track interaction in switches and crossings", *Vehicle System Dynamics*, vol. 61, no. 3, pp. 698–724, 2023.

[5] Y. Bezin and B. A. Pålsson, "Multibody simulation benchmark for dynamic vehicle-track interaction in switches and crossings: Modelling description and simulation tasks", *Vehicle System Dynamics*, vol. 61, no. 3, pp. 644–659, 2023.

[6] A. R. M. Wolfert, *Open Design Systems*. IOS Press, 2023.

ACKNOWLEDGEMENTS

It has been a challenging and transformative journey, which has not only shaped me as a researcher but also as an individual. The ups and downs, the lessons and blessings, have all contributed to this enriching experience that I deeply appreciate. As I bring my five-year-long PhD journey to an end, it is time to reflect on these years and express my gratitude towards those who have accompanied me in this marathon.

I would like to express my deepest appreciation to my supervisors, Rogier Wolfert and Maria Nogal, for their constant support, guidance, and encouragement throughout the process. They gave me full freedom to explore the topic, while always being available for any queries or discussions I needed. Rogier, thank you for bringing me on board. I am deeply grateful to you for guiding and challenging me to grow at every critical point of this trajectory. A particular memory that stands out is our discussion during the Covid-19 pandemic, which steered me into a new direction and reshaped my research path. Maria, thank you for sharing your knowledge and insights and supporting me to shape my academic growth. I am deeply grateful for the immeasurable contributions you made to my development. Your enthusiasm and upbeat character inspired me and made this journey a truly enjoyable experience.

I would also like to express my gratitude to the independent members of my doctoral committee: Prof. Lóri Tavasszy, Prof. Eugene OBrien, Dr. Alfredo Nunez Vicencio, Dr. Daniel Cantero, and Dr. Amy de Man. Thank you for your time, dedication, and invaluable insights. It is an honor to present my work for your discerning evaluation.

My sincerest thanks go to my mentors, Martine van den Boomen and Amy de Man. Martine, your unconditional support and insightful advice have profoundly shaped my perspective on life. Your commitment to education and genuine care for students truly make you a role model. Amy, I would like to thank you again for offering me the internship opportunity back in 2018 and providing me with system data to make this research possible. I am deeply grateful for all the insightful discussions we had along this journey.

I also had the great pleasure of working with my co-authors, Rui Teixeira and Haoyu Wang. Rui, I deeply appreciate your guidance in the field of metamodeling and your invaluable support to help me go through difficult times. Haoyu, I extend my appreciation for your support during the initial phase of my PhD journey, our insightful discussions, and your constructive feedback.

I must also thank my friends, Yuanyuan Pan and Erica Arango. Yuanyuan, we have known each other for more than a decade. Despite being situated in different corners of the world, we have always maintained a close connection. Erica, thank you for always being there to encourage me and lessen my stress. I also thank all the friends who accompanied me along this journey, Shi Xu, Yifei Ren, Le Zhu, Lizzy Zhou, Tianxiang Wang, Dawei Fu, Mingyang Zhang, Siqi Zhao, Xiuli Wang, Hongjin Du, Xiaoling Zhang, Yin Xie, and Xibin Qiao.

Furthermore, I would like to thank my colleagues in the IDM section. My PhD and postdoc peers Dmitry Zhilyaev, Zhaowen Liu, Yara Kharoubi, Yirang Lim, Susana Toboso Chavero, Yan Liu, Xinyu Liu, Dominika Teigiserová, Quirien Reijtenbagh, and Sana Lateef, it has always been my pleasure to exchange ideas and share moments with you. A special appreciation goes to Sandra Schuchmann-Hagman, our secretary. Thank you very much for always being supportive and kind.

My deepest gratitude goes to my husband and my family. Pei, we have been together throughout our entire twenties. You are my friend and teammate. Without your understanding and support, I would not have reached this milestone. Mom and Dad, our physical distance over the past eight years has been challenging. However, your unconditional love, support, respect, and belief in my capabilities have shaped my perspective and trajectory in life. I dedicate this thesis to you all.

Yue Shang
Delft, December 2023

CURRICULUM VITÆ

Yue SHANG

02-02-1994 Born in Xi'an, China.

EDUCATION

2012–2016 B.Eng. in Highway & Bridge Engineering
Chang'an University, Xi'an

2016–2018 M.Sc. in Construction Management & Engineering
Delft University of Technology, Delft
Thesis: Uncertainty Modelling in Life Cycle Costing Analysis for Rail Level Crossing System
Supervisors: Dr. M. van den Boomen
Dr. M.T.J. Spaan
Prof. dr. ir. A.R.M. Wolfert

2019–2023 Ph.D. research
Delft University of Technology, Delft
Thesis: Design Optimization for Railway Transition Zones
Promotors: Prof. dr. ir. A.R.M. Wolfert
Dr. ir. M. Nogal

RELATED WORKING EXPERIENCE

2018–2018 Graduate intern at edilon)(sedra B.V., Haarlem

2021–2023 Teaching assistant at Delft University of Technology
- Engineering Systems Optimization 2021/22 Q4; 2022/23 Q4
- Managing Uncertainty and Data 2022/23 Q1
- Modelling, Uncertainty, and Data Analysis for Engineers 2023/24 Q2

2022–2023 M.Sc. thesis supervisor at Delft University of Technology

- Sub-components' lead time optimization in precast concrete house-building industry (2022)

- Construction schedule optimisation: Optimisation of BIM-based, component-level construction schedule for building structural and MEP systems considering parallel working zones (2022)

- Evaluation of monopile logistical processes through discrete-event simulation and sensitivity analysis techniques (2023)

LIST OF PUBLICATIONS

JOURNAL PUBLICATIONS

7. **Y. Shang**, M. Nogal, R. Teixeira, A.R.M. Wolfert. *Extreme-oriented sensitivity analysis using sparse polynomial chaos expansion. Application to train-track-bridge systems*, Reliability Engineering & System Safety, 243, 109818 (2024).

6. **Y. Shang**, M. Nogal, R. Teixeira, A.R.M. Wolfert. *Optimal design of rail level crossings and associated transition zones using adaptive surrogate-assisted optimization*, Engineering Structures, 282, 115740 (2023).

5. **Y. Shang**, M. Nogal, H. Wang, A.R.M. Wolfert. *Systems thinking approach for improving maintenance management of discrete rail assets: a review and future perspectives*, Structure and Infrastructure Engineering, **19**(2), 197-215 (2021).

4. Y. Pan, **Y. Shang**, G. Liu, Y. Xie, C. Zhang, Y. Zhao. *Cost-effectiveness evaluation of pavement maintenance treatments using multiple regression and life-cycle cost analysis*, Construction and Building Materials, 292, 123461 (2021).

3. Y. Pan, T. Yang, **Y. Shang**, M. Lin, G. Liu, Y. Xie, Y. Zhao. *Field and laboratory evaluations of maintenance treatments for semi-rigid base asphalt pavement*, Construction and Building Materials, 258, 119726 (2020).

2. M. van den Boomen, M.T.J. Spaan, **Y. Shang**, A.R.M. Wolfert. *Infrastructure maintenance and replacement optimization under multiple uncertainties and managerial flexibility*, Construction management and economics, **38**(1), 91-107 (2020).

1. **Y. Shang**, M. van den Boomen, A.P. de Man, A.R.M. Wolfert. *Reliability-based life cycle costing analysis for embedded rails in level crossings*, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, **234**(8), 821-833 (2019).

PEER-REVIEWED CONFERENCE PROCEEDINGS

2. **Y. Shang**, R. Binnekamp, A.R.M. Wolfert. *Multi-stakeholder service life design for rail level crossings*, in Life-Cycle of Structures and Infrastructure Systems (pp. 949-956). CRC Press (2023).

1. **Y. Shang**, M. Nogal, A.R.M. Wolfert. *A co-simulation solution for vehicle-track interaction dynamics problems*, in The Fifth International Conference on Railway Technology: Research, Development and Maintenance (pp. 31-15). Civil-Comp Press (2022).