DISSERTATION

Benchmarking Cellular Wireless Connectivity in Railway Scenarios

---

The dissertation is submitted for the degree of

Doktorin der Technischen Wissenschaften

by

Dipl.-Ing. Sonja Tripković

Matriculation Number: 01227403

Institute of Telecommunications

Faculty of Electrical Engineering and Information Technology

Technische Universität Wien

Advisor: Univ. Prof. Markus Rupp

Co-advisor: Dr. Philipp Svoboda

Vienna, Austria

May 2025

Erklärung zur Verfassung der Arbeit

Dipl.-Ing. Sonja Tripković

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe. Ich erkläre weiter, dass ich generative KI-Tools im Rahmen dieser Dissertation ausschließlich zu Korrekturzwecken verwendet habe, um sprachliche und grammatikalische Anpassungen vorzunehmen, ohne den ursprünglichen Inhalt zu verändern.

Wien, 5. Mai 2025

Sonja Tripković

Abstract

The increasing demand for reliable and high-speed wireless connectivity in railway transportation has driven research into optimizing mobile network performance in train environments. This dissertation presents a comprehensive benchmarking methodology for evaluating cellular network performance on moving trains, addressing the unique challenges posed by high mobility, vehicle penetration loss, and trackside infrastructure limitations.

A structured approach is developed for data collection, processing, and analysis, utilizing controlled measurement campaigns on various train types. Network key performance indicators are examined to assess the impact of train design, network infrastructure, and environmental factors on connectivity.

Furthermore, the study compares different onboard connectivity enhancement solutions, such as repeater systems and frequency-selective surface window coatings, analyzing their effectiveness in mitigating signal attenuation. Advanced statistical methods and geospatial modeling techniques are applied to improve measurement strategies and ensure accurate evaluations. Additionally, this research integrates the digital twin concept by modeling railway environments, cellular network deployments and train properties in a virtual framework. This approach enables a comprehensive evaluation and optimization of mobile connectivity onboard trains.

The findings contribute to a better understanding of mobile network performance in railway scenarios and provide insights for optimizing future deployments. The methodologies developed in this thesis can serve as a foundation for railway operators and mobile network providers to enhance passenger connectivity and operational efficiency.

Kurzfassung

Die zunehmende Nachfrage nach Konnektivität im Schienenverkehr hat die Forschung im Bereich zuverlässiger und schneller Mobilfunknetze in diesem Kontext maßgeblich vorangetrieben. In der vorliegenden Arbeit wird eine umfassende und skalierbare Benchmarking-Methode präsentiert, die dazu dient, die Leistungsfähigkeit zellularer Netzwerke in fahrenden Zügen zu evaluieren. Ein besonderes Augenmerk liegt dabei auf den spezifischen Herausforderungen, die durch hohe Mobilität, Fahrzeugdurchdringungsverluste (Vehicle Penetration Loss, VPL) sowie die Einschränkungen der streckenseitigen Infrastruktur entstehen.

Zu diesem Zweck wird ein strukturierter Ansatz zur Datensammlung, -verarbeitung und -analyse entwickelt, der auf kontrollierten Messkampagnen verschiedener Zugtypen basiert. Im Rahmen der Untersuchung werden wesentliche Netzwerkschlüsselkennzahlen (Key Performance Indicators, KPIs) systematisch analysiert, um den Einfluss von Zugdesign, Netzwerkinfrastruktur sowie umgebungsbedingten Faktoren auf die Konnektivität zu bewerten.

Darüber hinaus vergleicht die Studie verschiedene Lösungen zur Verbesserung der Konnektivität an Bord, wie Repeater-Systeme und frequenzselektive Oberflächenbeschichtungen (Frequency Selective Surfaces, FSS) an Fenstern, und analysiert deren Wirksamkeit bei der Minderung von Signalabschwächung. Fortschrittliche statistische Methoden und geospatiale Modellierungstechniken werden angewendet, um die Messstrategien zu verbessern und genaue Bewertungen sicherzustellen. Zusätzlich integriert diese Dissertation das Konzept des Digitalen Zwillings (Digital Twin, DT), bei dem Schienenumgebungen, zellulare Netzwerke sowie spezifische Zugmerkmale in einem virtuellen Rahmen modelliert werden. Dieser Ansatz gestattet eine umfassende Evaluierung und gezielte Optimierung der mobilen Konnektivität an Bord von Zügen.

Die gewonnenen Erkenntnisse tragen zu einem vertieften Verständnis der Mobilfunknetzleistung im Schienenverkehr bei und liefern wichtige Einblicke zur Optimierung zukünftiger Netzbereitstellungen. Die in diesem Rahmen entwickelten Methoden stellen eine Grundlage für Bahnunternehmen und Mobilfunkanbieter dar, um sowohl die Passagierkonnektivität als auch die betriebliche Effizienz nachhaltig zu verbessern.

Acknowledgments

I want to thank my advisor, Prof. Markus Rupp, for recognizing my potential and offering me this opportunity. His invaluable guidance, insightful suggestions, and exceptional responsiveness have been instrumental throughout my research. I am endlessly grateful to my co-advisor, Philipp Svoboda, whose unwavering encouragement extended far beyond research. His enthusiasm, countless ideas, and tireless commitment — whether in drive-testing, brainstorming, or shaping research directions — have been truly inspiring. I also extend my sincere thanks to Prof. Ke Guan and Assoc. Prof. Marek Galinski for reviewing my thesis.

This research would not have been possible without the generous support of several industry partners. I am grateful to ÖBB-Infrastruktur AG and ÖBB-Personenverkehr AG for facilitating access to the railway environment and supporting the organization of multiple onboard measurement campaigns across various train types and tracks. Their contributions were essential for conducting realistic field measurements under operational conditions. I also thank A1 Telekom Austria AG for providing anonymized crowdsourced network monitoring data and crucial base station information, which significantly enriched the analysis and modeling aspects of this work. Furthermore, I am thankful to Focus Infocom GmbH for organizing and conducting a dedicated measurement campaign and for granting access to the collected data. I sincerely appreciate the technical expertise, collaboration, and continued engagement of all colleagues from these organizations in addressing research questions of practical relevance.

I deeply appreciate my coauthors, especially Lukas Eller and Vaclav Raida, whose encouragement has been invaluable. Lukas, in particular, deserves special thanks for choosing to embark on and complete his PhD journey alongside me. I am also grateful to my colleagues, Martin Lerch, Gerfried Krainz and Daniel A. Maierhofer, for their continued support in organizing measurement campaigns and for providing the Nemo file format parser. Beyond research, I deeply appreciate the collaborative and welcoming environment at the Institute of Telecommunications. To my colleagues, thank you for the support, stimulating discussions, and many memorable lunch breaks and social events.

To my friends in Serbia, my longest and dearest friendships — thank you for always being there. Despite the distance, your support, transmitted over the wire, has remained a constant source of strength and joy. To those who have supported me along the way, in ways big and small, including past relationships that shaped my journey — thank you. To my friends in Vienna, you have been my support system, my second family. Whether we studied together or not, you shared in my daily struggles and joys and made Vienna feel like home.

I also want to thank my extended family for their love and encouragement over the years. A special thanks to my cousin Bogdan — for calling me the smartest grandchild, and showing me that studying abroad can be fun. To my dear nieces, Maša, Sofija, and Jana — you inspire me every day. I hope I have been a role model for you, showing that perseverance, curiosity, and dedication can take you far, I hope to continue making you proud.

To my love, Leonard — for believing in me when I couldn't, and standing by me when it counted most — thank you. If it weren't for this PhD journey, I might never have crossed paths with you, and for that, I'm forever grateful. Your love and support have made all the difference, and I cherish every day with you by my side.

And finally, I want to express gratitude to my parents, Sladjana and Miroslav, and my brother Uroš for their sacrifices — both emotional and financial — that made my studies abroad possible. Their unwavering love, support, and belief in me have been the foundation of everything I have achieved. I feel incredibly lucky to have grown up in a family where love and support were constant and unconditional. I hope they know how deeply I value this — and how much it has shaped me into the person I am today. No matter the distance, they have always been there, and without them, this journey would not have been possible. I love you dearly.

Contents

List of Acronyms

3GPP 3rd Generation Partnership Project
5G Fifth Generation
5G NR 5th Generation New Radio
6G Sixth Generation
AAF Amplify-and-Forward
AI Artificial Intelligence
AM Acknowledged Mode
ATO Automatic Train Operations
AoA Angle of Arrival
BPL Building Penetration Loss
BS Base Station
CEPT European Conference of Postal and Telecommunications Administrations
CLC Corine Land Cover
CL Cell Load
CQI Channel Quality Indicator
CRS Cell-specific Reference Signal
CP Cyclic Prefix
CDF Cumulative Distribution Function
DL Downlink
DLNP Deep Learning Network Planner
DT Digital Twin
DTM Digital Terrain Model
ECC Electronic Communications Committee
ECDF Empirical Cumulative Distribution Function
eNB Evolved Node B
ETCS European Train Control System
FSPL Free-Space Path Loss
FSS Frequency Selective Surface
FRMCS Future Railroad Mobile Communication System
GNSS Global Navigation Satellite System
GPS Global Positioning System
GP Gaussian Process

GPR Gaussian Process Regression
GSW Geodatenviewer der Stadtvermessung Wien
GU Geographical Unit
GSM-R Global System for Mobile Communications - Railway
HARQ Hybrid Automatic Repeat Request
ICIC Inter-Cell Interference Coordination
Indoor PL Indoor Path Loss
IoT Internet of Things
IP Internet Protocol
ISI Inter-Symbol Interference
ITU International Telecommunication Union
KPI Key Performance Indicator
KS Kolmogorov-Smirnov
LOS Line-of-Sight
LTE Long Term Evolution
LSF Large Scale Fading
MAC Medium Access Control
MAE Mean Absolute Error
MFCN Mobile/Fixed Communication Networks
MCS Modulation and Coding Scheme
MDT Minimization of Drive Test
MIMO Multiple-Input Multiple-Output
MNO Mobile Network Operator
NLOS Non-Line of Sight
OFDM Orthogonal Frequency-Division Multiplexing
OFDMA Orthogonal Frequency-Division Multiple Access
OSM Open Street Maps
PCell Primary Serving Cell
PCI Physical Cell Identifier
PCC Policy and Charging Control
PDCP Packet Data Convergence Protocol
PHY Physical layer
PL Path Loss
PRB Physical Resource Block
QAM Quadrature Amplitude Modulation
QoE Quality of Experience
QoS Quality of Service
RE Resource Element

RLC Radio Link Control
RMR Railway Mobile Radio
RMSE Root Mean Squared Error
RS Reference Signal
RSRP Reference Signal Received Power
RSRQ Reference Signal Received Quality
RSSI Received Signal Strength Indicator
RTR Regulatory Authority for Broadcasting and Telecommunications
RT Ray Tracing
RRC Radio Resource Control
SCell Secondary Cell
SD Standard Deviation
SDK Software Development Kit
SF Shadow Fading
SIM Subscriber Identity Module
SINR Signal-to-Interference-plus-Noise Ratio
SIR Signal-to-Interference Ratio
SRS Simple Random Sampling
SS Stratified Sampling
SSF Small-Scale Fading
T Throughput
TA Timing Advance
TDD Time Division Duplex
TTI Transmission Time Interval
UE User Equipment
UL Uplink
UM Unacknowledged Mode
UPS Uninterruptible Power Supply
URLLC Ultra-Reliable Low-Latency Communication
UV Ultraviolet
VPL Vehicle Penetration Loss
WLAN Wireless Local Area Network

Chapter 1

Introduction

The demand for reliable, high-speed wireless connectivity in railway transportation has surged in recent years, driven by growing passenger expectations and the increasing integration of digital services. Trains play a crucial role in modern transportation infrastructure, providing efficient and sustainable mobility across regions. Today's passengers expect seamless mobile connectivity not only for communication but also for work, streaming, and entertainment during their journeys. Simultaneously, railway operators rely on robust wireless networks for safety-critical applications, real-time infrastructure communications, and remote diagnostics. As a result, ensuring stable and high-quality mobile connectivity inside trains has become a priority for both railway operators and mobile network providers.

Delivering consistent wireless performance along rail tracks remains a significant challenge. Trains travel through diverse geographical environments with varying network coverage, making uninterrupted connectivity difficult to maintain. To address this, mobile network operators are continuously upgrading infrastructure, including Base Station (BS) deployments along railway corridors, to enhance coverage and capacity.

Global System for Mobile Communications - Railway (GSM-R) has long been the standard for providing circuit-switched voice and packet-switched data services in railway connectivity [1]. Currently, 75% of Österreichische Bundesbahnen (ÖBB) - Infrastruktur AG's railway network in Austria is covered by GSM-R, though it is set to be gradually phased out by 2030 [2]. This transition highlights the increasing importance of advanced wireless connectivity in railway environments, leading to the development of the Future Railroad Mobile Communication System (FRMCS) [3]. Designed to fully digitalize railway operations, support higher levels of Automatic Train Operations (ATO), and leverage Fifth Generation (5G) technology, FRMCS is set to play a pivotal role in ensuring reliable and efficient wireless communication for modern railway systems.

Furthermore, as outlined in International Telecommunication Union (ITU) Report M.2442 [4], radiocommunication systems in railways are increasingly essential for safety-critical applications such as real-time diagnostics, emergency communications, and control systems. As railways transition to next-generation communication systems, the demand for enhanced onboard wireless performance will continue to grow, driving the need for fair and efficient benchmarking and optimizing cellular connectivity in railway scenarios.

1.1 Challenges in Railway Connectivity

Ensuring reliable cellular network connectivity onboard trains presents challenges that distinguish it from typical deployments in urban or rural areas. Unlike stationary or pedestrian environments, trains require specialized solutions due to their high mobility, metal structures, and varying environments through which they travel.

One of the critical issues affecting wireless communication in trains is high Vehicle Penetration Loss (VPL) or train cabin attenuation[1]. VPL refers to the attenuation of wireless signals as they pass through the materials and structures of a vehicle, leading to significant degradation in signal quality. This phenomenon is particularly pronounced in railway environments, as trains have mostly metallic bodies and specialized window coatings designed for energy efficiency and passenger comfort. These coatings are designed to reduce heat dissipation, which significantly lowers heating and cooling costs, while also providing protection against Ultraviolet (UV) radiation. However, the very properties that make these materials beneficial also substantially hinder wireless signals, necessitating effective mitigation solutions.

Traditional cellular network deployments in cities and rural areas rely on fixed BSs and relatively predictable user mobility patterns. In contrast, trains traverse diverse environments with fluctuating coverage, including tunnels, remote locations, and densely urbanized areas. Therefore, the cellular infrastructure designed for trains must account for both the mobile nature of the vehicle and the confined environment within the train cabin. Furthermore, the varying passenger load introduces dynamic Cell Load (CL), significantly affecting network performance. In Austria, a typical intercity train can accommodate up to 400 passengers during peak times, while in some countries, this number can be tripled. Every Mobile Network Operator (MNO) must consider this high passenger density when deploying infrastructure along railway corridors, as each connected user shares the available bandwidth and capacity within a sector. This shared nature of network resources can lead to congestion, affecting the quality of service for all users on board.

Given that trains travel through diverse environments with varying network coverage, it is essential to analyze these use cases separately due to the distinct propagation conditions they present. The propagation of wireless signals can be significantly affected by the environment, with urban areas often experiencing multipath propagation due to buildings, while rural areas might have less obstruction but suffer from longer distances between BSs. Tunnels, on the other hand, can create complete signal loss, necessitating specialized solutions for maintaining connectivity. By examining each environment individually - urban, rural, tunnels, and stations - we can better understand the specific challenges and develop targeted solutions for optimizing wireless communication in railroad scenarios. This approach allows for tailored strategies that consider unique propagation characteristics of each scenario, ensuring more reliable connectivity for passengers and operational efficiency for train operators.

[1] These two terms are used interchangeably throughout this thesis.

1.2 Motivation for Research

Investigating cellular network connectivity on board trains presents a distinct set of limitations compared to typical deployments in urban, rural, or highway environments. The fundamental difference lies in the operational and structural constraints imposed by trains, which make traditional network performance measurement approaches, such as drive tests, significantly harder to implement.

Unlike drive tests in cars, where continuous measurements can be collected relatively easily and cost-effectively, conducting tests inside trains is a more complex and resource-intensive process. Additionally, drive tests benefit from widespread accessibility - anyone can purchase a car and freely conduct measurements along highways or urban streets. In contrast, trains represent highly controlled environments with restricted access, making independent or frequent measurement campaigns impractical. The use of dedicated rooftop antennas for network measurements, which is typical in car-based tests, is not feasible on trains. Instead, measurements must rely on existing onboard infrastructure, which is primarily designed to provide internet access to passengers rather than facilitate controlled network evaluations. This reliance on shared infrastructure introduces logistical constraints, as measurements must be conducted when the train is empty to avoid disrupting passenger services. However, opportunities to access an empty train dedicated solely to experimentation are rare and often limited to one or two test runs, making data collection a one-time opportunity with little room for error.

Moreover, the unidirectional nature of train routes makes repeating measurements in the same location exceedingly difficult. Unlike cars, trains cannot simply reverse direction to revisit specific areas of interest, limiting the ability to capture repeated data and conduct comprehensive comparisons across different sections of track. These constraints highlight the need for efficient methodologies that allow for meaningful comparisons between different train configurations, even when measurements cannot be conducted under identical conditions.

Despite advancements in cellular network evaluation, a standardized approach for benchmarking connectivity on trains remains absent. Existing methods, such as crowd-sourced data collection and simulations, are informative, but often fail to capture the full complexity of real-world conditions onboard moving trains. These limitations underscore the need for dedicated measurement campaigns to systematically assess different connectivity solutions.

The primary motivation for this research is to provide actionable guidance for both network operators and railway companies seeking to enhance onboard connectivity. For network operators, understanding connectivity challenges along train routes can inform decisions on infrastructure improvements, such as optimizing cell tower placements or modifying network configurations. For railway operators, understanding how train design, seating arrangements, and window materials impact connectivity can guide decisions on fleet upgrades and new train designs to improve passenger experience.

Train operators, in particular, are eager to evaluate how different wagon designs perform in terms of mobile connectivity before making significant capital investments. Additionally, understanding the impact of cabin loss - the attenuation of signals caused by train materials such as walls, windows, and coatings - is crucial for accurately

assessing onboard connectivity. Since testing all potential train configurations on the same tracks is impractical, it is essential to develop a benchmarking methodology that allows for fair and reliable comparisons across diverse environments. This is especially relevant given that train characteristics - such as window coatings, materials, and onboard antenna placements - play a critical role in determining connectivity quality.

Recent advancements in onboard connectivity solutions, such as passive frequency-selective window coatings, have introduced new challenges in evaluating their performance across different train types and routes. Older solutions, such as active repeaters, offer internal signal enhancement but involve higher maintenance requirements and present operational limitations. In contrast, emerging passive solutions modify train window coatings to improve signal penetration without ongoing maintenance. These different solutions directly influence the behavior of User Equipment (UE) once inside the cabin. Depending on the solution in place, the UE may switch between relying on external macro networks or onboard signal relays, affecting its performance, power consumption, and handover behavior. This dynamic is crucial for operators to understand, as it has implications not only for coverage but also for network planning and resource allocation. However, the effectiveness of these technologies varies depending on external signal conditions, highlighting the need for reliable methods to compare their performance across different scenarios.

We seek to address these challenges by developing systematic methodology for benchmarking cellular network performance onboard trains. Furthermore, we apply the proposed methodology to compare existing connectivity solutions in the Austrian railway fleet, enabling a data-driven evaluation of different approaches. To further support this research, we investigate strategies for collecting these measurements, ensuring that data is gathered efficiently and in sufficient quantity to make informed decisions about where further measurements are needed. By leveraging sampling theory, we aim to determine the necessary amount and distribution of data points, allowing for targeted and resource-efficient measurement campaigns. By exploring how train characteristics - such as window materials, antenna placements, and route conditions - affect connectivity, this study aims to provide a foundation for improving onboard mobile service. The insights gained will support railway operators in selecting appropriate connectivity solutions and assist MNOs in optimizing their network deployments along railway routes, contributing to a broader understanding of network performance in complex and dynamic transportation environments.

1.3 Digital Twin in Mobile Communications for Railways

Digital Twin (DT) technology enables the creation of virtual replicas of railway environments, offering a powerful tool for simulating and analyzing cellular network performance across diverse real-world conditions. This reduces the need for extensive physical testing and supports data-driven optimization strategies in rail communications. Studies have demonstrated the effectiveness of model-based DT platforms in integrating railway infrastructure components - such as trackside deployments, trains, and environmental

factors - into unified systems that support monitoring, diagnostics, and predictive maintenance [5–7].

While DTs are still in the early stages of adoption within the railway industry, a comprehensive survey has highlighted the potential of DTs in real-time operational applications [8]. The authors outline the foundational technologies and discuss potential applications, challenges, and the future directions for DT development in railway systems. This thesis builds on that perspective by applying the DT concept to better understand network behavior in complex railway scenarios and to support the optimization of deployment strategies.

DTs have proven to be versatile beyond communications and operational monitoring. For example, a DT for railway station buildings has been developed to help manage maintenance and assess environmental impacts like carbon emissions [9]. This shows the broader applicability of DTs across different railway applications - from boosting network performance to supporting sustainability efforts in station management.

One of the primary advantages of DTs lies in their capacity to dynamically model interactions between trackside infrastructure, train cabins, and surrounding environments, enabling real-time understanding of how these factors impact cellular network performance. While DTs in railway networks have traditionally focused on infrastructure and operational efficiency, recent work has also explored the use of augmented DTs for safety-critical applications. An augmented DT has been introduced to assess derailment risk in real-time using machine learning and multibody dynamics in [10]. Recent developments also highlight the opportunity to extend DT capabilities by incorporating service models that describe user experience metrics such as throughput, latency, and reliability. This approach highlights the broader potential of DTs in optimizing various aspects of railway operations, including safety and operational efficiency, which could be relevant when integrating network performance monitoring. Ultimately, the objective of using DTs in this domain is to improve the design, testing, and maintenance of railway communication systems, achieving greater accuracy, resilience, and efficiency in network performance.

Fig. 1.1 illustrates our concept of the DT for railway connectivity, composed of three components: Environment and Rail Track DT, Cellular Network DT, and Train DT.

Environment and Rail Track DT

The environmental conditions along a train's route, including topography, vegetation, and urban infrastructure, significantly influence wireless connectivity. Furthermore, the characteristics of the railway track itself, such as curves, tunnels, noise barriers, or other physical obstacles, influence signal propagation.

A DT of the environment and track integrates topographical and environmental data to simulate the effects of these factors on signal quality. For instance, tunnels and mountainous regions are known to cause significant signal loss. The DT can also include dynamic environmental features like CL changes around residential areas or villages along the route, which impact available capacity and service quality. By replicating the physical environment in a DT, railway operators can simulate how the network will perform in these challenging conditions. This model ensures that potential solutions are

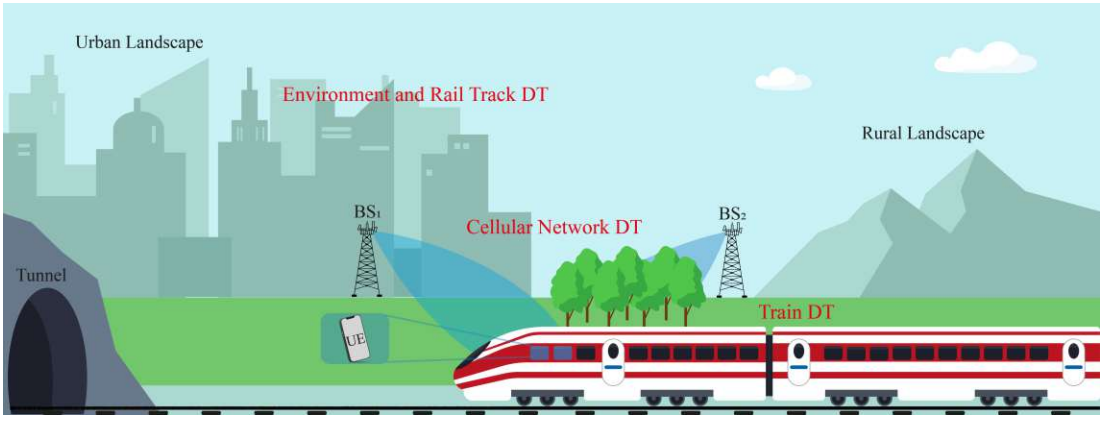


Figure 1.1: Components of DT for Railway Connectivity.

tested and optimized specifically for the railway environment, rather than applying generalized network configurations. The integration of Internet of Things (IoT) technology has been proposed as a key enabler for the development of DTs in the railway sector, supporting predictive maintenance and intelligent data acquisition for onboard systems [11].

Cellular Network DT

Cellular network trackside deployment refers to the cellular infrastructure, such as BSs signal towers, positioned along railway routes to provide consistent connectivity. This deployment is crucial for ensuring uninterrupted service, particularly in high-speed rail environments where trains must maintain continuous connections despite rapidly moving between cells.

A DT of the trackside deployment digitally replicates BS locations, antenna orientations, tilts, transmit powers and handover mechanisms. It allows network engineers to simulate the impact of changes in BS placement, antenna tilt, and network configuration. Additionally, real-time data from BSs can be incorporated into the DT to model the effects of terrain, weather, or obstacles such as tunnels on signal propagation. This virtual network model can also simulate how different traffic distributions, including the number of users per cell or temporary spikes in commuter density, affect quality of service. Through this virtual model, railway operators can evaluate the efficiency of current network setups and predict how future infrastructure modifications will influence network performance [7, 12].

Train DT

The train itself presents distinct challenges for mobile connectivity due to its physical properties, including its metal frame, window materials, and interior layout. These characteristics contribute to high VPL, which reduces the quality of the mobile signals inside the train.

A DT of the train would digitally replicate various parameters such as the train's body structure, window materials (e.g., coated or uncoated), and the configuration of

onboard antennas or repeaters. Simulating these parameters helps engineers understand how external signals are influenced by different train design choices. Furthermore, the DT could simulate the effect of different user densities inside the cabin - e.g., simulating peak vs. off-peak hours - on internal signal quality and traffic management. Moreover, train modifications - such as changing window coatings or repositioning antennas - can be tested virtually before implementing physical changes, saving time and resources. This DT allows engineers to experiment with different configurations to identify the optimal combination for reducing VPL and improving connectivity [7, 8].

Towards a Holistic DT for Railways

A comprehensive DT for railways would integrate the DTs of trackside deployments, train characteristics, and the surrounding environment. This holistic approach offers a complete understanding of how various elements interact to influence the onboard mobile experience. More importantly, this integration would allow virtual reconfiguration of the system - such as adjusting antenna placements, testing new onboard technologies, or reallocating network resources based on simulated commuter density or traffic demand. Such a model enables railway operators to simulate and evaluate network performance under different conditions and identify potential issues before they arise. It also would also facilitate iterative testing of "what-if" scenarios, supporting long-term planning and optimization for both service quality and operational efficiency.

This research aims to contribute to the development of such a holistic DT by exploring how trackside BS configurations, train design, and environmental factors interact to influence mobile connectivity. By developing innovative methodologies for benchmarking cellular network performance in railway environments, this work aims to support improvements in wireless communication systems for high-speed rail.

1.4 Objectives of the Research

This thesis provides solutions to key challenges and complexities associated with benchmarking cellular connectivity in high-speed railway environments. The following objectives guide the study:

- *Design and Validate a Scalable and Repeatable Benchmarking Methodology for Wireless Connectivity in Railway Environments*: A key objective is to create and validate a scalable framework for measuring and evaluating wireless performance in railway scenarios. Despite the inherent limitations of train-based measurements - such as differences in railway tracks, varying network deployments, fluctuations in passenger load and train speed, and environmental factors like terrain and weather - this benchmarking framework is designed to be repeatable, transparent, and adaptable across diverse contexts. By deliberately incorporating these features, we enable consistent evaluation of different connectivity solutions in trains, ensuring that assessments are meaningful, data-driven, and applicable to real-world deployments.

- *Assess the Influence of Trackside BS Deployment Strategies on Signal Strength and Connectivity Performance*: We design a framework for calculating the Angle of Arrival (AoA) of signals on moving trains under real-world conditions, deliberately omitting controlled laboratory measurements. As full-train lab setups are impractical due to the size and complexity of railway environments, the framework is developed for direct application across different train types and tracks. Using this framework, we assess how the design and placement of trackside infrastructure influence network performance. We focus on AoA analysis because, in newer onboard connectivity solutions such as modified windows, external signals primarily enter through the windows, and signal strength strongly depends on the AoA and the BS position relative to the track.

- *Evaluate the Role of Train Design in VPL and its Effect on Cellular Connectivity*: We examine how VPL impacts wireless communication inside trains, with a particular focus on window coatings, patterns, and design elements. By quantifying signal degradation due to VPL, we aim to identify strategies to mitigate these losses and improve connectivity. Understanding the influence of train features on cellular connectivity is essential for optimizing DT representations of trains and enhancing overall network performance.

- *Optimize Measurement Strategies for Enhanced Data Accuracy and Benchmarking Reliability*: We investigate various sampling strategies, including Stratified Sampling (SS) and its application to crowdsourced data analysis, to improve the granularity and accuracy of collected data and estimated Key Performance Indicator (KPI) of interest. By integrating optimized measurement strategies informed by sampling theory, this approach ensures that the collected data accurately reflects real-world conditions, ultimately contributing to more reliable benchmarking.

The following publications document the work we conducted towards these four objectives, respectively:

- **S. Tripkovic**, P. Svoboda, M. Rupp, *Benchmarking of Mobile Communications in High-Speed Scenarios: Active vs. Passive Modifications in High-Speed Trains* [13].

- **S. Tripkovic**, P. Svoboda, M. Rupp, *Measuring the Effects of AoA on Vehicle Penetration Loss in Cellular Networks* [14].

- **S. Tripkovic**, P. Svoboda, M. Rupp, *Enhancing Mobile Communication on Railways: Impact of Train Window Size and Coating* [15].

- **S. Tripkovic**, L. Eller, P. Svoboda, M. Rupp, *Unbiased Benchmarking in Mobile Networks: The Role of Sampling and Stratification* [16].

By achieving these objectives, this work makes significant contributions to the field of railway connectivity, supporting improvements in both passenger experience and operational efficiency in train systems.

1.5 Structure of the Thesis

This thesis is structured to systematically address the research objectives outlined above. The document is organized as follows:

- **Chapter 1: Introduction** - In this chapter, after discussing the challenges of cellular communication in high-speed railway environments, we introduced DTs as a promising tool for addressing these challenges and outlined its key components. While a fully integrated, holistic DT for railway connectivity is beyond the scope of this thesis, several elements - such as environment modeling, train characteristics, and network deployment parameters - are used to inform and contextualize the benchmarking approach developed in this work.
- **Chapter 2: Metrics, Measurements, and Methodologies in Network Optimization** - This chapter presents the measurement framework and data processing methodology for evaluating wireless connectivity in railway environments. It begins by introducing key performance metrics and then addresses the shortcomings of existing approaches, such as the limited accuracy of Ray Tracing (RT) simulations, thereby motivating the need for dedicated measurement campaigns. Key aspects include measurement design, equipment selection, and geospatial data integration to precisely align data with railway tracks. Techniques such as Global Positioning System (GPS) interpolation, data segmentation and stratification ensure ensure repeatability, transparency, and contextual relevance across different train configurations and network conditions. By applying this structured approach, the research enables reliable benchmarking and optimization of railway connectivity.
- **Chapter 3: Railway Measurement Datasets** - This chapter provides an overview of the datasets collected on board moving trains in Austrian railways and used throughout this thesis. Measurement campaigns were conducted on three distinct ÖBB train types - Talent, Railjet, and Nightjet - to assess the impact of various connectivity solutions and cabin configurations on mobile network performance. These datasets form the foundation for the subsequent analyses in this dissertation, enabling a comparative evaluation of different window coating modifications and signal enhancement strategies.
- **Chapter 4: Evaluation of Mobile Connectivity Solutions in Trains** - This chapter assesses two connectivity solutions - Amplify-and-Forward (AAF) repeaters and Frequency Selective Surface (FSS) window coatings - using real-world railway measurements. We evaluate their effectiveness under different train configurations and highlight practical trade-offs between active signal amplification and passive enhancement strategies. By leveraging rail, environment, and cellular datasets, we provide a comparative assessment of these solutions, highlighting their strengths and limitations. This analysis serves as a foundation for optimizing onboard connectivity strategies in high-speed rail networks.

- **Chapter 5: Impact of Train Characteristics on VPL** - This chapter investigates how various train characteristics impact VPL, focusing on factors such as AoA, window size and coating, and the position dependency of devices within the train. The findings from these measurements contribute to models that can be applied for building the train DT, and offer guidance for optimizing wireless connectivity in railway systems.
- **Chapter 6: Optimizing Measurement Strategies through Sampling** - This chapter explores the application of SS to improve measurement strategies in the context of big data and crowdsourcing. It introduces the theory behind stratification, proposes its application in determining minimal measurement requirements, and presents case studies involving simulated and Minimization of Drive Test (MDT)-based datasets. These methods help identify optimal data collection strategies while minimizing campaign effort.
- **Chapter 7: Future Directions** - This chapter outlines opportunities for further research, including the integration of crowdsourced data into railway benchmarking workflows and the use of the developed tools to support simulation calibration and coexistence studies for emerging technologies like FRMCS. It emphasizes the need for ongoing evaluation as networks and train technologies evolve.
- **Chapter 8: Conclusion** - This chapter summarizes key contributions to evaluating and optimizing cellular connectivity on trains through measurements, simulations, and scalable methodologies.

Chapter 2

Metrics, Measurements, and Methodologies in Network Optimization

In this chapter, we begin by examining the KPIs used to evaluate cellular network performance within railway environments. We discuss their relevance and application in capturing the spatiotemporal variability of network conditions. Subsequently, we explore the role of simulations in evaluating cellular network performance within railway environments, highlighting their strengths and limitations, particularly in comparison to real-world measurements. Subsequently, we introduce a proposed measurement methodology along with a post-processing framework designed to facilitate accurate and fair benchmarking of various train types, cabin configurations, track segments, and environmental conditions. This methodology ensures that network evaluations are rooted in real operational conditions, supporting meaningful analysis and comparison of connectivity solutions.

2.1 Cellular Network Performance: Technical Background and Propagation Effects

In railroad scenarios, evaluating network performance is particularly critical due to the unique challenges posed by high-speed mobility, frequent cell handovers, and variable user density. Trains carrying hundreds of passengers impose a substantial load on the serving BS, while continuous transitions between coverage areas introduce delays, increased signaling overhead, and potential connectivity disruptions [17].

2.1.1 LTE and 5G: The Role of OFDM in Performance Metrics

Railway environments in Austria currently rely on Long Term Evolution (LTE) networks, with future expansion toward 5G networks. Both LTE and 5G, which includes the 5th Generation New Radio (5G NR) standard, employ Orthogonal Frequency-Division Multiplexing (OFDM) as their primary modulation scheme, ensuring efficient spectrum utilization and robustness against multipath fading, which is particularly beneficial in high-mobility scenarios like train-based connectivity.

OFDM divides the available spectrum into multiple orthogonal subcarriers, minimizing Inter-Symbol Interference (ISI) and providing high spectral efficiency. In the Downlink (DL), LTE utilizes Orthogonal Frequency-Division Multiple Access (OFDMA), where the fundamental scheduling unit is the Physical Resource Block (PRB), consisting of 12 subcarriers over a 0.5 ms time slot. In 5G NR, a flexible subcarrier spacing system enables scalability across various deployment scenarios.

A critical enabler of LTE and 5G's OFDM-based transmission is the Cell-specific Reference Signal (CRS). The CRS is embedded within the OFDM time-frequency grid, providing a dedicated signal for channel estimation, mobility management, and computation of KPIs.

Following 3rd Generation Partnership Project (3GPP) TS 36.214 guidelines [18], the CRS enables UEs to:

- Measure Reference Signal Received Quality (RSRQ) to assess interference and congestion.
- Estimate Reference Signal Received Power (RSRP) for cell selection and handover decisions.
- Derive Signal-to-Interference-plus-Noise Ratio (SINR), which impacts link adaptation and scheduling.

In practical measurement campaigns, such as drive tests and scanner-based evaluations, the CRS serves as a stable reference signal that enables consistent estimation of KPIs like RSRP and RSRQ, making it fundamental to the assessment of LTE and early 5G networks.

2.1.2 Key Performance Indicators in Cellular Networks

Evaluating cellular network performance in railway environments requires analyzing several KPIs that reflect signal quality, interference, and network capacity.

1. **Received Signal Strength Indicator (RSSI)** - measure of total received power within the measurement bandwidth:

$$\text{RSSI (dBm)} = 10 \log_{10} \left(P_{\text{signal}} + P_{\text{interference}} + P_{\text{control}} + P_{\text{noise}} \right). \tag{2.1}$$

   It includes
   - power from the serving CRS,
   - power from other interfering cells,
   - power from control channels and data transmissions,
   - thermal noise and background interference [18].

   All power-related quantities P_{X} in this context are expressed in Milliwatts (mW). Since it therefore represents an absolute power level referenced to 1 mW, it is typically expressed in dBm.

2. **Reference Signal Received Power (RSRP)** - defined as the linear average of the power contributions of Resource Elements (REs)[1] carrying CRS within the measurement bandwidth:

$$\text{RSRP (dBm)} = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^{N} P_i \right), \tag{2.2}$$

where

- P_i is the received power of the i-th reference signal resource element (in mW),
- N is the number of REs used for RSRP computation.

Unlike RSSI, which includes power from all sources (e.g., interference and noise), RSRP focuses only on the power of reference signals, making it a more precise indicator of cell selection, handovers, and signal quality in LTE networks. Analogously to RSSI, it is typically expressed in dBm.

3. **Reference Signal Received Quality (RSRQ)** - defined as the ratio of RSRP to the RSSI, measured over the same bandwidth [18]:

$$\text{RSRQ (dB)} = 10 \log_{10} \left(\frac{N \cdot 10^{\text{RSRP (dBm)}/10}}{10^{\text{RSSI (dBm)}/10}} \right), \tag{2.3}$$

where

- N represents the number of PRBs over which RSSI is measured,
- RSSI is the total power received, including reference signals, control signals, interference, and thermal noise.

RSRQ is essential for assessing signal quality in loaded networks, particularly in scenarios where congestion, co-channel interference, or poor scheduling impact performance. As it represents the ratio between two power levels it is typically expressed in dB.

4. **Signal-to-Interference-plus-Noise Ratio (SINR)** - quantifies the strength of the desired signal relative to noise and interference, defined as:

$$\text{SINR (dB)} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{interference}} + P_{\text{noise}}} \right), \tag{2.4}$$

and is typically expressed in dB.

Unlike RSRP and RSSI, which focus primarily on signal strength, SINR provides a more comprehensive indication of signal quality by accounting for interference and noise. As such, it is a crucial metric for evaluating network performance, handover decisions, and overall user experience in LTE networks. SINR is a key parameter for channel quality assessment, influencing the Channel Quality Indicator (CQI)

[1] In LTE and 5G NR, a Resource Element (RE) is the smallest unit of time-frequency resource that can carry data or control information.

reported by the UE to the Evolved Node B (eNB). This feedback helps optimize link adaptation and resource scheduling, ensuring efficient utilization of network capacity.

5. **Throughput (T)** - refers to the actual data transmission rate achieved by a user [19].

   In 3GPP LTE networks, T is measured at multiple protocol layers, each representing different aspects of data transmission as defined in 3GPP specifications (TS 36.321 for Medium Access Control (MAC) [20], TS 36.322 for Radio Link Control (RLC) [21], TS 36.323 for Packet Data Convergence Protocol (PDCP) [22], and TS 36.331 for Radio Resource Control (RRC) [23]).

   At the Physical layer (PHY), T is determined by the number of PRBs allocated per Transmission Time Interval (TTI), the Modulation and Coding Scheme (MCS), and the Multiple-Input Multiple-Output (MIMO) configuration. The achievable PHY T is directly dependent on CQI reports, which guide the eNB scheduler in resource allocation. The theoretical peak T is specified in 3GPP TS 36.213 [24] and depends on factors such as carrier aggregation and higher-order MIMO.

   At the MAC layer, T is influenced by scheduling algorithms (Proportional Fair, Round Robin, Max C/I), Hybrid Automatic Repeat Request (HARQ) retransmissions, and Quality of Service (QoS) prioritization. While PHY T represents raw transmission capacity, MAC T reflects effective data delivery, accounting for error recovery and resource scheduling, as defined in 3GPP TS 36.321 [20].

   The RLC layer further impacts T through segmentation, reassembly, and retransmission mechanisms, using Acknowledged Mode (AM) for reliability or Unacknowledged Mode (UM) for low-latency applications. Overhead is introduced at the PDCP layer, which provides header compression, ciphering, and integrity protection, as defined in 3GPP TS 36.323 [22].

   At the Internet Protocol (IP) layer, T represents the actual end-user data rate, shaped by network congestion, backhaul limitations, and QoS policies (as per 3GPP TS 23.203 [25] on Policy and Charging Control (PCC)). While PRB allocation directly influences PHY and MAC T, factors such as HARQ retransmissions, protocol overhead, and scheduling efficiency determine the final user-perceived T. Effective LTE network optimization requires balancing these factors to maximize spectrum efficiency and Quality of Experience (QoE), following 3GPP guidelines for LTE and LTE-Advanced.

6. **Cell Load (CL)** - defined as the fraction of available PRBs that are occupied by active users in a given time interval:

$$L = \frac{N_{\text{used PRBs}}}{N_{\text{total PRBs}}}, \tag{2.5}$$

   where

   - $N_{\text{used PRBs}}$ is the number of PRBs allocated to active UEs,

- $N_{\text{total PRBs}}$ is the total PRBs available in the cell.

A CL of one indicates that all available PRBs are currently in use, meaning the cell is fully utilized at that moment. However, new users may still be scheduled depending on priority and traffic demand, although overall capacity for additional data transmission is limited [17].

In 3GPP LTE networks, CL has a significant impact on key T parameters across different protocol layers, influencing overall network efficiency and QoS. According to 3GPP TS 36.213, TS 36.321, and TS 36.331 [20, 23, 24], increasing CL leads to reduced PRB availability per user, directly affecting PHY layer T. As more users compete for resources, the eNB scheduler (TS 36.321 [20]) must dynamically allocate PRBs. Overall high CL in neighboring cells causes interference, often resulting in the selection of lower MCS due to reduced SINR.

At the MAC Layer, a high CL results in increased HARQ retransmissions, as interference and congestion lead to packet errors, reducing effective T. Additionally, scheduling algorithms (e.g., Proportional Fair, Round Robin, TS 36.321 [20]) must prioritize users based on CQI feedback (TS 36.213 [24]), leading to potential latency variations and reduced T for edge users.

At the RLC Layer, congestion-induced delays impact AM retransmissions, further lowering effective T and increasing latency, as specified in 3GPP TS 36.322 [21]. Similarly, at the PDCP layer, increased load results in additional processing overhead due to higher encryption and integrity protection demands (TS 36.323 [22]), affecting T efficiency.

At the IP Layer, a high CL exacerbates packet queuing delays and packet loss, reducing user-perceived T. Backhaul constraints also become critical under heavy load, as defined in 3GPP TS 23.203 [25], impacting QoS enforcement and limiting guaranteed bit rate traffic performance.

In summary, increased CL reduces PRB availability, lowers MCS selection, increases retransmissions at MAC/RLC layers, and degrades IP T due to congestion and queuing delays. Effective load balancing, Inter-Cell Interference Coordination (ICIC) (TS 36.423 [26]), and QoS-aware scheduling are essential for mitigating these effects in LTE and LTE-Advanced networks.

Metrics such as RSRP and SINR are key indicators of link quality and play a critical role in various radio resource management tasks, including mobility management, scheduling, and link adaptation. In the context of handovers they help determine the optimal time and target cell for transition to maintain service continuity. Additionally, information about the number of active UEs and CL - both in the serving and neighboring cells - supports analysis of traffic distribution, aiding in load balancing and resource optimization.

Spatiotemporal variability in network conditions along the railroad route - affected by terrain, interference, and fluctuating user activity - further underscores the importance of combining KPIs, such as RSRP, T and CL, to adaptively evaluate and maintain service quality. Passengers' increasing expectations for reliable, high-speed internet for

work and entertainment add to the necessity of accurate T estimation and proactive optimization.

2.2 Focus on Path Loss and Reference Signal Received Power in Railroad Scenarios

Our objective is to model network KPIs, denoted by $I(t, x, T, S)$, where t denotes time, x represents geographic location, T characterizes the train type, and S corresponds to the seating position within the train.

A comprehensive evaluation of network performance requires a focus on two fundamental indicators: Path Loss (PL) and signal strength. PL characterizes the attenuation of the signal as it propagates through the environment, revealing the spatial factors contributing to connectivity degradation. Signal strength, commonly represented by the RSRP, is critical for assessing connectivity quality. Due to the single-pass nature of train-based measurement campaigns, we have limited control over dynamic network-side variables like CL, which fluctuate throughout the day. As it is not feasible to capture identical track conditions under varying loads, we prioritize RSRP as a stable and reproducible metric, which excludes noise and interference terms and is invariant to traffic patterns [27]. This makes RSRP particularly suitable for isolating the effects of large-scale propagation. Its independence from traffic conditions makes it a reliable baseline for analyzing large-scale propagation effects and for comparing different train configurations and deployment environments. The relationship between RSRP and PL is crucial - as PL increases due to factors like distance, environmental obstructions, or the train's physical structure, the RSRP decreases, indicating a weaker signal at the receiver. This connection can be mathematically expressed as:

$$\text{RSRP} = P_t - \text{PL} + G_t + G_r, \tag{2.6}$$

where

- P_t is the transmitted power of the reference signal,
- PL is the signal attenuation over the distance and through obstacles, and in indoor train scenario it consists of indoor and outdoor PL, as well as the VPL

$$\text{PL} = \text{PL}_{\text{outdoor}} + \text{PL}_{\text{indoor}} + \text{VPL}, \tag{2.7}$$

- G_t is the transmitter antenna gain,
- G_r is the receiver antenna gain.

Therefore, to measure the influence of the cabin structure we can measure RSRP inside and outside the cabin, where their difference provides a direct measure of how much the signal is attenuated as it passes through the train's cabin materials, including windows, walls, and other structural elements. Understanding VPL is essential for evaluating the performance of the network inside the train and optimizing solutions to mitigate this additional signal degradation.

2.2.1 Link Budget and Key Metrics

The link budget in a railroad scenario accounts for the cumulative gains and losses that determine the signal strength received by the UE inside a train. It provides a structured approach to quantifying how various propagation effects impact the signal from the BS to the UE. Signal propagation is influenced by multiple factors that can be broadly categorized into large-scale and small-scale effects:

- **Large-Scale Effects:**
  - Path Loss (PL): The deterministic component of signal attenuation caused by the spreading of electromagnetic waves as they propagate through free space or obstructed environments. This is typically modeled using models like Free-Space Path Loss (FSPL), COST Hata, or other empirical models [28].
  - Shadow Fading (SF): The stochastic component of large-scale fading, caused by obstacles such as buildings, trees, or terrain blocking or scattering the signal. SF is often modeled as a log-normal random variable with a specific standard deviation σ and spatial correlation distance d_{corr} [27].
- **Small-Scale Effects:**
  - Small-Scale Fading (SSF): Caused by multipath propagation, where the signal reaches the receiver through multiple paths, leading to constructive or destructive interference. This type of fading is highly dependent on the surrounding environment and the relative motion of the transmitter, receiver, and scatterers [29]. In dynamic scenarios like a moving train, Doppler effects may also contribute to small-scale variations, depending on the velocity and scatterer distribution [28]. However, in our analysis, we focus on PL as the dominant factor, as SSF effects are averaged out over large spatial scales.

In wireless communications, understanding and modeling both large-scale and small-scale effects are critical for accurate network analysis [30]. However, the relevance of each component depends on the measurement conditions and objectives.

We focus on large-scale PL, leveraging the positional diversity of the moving train to ensure that SF effects are averaged out across the measurements. SF, the random variation in signal strength caused by obstructions such as buildings and terrain, naturally averages out in this scenario due to the extensive range of environments traversed by the train. Hence, the primary components of the link budget include:

- Free-Space Path Loss (FSPL): Represents the attenuation of the signal as it propagates over distance in free space from BS to the train exterior (or train rooftop antenna), modeled as in [31]:

$$\text{FSPL (dB)} = 20\log_{10}\frac{4\pi}{\mathrm{c}} + 20\log_{10}\mathrm{f} + 20\log_{10}\mathrm{d}, \tag{2.8}$$

  where f is the carrier frequency in Hz, d is the distance in meters, and c is the speed of light in vacuum.

- Vehicle Penetration Loss (VPL): Accounts for the loss as the signal penetrates the train's exterior, influenced by window properties, materials, and geometry. Expressed as the difference between RSRP measured outside and RSRP measured inside the train:

$$\text{VPL (dB)} = \text{RSRP}_{\text{outside}} \text{ (dBm)} - \text{RSRP}_{\text{inside}} \text{ (dBm)}. \tag{2.9}$$

- Indoor Path Loss (Indoor PL): Reflects the signal degradation as it travels within the train cabin, often dependent on passenger density, internal reflections, and obstructions [32].

Throughput Estimation: Higher SINR allows for higher-order modulation and more efficient coding, thereby increasing spectral efficiency and the achievable throughput, as depicted in Fig. 6.3 of [17]. The impact of SINR on MCS is reflected through the CQI, which directly influences the achievable T. Fig. 18 in [33] illustrates the relationship between Reference Signal (RS)-SINR and the reported CQI in crowdsourced data. Since CQI determines the selection of MCS, which in turn dictates the spectral efficiency and data rate, understanding this relationship is crucial for T estimation. While for a UE in an empty train connected to an unloaded cell, the Shannon-Hartley Theorem [34] provides a theoretical upper bound on channel capacity for a channel of bandwidth W:

$$T_{\max} = W \cdot \log_2(1 + \text{SINR}), \tag{2.10}$$

in practical LTE deployments, T is constrained by standardized modulation and coding schemes. For example, with 64-Quadrature Amplitude Modulation (QAM), 2×2 MIMO, and a 20 MHz bandwidth, the DL peak T reaches approximately 150 Mbps. Thus, Eq. (2.10) serves only as a broad upper bound that scales with SINR, which is itself limited by propagation conditions, interference, and device capabilities.

Under loaded conditions, T scales with the available PRBs, which are determined by the CL:

$$T = T_{\max} \cdot (1 - L), \tag{2.11}$$

where $L \in [0, 1]$ represents the CL. Consequently, CL is a critical factor in determining the actual T available to users. In addition to reducing available resources, a higher CL also increases interference for neighboring cells, as more users and transmissions contribute to inter-cell interference, degrading signal quality and overall network performance.

In our measurements, we focus on the RSRP due to its temporal stability and independence from traffic patterns. Conducting measurements in an empty test train is essential as it provides a baseline reference for T under ideal conditions. This reference can then be adjusted to reflect the expected CL during typical operations, enabling accurate predictions of T in real-world scenarios.

Modeling Pathloss: In the context of modeling PL for railroad scenarios, the COST Hata model and the FSPL model are frequently utilized due to their foundational roles in predicting signal attenuation in various environments.

FSPL Model: The FSPL model calculates the idealized PL assuming a clear, unobstructed Line-of-Sight (LOS) between the transmitter and receiver [35]. It serves as a baseline, representing the minimum possible PL in the absence of obstacles and environmental factors. While actual environments introduce additional losses due to obstructions, reflections, and diffractions, the FSPL model is essential for understanding the fundamental limits of signal propagation. The formula for FSPL, where d is the distance in km and f is the frequency in MHz, is given by:

$$L_{\text{FSPL}} = 32.45 + 20\log_{10}\frac{\text{f}}{\text{MHz}} + 20\log_{10}\frac{\text{d}}{\text{km}}, \tag{2.12}$$

where the constant 32.45 accounts for free-space propagation at a frequency of 1 MHz and a reference distance of 1 km. This model serves as a lower bound for PL in ideal propagation conditions.

COST Hata Model: COST 231 [36] extended the Hata model [37], based on Okumura's correction functions [38], to accommodate frequencies up to 2 GHz, making it applicable for modern cellular communications. It is particularly suited for urban and suburban environments, providing a practical estimation of PL by considering factors such as frequency, BS height, mobile station (or UE) height, and the distance between them. The model's empirical nature, derived from extensive measurements, offers a balance between complexity and accuracy, which is beneficial for initial network planning and assessments. It considers a variety of urban characteristics, making it a realistic estimation for Non-Line of Sight (NLOS) propagation scenarios. In original Hata model, the basic transmission loss is defined for frequencies [150-1000] MHz as:

$$\begin{aligned} \text{L}_\text{b} =& 69.55 + 26.16 \cdot \log_{10}\frac{\text{f}}{\text{MHz}} - 13.82 \cdot \log_{10}\frac{\text{h}_\text{B}}{\text{m}} - \text{a}\left(\text{h}_\text{M}\right) \\ &+ \left(44.9 - 6.55 \cdot \log_{10}\frac{\text{h}_\text{B}}{\text{m}}\right) \cdot \log_{10}\frac{\text{d}}{\text{km}}. \end{aligned} \tag{2.13}$$

COST 231 extended this model to the frequencies [1500-2000] MHz as:

$$\begin{aligned} \text{L}_\text{b} =& 46.3 + 33.9\log_{10}\frac{\text{f}}{\text{MHz}} - 13.82\log_{10}\frac{\text{h}_\text{B}}{\text{m}} - \text{a}\left(\text{h}_\text{M}\right) \\ &+ \left(44.9 - 6.55\log_{10}\frac{\text{h}_\text{B}}{\text{m}}\right)\log_{10}\frac{\text{d}}{\text{km}} + \text{C}_\text{m}, \end{aligned} \tag{2.14}$$

where

- h_B: height of the BS antenna (in meters),
- h_M: height of the mobile station antenna (in meters),
- $\text{a}\left(\text{h}_\text{M}\right)$: a correction factor for mobile antenna height, which is defined depending on frequency as

$$\text{a}\left(\text{h}_\text{M}\right) = \left(1.1 \cdot \log\frac{\text{f}}{\text{MHz}} - 0.7\right)\frac{\text{h}_\text{M}}{\text{m}} - \left(1.56 \cdot \log\frac{\text{f}}{\text{MHz}} - 0.8\right), \tag{2.15}$$

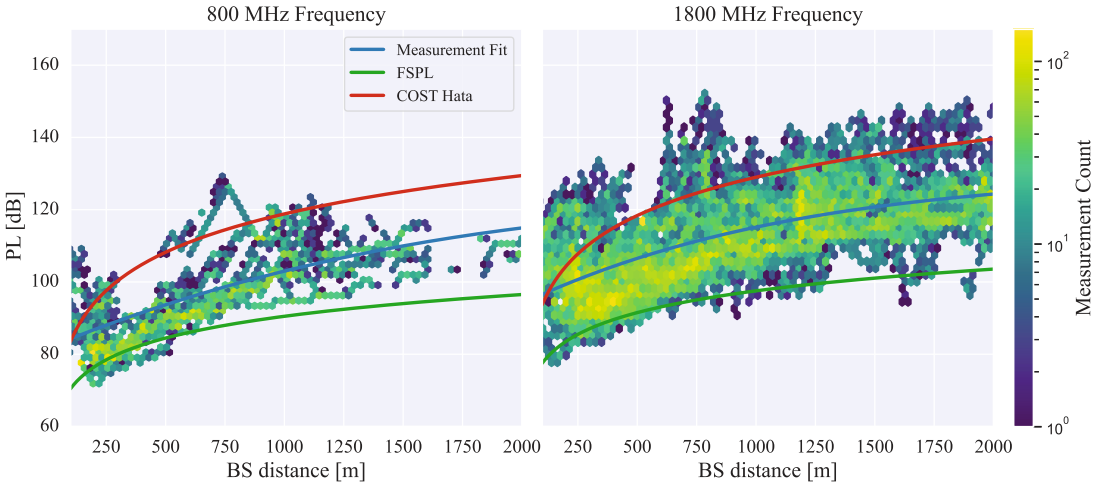


Figure 2.1: Comparison of PL models and measurements in a railroad scenario.

- C_m: a correction factor for the propagation environment, with values of 0 for medium cities and suburban areas, and 3 dB for metropolitan areas.

Several studies have evaluated the applicability of the COST Hata and FSPL models in railroad scenarios. These models are often used as benchmarks due to their foundational roles in predicting signal attenuation. However, research indicates that these traditional models, that have been used for their simplicity and standardization, may not accurately capture the unique propagation characteristics of railway environments [39], necessitating the development of specialized models.

For instance, in [40] authors conducted measurements in high-speed railway environments and found that both the FSPL and COST Hata models did not align well with empirical data. They proposed tuned PL models specifically for viaduct and plain scenarios, demonstrating improved accuracy over traditional models.

Similarly, in [41] introduced an empirical PL model tailored for high-speed railway viaduct scenarios. Their study highlighted the limitations of conventional models and emphasized the need for models that consider factors such as viaduct height and BS distance to accurately predict PL in these environments.

Figure 2.1 compares the measured PL values with predictions from the COST Hata and FSPL models, based on data collected along a suburban railway track. During the campaign, a measurement scanner connected to the train's rooftop antenna captured signal strength at 800 MHz and 1800 MHz. The PL was calculated using the known antenna gains of both the BS and the train antenna, together with the calculated distance to each BS.

To contextualize the results, we simulated both propagation models under appropriate conditions. For the COST Hata model, we set $C_m = 0$ (representing a suburban/urban environment) and applied the frequency-dependent correction factor $a(h_M)$ for

both frequencies[2]. A polynomial fit was also introduced to the measurement data to approximate the large-scale trend while smoothing out SSF effects.

Rather than displaying a standard scatter plot - which is less informative due to significant overlap of measurement points - we employ a density plot in Fig. 2.1 to better visualize the distribution of measurements. The density plot reveals dominant trends in the PL with respect to BS distance while mitigating visual clutter and emphasizing discrepancies between theoretical models and empirical observations.

From the figure, it is evident that both the FSPL and COST Hata models show substantial deviation from the measured values. The FSPL model significantly underestimates the PL, as it assumes free-space conditions without obstructions. The COST Hata model, while more realistic, still fails to capture the propagation characteristics accurately - particularly in mid-to-long range distances along the suburban rail environment.

These discrepancies underscore the limitations of existing PL models in dynamically varying railway environments, where spatiotemporal effects play a crucial role. As an alternative, RT simulations can be employed to model complex propagation more accurately and to predict expected RSRP and achievable T under ideal conditions.

In Section 2.3, we will delve into one such simulation approach, discussing its methodology, advantages, and limitations.

2.2.2 Small-Scale Fading in Train

To analyze the impact of SSF, we conducted indoor measurements at 12 distinct seats within the train cabin, in a static scenario. At each seat, measurements were performed twice by placing the antenna on the fold-down table, once shifted toward the left side and once toward the right side, as illustrated in Figure 2.2. This figure also shows the scanner used for the measurements - a PCTEL IBflex scanner [42]. At each position, measurements were conducted for several minutes, enabling repeated sampling to average out SSF effects. This study focuses on a single BS sector and a single frequency band, under the assumption that the device remains connected to the same BS for all positions. This consistency allows us to evaluate the influence of measurement position within the cabin. Such analysis would not be feasible if connections to different BSs or frequency bands occurred across positions, as it would introduce additional variability unrelated to the positional effects.

To process the collected data, we applied Gaussian Process Regression (GPR), a non-parametric statistical method used for interpolation and prediction. GPR models the data by assuming a prior distribution over functions, which is updated with observed data to produce posterior predictions, for more details, see Appendix A1. The steps involved in applying GPR are as follows:

- Measurements were conducted at 24 positions, each corresponding to one of two sides of 12 seats.

[2]Simulations were conducted for both 800 MHz and 1800 MHz frequencies, corresponding to the measurements taken on this track.

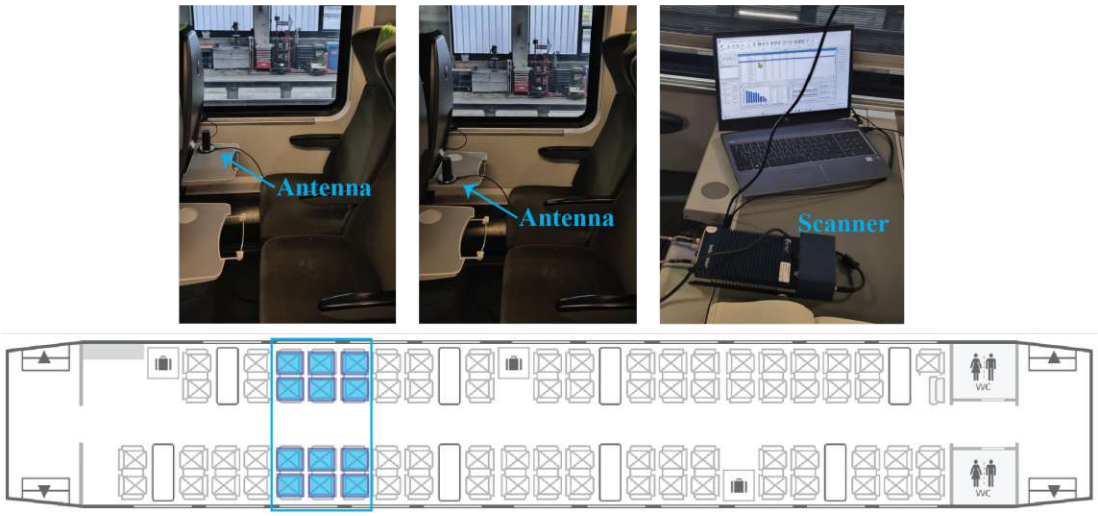


Figure 2.2: Measurement setup in a static train scenario, where the antenna was mounted on the fold-down table. The sketch illustrates the train plan, with the measurement positions highlighted in blue. Measurements were conducted across 12 designated seats, with the antenna positioned at two distinct locations on each table: once closer to the window and once farther from it. This configuration resulted in a total of 24 measurement positions. The antenna was connected to a scanner to perform network measurements.

- At each position, multiple RSRP values were recorded over several minutes.
- For each position, the mean μ_i and standard deviation σ_i of the measured RSRP values were calculated:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} \mathrm{RSRP}_{ij}, \qquad \sigma_i = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N} (\mathrm{RSRP}_{ij} - \mu_i)^2}, \tag{2.16}$$

  where N is the number of measurements at position i, and RSRP_{ij} is the j-th measurement at position i.
- These μ_i and σ_i values were used as inputs to the GPR model, which generated a reconstructed map by interpolating both the mean RSRP values and their associated variability, while also providing confidence intervals.

GPR is particularly well-suited for this application as it provides both predicted mean values and a measure of prediction uncertainty, ensuring a robust representation of the spatial signal distribution.

Fig. 2.3 and Fig. 2.4 show the GPR interpolation results for 800 MHz and 1800 MHz bands respectively. Several observations were made from the GPR analysis and measurements:

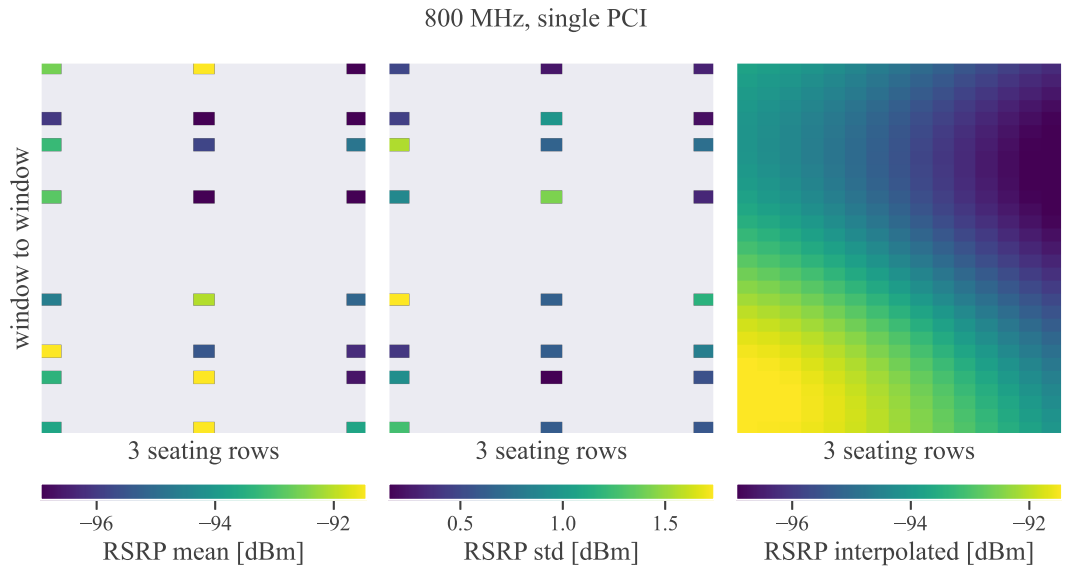


Figure 2.3: RSRP measurement mean μ_i (left), standard deviation σ_i (middle) at each position i, and GPR interpolation using the mean and standard deviation of measurements (right) for single BS in 800 MHz frequency band. Root Mean Squared Error (RMSE) between interpolated mean RSRP and measured mean RSRP is 2.03 dB.

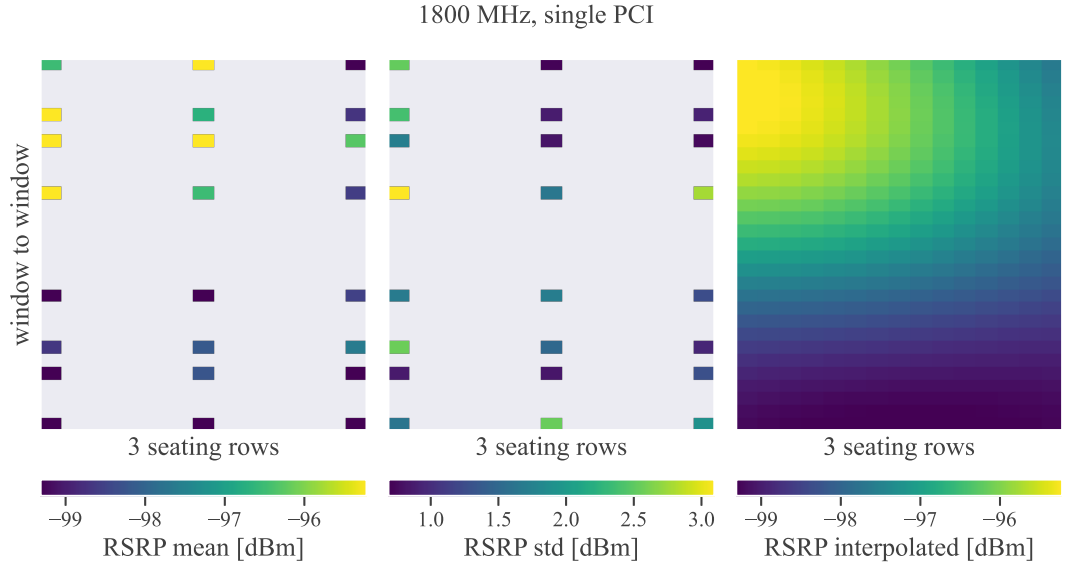


Figure 2.4: RSRP measurement mean μ_i (left), standard deviation (middle) σ_i at each position i, and GPR interpolation using the mean and standard deviation of measurements (right) for single BS in 1800 MHz frequency band. RMSE between interpolated mean RSRP and measured mean RSRP is 2.06 dB.

- **SSF Effects:** Rapid signal fluctuations due to multipath propagation were observed, caused by reflections from cabin walls, seats, and other obstructions - a typical characteristic of indoor environments with dense multipath conditions. However, the GPR-generated signal map exhibited a smooth and continuous spatial variation with modest variation (on the order of 3 dB) across all 24 positions. This indicates that while individual measurements are impacted by SSF, averaging multiple samples at each position effectively suppresses small-scale fluctuations in this static scenario, allowing GPR to capture primarily the Large Scale Fading (LSF) effects.
- **Smoothing via Measurement Averaging:** By averaging repeated measurements at each position, the mean impact of SSF fluctuations was reduced, while the remaining variability, attributed to multipath propagation, was still captured through the standard deviation (σ_i), which was incorporated into the GPR model. This highlights the value of repeated measurements for improving signal stability estimates.
- **Minimal Spatial Variability:** The reconstructed signal map shows limited but noticeable spatial variability, suggesting that while SSF effects exist at individual measurement points, their impact is reduced when averaged over time and interpolated. This indicates that localized differences in measurement positions have a limited influence on the overall signal strength distribution within the cabin.
- **Implications for Coverage Planning:** The smooth RSRP variation in the reconstructed map suggests that, at a static level, signal strength inside the cabin is relatively stable across small positional changes. In a dynamic scenario, SSF effects would average out over time, potentially leading to an even more stable RSRP distribution, though larger-scale variations due to mobility effects would need further investigation.

The results confirm that while SSF introduces significant local variations, these effects are not critical for overall performance evaluations when measurements are spatially averaged. Furthermore, the negligible variability in the GPR results suggests that the precise measurement location within a seating cabin does not significantly impact the quality of signal strength.

2.3 Ray Tracing Simulations for Railway Connectivity Modeling

RT simulations play a key role in modeling and optimizing cellular network performance in railway environments. While train connectivity is often challenged by signal degradation in tunnels, stations, and dense urban areas, the majority of railway tracks - especially in Austria - run through rural and open areas where LOS conditions dominate. This characteristic makes RT particularly well suited for simulating signal propagation and assessing network coverage along railway corridors.

RT models the propagation of electromagnetic waves as rays and accounts for physical phenomena such as reflection, diffraction, and transmission [43]. This enables accurate predictions of signal behavior, especially in geometrically complex areas. Prominent RT tools include Sionna RT, CloudRT, RemCom, and MATLAB's 5G toolbox [44–47]. In the context of DT use cases for railways, RT simulations enable:

- **Network coverage planning:** Identifying coverage gaps and optimizing the placement of BSs along railway tracks.
- **Signal interference management:** Predicting signal scattering and reflection in urban environments to mitigate interference.
- **Tunnel and station optimization:** Determining optimal antenna and signal booster placement to maintain connectivity in confined spaces.

While RT simulations provide significant benefits for the optimization of cellular networks in railway systems, one key challenge remains: the accuracy of RT simulations is highly dependent on the calibration of their underlying parameters. These parameters, such as material properties, signal scattering, reflection, diffraction, and transmission characteristics, can vary greatly depending on the specific environment being modeled. Without accurate calibration, simulations may fail to reflect real-world conditions, resulting in suboptimal network performance and coverage predictions.

To ensure simulation accuracy, these parameters must be calibrated using real-world measurements taken in the specific environment. For instance, signal behavior in an urban area with dense steel and concrete structures will differ from that in a rural agricultural or mountainous setting. By using real-world measurements, these parameters can be adjusted to allow the RT model to accurately predict signal propagation in the environment.

This calibration process is essential for generating precise results. Without it, RT simulations may produce inaccurate outcomes, leading to wrong assumptions about network coverage, signal strength, and interference. Measurements form the foundation upon which RT simulations are calibrated, ensuring that they accurately reflect the unique characteristics of each environment and are, therefore, reliable for practical applications.

We investigate an example scenario using the Sionna RT simulator [48] to assess the impact of RT parameter calibration on the accuracy of signal predictions. The measurements were conducted in an urban environment within the city of Vienna, Austria. To construct the underlying 3D model, we utilized the 3D city map of Vienna from Geodatenviewer der Stadtvermessung Wien (GSW), which provided the necessary data to generate a detailed 3D representation of the area of interest. Specifically, we obtained the generalized roof model (LOD2.1) and the terrain model from the City of Vienna's official platform [49]. These models were subsequently imported into Blender [50], where we leveraged several predefined materials available within the Sionna RT [48] framework. The material `itu_brick` was assigned to the building structures, `itu_medium_dry_ground` to the terrain, and `itu_meta` to the train. A top-down view of the scene, showing the layout of the measurement and BS locations, is presented in Fig. 2.5 (bottom).

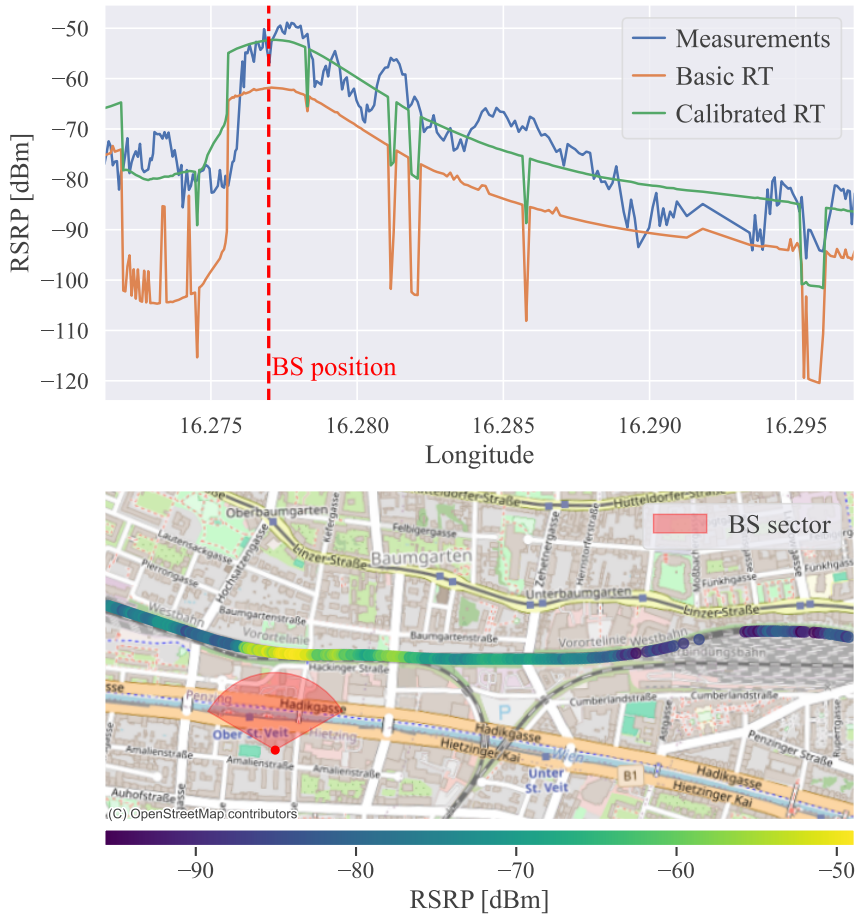


Figure 2.5: Comparison of measurements and RT simulations for a single BS operating in the 800 MHz band. Top: Observed measurements contrasted with simulated results. Bottom: Visualization of rail tracks with measurement data used for parameter training.

Fig. 2.5 (top) compares RSRP measurements on the rail track with simulated RSRP values at the same locations, using both an untrained RT model (Basic RT) and its calibrated version (Calibrated RT). As shown in Table 2.1, the calibration significantly reduces the MAE across all locations, particularly in NLOS conditions, where the error decreases from 24.2 dB to 6.77 dB.

Beyond the reduction in MAE, calibration also improves the correlation between measurements and simulations, indicating a stronger alignment between the modeled and real-world propagation characteristics. The correlation coefficient increases from 0.747 to 0.827 overall, with the most notable improvement in NLOS conditions (from 0.459 to 0.509). A higher correlation coefficient in these scenarios suggests that the calibrated model better captures variations in the measured data, though discrepancies remain.

Table 2.1: Mean Absolute Error (MAE) and correlation coefficient values for trained (calibrated) and untrained (basic) RT simulator.

	MAE [dB]			Correlation Coefficient		
	All Locations	LOS	NLOS	All Locations	LOS	NLOS
Basic RT	13.2	9.2	24.2	0.747	0.869	0.459
Calibrated RT	5.1	4.6	6.77	0.827	0.871	0.509

The NLOS scenarios particularly influence the high error observed in the untrained model. These scenarios are characterized by complex signal propagation due to obstacles blocking the direct line of sight between the transmitter and receiver. The NLOS error significantly decreases after calibration, demonstrating the importance of accurately modeling the environment to improve simulation accuracy. For instance, MAE for NLOS locations drops from 24.2 dB in the untrained model to 6.77 dB in the calibrated model, illustrating the substantial improvement calibration brings to RT simulation accuracy in these challenging conditions.

However, the remaining high errors highlight the critical need for real-world measurements and highly accurate 3D environment models for the calibration of simulation parameters. Despite the improvements made through calibration, the RT simulations still fall short of fully replicating real-world conditions. While these simulations are informative, they do not yet provide a complete picture of the network's behavior under actual operating conditions. To truly capture the complexities of real-world networks, large-scale measurement campaigns remain indispensable. In the following section, we propose a methodology for collecting and post-processing such measurements, as well as for analyzing network performance based on real-world data, which will help bridge the gap between simulation and reality.

2.4 Integrated Framework for Data Collection and Analysis

In this section, we present the data collection and postprocessing framework that we developed and applied across multiple measurement campaigns to benchmark cellular connectivity performance in railway environments. The framework covers the design and planning of the campaigns, the selection and configuration of measurement equipment, and the postprocessing of collected data. A key aspect of our approach is the systematic combination of different datasets during the analysis phase, enabling a more comprehensive and accurate evaluation of connectivity performance.

2.4.1 Designing the Measurements and Measurement Equipment

Collecting measurements onboard trains is a time- and cost-intensive process. Train operators cannot afford to run trains empty along the tracks solely for measurement

purposes, as this would disrupt their operations and increase costs. For network performance assessments, it is critical to conduct measurements in an empty train, as this provides a baseline of the network's capacity without passenger-induced resource sharing. These measurements can later be scaled to account for real-world network load conditions, where passengers onboard share the available resources.

To address these challenges, we leverage the time during train operator's mandatory safety tests, tagging along on test drives in empty trains to perform cellular network measurements in parallel. This is made possible through our industry collaboration projects with ÖBB (Austrian Federal Railways), enabling access to these test runs while minimizing additional operational disruptions.

Modern trains utilize multiple cellular modems to maintain continuous connectivity with local mobile networks, supporting various onboard systems. These modems are typically mounted on the train roof and connected to pre-installed antennas designed for reliable signal reception, even in rural or weak coverage areas. For our measurements, we connect directly to the output of these existing antennas, as we are not permitted to install our own antennas on trains in operation. This setup requires authorization from the train operator and means that we must work with the technical specifications of the pre-installed antennas rather than selecting our own.

The measurement setup leverages both outdoor and indoor configurations to capture a comprehensive picture of network performance under real-world conditions.

For outdoor measurements, the train's rooftop antenna, originally used to provide Wireless Local Area Network (WLAN) connectivity onboard through cellular networks, can be repurposed. During measurement campaigns, one of the WLAN modems is bypassed, and a network scanner is directly connected to the antenna's output. Alternatively, a UE with an external antenna can be used for network measurements. This setup allows us to measure the cellular network coverage along the track as provided by the MNO, capturing signal conditions along the route. To ensure continuous operation, the network scanner is powered through an Uninterruptible Power Supply (UPS), preventing interruptions in case of short power outages, which are common on trains. Additionally, the network scanner is connected to both the train's GPS antenna and an external u-blox Global Navigation Satellite System (GNSS) module [51] for enhanced GPS accuracy. The measurement setup is presented in Fig. 2.6. By combining reported GPS data with network measurements through timestamp alignment in postprocessing, we ensure accurate mapping of network performance relative to the train's location along the track.

For indoor measurements, which focus on the signal quality experienced by passengers, we deploy additional equipment. Specifically, we can use either a network scanner or UEs inside the train. When using a scanner for indoor measurements, it is connected to an external antenna that is placed on a fold-down table, seat, or another suitable surface inside the train. This setup allows the scanner to passively capture detailed network information across multiple frequency bands simultaneously. Alternatively, we use UEs to actively generate user-like traffic by performing constant downloads inside the train. Our measurement setup, shown in Fig. 2.7, consists of multiple smartphones running Nemo Handy Handheld Measurement Solution by Keysight [52], which records

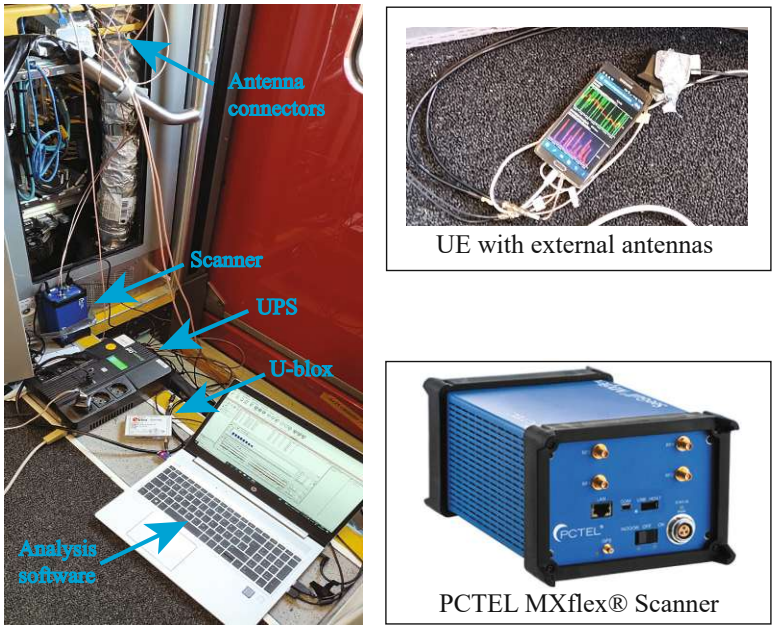


Figure 2.6: Measurement setup for outdoor train measurements. The left side illustrates the connection of a scanner to the train's rooftop antenna. The system is supported by UPS to prevent data loss during short power outages and utilizes a u-blox GNSS module for improved GPS accuracy. The right side (top) shows a UE with an external antenna, which can be used as an alternative to the scanner, while the right side (bottom) displays the PCTEL MXflex scanner.

Figure 2.7: Measurement setup for indoor train testing. On the left side, a tablet acts as the master unit, automatically triggering measurement scripts on all connected UEs. On the right side, an external antenna is positioned as an alternative setup option and used for scanner-based measurements (though the scanner itself is not visible in the image).

signal metrics and network performance data. A central tablet serves as the master unit, automatically triggering the measurement scripts on all connected UEs. This ensures synchronization and minimizes manual intervention during data collection. Our measurement campaigns typically involve up to six smartphones operating in parallel. This approach ensures measurement repeatability under realistic conditions, since train-based measurements, unlike car-based drive tests, cannot easily be repeated over the same track segment. By capturing data from multiple devices simultaneously, we enhance the reliability and robustness of the collected data. The measurement campaigns aim to collect KPIs such as RSRP, T, and voice quality metrics, alongside geographical positioning data and timestamps. This data forms the basis for subsequent analysis. The choice of measurement equipment significantly impacts the type and quality of collected data. Network scanners, such as the PCTEL MXflex, and UEs (smartphones) each have distinct advantages and limitations, making them suitable for different aspects of cellular network analysis.

Network scanners provide a high level of granularity and accuracy in signal measurements. They can simultaneously scan multiple frequency bands, including those not actively used by a particular UE, offering a full picture of the available network spectrum. For example, the PCTEL MXflex[3] can monitor up to 48 channels across different frequency bands and perform blind scans to detect all available signals. Additionally, scanners support advanced antenna configurations such as 4×4 MIMO and deliver high-precision measurements, with an RSRP accuracy of ± 1 dB and a minimum detection level as low as -140 dBm. These characteristics make scanners particularly useful for benchmarking network coverage and detecting interference or weak signal areas with high accuracy. Since scanners operate passively and do not generate traffic, they are unable to capture performance metrics that depend on active user behavior, such as T or latency.

In contrast, UE-based measurements focus on the user experience. Smartphones operate under real-world conditions, connecting to a BS and dynamically selecting frequency bands based on network conditions and resource allocation strategies. With carrier aggregation, UEs can be connected to one Primary Serving Cell (PCell) and more than one Secondary Cell (SCell) simultaneously, enabling higher T and increased bandwidth utilization. This provides a more realistic representation of network performance for mobile users. While UEs reflect real-world performance, including signal strengths and T, they have significant limitations in measurement accuracy compared to scanners. As shown in Table 2.3, UEs have an RSRP accuracy of ± 8 dB, much lower than the ± 1 dB precision of scanners. Additionally, UEs are limited by their chipset capabilities, supporting fewer simultaneous frequency bands and typically only lower-order MIMO configurations (e.g., 2×2 or 4×4). Moreover, the GPS accuracy of UEs, ranging between [10-50] m, is considerably worse than the 2.5 m accuracy achieved by the scanner, which affects location-based analyses.

Scanners can measure signals from multiple networks with high accuracy, making them valuable for coverage mapping and interference analysis. However, they are expensive, require specialized software, and do not reflect real user experience. In

[3] Table 2.2 shows LTE Measurement Parameters in PCTEL MXflex scanner.

Table 2.2: LTE Measurement Parameters in PCTEL MXflex scanner [53].

Measurement Modes	Top N Sync Channel RS (P-SCH/S-SCH), Resource Block (Wideband, Subband), Dynamic Spectrum Sharing (DSS), Blind Scan, eMBMS Multicast RS
Data Modes	RSRP (Received Power), RSRQ (Received Quality), CINR (Carrier-to-Interference Ratio), Cyclic Prefix, Time Offset, Delay Spread, Layer 3 Analysis, LTE MIMO (ECQI, Est. Throughput)
Channel Bandwidths	[1.4, 3, 5, 10, 15, 20] MHz
Max Channels	48 total (FDD + TD-LTE)
Antenna Techniques	SISO, MISO, MIMO (4×4, 4×2, 2×2)
Measurement Rate	LTE FDD: 48/sec, 2×2 MIMO: 24/sec, 4x4 MIMO: 3/sec, TD-LTE: 19/sec, eMBMS: 2/sec
RSRP Min Detection Level	-140 dBm @ 15 kHz
RSRP Accuracy	± 1 dB
GPS Accuracy	2.5 m
Max PCIs	16

Table 2.3: LTE measurement parameters for a typical UE [54].

Measurement KPIs	RSRP, RSRQ, CINR, throughput Timing Advance, Layer 3 Signaling, etc.
Channel Bandwidths	[1.4, 3, 5, 10, 15, 20] MHz
Antenna Techniques	SISO, MIMO (2×2, 4×4)
Measurement Rate	Depends on UE sampling rate and network conditions (typically 2/sec in connected state)
RSRP Range	-140 dBm to -44 dBm
RSRP Accuracy	± 8 dB
GPS Accuracy	10 m-50 m
Max PCIs	Limited by UE chipset capabilities

contrast, commercial phones or test UEs are cost-effective, easy to deploy at scale, and capable of generating traffic like actual users, but they are less accurate and limited in the frequency bands they can observe. In our setup, we use scanners mounted on

the train roof to measure outdoor coverage, while UEs are placed inside the train to assess performance from a passenger perspective. This hybrid approach provides a comprehensive view of network performance, combining coverage potential with real-world user experience.

2.4.2 Geospatial Data Integration for Environment Digital Twin Development

Accurate geographical data is fundamental for mapping measurements to their respective locations, ensuring precise contextual analysis. To apply the DT concept in practice, we built an Environment and Rail Track DT model by fusing GPS, topography, and land cover data - serving as the foundation for realistic simulation of signal propagation scenarios along the rail tracks.

Open Street Maps (OSM) [55] is a collaborative, open-access geographic database that provides detailed, up-to-date information on transportation networks, including rail tracks, roads, and other infrastructure. Known for its flexibility and community-driven accuracy, OSM offers a comprehensive and continuously updated representation of transportation infrastructure. This makes it a crucial resource for modeling the railway network on which the DT of the train operates. We use OSM data to define the network layout, including the positioning of rail tracks, tunnel data and the broader urban environment, providing the foundational geographic framework for simulating wireless signal propagation in the railway system.

Geodatenviewer der Stadtvermessung Wien (GSW) is a platform that offers access to 3D building models, specifically for the city of Vienna, Austria [49]. These models are provided in shapefile format and include detailed representations of buildings and other urban features. This data is invaluable for simulating signal behavior in complex urban environments, where factors such as diffraction, reflection, and shadowing play a significant role in wireless communication. The 3D models allow for precise environmental modeling, which is necessary to simulate the propagation of wireless signals in dense urban areas, such as those found in Vienna. These 3D models contribute to the digital representation of the railway environment used in this work.

Digital Terrain Model (DTM) [56] is a high-resolution model that provides detailed elevation data for the entire Austria. Produced through airborne laser scans as part of the Geoland project, the DTM offers a 10 m resolution, capturing terrain features, such as mountains, valleys, and other natural geographical structures, which significantly impact the propagation of wireless signals. In rural areas, where urban 3D models are unavailable, the DTM serves as the primary source of geographic data, helping to model how terrain features influence wireless signal behavior.

Corine Land Cover (CLC) [57] is a dataset provided by the Copernicus Land Monitoring Service, which classifies land cover across Europe into 44 categories, including forests, agricultural land, wetlands, and urban areas. Based on satellite imagery and field data, the CLC dataset helps to understand the influence of land use patterns on signal propagation. In rural areas, where the DTM provides terrain elevation data, the CLC dataset adds essential context regarding land cover, helping to model how different land types - such as forests or agricultural zones - affect signal strength and propagation.

Use Cases for Jointly Using These Datasets

The combination of these datasets creates a comprehensive framework for simulating wireless signal propagation in both urban and rural environments. In urban scenarios, the integration of the GSW 3D building models, OSM data, and the DTM allows for the creation of a highly accurate Environment and Rail Track DT of the railway system. This DT represents a precise model of the physical environment, including terrain features and infrastructure, which can be used to simulate the real-world propagation of wireless signals. Additionally, railway network details, including rail tracks and tunnel data obtained from OSM, are mapped to train locations using GPS and interpolation techniques, ensuring accurate representation of train movements within the DT. By merging this DT with Cellular Network DT, containing network information-such as BS positions, heights, transmission power, and other network parameters-obtained from the MNO, we can build a complete environment for simulating cellular signal propagation. This combined model can then be used in RT simulations, as demonstrated in Section 2.3.

The data processing workflow applied for creating DT for Railway Connectivity is summarized in Fig. 2.8. The flowchart illustrates the sequential integration of geospatial datasets, starting from collecting raw train measurements to embedding the environment. It highlights key processes such as GPS mapping, network data association, land cover classification, and embedding of 3D surroundings. These steps ensure that both train connectivity and surrounding environmental conditions are accurately represented in the model.

In rural environments, the DTM and CLC datasets are combined to account for terrain elevation and land use, providing sufficient information to evaluate the effects of terrain and land cover on signal propagation. By integrating these datasets, the methodology ensures that the simulation of signal propagation accounts for both the physical terrain and the land cover, which are critical factors in determining signal strength and coverage.

Furthermore, the OSM and CLC datasets are employed to classify the measurement data into urban and rural categories, which helps optimize the measurement strategy. The goal of this classification is to tailor the measurement approach to the specific characteristics of each environment, as will be discussed in Chapter 6.

In the context of AoA calculations, the DTM, OSM, and Cellular Network DT providing the network layout are used to model the environment in such a way that the direction of the incoming signal can be accurately determined. This is particularly important for network optimization, where precise knowledge of signal angles is crucial for effective beamforming and antenna placement.

By leveraging the strengths of each dataset, this work aims to create a robust and flexible framework for simulating wireless connectivity across a wide range of geographical environments. The combination of detailed terrain data, land use information, and railway infrastructure allows for more accurate simulations of signal propagation, ultimately improving the understanding and optimization of wireless networks in both urban and rural railway scenarios.

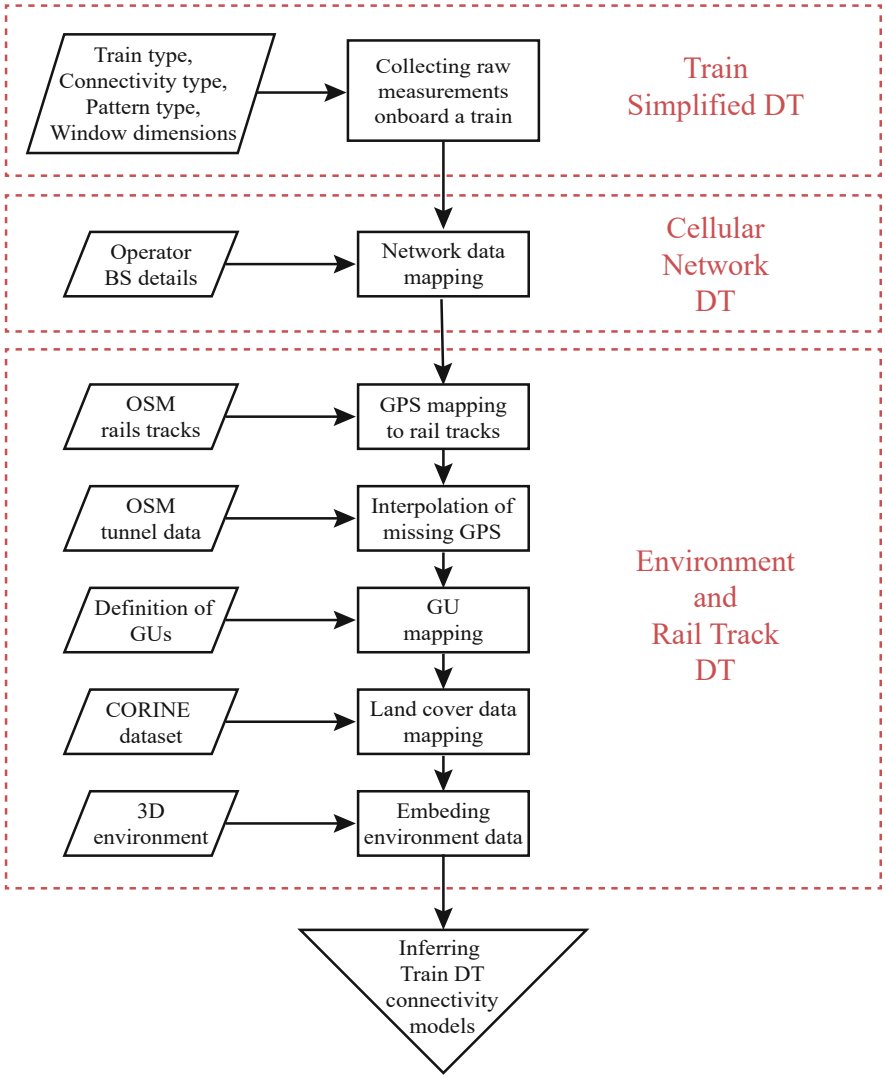


Figure 2.8: Workflow for constructing the DT for Railway Connectivity, integrating geospatial, network, and environmental data for signal propagation modeling.

2.4.3 Data Fusion and Spatial Interpolation Using Rail Track Proximity Mapping

The raw GPS data collected during measurement campaigns often contains inaccuracies or missing values, particularly in challenging environments such as tunnels, where satellite signals are obstructed. To address this issue, the raw measurements are mapped to a known reference path using proximity mapping techniques. This approach is conceptually similar to dead reckoning [58], where position estimation relies on previous location data and movement assumptions. For rail tracks, the exact location of each measurement track is queried from OSM data, allowing for precise alignment of measurement points with the rail track. This alignment is achieved by projecting each measurement point location, as reported by the device's GPS, to the closest point on the reference rail track, ensuring that the spatial accuracy of the GPS data is improved. Unlike general dead reckoning methods, this approach benefits from the fixed-path nature of railways, simplifying the correction process.

While proximity mapping improves GPS accuracy, it's important to acknowledge that no solution is perfect. The effectiveness of mapping techniques depends on specific measurement scenarios, and there will always be cases where inaccuracies persist due to environmental constraints, interpolation limitations, or insufficient reference data [59].

Let $P_i = (x_i, y_i)$ represent the GPS coordinates of a measurement point at timestamp t_i, where $i = 1, 2, \ldots, n$. The rail track is represented as a line geometry, $R(s)$, parameterized by the track traversed distance s from the track starting point $P_0 = (x_0, y_0)$. Each GPS point P_i is projected onto the track by finding the point $R(s_i)$ that minimizes the orthogonal distance:

$$s_i = \arg\min_s \|P_i - R(s)\|. \tag{2.17}$$

This ensures all GPS points are aligned to the closest positions on the track.

For intervals where GPS data is missing (e.g., in tunnels), the positions and timestamps are interpolated along the track geometry based on the time elapsed between measurement points and the traversed distance along the track, assuming constant speed. This approach simplifies the interpolation process by reducing the problem to a one-dimensional interpolation, as opposed to interpolating separately for latitude and longitude coordinates. Given that all measurements are associated with timestamps, the traversed distance along the rail track can be computed based on the GPS positions and timestamps. The following steps are undertaken in the process:

- Identification of gap boundaries:
  - **Start point** - defined as the last known GPS position immediately preceding the gap, characterized by its timestamp t_1 and the distance traversed along the track d_1 (from the track start to that point).
  - **End point** - defined as the first known GPS position immediately following the gap, characterized by its timestamp t_2 and the distance traversed along the track d_2 (from the track start to that point).

The gap is thus defined by these two points, encompassing all missing positional data between them.

- Let:
  - $\Delta t = t_2 - t_1$ represent the elapsed time between the start and end points of the gap.
  - $\Delta d = d_2 - d_1$ denote the distance along the track between the start and end points of the gap.
  - t_i indicate the timestamp of the measurement within the gap.

  The interpolated distance from the track start, denoted as d_i, is then computed using the following formula:

$$d_i = d_1 + \frac{(t_i - t_1)}{\Delta t} \cdot \Delta d. \tag{2.18}$$

- Once the distance d_i for each measurement point is determined, the corresponding longitude and latitude coordinates along the track can be calculated as a function of the track geometry and the traversed distance. This is given by:

$$(\text{lon}_i, \text{lat}_i) = f(R(s), d_i). \tag{2.19}$$

By following these steps, we interpolate missing GPS measurements along the railway track using the known boundaries of the gap and the time-distance relationship. Our approach assumes a constant velocity between consecutive mapped points, a reasonable assumption based on existing literature. This method ensures a structured reconstruction of missing positions by leveraging both temporal and spatial track information, though further validation could be conducted to quantify its accuracy under varying train dynamics. Once the traversed distances are calculated, the corresponding geographic coordinates are derived, allowing for a precise reconstruction of the missing GPS data, which can then be used for further analysis or modeling of the train's trajectory.

2.4.4 Mitigating Spatial Confounding in Measurement Campaigns

Train speed and irregular stops can lead to spatial confounding, with varying numbers of measurements collected along the track. In measurement campaigns with constant measurement sampling intervals, the impact of train speed becomes particularly pronounced. The measurement frequency remains fixed, so for sections where the train is moving slowly, the space intervals between samples are smaller, resulting in higher spatial density of data points. This can lead to an overrepresentation of such track segments in the dataset, potentially skewing the analysis if not properly addressed. In contrast, track segments in which the train is moving with higher speed will have fewer data points, which may underrepresent the true signal behavior in these areas. To mitigate this, our framework segments the measurement path into fixed-length GUs, such as the 100 m intervals recommended by ITU guidelines in [60] for measurement aggregation in terrestrial mobile networks. While ITU specifically applies this approach

to environmental field strength predictions and coverage assessments, the analogy is valid for rail tracks, where consistent measurement intervals allow for more reliable analysis of connectivity.

The ITU suggests the use of 100 m × 100 m segments for aggregating measurements, particularly in urban and rural settings, to ensure consistency and comparability across measurements collected in different locations and conditions. In line with these recommendations, we apply 100 m length segments to the rail tracks for this study, leveraging their practical utility in minimizing the effects of train speed variations and enabling uniform data collection. Alternatively, we also incorporate 100 m × 100 m segments from the Statistik Austria raster [61], which is a widely used grid for mapping statistical data across Austria. These standardized grid-based segments align well with the ITU's suggestion for measurement aggregation.

By averaging measurement samples within each Geographical Unit (GU), the influence of the train speed is neutralized, resulting in a spatially uniform dataset. This process facilitates cross-campaign comparisons and ensures that the data is representative across different landscape types, network conditions, and infrastructure environments, such as rural and urban rail sections. In doing so, it also accommodates the temporal and spatial dynamics of train operations, ensuring that data is captured and analyzed in a way that reflects real-world conditions without introducing bias from measurement discrepancies caused by speed fluctuations. In addition, this approach provides confidence intervals for each segment, offering a measure of uncertainty and enhancing the reliability of the dataset.

The use of GUs as a basis for aggregation aligns with the ITU's approach, which emphasizes consistency in measurement methods to enhance the comparability and accuracy of data used for predicting field strength and mobile network coverage. Although ITU's guidelines were initially developed for more general mobile communications rather than rail tracks, the principles of segmentation and aggregation can be applied to rail environments, where factors such as terrain, infrastructure, and train movement affect signal propagation.

By implementing this approach, we ensure that the collected measurement data maintains consistency, reduces the impact of irregular train operations, and aligns with internationally recognized standards for geographic data aggregation.

2.4.5 Spatiotemporal Stability and Measurement Aggregation

When analyzing measurements in train environments, it is crucial to consider how data can be aggregated over multiple days and/or opposite driving directions to increase confidence in the results. This is particularly relevant for metrics such as RSRP, which is known to exhibit high stability over time. To validate this approach, we conducted a test drive over two consecutive days along the same track between Vienna and Bregenz, with the second day's drive performed in the opposite direction.

For consistency, we used the same UE for measurements on both days. The results were compared by evaluating the Empirical Cumulative Distribution Function (ECDF) of the RSRP values. To ensure a meaningful comparison, we filtered the measurements to include only those BSs sectors that were registered on both days. This filtering

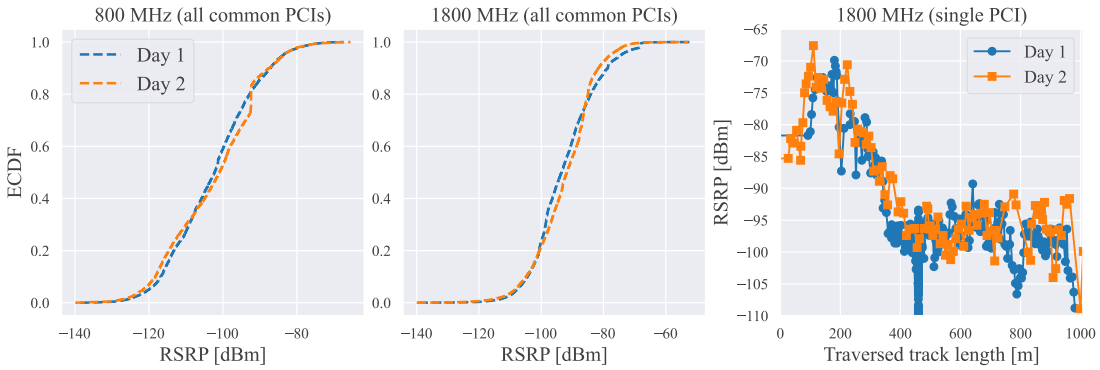


Figure 2.9: RSRP distribution from drive tests conducted in forward and backward driving directions on two consecutive days for two frequency bands: 800 MHz (left) and 1800 MHz (middle). The RSRP along the track is shown for a single BS sector in the forward direction on the first day and the backward direction on second day (right).

step was essential to focus the analysis on the relevant data and avoid the inclusion of unrelated measurements that could introduce noise.

The comparison of the ECDFs, as illustrated in Fig. 2.9, provides further validation of the measurement repeatability:

- The left two plots in Fig. 2.9 depict the ECDF of RSRP for the 800 MHz and 1800 MHz frequency bands, respectively. These results represent all common PCIs observed across both days, comparing the test drives conducted in forward and backward directions. A single UE was used for measurements, positioned at the same location in the train on both days to ensure consistency.
- The third plot in Fig. 2.9 (right) demonstrates an example BS sector, showing how the RSRP curve evolves over 1 km of the traversed track. The general trend of RSRP remains stable over time, with minor fluctuations likely caused by SSF within the train cabin. Importantly, the forward and backward directions exhibit excellent alignment.

The ECDF results confirm three key observations:

1. **Directional Consistency**: The measurements are not significantly affected by the direction of travel. Whether traveling from Vienna to Bregenz or in the reverse direction, the RSRP values remain stable.
2. **Feasibility of Aggregation**: The high degree of overlap indicates that measurements from multiple days can be aggregated to improve statistical confidence without compromising the accuracy of the results.
3. **Significance of Spatial Confounding**: On Day 2, both the 800 MHz and 1800 MHz curves in the ECDFs exhibit a noticeable jump, which can be attributed

to the train remaining stationary for an extended period while measurements continued. This highlights the importance of mitigating such artificial effects through proper data segmentation and averaging, as previously discussed, to ensure a more accurate representation of network performance.

To verify the validity of aggregating measurements from forward and backward directions for future analyses, we performed a two-sample Kolmogorov-Smirnov (KS) similarity test (see Appendix A2 for a detailed explanation). This test was applied to segment-averaged RSRP distributions to assess whether they originate from the same underlying distribution. The resulting p-values were 0.409 for the 1800 MHz band and 0.956 for the 800 MHz band, indicating no significant differences between the distributions collected in opposite driving directions. A p-value represents the probability of observing a test statistic as extreme as, or more extreme than, the one calculated from the data, under the assumption that the null hypothesis H_0 is true. When the p-value is greater than 0.05, it suggests that there is insufficient evidence to reject the null hypothesis. In this context, the high p-values indicate that the RSRP distributions for the forward and backward directions are statistically similar. These findings confirm that aggregating RSRP measurements from both directions and over different time periods is statistically sound and introduces no bias in the analysis.

We have shown that multiple test drives can be aggregated to enhance the robustness of evaluations and modeling in train scenarios. This is possible because the distributions of key metric, i.e.RSRP, does not change significantly over time. By aggregating multi-day data, we enhance the robustness of evaluations and modeling in train scenarios, reducing variability and providing a stronger basis for identifying trends and making informed decisions.

2.4.6 Filtering the Dataset

The data is filtered to focus on specific KPIs of interest, such as RSRP, T, and other network performance metrics, while ensuring that measurements are collected under relevant conditions. Filtering criteria are applied to exclude irrelevant or secondary data that could obscure meaningful patterns. Specifically, we can choose to retain only measurements from PCell to ensure that the analysis accurately reflects the end-user experience.

This approach is particularly relevant because the PCell serves as the main communication link between the UE and the network, providing a stable and continuous dataset for analysis. In contrast, SCells, which are utilized in carrier aggregation, are activated dynamically based on network conditions and may not always be present. Since their availability depends on scheduler decisions, network load, and signal conditions, including them in general performance analysis could introduce inconsistencies.

Given the diverse railway environments - urban, rural, and tunnels - filtering for PCells helps maintain propagation consistency and prevents measurements from being skewed by transient SCell activity. This filtering improves the reliability of the data for propagation analysis and enables more accurate assessment of network performance across different railway scenarios.

2.4.7 Stratifying the Dataset

Stratification is a systematic method for partitioning a dataset into meaningful subsets based on predefined criteria. In railroad measurement campaigns, this typically involves categorizing environments - such as urban, rural, forest, water, and tunnel areas - using landscape classes derived from databases like the CLC. This approach enables precise comparisons within similar environmental settings and ensures that key factors affecting wireless communication performance are accounted for. By isolating the unique propagation characteristics of each environment, stratification enhances the reliability and relevance of the analysis.

Stratification is particularly important in railroad contexts due to the diverse and challenging environments rail corridors traverse. Urban areas, for example, are characterized by dense infrastructure and multipath reflections, while tunnels require specific solutions for severe signal attenuation. Rural and forested areas, on the other hand, face challenges such as sparse infrastructure and vegetation-induced losses. Stratifying measurements helps ensure that performance is evaluated comprehensively across these varied conditions, enabling environment-specific evaluations and optimization strategies.

Moreover, stratified sampling improves efficiency by reducing the number of measurements required while preserving representativeness. Instead of exhaustive data collection, well-defined strata allow researchers to focus on key areas, ensuring balanced coverage of different environments. This is crucial for benchmarking performance in diverse scenarios and optimizing resource allocation during measurement campaigns.

In Chapter 6, we will explore how stratification can optimize measurement collection strategies by identifying the minimum number of measurements or areas needed to ensure accurate performance estimates. This approach not only streamlines data collection but also ensures robust evaluations of wireless communication systems.

2.4.8 Summary

The framework discussed in this chapter offers a systematic and comprehensive approach to the design, execution, and analysis of cellular connectivity measurement campaigns. By incorporating tools and datasets for precise mapping and spatial context, the framework ensures accurate alignment of measurements with real-world geography. It addresses spatial confounding by employing GU-segmentation, which enables localized analyses and reduces biases caused by environmental variability. Furthermore, the integration of data filtering and stratification techniques allows for focused analyses by isolating key KPIs, environments, or frequency bands, ensuring the data remains relevant and representative. These elements collectively enhance the robustness, reproducibility, and scalability of the methodology, making it suitable for a wide range of benchmarking scenarios, from urban performance comparisons to evaluations in complex environments like tunnels or rural areas. This structured approach not only optimizes resource use but also ensures that the results support effective network optimization and future deployments.

Chapter 3

Railway Measurement Datasets

Ensuring reliable mobile connectivity aboard trains is essential for both passengers and railway operators. Modern trains are typically equipped with rooftop antennas to facilitate mobile communication services, including eCall functionality [62]. These antennas receive mobile signals, which are then distributed within the cabin via onboard WiFi infrastructure[1]. However, the quality of the connectivity experienced by passengers is influenced by multiple factors, such as high VPL, frequent handovers due to mobility, and fluctuating network coverage along railway routes.

To capture and understand these challenges, we conducted a series of controlled measurement campaigns aboard moving trains under realistic conditions. These campaigns enabled the systematic collection of both indoor and outdoor RSRP measurements using a combination of professional scanners and UEs. The resulting datasets were then cleaned and postprocessed using the framework outlined in Section 2.4 to ensure consistency and usability for further evaluation.

Compared to simulation-based studies or purely crowdsourced approaches, controlled campaigns offer higher reliability and precision. While simulation tools, such as data-driven methods [63] or deterministic RT models [64], are valuable for estimating outdoor signal strength, they cannot fully capture real-world complexities - especially in outdoor-to-indoor propagation scenarios. Crowdsourced data via MDT [65] offers scale, but suffers from limitations such as lower localization accuracy and difficulty in differentiating between indoor and outdoor reception. These limitations are further discussed in Section 7.1. We prioritize precise data collection under known conditions, which is essential for evaluating and comparing connectivity solutions onboard trains.

One key focus of our measurement efforts was VPL, which significantly affects the signal levels observed inside the train. While the most accurate way to analyze VPL would involve full-scale anechoic chamber testing - potentially with 3D scanning to account for complex train geometries - this approach is infeasible in practice. Instead, we assessed penetration loss directly over operational cellular networks by comparing signal strength at the rooftop antenna and within the cabin, offering a realistic and scalable alternative.

Data collection was made possible through collaboration with ÖBB, which granted access to a variety of train types and enabled efficient campaign execution. Since installing custom rooftop antennas was not permitted due to safety regulations, we

[1] https://nomad-digital.com/solutions/connect/broadband-wifi-rail-connectivity

utilized the preinstalled antennas on each train. As a result, rooftop measurement conditions varied across train types.

To enable geospatial and propagation-related analysis, access to BS location data was critical. This was made possible through collaboration with A1 Telekom Austria, who provided anonymized BS location data. Although this information is not publicly shared, it was essential for enriching the datasets - for example, by allowing us to compute the distance from each measurement point to the serving BS.

The datasets were collected on three distinct ÖBB train types - Talent, Railjet, and Nightjet - across different railway routes and cabin configurations. Our measurements cover a range of connectivity solutions, including standard and modified windows as well as onboard repeaters. The resulting datasets are not only used throughout this thesis but also represent a valuable contribution to the research community. A subset of the data has been made publicly available, facilitating reproducibility and further studies on mobile connectivity in railway environments.

A summary of all datasets used in this thesis is provided in Table 3.5, including information on train types, tracks, measurement equipment, connectivity solutions, and sample counts. An illustration of the tested trains and their respective window coating types is shown in Fig. 3.6.

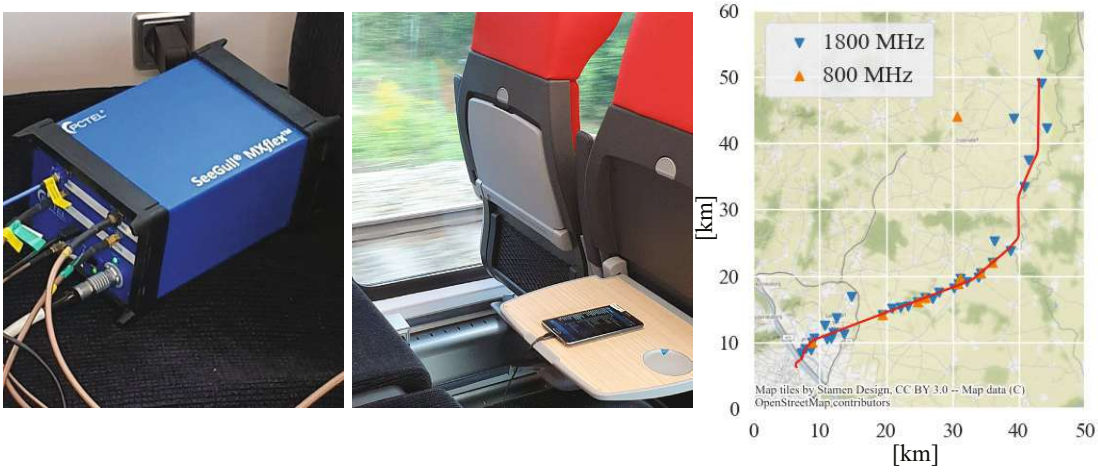
3.1 Dataset 1: ÖBB Talent

The first dataset was collected on an ÖBB Talent railcar (pictured in Fig. 3.6a) operating along the 65 km-long Nordbahn track, as shown in Fig. 3.1c. This track was chosen due to the absence of repeater chains along its path, ensuring that the reported Physical Cell Identifier (PCI) values corresponded directly to distinct BS antennas. This was essential for accurately computing the AoA from the BS antenna to the train antenna and the user, as detailed in Section 5.1.1, and based on the work originally published in [14]. The train reached speeds of up to 230 km/h. Its cellular connectivity setup did not include a repeater but featured prototypical Modified Windows Type 1 (MW1), as depicted in Fig. 3.6d.

This dataset captures detailed signal strength measurements collected from four UEs strategically positioned inside the moving train. It also includes signal strength measurements obtained from an external reference source using the rooftop train antenna, systematically interpolated on a cell and frequency basis. In addition, it contains calculated elevation and azimuth AoA, along with all supporting variables necessary for their computation, including UE location data, frequency bands, BS identification, and geometric relationships between the train and BSs. Environmental factors such as BS antenna height, ground elevation differences, and distance metrics are also included to enhance the accuracy of the analysis. We have made this dataset publicly available in [66] to support further research on railway connectivity and to enable reproducible analysis.

Table 3.1: BS count and sample share per frequency band in ÖBB Talent.

Frequency Band	BS Count	Sample share (%)
1800 MHz	63	83
800 MHz	13	17



(a) (b) (c)

Figure 3.1: Measurement setup in ÖBB Talent with MW1. (a) The scanner was linked to the external rooftop-mounted railcar antenna. (b) Inside the railcar, four UE devices were strategically positioned on seats and fold-down tables. (c) Traversed Nordbahn track with BS locations.

3.1.1 Measurement Setup

Outdoor Measurements: Conducted using a PCTel MXflex scanner (Fig. 3.1a) connected to the preinstalled rooftop antenna, capturing a variety of network parameters in LTE B20 (800 MHz) and B3 (1800 MHz) bands every 200 ms.

Indoor Measurements: Recorded using four Samsung Galaxy Note 4 UEs running Keysight NEMO software [52], placed on fold-down tables to simulate real passenger conditions (Fig. 3.1b). UEs ran a scripted download using NEMO software, ensuring sustained traffic throughout the measurements, while collecting a broad range of KPIs.

GPS Measurements: The phones recorded GPS measurements along the drive. To ensure accurate localization, railway track coordinates were extracted from OSM, mapping each measurement to the nearest track location and interpolating missing GPS data, as described in Section 2.4.3.

The measurements were conducted across multiple BSs operating in two frequency bands. Table 3.1 provides an overview of the BS count and the distribution of measurement samples between the 1800 MHz and 800 MHz bands. The majority of samples (83%) were recorded in the 1800 MHz band, which also had a significantly higher number of BSs compared to the 800 MHz band.

3.2 Dataset 2: ÖBB Railjet

The second dataset was collected on the 490 km-long the Vienna-Innsbruck railway track (Fig. 3.2) using an ÖBB Railjet train (Fig. 3.6b) to analyze the impact of different wagon configurations on mobile network performance [13]. To fairly compare the connectivity solutions, two measurement positions were selected per wagon: one on a fold-down table and another at head height. Positional dependency was previously analyzed in Section 2.2.2 in a static scenario, indicating minimal positional differences inside the same wagon. The train consisted of three different types of railroad cars: one with standard windows and no AAF repeater, a second with modified windows [67] (Fig. 3.6e) and no AAF repeater, and a third with standard windows equipped with an AAF repeater and a leaky feeder installed on the ceiling. Throughout this thesis, these configurations characteristic for the Railjet train are referred to as Standard Windows (SW), Modified Windows Type 2 (MW2), and Repeater, respectively. By maintaining identical indoor placements across all three wagon types and comparing signal quality under real-world conditions, this dataset enables a systematic evaluation of different window modifications and signal enhancement strategies.

The measurements were collected within a single day, with the train operating from Vienna to Innsbruck in the morning and returning from Innsbruck to Vienna in the afternoon. Measurements were conducted in both travel directions at identical indoor locations to ensure repeatability. The number of collected measurements across two indoor positions in each wagon, combined with the outdoor antenna measurements, amounted to $\approx 330\,000$ per trip, resulting in a total of $\approx 660\,000$ measurement samples. Table 3.2 summarizes the train speed profile and landscape statistics as defined in the CORINE project. Table 3.3 presents the frequency band distribution of the dataset.

3.2.1 Measurement Setup

Data collection was carried out using the ACT Rail system, operated by Focus Infocom GmbH, which served as the UE testbed for the measurement campaign. The UE fleet included the following phone models: Sony Xperia XZ, Sony Xperia XZ2, and Samsung Galaxy Note 10+. The phones were connected to a centralized remote controller, allowing the tester to start and stop all measurement scripts simultaneously. The testbed setup is depicted in Fig. 3.3. The measurement script executed a continuous sequence of data transmissions and voice calls to maintain active traffic throughout the experiment.

Outdoor measurements: The outdoor antenna mounted on the train's rooftop was connected to an indoor distribution unit, to which the measurement setup, i.e., another UE, was connected.

Indoor measurements: To ensure comparability between configurations, testbeds with UEs were installed at predefined, identical locations inside each of the three railroad cars. Figure 3.4 illustrates the testbed positions inside the railroad cars.

Table 3.2: Speed and landscape statistics of the dataset in ÖBB Railjet.

Rail Length	v_{max}	v_{mean}	Urban	Rural	Forest & Water	Tunnel
490 km	230 km/h	115 km/h	47%	40%	7%	6%

Table 3.3: Frequency band statistics of the dataset in ÖBB Railjet.

Band 3 (1800 MHz)	Band 20 (800 MHz)	Band 1 (2100 MHz)	Bands: 7, 8, 28, 32 (combined)
45%	22%	23%	10%

Figure 3.2: 490 km Vienna-Innsbruck track (ÖBB Railjet).

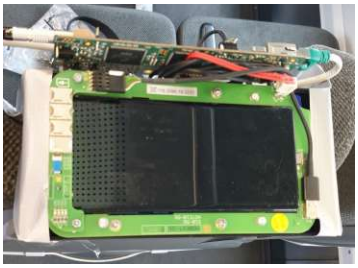


Figure 3.3: UE testbed (ÖBB Railjet).

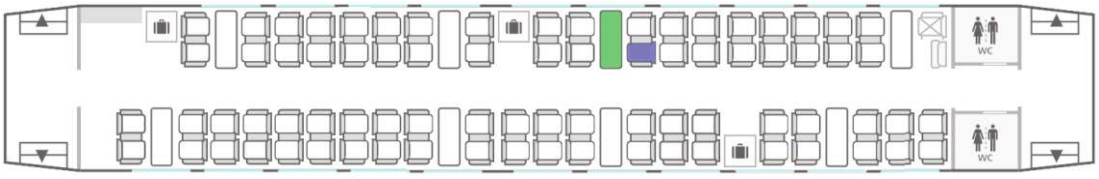


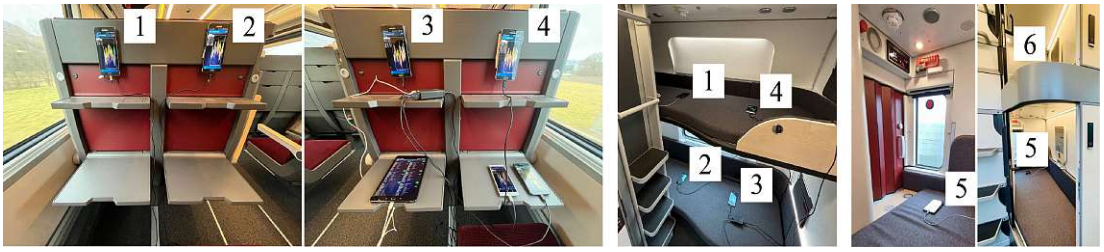
Figure 3.4: Testbeds deployed at identical positions inside three separate wagons of the ÖBB Railjet, each equipped with a different configuration: SW, MW2, or Repeater. One UE is located on the seat headrest (blue) and another UE is located on the table (green).

3.3 Dataset 3: ÖBB Nightjet

The ÖBB Nightjet train was selected to investigate the influence of window size on signal strength inside cabins. The Nightjet features multiple cabin layouts with varying window dimensions, making it an ideal test case. Additionally, it employs the latest generation of window modifications, Modified Windows Type 3 (MW3), allowing comparisons with the MW1 and MW2 modifications used in the Talent and Railjet trains, respectively. The seating coach layout across all three trains is similar, facilitating direct performance comparisons, which are further analyzed in Section 5.2.4.

Table 3.4: Window sizes and VPL in ÖBB Nightjet cabin types.

	Window Size	Window Area / Wagon	Mean VPL
Seating Coach	1.218 m^2	$\approx$ 22 m^2	14.3 dB
Deluxe Cabin	1.044 m^2	$\approx$ 22 m^2	17.4 dB
Mini-suite Cabin	0.583 m^2	$\approx$ 17.5 m^2	18.4 dB



(a) Seating Coach (left and right) (b) Deluxe Cabin (c) Mini-Suite Cabins

Figure 3.5: Measurement setup in ÖBB Nightjet with MW3.

All cabin types in the train were equipped with MW3 pattern, as shown in Fig. 3.6f, while the specific window sizes for each cabin type are listed in Table 3.4. The train contained three different cabin configurations, illustrated in Fig. 3.5:

Seating Coach Seating layout similar across the Talent, Railjet, and Nightjet trains, with window sizes comparable to the Railjet.

Deluxe Cabin Premium cabin layout featuring a door and moderately sized windows.

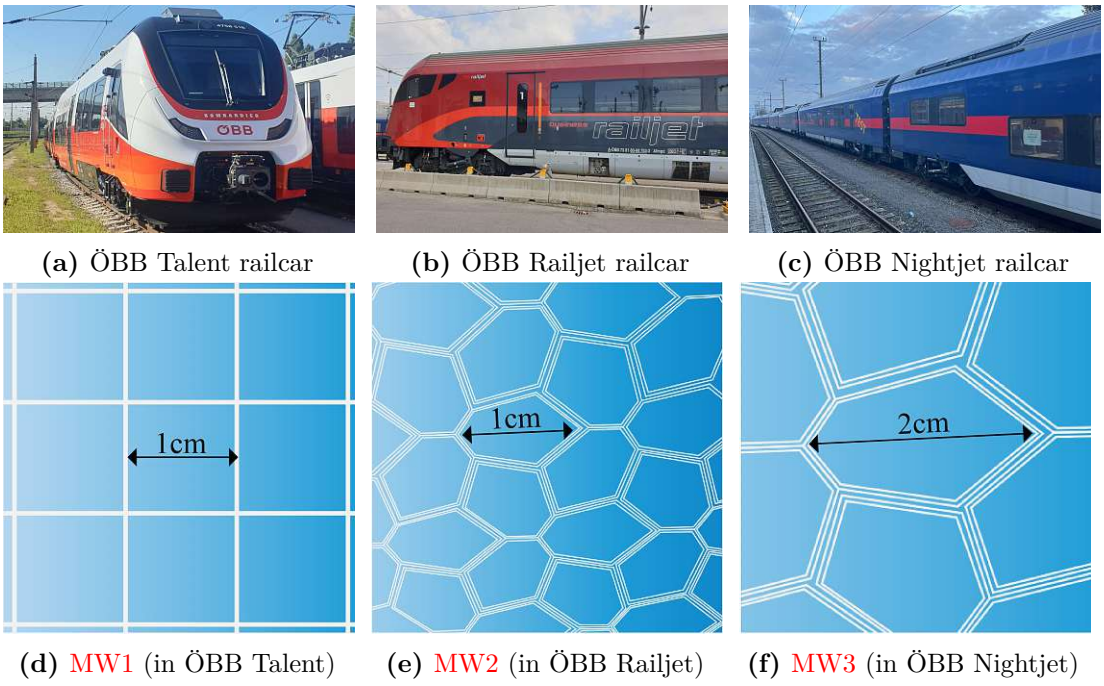
Mini-suite Cabin Compact sleeping box with a closed door and significantly smaller windows.

We conducted measurements over two consecutive days, covering the route from Vienna to Bregenz on the first day and the return journey from Bregenz to Vienna on the second. Following postprocessing, we made the resulting dataset publicly available for download at [68] to support further research on railway connectivity.

3.3.1 Measurement Setup

To evaluate cellular network performance inside the train, measurements were performed using Keysight NEMO software [52] running on six Samsung Galaxy Note 4 smartphones, each equipped with a Subscriber Identity Module (SIM) card from a major Austrian operator. Each device was configured specifically for LTE cellular technology.

On the first measurement day, UEs 1-4 were placed inside the Deluxe Cabins, while UEs 5 and 6 were positioned in the Mini-suite Cabins (Fig. 3.5b and Fig. 3.5c). Phones were arranged on both upper and lower racks to analyze the impact of height on signal attenuation. On the second day, UEs 1-4 were relocated to the Seating Coach, with the phones evenly distributed between the left and right sides of the cabin (Fig. 3.5a).



(a) ÖBB Talent railcar **(b)** ÖBB Railjet railcar **(c)** ÖBB Nightjet railcar

(d) MW1 (in ÖBB Talent) **(e)** MW2 (in ÖBB Railjet) **(f)** MW3 (in ÖBB Nightjet)

Figure 3.6: Top: Three types of trains under test. Bottom: Patterns of metal coating in modified windows. The white lines indicate perforations in the metallic shield coating, forming the frequency selective structure.

In addition to the smartphones measuring indoor performance, a reference measurement was obtained using a Keysight NEMO smartphone equipped with external antennas. This device simultaneously recorded signals from the rooftop antenna at the MAIN port and a cabin antenna at the MIMO port. Through data preprocessing - including track alignment and segmentation [13] - these reference signals were analyzed to assess the attenuation experienced by the distributed test UEs within the cabin. The impact of UE placement on attenuation is further discussed in Section 5.2.2.

Table 3.5: Summary of datasets collected on ÖBB trains.

Dataset 1: ÖBB Talent	
Track Name	Nordbahn (Wien - Breclav)
Track Length	65 km
Wagon Type	Wagon 1: Seating Coach
Connectivity Solutions	Wagon 1: Modified Windows (MW1)
Frequency Bands	LTE B3, B20
Outdoor Measurement Equipment	PCTel MXflex scanner with rooftop antenna
Indoor Measurement Equipment	4 × Samsung Galaxy Note 4 (Keysight NEMO)
Total Data Points	≈ 150 000
Publicly Available	Yes, at [66]
Dataset 2: ÖBB Railjet	
Track Name	Westbahn (Vienna - Innsbruck)
Track Length	490 km (980 km round trip)
Wagon Type	Wagon 1: Seating Coach Wagon 2: Seating Coach Wagon 3: Seating Coach
Connectivity Solutions	Wagon 1: Standard Windows (SW) Wagon 2: Modified Windows (MW2) Wagon 3: AAF Repeater (Repeater)
Frequency Bands	LTE B1, B3, B7, B8, B20, B28, B32
Outdoor Measurement Equipment	PCTel MXflex scanner with rooftop antenna
Indoor Measurement Equipment	Sony XZ, XZ2, Samsung N10+
Total Data Points	≈ 660 000
Publicly Available	No (Focus Infocom GmbH ownership)
Dataset 3: ÖBB Nightjet	
Track Name	Westbahn (Vienna - Bregenz)
Track Length	750 km (1 500 km round trip)
Wagon Type	Wagon 1: Seating Coach Wagon 2: Deluxe Cabins Wagon 3: Mini-Suite Cabins
Connectivity Solutions	Wagon 1: Modified Windows (MW3) Wagon 2: Modified Windows (MW3) Wagon 3: Modified Windows (MW3)
Frequency Bands	LTE B1, B3, B7, B8, B20
Outdoor Measurement Equipment	NEMO smartphone with external antennas
Indoor Measurement Equipment	6 × Samsung Galaxy Note 4 (Keysight NEMO)
Total Data Points	≈ 3 800 000
Publicly Available	Yes, at [68]

Chapter 4

Comparative Performance Analysis of Mobile Connectivity Solutions on Railways

In recent years, Austrian railways have undergone significant reconstruction and modernization to meet climate neutrality goals outlined in [69]. This transformation has coincided with a steady increase in the number of commuters using trains not only as a means of transportation but also as mobile offices, necessitating reliable internet access [70, 71]. To meet this demand, train operators and MNOs have focused on improving network services, including optimizing BS deployments along railway routes. Improving BS deployment along railways requires analyzing current KPIs, e.g., RSRP, along particular routes, thus allowing us to tackle problematic signal gap areas. However, even with railway-dedicated BS deployments, maintaining consistent in-train connectivity remains a formidable challenge, due to the high VPL. A survey of the challenges associated with delivering high capacity connectivity to train users is provided in [72].

Train metal structure combined with the metal coating of the windows, used mainly for isolation to the infrared sunlight, act as a Faraday cage for the electromagnetic signals. Our measurements demonstrate that this metal structure typically attenuates mobile signals $\approx 20\,\mathrm{dB}$ at the minimum. To address this challenge, two categories of VPL reduction solutions are explored: active and passive. Active solutions involve systems that dynamically amplify and redistribute signals, requiring external power and frequent configuration to adapt to network requirements. Passive solutions, in contrast, rely on structural modifications, such as specialized coatings on train windows, to inherently improve signal penetration without the need for external power or ongoing adjustments.

In this chapter, we analyze these VPL mitigation strategies based on our earlier work published in [13]. In Section 4.1, we examine active repeater systems, while Section 4.2 focuses on state-of-the-art window modifications. In Section 4.3, we describe the experimental setup designed for a fair comparison of both solutions, building on the framework introduced in Chapter 2. We then present the comparison results and draw conclusions regarding the effectiveness and practical feasibility of these approaches for improving in-train connectivity.

4.1 Repeater Systems in Trains

AAF repeater systems are vital for improving mobile connectivity in trains, where high VPL hinders signal quality. In recent years, cellular repeaters have become standard in most intercity trains across Europe. So far, there has been extensive research focusing on expanding the coverage for cellular users onboard trains using static repeater deployments [73], [74]. Operated by railway companies, these repeaters significantly mitigate the signal attenuation caused by the metallic structure of trains, thereby improving connectivity for both passengers and railway operations. They are particularly effective in rural areas, where sparse cellular deployment and higher path loss typically constrain data rates.

The physical architecture of a train repeater system typically consists of a rooftop pick-up antenna, an amplifier and an interior distribution system. By capturing cellular signals received via a roof-mounted pick-up antenna and actively amplifying them, repeater systems overcome the high attenuation problem posed by the metal structure and metal-coated windows in trains. The amplified signals are then distributed throughout the train via a leaky feeder cable mounted inside the train, typically on the car ceiling or floor, as depicted in Fig. 4.1. In the Uplink (UL), signals from passenger UEs are similarly amplified and sent back to the cellular network via the roof antenna. This configuration minimizes overall PL while maintaining the transparency required by mobile operators [75, 76].

Despite their benefits, train-mounted repeaters must operate across multiple countries, accommodating diverse cellular operators and configurations. To ensure seamless integration, these systems are generally implemented as simple AAF devices. This approach avoids advanced functionalities like Integrated Access Backhaul or Non-Contiguous Radio, which require direct operator control. Instead, AAF-based repeaters maintain transparency to cellular networks by employing uniform gain in both the DL and UL directions.

Research into vehicular repeater deployments in trains is limited, as it requires collaboration between railway and cellular operators and is highly time and cost demanding. Nevertheless, a growing body of work provides insights into system models [75, 76], deployment practices [77, 78], and the non-linear effects associated with AAF repeaters in railway environments [79–83].

The following sections outline repeater advantages, disadvantages, technical description, and general reasons for industry's shift toward passive solutions.

4.1.1 Repeater Systems: Benefits and Limitations

Repeater systems offer a promising solution for improving mobile connectivity within train environments, particularly by addressing the issue of high VPL. One of their key advantages is the ability to extend network coverage into areas where direct signal penetration is weak or nonexistent - such as tunnels, remote routes, deep cuttings or sound barriers - ensuring more continuous service throughout the journey. By amplifying signals received by rooftop antennas, repeaters mitigate cabin attenuation and help maintain sufficient indoor signal levels, even under weak outdoor coverage. This signal

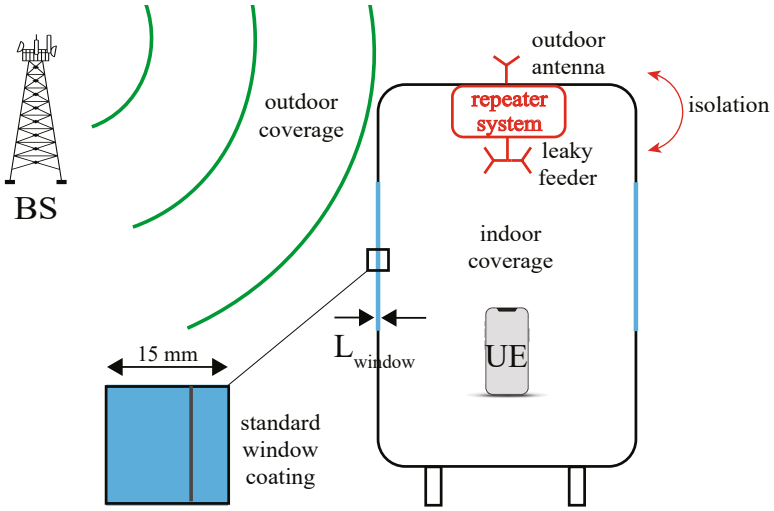


Figure 4.1: Schematic representation of a in-train repeater system.

boosting improves both voice call quality and data throughput, leading to a more stable and satisfactory passenger experience.

Modern repeater systems also offer multi-operator support, enabling simultaneous service for passengers using different mobile network operators. This makes them a viable shared infrastructure solution. Furthermore, their scalability allows the signal to be distributed across multiple cars, making repeaters adaptable to varying train configurations and service requirements.

However, these systems are not without drawbacks. Repeaters require an active power supply, and their energy demand increases with higher output power. Given that trains must allocate power across many systems - propulsion, climate control, passenger services - energy efficiency is a key constraint. Additionally, repeaters have a maximum output power limit. In areas with strong external coverage, the repeater may saturate, inadvertently degrading the signal quality instead of enhancing it.

Repeaters also demand ongoing maintenance, including recalibration and occasional hardware replacements, which can increase operational costs. Functionally, they operate at the physical layer and cannot manage network congestion or allocate bandwidth. As a result, during peak usage scenarios, passengers may experience degraded performance due to the lack of traffic-aware optimization.

Their long-term compatibility is another concern. While trains remain in service for decades, mobile network technologies evolve every ten years. This disparity can lead to compatibility issues, requiring repeaters to be updated or replaced to support future technologies and frequency bands.

From a signal-processing standpoint, repeaters introduce a fixed delay (typically around 4.7 μs), which can exceed the Cyclic Prefix (CP) duration in LTE systems. This delay introduces ISI, reduces T, and forces UEs to transmit at higher power without improving the Signal-to-Interference Ratio (SIR). Mitigating this requires costlier design

measures such as tighter processing delays, increased output power, or improved window isolation [83, 84].

Finally, repeaters amplify not only the intended signal but also background thermal noise, particularly in the UL. When the repeater's gain exceeds the outdoor PL, this noise can degrade performance not only for onboard users but also for nearby UEs outside the train. Imperfect window shielding can further allow direct signal leakage and interference. In such cases, the maximum repeater gain must be reduced to stay below the isolation threshold - typically by 15 dB - to prevent instability, where isolation refers to the attenuation between the rooftop pick-up antenna and the internal leaky feeder system.

In summary, repeater systems can significantly enhance onboard connectivity when deployed carefully and within operational constraints. However, their energy demand, limited intelligence, noise amplification, and long-term compatibility challenges require careful consideration to ensure long-term effectiveness.

4.1.2 Repeater Systems: Architecture and Functionality

The repeater model, as presented in [84] and based on a typical commercial multi-band repeater [85], is depicted in Fig. 4.2. Like most repeaters, it consists of separate DL and UL paths, characterized by nominal gain ($G_{\mathrm{X},0}$), maximum output power ($P_{\mathrm{X,out,max}}$), and additive noise ($P_{\mathrm{n,X}}$) [1]. While this architecture is common across repeater designs, the chosen model serves as a representative example for analyzing performance in the railway environment. The actual gain and output power are expressed as:

$$G_{\mathrm{X}} = \min(G_{\mathrm{X},0}, P_{\mathrm{X,out,max}} - P_{\mathrm{X,in}}), \tag{4.1}$$

$$P_{\mathrm{X,out}} = \min(P_{\mathrm{X,out,max}}, P_{\mathrm{X,in}} + G_{\mathrm{X},0}). \tag{4.2}$$

As illustrated in Fig. 4.3, when the input power does not saturate the amplifier ($P_{\mathrm{X,in}} + G_{\mathrm{X},0} \leq P_{\mathrm{X,out,max}}$), the actual gain equals the nominal gain. Otherwise, the gain is reduced to ensure $P_{\mathrm{X,out}} = P_{\mathrm{X,out,max}}$. Therefore for strong coverage on the track (or equivalently low PL) the repeater adjusts its gain in the DL direction to stay withing its maximum output power. To minimize UL noise, gain trailing dynamically adjusts the UL gain based on the DL gain:

$$G_{\mathrm{UL}} = \min(G_{\mathrm{DL}}, P_{\mathrm{UL,out,max}} - P_{\mathrm{UL,in}}), \tag{4.3}$$

$$G_{\mathrm{DL}} = \min(G_{\mathrm{X},0}, P_{\mathrm{DL,out,max}} - P_{\mathrm{DL,in}}). \tag{4.4}$$

Considering a single BS with a transmit power of P_{TX}, the repeater experiences saturation when the outdoor PL (from the BS to the train rooftop antenna), is below the threshold $P_{\mathrm{BS}} - P_{\mathrm{DL,out,max}} + G_{\mathrm{DL}}$. This typically occurs in areas of strong signal coverage. Saturation, along with other limitations such as signal delay, additive noise, and compatibility issues, can significantly diminish the effectiveness of repeaters in certain scenarios.

[1] X stands either for DL or UL, depending on the path taken.

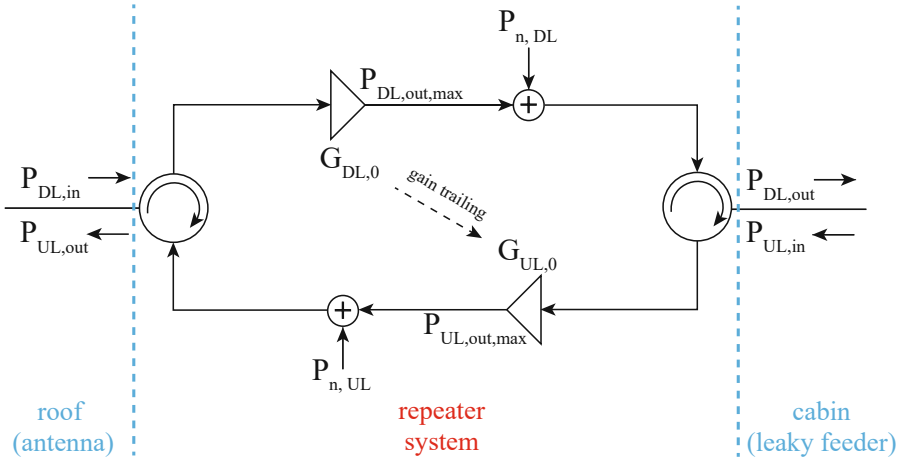


Figure 4.2: Block diagram of the in-train repeater.

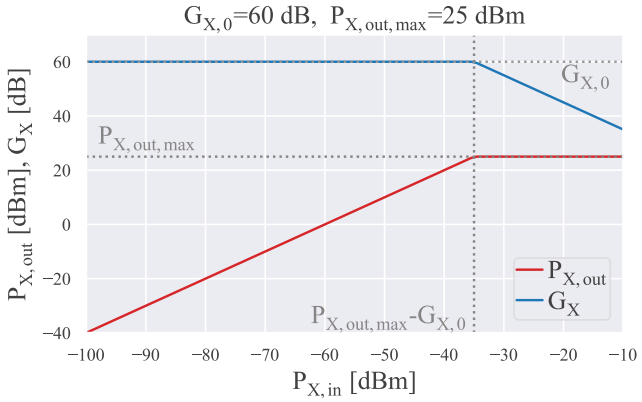


Figure 4.3: Gain control of the repeater shows a region of fixed gain at low and medium input power, and a region of fixed output power at high output power [84].

Due to these drawbacks, the railway industry is increasingly exploring passive alternatives. One promising approach is the implementation of frequency-selective window patterns, which enhance signal penetration into train cabins without the added complexity and potential inefficiencies of active systems. These passive solutions will be discussed in greater detail in the Section 4.2.

Additionally, repeater hardware constraints, including output power limitations and susceptibility to saturation, are further analyzed in Section 4.3. This analysis is based on real-world measurements we conducted in a live network with a train in operation. The results provide valuable insights into how these hardware limitations impact overall system performance and passenger connectivity. Moreover, they offer a comparative perspective on the benefits of passive window solutions over traditional active repeaters.

4.2 Frequency Selective Window Coatings

A train can be thought of as a metal box where the primary entry points for wireless signals are its glass windows. These windows, however, are often coated with metallic layers to prevent heat dissipation from the cabin and to block harmful UV radiation. While this design serves energy efficiency and passenger comfort, it introduces a significant drawback: the strong attenuation of cellular signals. To address this issue, manufacturers adopted the concept of patterned coatings, an idea originally developed for buildings and later applied to trains. This approach involves introducing periodic patterns into metallic coatings, effectively creating a FSS. These patterns allow specific frequencies of electromagnetic waves to pass while maintaining thermal and UV protective properties. Fig. 4.4 illustrates the impact of different types of coatings - plain glass (no coating), fully coated glass, and glass with a frequency-selective patterned coating - on cellular signal transmission, UV radiation blocking, and heat dissipation. Metallic coating is a requirement in terms of energy efficiency for railroad operators. However, it causes cellular signal attenuation larger than 20 dB [86, 87].

Early studies have demonstrated the feasibility of FSS coatings in reducing signal attenuation while maintaining energy efficiency in architectural applications [88, 89]. Additionally, the use of FSS in indoor wireless environments to control radio wave propagation by selectively increasing transmission loss or channeling signals has been investigated in [90]. Building on these findings, the concept was extended to train windows, where the challenges of maintaining signal transmission at higher frequencies became even more pronounced, due to abrupt changes in channel conditions [91]. Studies on FSS for railroad deployment have been increasing in recent years, with the development - in [92], the authors summarize the representative studies on propagation loss of rail vehicle window glass. Ongoing research is focused on improving window modification techniques, specifically by incorporating FSS to optimize signal penetration [67, 93]. By selectively removing 4 to 5% of the metallic window coating in FSS-based patterns, this approach enhances wireless signal transmission at specific frequencies while preserving thermal insulation [88, 94]. In [95], researchers investigated the laser ablation patterning of glass coatings using both measurements and simulations, with patterns ranging from 2 to 40 mm grid size. Results demonstrated that a mere 4% removal of the coating results in a twofold increase in transmission within the infrared domain. In [96] a theoretical study has been carried out and compared with experimental measurements of commercial windows.

Depending on the geometry and arrangement of the FSS, these patterns can function as low-pass, high-pass, band-pass, or band-stop filters [97] - some examples are depicted in Fig. 4.5. By using more complex designs, combinations of these characteristics can also be achieved. Even simple structures can be finely tuned for specific frequencies by adjusting two critical parameters: the periodicity of the pattern, D, and the width of the pattern elements, w (both denoted in Fig. 4.4d). Smaller D values enable operation at higher frequencies, while narrower w values minimize the area ablated from the coating, preserving the optical and thermal properties of the window. The low-pass case in Fig. 4.5a is one of the patterns we had the opportunity to measure (MW1). This type of pattern is often used in trains, optimized for frequencies below 5 GHz to act as a

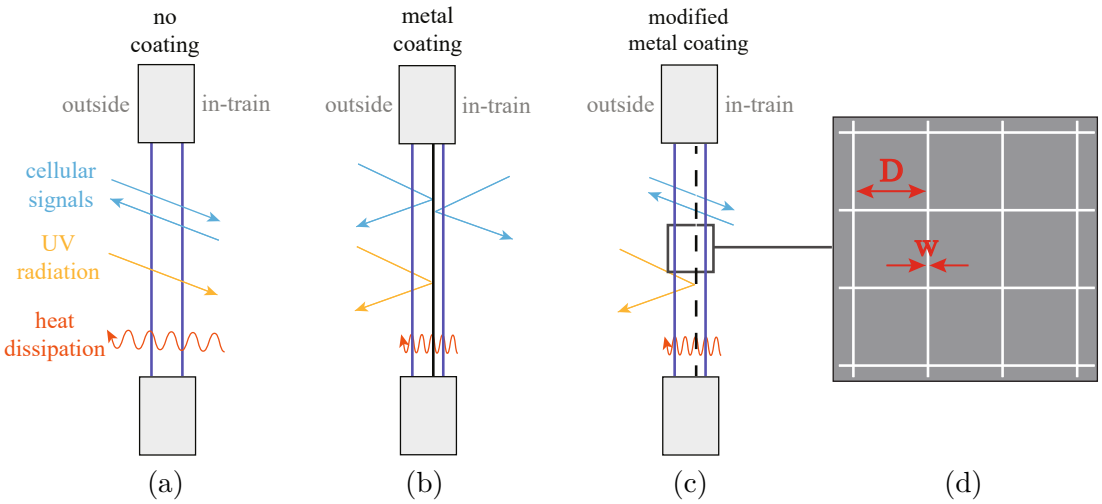


Figure 4.4: Double glass pane without metallic coating (a), with low-e coating (b), with structured low emissivity coating (c). (d) Coating periodicity D and width w of the pattern elements.

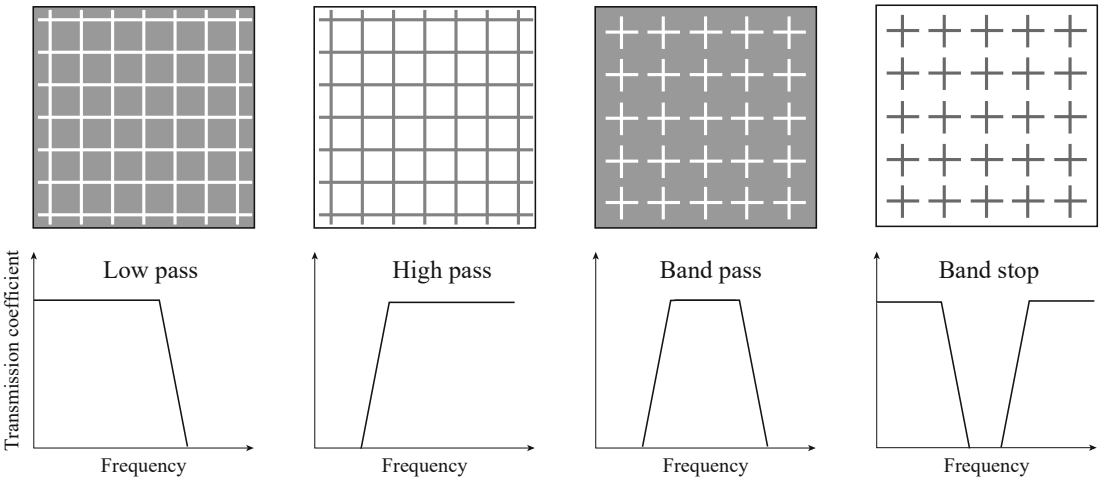


Figure 4.5: Frequency response characteristics of the FSS.

low-pass filter, allowing signals within this range to pass while effectively attenuating higher-frequency noise and interference.

Research on FSS window coatings has focused extensively on the development and simulation of different patterns, including hexagonal structures [88, 94]. Hexagonal designs, in particular, have demonstrated superior performance in achieving low attenuation across wide frequency ranges while maintaining thermal efficiency. Experimental studies show that hexagonal patterns, with D = 2 mm and w = 0.035 mm, can reduce signal attenuation from 20 dB to as low as 1 dB without significantly affecting visible light transmittance [88]. These findings highlight the potential of FSS coatings in meeting the dual requirements of energy efficiency and enhanced wireless connectivity. However, in the context of rail transportation, the effectiveness of FSS window coatings is limited by the structural characteristics of trains. Unlike buildings, where windows constitute a significant portion of the facade, a train is largely a metal enclosure, with only a small fraction of its surface composed of windows. As a result, the potential benefits of FSS window coatings are inherently restricted to localized areas, primarily around passenger seating zones. Signal penetration through windows may improve, but overall connectivity remains constrained by the metallic structure of the train car, which causes substantial attenuation.

Additionally, BS masts are often placed low and close to the tracks, creating challenges for coverage and increasing obstruction sensitivity. This placement exacerbates signal degradation, as the train's metallic structure obstructs LOS paths, which is particularly relevant for high-frequency bands used in 5G and future Sixth Generation (6G) networks.

While previous studies primarily focused on the design and performance of FSS coatings using simulations and measurements, this thesis shifts the focus toward a comparative evaluation on trains in operation. Specifically, the proposed benchmarking methodology compares the performance of window patterns with traditional repeater solutions in a live train network environment. This approach aims to provide a comprehensive assessment of the practical benefits and limitations of FSS coatings, contributing to the broader adoption of this technology in real-world applications.

4.3 Comparative Performance Analysis

The body of literature on measurements conducted inside railroad cars remains sparse, with several notable works addressing specific aspects. In particular, [98] analyzed KPI quantiles in train measurements using a modified Harel-Davis method and nonparametric bootstrap, focusing on AAF repeaters and FSS window technologies. Despite these efforts, there is a notable gap in studies addressing benchmarking methodologies tailored to evaluate cellular network performance specifically in train environments.

To address this gap, in [13] we introduced a methodology designed to assess and compare the performance of repeater, standard metal-coated windows, and windows featuring coating patterns under three distinct railroad car configurations. By controlling for measurement confounding, particularly the varying speeds of trains, we ensure that the resulting dataset reflects an unbiased representation of real-world conditions. This is achieved by segmenting measurements into 100-meter track segments and utilizing

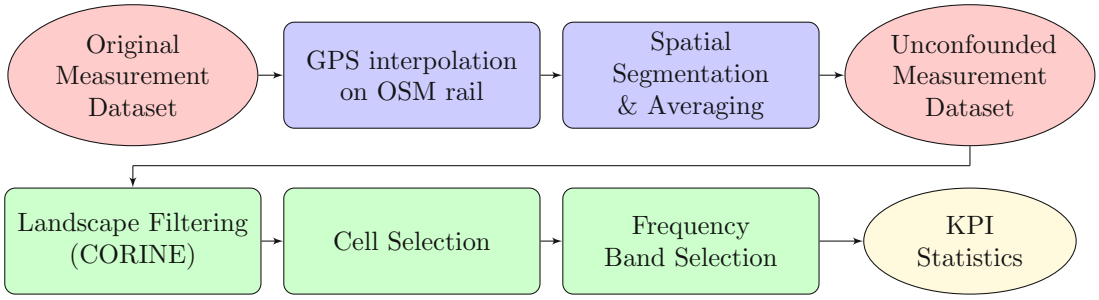


Figure 4.6: Processing flow chart: From raw measurement data to KPI statistics of the unconfounded dataset.

CLC land use classes to define independent strata. These methodological approaches enable a robust analysis of cellular performance while mitigating potential sources of variability and ensuring replicability.

4.3.1 Measurement Methodology

In this section, we apply the proposed Integrated Framework for Data Collection and Analysis from Section 2.4, encompassing the entire process from designing the measurement setup to post-processing, spatial confounding removal, stratification, and filtering, before proceeding to the results analysis.

Given the challenges discussed in Section 2.4, we propose the data processing flow chart illustrated in Fig. 4.6. Starting with the original raw measurement data, we perform OSM rail-incorporated location interpolation and apply spatial segmentation into GUs. Depending on the train speed, some GUs contain significantly more samples than others, leading to a confounded dataset that is not informative for analysis. By averaging over the spatial segments, the influence of train speed is removed, resulting in a spatially unconfounded dataset.

When analyzing signal strengths, the PCell is the most relevant as it dictates the UE's connection stability, handover performance, and overall network experience, making it the primary focus for performance evaluations, coverage analysis, and network optimizations. Therefore, in the subsequent step, we filter the measurements to retain only the PCell and then apply cell and frequency band filtering before computing the statistics of the KPI of interest, such as RSRP.

Measurement Setup

Ensuring reliable mobile in-train connectivity remains challenging due to factors such as location-dependent VPL, high train speeds, and dynamic network load driven by passenger density. The Railjet dataset, introduced in Section 3.2, was collected on an ÖBB-operated train along the Vienna-Innsbruck route. To evaluate the impact of different connectivity solutions, the train included three distinct wagon configurations - SW, MW, and Repeater - as detailed in Section 3.2.

To mitigate the impact of spatial confounding caused by high train speeds, the dataset was processed using a rail track segmentation approach, ensuring a consistent spatial

representation of the measured data. Additionally, measurements were performed in both travel directions and aggregated to improve robustness, as justified in Section 2.4.5. Since cellular networks must share available bandwidth among connected users, all measurements were conducted on an empty train, establishing a baseline independent of passenger-induced network load. This setup provides a scalable reference for MNOs when assessing network performance under varying occupancy conditions.

Spatial Interpolation and OSM Mapping

We use GPS for UE positioning despite its moderately high inaccuracy. In addition to the GPS inaccuracy we face in open regions, many railroad tunnels represent even more problematic areas. The GPS signal inside the tunnel drops, and the reported GPS values inside the tunnel are either lost or mapped back to the last known point - the tunnel start.

Each measurement consists of reported longitude, latitude, GPS speed, various network parameters, and a timestamp. Accordingly, a significant number of missing GPS measurements in the recorded dataset requires reconstruction solutions. In [99], authors used dynamic time warping given a reference path to map the measurements to the correct locations. In the case of rails, we know the exact rail track location from our Environment and Rail DT, thus allowing us to map the inaccurate GPS locations to the closest point on the rail track, see Fig. 4.7. A typical GPS interpolation consists of performing linear interpolation for longitude and latitude separately based on the reported timestamp. The linear interpolation formula is given in Eq. (4.5), where x_1 and x_2 are the timestamps at which the longitude/latitude y_1 and y_2, respectively, are reported.

$$y = y_1 + (x - x_1) \cdot \frac{\Delta y}{\Delta x} = y_1 + (x - x_1) \cdot \frac{y_2 - y_1}{x_2 - x_1}. \tag{4.5}$$

However, mapping the points along the track and then performing the longitude and latitude interpolation introduces further inaccuracies. The interpolated points no longer lie on the rail track as the OSM rail information is not considered in this approach. Hence, we interpolate the traversed distance along the given rail based on the reported timestamps. Interpolating a single parameter, e.g., traversed distance, reduces the inaccuracies since all reconstructed points lie on the reference path. We then obtain the reconstructed GPS locations by mapping the traversed distance back into coordinates as described in Section 2.4.3.

Spatial Confounding - Track Segmentation

Variable train speed and occasional stops during the measurement campaign can distort the measurement distribution along the rail track, as discussed in Section 2.4.4. To mitigate this, we divide the rail track into 100 m segments, or GUs, as proposed by ITU in [100]. Mapping our measurements to each GU and averaging allows us to achieve a uniform measurement distribution for each UE and frequency band along the track.

Landscape Specific Stratification

Clustering according to landscape allows us to benchmark the rails in different country areas within the same class. This enables us to compare the performance of different

Table 4.1: Frequency bands statistics containing only PCell.

Band 3 (1800 MHz)	Band 20 (800 MHz)	Band 1 (2100 MHz)	Bands: 7, 8, 28, 32 (combined)
68%	16%	7%	8%

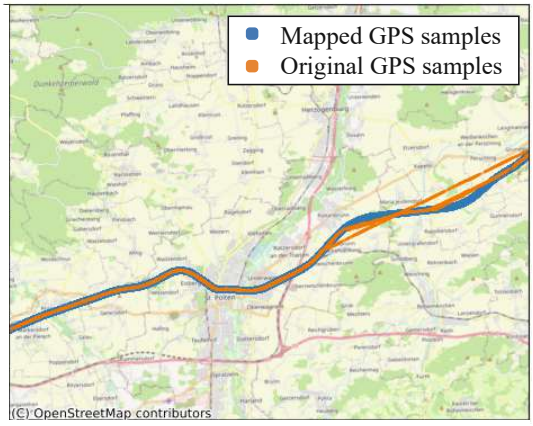


Figure 4.7: GPS mapping example.

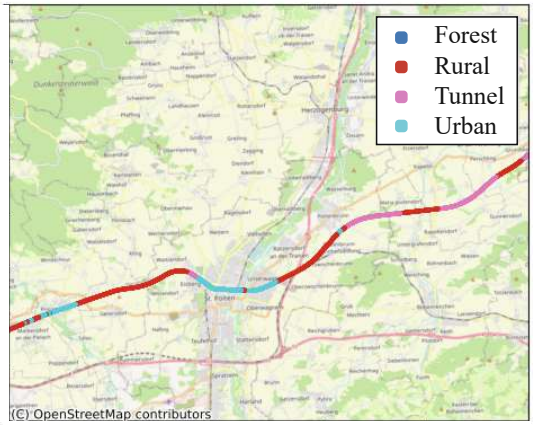


Figure 4.8: CLC mapping example.

configurations in geographically similar areas, e.g., given two different cities with a comparable landscape. The BS location and its proximity to the connected user [101], AoA [102], and multipath propagation properties strongly correlate with the landscape. To capture these dependencies, we split our measurement dataset into landscape classes according to the CLC database [57] - see mapping example in Fig. 4.8. We group the 44 CLC classes into four broad landscape classes: City, Land, Forest, and Water. Tunnels have very different wireless propagation properties and BS rail side deployments. Therefore, we define an additional landscape class for tunnels.

Network Specific KPI Filtering

The automatized scripts used in this measurement campaign combine voice calls and data transmissions. All UEs in our fleet use carrier aggregation to increase data rate during data transmission. However, many measurements originate from voice calls, where carrier aggregation is not supported. Hence, we have to separate different views before proceeding to the analysis. On one side, we could analyze all serving cells combined to get a bigger picture from the operator's perspective. However, we choose to focus exclusively on PCells as we are interested in the actual signal quality experienced by the user. We filter out all measurements originating from the SCells. The original frequency band statistics in Table 3.3 now changes as shown in Table 4.1.

Table 4.2: Mean absolute difference between the ECDF curves of the two routes. The displayed results are for a single UE positioned on the table in front of seat 53 or the green position in Fig. 3.4.

	SW	MW2	Repeater
Band 3	0.2 dB	0.9 dB	1.5 dB
Band 20	0.6 dB	0.3 dB	0.6 dB

4.3.2 Data Analysis

In this section, we evaluate the RSRP statistics of the unconfounded dataset to compare the Repeater, MW2, and SW configurations.

We choose to examine the RSRP as it is the least volatile KPI parameter [103] that shows no time of day effects, meaning we can combine the measurements originating from different driving directions that are occurring in the same segment. We calculate the ECDF for two frequency bands where most of our samples are reported, i.e., Band 3 and Band 20, and all three rail car configurations in order to justify this premise. The mean absolute difference between the ECDF curves of the Vienna-Innsbruck and Innsbruck-Vienna train runs for a single UE is given in Table 4.2. Given that the measurements were conducted using commercial grade UEs, we conclude that the observed difference is insignificant and occurs mainly due to measurement resolution, thereby allowing us to combine both train runs, concurring with our previous conclusions from Section 2.4.5.

We compute the ECDF confidence bounds using the Dvoretzky-Kiefer-Wolfowitz inequality [104]. The inequality itself states that

$$P\left(\sup_x |F(x) - \hat{F}(x)| > \epsilon\right) \leq 2\exp\left(-2n\epsilon^2\right) = \alpha, \tag{4.6}$$

where $F(x)$ is the true population Cumulative Distribution Function (CDF), $\hat{F}(x)$ is our sample CDF, and n is the number of data points. To get the region that contains the whole CDF with probability $1-\alpha$, we set the right-hand side of Eq. (4.6) equal to α and then rearrange, thus yielding the lower and upper bounds $L(x)$ and $U(x)$, respectively:

$$\begin{aligned} L(x) &= \max\{\hat{F}(x) - \epsilon, 0\}, \\ U(x) &= \min\{\hat{F}(x) + \epsilon, 1\}, \end{aligned} \tag{4.7}$$

where

$$\epsilon = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}. \tag{4.8}$$

The bounds represented in Fig. 4.9 are computed for $\alpha = 0.9$, and represent the $[5-95]\%$ confidence interval.

The weaker signal components receive maximum available gain through the repeater, whereas the repeater's maximum output power bounds the amplification of stronger

Table 4.3: Two-sample KS test results comparing mean RSRP distributions for different wagon configurations. The null hypothesis (H_0) assumes that the two sample sets are drawn from the same distribution. If $p > 0.01$, H_0 is accepted.

UE Location	SW vs. MW2		SW vs. Repeater	
	p-value	H_0 accepted	p-value	H_0 accepted
UE Head 53	0.62	✓	9.0×10^{-50}	×
UE Table 53	0.18	✓	2.7×10^{-62}	×

Table 4.4: Two-sample KS test results comparing mean RSRP distributions for different window configurations with the outside rooftop antenna reference. The null hypothesis (H_0) assumes that the two sample sets are drawn from the same distribution. If $p > 0.01$, H_0 is accepted.

UE Location	SW vs. Rooftop Ref.		MW2 vs. Rooftop Ref.	
	p-value	H_0 accepted	p-value	H_0 accepted
UE Head 53	7.3×10^{-6}	×	2.6×10^{-7}	×
UE Table 53	8.8×10^{-7}	×	3.8×10^{-5}	×

signal components. In Fig. 4.10c, where the inner signal strength is plotted against the outside antenna, we can clearly see the repeater bound of the stronger signals. Combined with nonlinear repeater behavior, it explains the nonlinear shift between the rooftop reference and the Repeater curve in Fig. 4.9.

In both SW and MW2 cases, the window configurations possess the same window size, number, and spacing of glass layers, thus providing the temperature isolation and causing possible total reflection for AoA larger than Brewster's angle. The main difference between these two windows lies in the modified windows' metal coating, which provides less attenuation for electromagnetic waves.

We conduct the two-sample KS test [105] (see Section A2) to prove the hypothesis: "The RSRP samples of the two given configurations originate from the same distribution with a mean shift." The test statistics are provided in Table 4.3. With significance level of $p > 0.01$, we cannot reject the given hypothesis in favor of the alternative [106]. Accordingly, the test rejects the hypothesis between the SW and Repeater configurations - the difference between these two distributions is also depicted in Fig. 4.9.

However, the MW2 and SW configurations originate from the same distribution after removing their means. We simply have to shift the distribution depending on the applied metal coating. In [102], the difference in the VPL of the same windows under test is 5 dB in a static reference measurement for 60° AoA. In our measurement campaign conducted under the real-world scenario of over 490 km railway, the average difference between the ECDF curves amounts to 5.5 dB. These results match well with the static measurements. The outside to inside RSRP difference is substantially higher than the static results, likely due to different antennas and positioning.

Table 4.5: Mean absolute difference between ECDF curves in Rural and Urban areas at 1800 MHz. Results are shown for both UEs.

UE Location	SW	MW2	Repeater
UE Head 53	2.63 dB	2.79 dB	0.42 dB
UE Table 53	2.18 dB	3.74 dB	0.48 dB

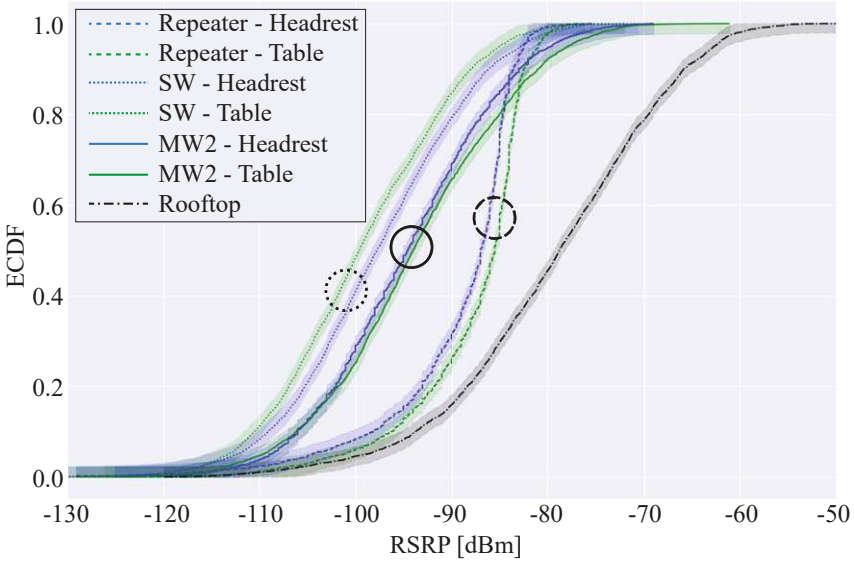


Figure 4.9: ECDFs for different wagon configurations at two indoor UE positions (blue and green in Fig. 3.4). The black line shows the reference from the rooftop-connected UE.

This leads to the second significant contribution of our measurement campaign. We conclude that performing measurements in a single location in one of the windows configurations is enough to understand the second configuration. Therefore, any improvement in the coating seems to shift the distribution. This then allows us to make predictions for future configurations based solely on the average reference attenuation measurements. The distribution shift between SW and MW2 can even be observed in Fig. 4.10a and Fig. 4.10b.

The KS test confirms that in the case of the MW2 and SW configurations, the different UEs inside a single wagon come from the same distribution - however, their means are different in this specific measurement campaign. Table 4.4 shows that the indoor and outdoor RSRP of both window configurations do not originate from the same distribution.

In the final step, we compare landscape strata. Table 4.5 shows the mean absolute difference between Urban and Rural ECDFs. Both window configurations exhibit a difference of approximately 2.5 dB, while the repeater compensates for this variation through amplification. This highlights the importance of separating scenarios to draw more definitive scenario-specific conclusions.

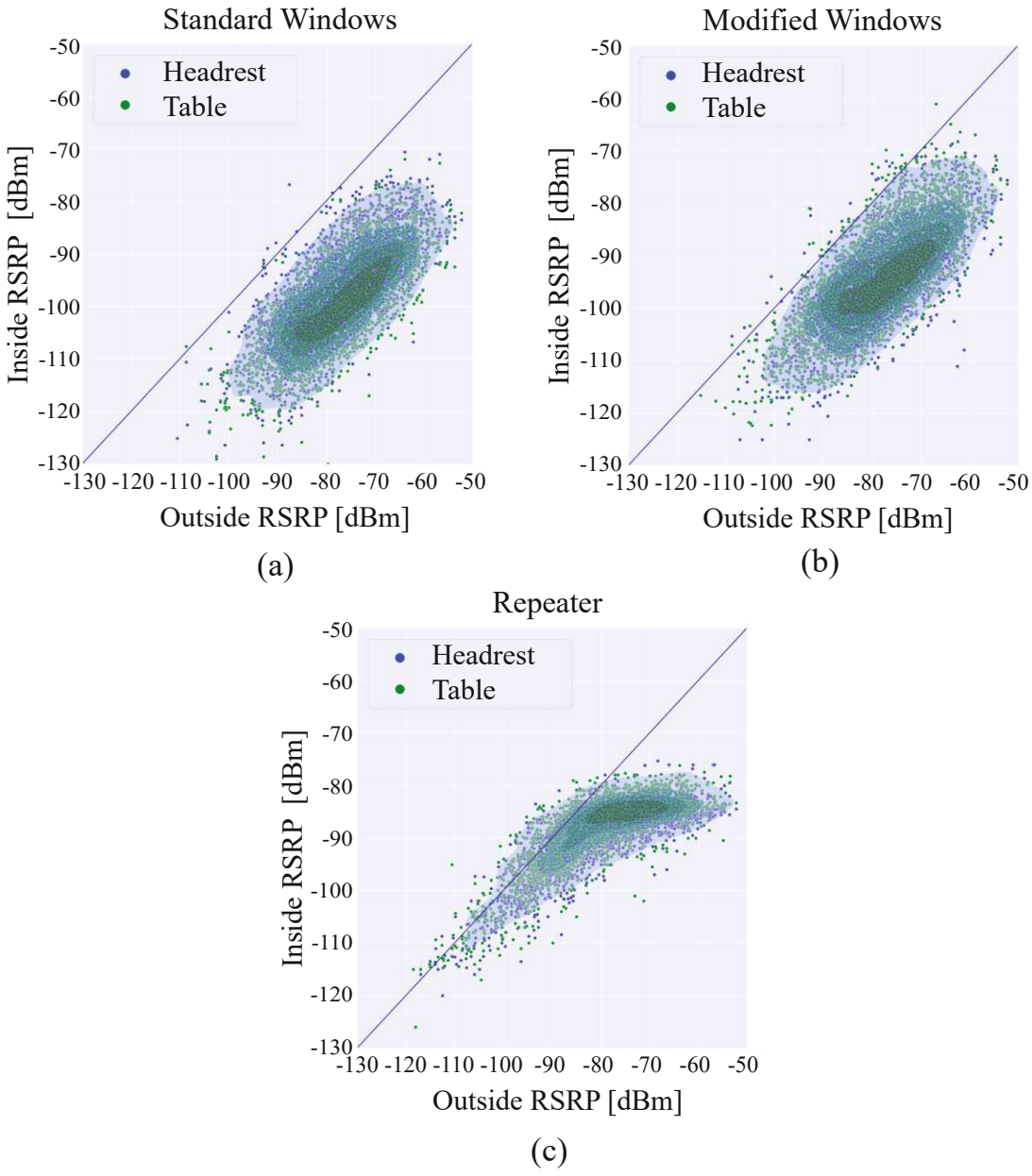


Figure 4.10: Each scatter point represents the segment mean RSRP measured inside the wagon, plotted against the mean RSRP at the external rooftop reference antenna, for three wagon configurations: (a) standard windows without repeater, (b) modified windows without repeater, (c) repeater with standard windows. Results correspond to the 1800 MHz band.

4.3.3 Summary of Comparative Findings and Conclusions

In this chapter, we benchmarked two primary strategies for mitigating penetration loss in trains: AAF repeater systems and FSS window coatings. Using the unified data processing framework described in Section 2.4, we corrected for spatial confounding through 100 m segment binning, mapped data onto the rail network, and applied landscape-based classification. Our key findings are summarized below:

- **Measurement Consistency:** The ECDFs from both train directions closely matched, confirming that RSRP is a stable metric unaffected by time-of-day or travel direction. This allowed for aggregating measurements across multiple runs, strengthening the robustness of the derived statistics.
- **Window Coating vs. Distribution Shape:** Hypothesis testing confirms that modifying the window coating does *not* alter the zero-mean distribution of the RSRP values-only the mean shifts. This finding enables a simplified benchmarking strategy: future window configurations can be evaluated by comparing against a baseline and analyzing the shift in mean attenuation. Full-system re-measurements become unnecessary, as the expected performance is inferred through a translation of the existing distribution.
- **Position Independence Within Cabin:** Analysis across multiple UE positions (e.g., headrest vs. table) shows that measurements come from the same distribution, confirming that measurement location within the wagon has minimal impact post-averaging and filtering.
- **Distinct Behavior of Repeaters:** Unlike passive coatings, the repeater configuration did not align with the same distribution under mean shift. Saturation effects and nonlinear gain response create fundamental differences in performance, necessitating separate evaluation.
- **Landscape Variation:** Comparing urban and rural environments revealed up to 3 dB difference in attenuation for window-based solutions, whereas the repeater compensated for these effects. This underscores the significance of evaluating different environments separately when analyzing connectivity performance.

In Section 5.2.4, we broaden the analysis by comparing different window coating patterns across multiple train types. The VPL-oriented benchmarking approach enables comparison across measurements conducted on different tracks, ensuring consistency and fairness based on the results presented in this chapter.

In Chapter 6, we build on the observed landscape variation by introducing a stratified measurement planning methodology. This approach enables a targeted allocation of measurement resources, focusing efforts on the most variable or underrepresented environmental classes to improve the representativeness and efficiency of future measurement campaigns.

Chapter 5

Influence of Train Window Properties on Vehicle Penetration Loss

Given the growing industry preference for passive window modifications, we set out to examine their performance in greater depth - specifically how they are influenced by network deployment conditions and by train-specific factors such as window size and internal layout. As these solutions are increasingly adopted in modern trains, it is important to better understand their practical limitations and effectiveness in real-world scenarios. This chapter integrates and refines two key studies we previously published in [14] and [15].

In [14], we introduced a practical framework to assess how signal angle of arrival (AoA) affects vehicle penetration loss (VPL). The method operates with UEs connected to live networks aboard moving trains, without requiring dedicated infrastructure, and is applicable even to crowdsourced datasets. It leverages the Environment and Rail DT as well as the Cellular Network DT, introduced in earlier chapters.

In [15], we evaluated modern permeable window coatings used in Austrian night trains, comparing their performance to earlier modifications and active repeater systems. Since the focus lies on VPL, the influence of antenna patterns and BS deployment variations are omitted. This allows for a fair comparison of connectivity solutions not only across different tracks, but also across different train types - reflecting how such evaluations are typically conducted in practice. This approach enables a fair comparison of connectivity solutions across different train types and tracks without requiring identical deployment conditions.

5.1 Angle of Arrival

The field of railroad car measurements remains underexplored in the literature. While some studies, such as [80] and [107], have analyzed radio channel properties both inside and outside railroad cars, others [108] have focused on wave-guiding effects within subway tunnels.

In contrast, research on indoor penetration loss in buildings is more extensive, encompassing diverse materials and scenarios. The ITU report [109] offers a comprehensive

collection of results on Building Penetration Loss (BPL), including elevation angle analyses. [110] outlines a measurement methodology for building wall penetration loss at 5 GHz, especially for materials like stone or glass. Similarly, [111] provides empirical results for VPL assessments in various car orientations. In [112], authors present results from 28 GHz channel sounding campaigns investigating outdoor-to-indoor penetration loss, showing a significant correlation with AoA. In [89], researchers explore window attenuation to microwaves of coated, uncoated, and patterned glass using measurements and simulations in the [1-10] GHz frequency range, while detailing the influence of the thickness of the air gap in a window, as defined in [113].

In [95] the authors measure the impact of the coated pattern's size on attenuation within the frequency range of [0.8-5] GHz . Their findings reveal approximately 6 dB attenuation at 800 MHz and around 7.5 dB attenuation at 1800 MHz, which exhibits a strong similarity to the outcomes observed in our own research at 0° incidence. Additionally, they conduct simulations to assess the impact of AoA on window attenuation at frequencies of 1, 2, and 3 GHz. Moreover, the impact of AoA on signal propagation characteristics was explored in a static train scenario, through a real-world measurement campaign conducted in Austria on a stationary train [102]. This involved a comparative analysis of two distinct window configurations spanning the frequency range [800-2600] MHz, focusing on azimuth AoA within [0-60]°. The examination of these configurations revealed a decrease in penetration loss for the prototype setup, attributed to alternative material treatments employed in the windows. However, unlike our proposed approach, this static measurement method requires considerable time, specialized equipment, and extended access to the train at a fixed measurement site - precisely the limitations our in-motion, network-based framework is designed to overcome.

To validate and verify our proposed methodology, we conduct a comprehensive measurement campaign that includes both interior and exterior train cabin measurements. Interpolation of RSRP values is used prior to VPL derivation to overcome SSF effects, as the measurements are collected simultaneously from multiple phones at different positions. This ensures that only large-scale propagation trends are captured when comparing indoor and outdoor signal levels. Precise track and BS coordinates - provided by the Environment and Rail DT and Cellular Network DT - are used for accurate AoA calculation.

5.1.1 Workflow for Vehicle Penetration Loss Calculation

For analyzing the influence of AoA on VPL in mobile communication for moving vehicles like trains, we utilize the Talent dataset from Section 3.1. This dataset includes indoor LTE measurements collected by four UEs, with a scanner connected to a rooftop antenna serving as an outdoor reference. The DT framework from Fig. 2.8 is employed to assess VPL by integrating indoor and outdoor measurements, as illustrated in Fig. 5.1. This approach enables a comprehensive evaluation of VPL caused by train structures and helps characterize wireless signal propagation dynamics.

Measurement alignment was performed to ensure consistency between indoor and outdoor data sources. This step is essential across all datasets, allowing accurate VPL computation on a sample-by-sample basis.

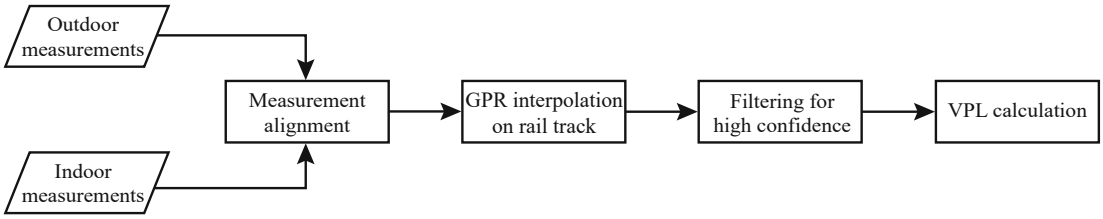


Figure 5.1: Workflow for VPL calculation from indoor and outdoor measurements, encompassing measurement alignment, GPR interpolation, and filtering.

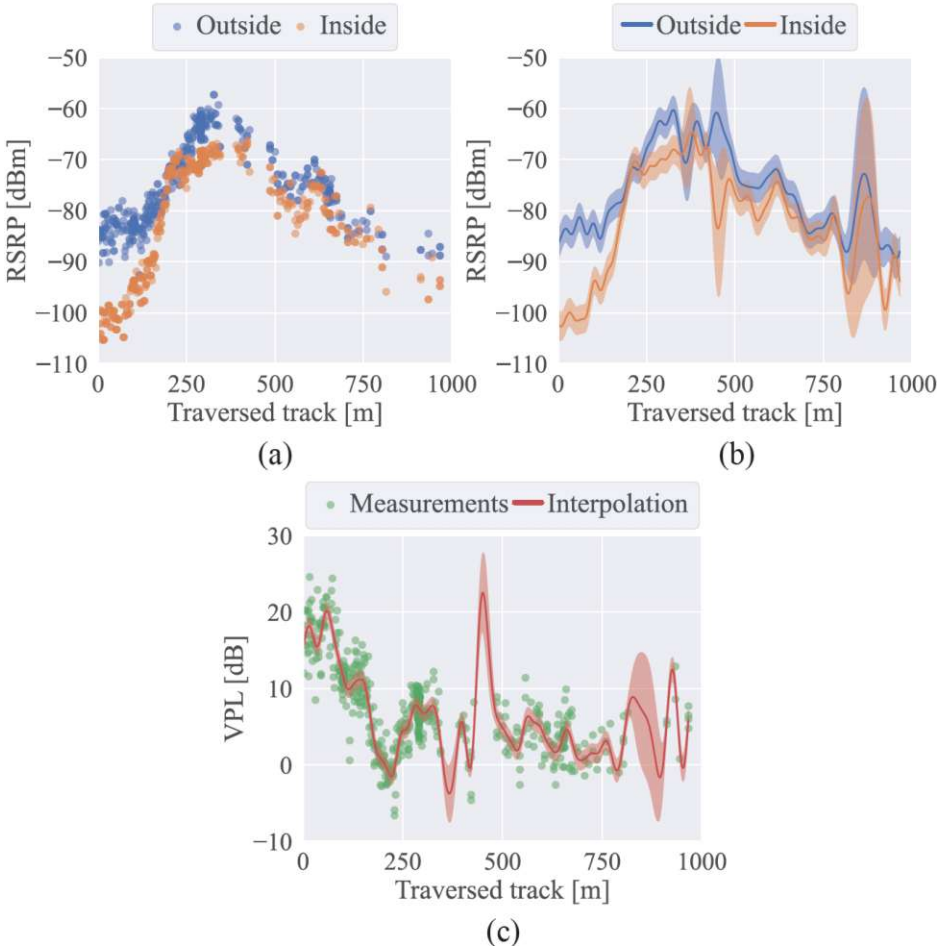


Figure 5.2: Visualization of (a) indoor and outdoor RSRP measurement points for a specific PCI in the 800 MHz band along a 1 km railway segment, (b) interpolated RSRP values with 95% confidence intervals, and (c) VPL calculated as the difference between indoor and outdoor RSRP. In (c), scatter points represent the difference between original indoor and outdoor RSRP measurements, while the interpolated values are shown with a 95% confidence interval.

As depicted in Fig. 5.1, the VPL computation workflow consists of measurement alignment, GPR interpolation, and filtering for high-confidence samples. GPR was selected for interpolation as it provides both the mean and standard deviation at each interpolation point. The standard deviation serves as an indicator of confidence in the estimated mean, enabling the retention of high-confidence samples while discarding potential outliers. Furthermore, interpolation ensures that RSRP values are available at identical positions for both indoor and outdoor measurements, a crucial requirement for analyzing the influence of AoA on VPL in Section 5.1.2. A theoretical overview of GPR is presented in Appendix A1.

To enhance the reliability of signal strength estimations, outdoor and indoor RSRP values were interpolated at 1 m intervals along the train track. The application of GPR resulted in smoothed RSRP estimates along with 95% confidence intervals, as depicted in Fig. 5.2b. To maintain accuracy, samples with a standard deviation exceeding 5 dB were excluded before computing the VPL:

$$\mathrm{VPL} = \mathrm{RSRP}_{\mathrm{out}} - \mathrm{RSRP}_{\mathrm{in}}, \tag{5.1}$$

depicted in Fig. 5.2c. This calculation captures the impact of the train cabin on signal attenuation, without dependency on the BS deployment, as both indoor and outdoor signals experience the same outdoor PL and arrive from the same direction.

5.1.2 Angle of Arrival Estimation

We infer the AoA from each connected BS to the train window closest to the indoor measurement by performing precise geometric calculations. These calculations are based on accurate geographic coordinates of the train track in OSM. Due to possible vertical and horizontal differences in the window pattern, we look at two separate cases: the azimuth AoA, denoted as α, and the elevation AoA, represented as ε. These scenarios are visually depicted in Fig. 5.3a and Fig. 5.3b through aerial and lateral views, respectively.

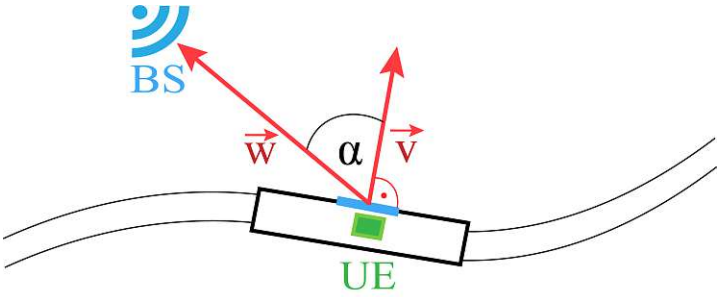
Azimuth AoA

By overlaying the measurements onto the precise track using the Environment and Rail DT, we establish the track vector, which functions as a tangent vector aligned with the track's orientation. Rotating this vector 90° produces the orthogonal vector $\mathbf{v}$, which is perpendicular to both the track and the window plane.

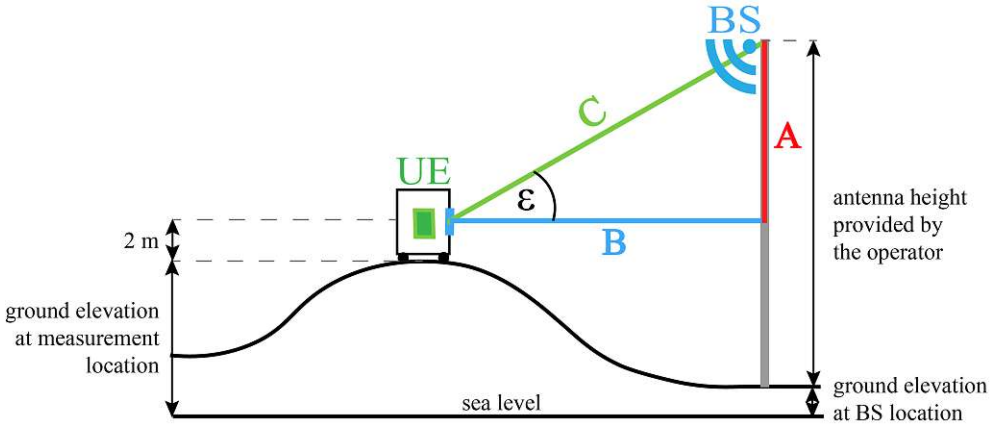
Furthermore, leveraging exact BS locations provided by the operator, we establish a vector connecting our measurement point to the connected BS location. In Fig. 5.3a, we denote this vector as $\mathbf{w}$. By assessing the alignment between the incoming signal direction $\mathbf{w}$ and the orientation that is perpendicular to the train window $\mathbf{v}$, we calculate the azimuth AoA (α) using the cosine theorem:

$$\alpha = \arccos\left(\frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|}\right)\frac{180^\circ}{\pi}. \tag{5.2}$$

W.l.o.g we consider both the train and the track to be 1D, having only the length dimension, and not the width, as one may assume from the Fig. 5.3a. Since the range



(a) Azimuth AoA



(b) Elevation AoA

Figure 5.3: (a) Azimuth AoA α: The rail tracks are depicted as parallel lines, and the train cabin as a black rectangle. The blue line represents the window plane from an aerial viewpoint, while the vector $\mathbf{v}$ is orthogonal to it and forms the angle α with the vector $\mathbf{w}$, indicating the direction towards the connected BS. (b) Elevation AoA ε: A lateral view of the same scenario, where the elevation angle is computed considering ground elevation at sender and receiver locations, antenna height, and the 2 m elevation of the train window.

of arccos is $[0, \pi]$rad (or $[0, 180]^\circ$ for α), we account for the possibility that the BS is located on the trackside opposite the direction of the vector $\mathbf{v}$ by:

$$\alpha = \begin{cases} \alpha & \alpha \leq 90^\circ, \\ 180^\circ - \alpha & \alpha > 90^\circ. \end{cases} \tag{5.3}$$

Ultimately, $\alpha = 0^\circ$ signifies that the incoming signal direction is perpendicular to the window plane. In contrast, a maximum value of $\alpha = 90^\circ$ indicates that the incoming signal aligns parallel to the window plane. An illustrative example of this mapping for a single BS is presented in Fig. 5.3b . In this representation, it is evident that at the point where the track is closest to the BS, the azimuth AoA measures 0°. Subsequently,

moving both forward and backward along the track from this point leads to a progressive increase in the angle, aligning with the anticipated behavior.

Elevation AoA

An alternative perspective is presented in the lateral orientation in Fig. 5.3b. The calculation of the elevation angle, under which the signal from the BS transmitter penetrates the train cabin window, necessitates consideration of multiple factors. Beyond the relative heights of the receiver (RX) and transmitter (TX), the ground elevation at these respective positions also play a crucial role. To address this, we use the terrain altitude data provided by the Environment and Rail DT, which incorporates elevation information from the DTM Austria [56].

We compute the length $A = (\mathrm{TX}_{\mathrm{height}} + \mathrm{TX}_{\mathrm{GE}}) - (\mathrm{RX}_{\mathrm{height}} + \mathrm{RX}_{\mathrm{GE}})$, where $\mathrm{RX}_{\mathrm{height}}$ approximates 2 m, representing the height above ground level for each UE situated within the train. The distance to the BS, denoted as B, is the Euclidian distance between the measurement coordinates and the BS position. Additionally, the distance to the BS transmitter antenna, $C = \sqrt{(A^2 + B^2)}$, is determined. The elevation AoA ε is subsequently calculated using the cosine theorem:

$$\varepsilon = \arccos\left(\frac{B^2 + C^2 - A^2}{2BC}\right)\frac{180^\circ}{\pi}. \tag{5.4}$$

Ultimately, $\varepsilon = 0^\circ$ signifies that the incoming signal direction is perpendicular to the window plane, or in other words that the TX is at the same elevation level as the RX. In contrast, higher values indicate that the BS $\mathrm{TX}_{\mathrm{height}}$ is significantly above the UE level. The maximum observed elevation AoA in our dataset is 58°.

Now that we have established the framework for AoA calculation aboard a moving train, we can apply it to analyze how AoA influences VPL - or any other KPI of interest. This framework is universally applicable across different train types and tracks, using either controlled measurement campaigns or crowdsourced data, without requiring a dedicated test train in a lab or fixed measurement site - both of which are impractical for large-scale evaluation.

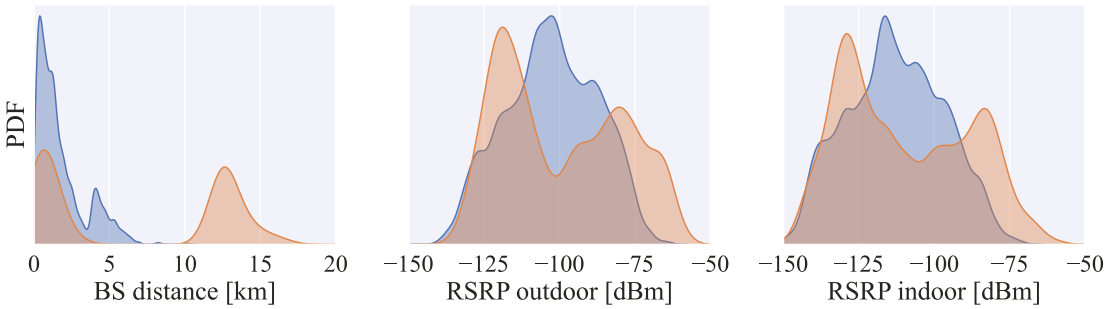
5.1.3 Dataset Description

The processed dataset, detailed in Section 3.1, used throughout this analysis, includes interpolated indoor and outdoor RSRP values, standard deviations, VPL, longitude, latitude, BS distance, azimuth AoA, and elevation AoA.

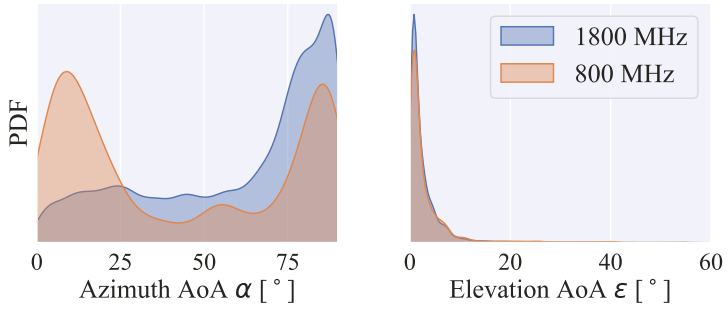
We separately analyze the attenuation in different frequency bands to understand how AoA impacts VPL across frequencies. We begin by studying the interpolated data statistics for each band before further investigation.

Statistics of the Interpolated Dataset

Fig. 5.4a and Fig. 5.4b present ECDFs encompassing BS distances, interpolated $\mathrm{RSRP}_{\mathrm{out}}$ and $\mathrm{RSRP}_{\mathrm{in}}$ values, alongside azimuth and elevation AoA for 800 MHz and 1800 MHz frequency bands.



(a) Measured statistics



(b) Derived statistics

Figure 5.4: Summary of statistics for the interpolated dataset across different channels, presented specifically for values where the standard deviation of both indoor and outdoor interpolated RSRP values was less than 5 dB.

Prominent trends emerge from the distribution of BS distances for different frequency channels, as shown in Fig. 3.1c. Specifically, the 1800 MHz channel exhibits a concentration of BS distances in close proximity to the track. Contrarily, the 800 MHz channel reveals a bimodal distribution, also connecting to BSs located further from the track due to its extended coverage range. This bimodal characteristic of BS distances in the 800 MHz channel is mirrored in the distribution of the RSRP within the same band, while the RSRP distribution in the 1800 MHz band exhibits reduced variability.

Turning to azimuth AoA, the distribution in 1800 MHz band reveals a fairly uniform spread below 60°, with the majority of samples positioned above this threshold. The dual-peak property of the BS distances within the 800 MHz band imparts a similar bimodality to the azimuth AoA distribution. The peak at 10° in the azimuth AoA distribution is a direct result of BSs located further away from the track, particularly pronounced due to the 800 MHz band's BS distance peak at 12 km.

The elevation AoA is mainly driven by the special topography of the area where the measurements were taken. It is worth noting that the relatively flat nature of the landscape had a major effect on this distribution. For both bands, a prominent peak in the elevation AoA emerges at $\approx 2°$, reflective of the predominantly flat landscape. An interesting observation arises when examining BSs positioned atop taller building rooftops in close proximity to the track. This scenario results in the maximum elevation

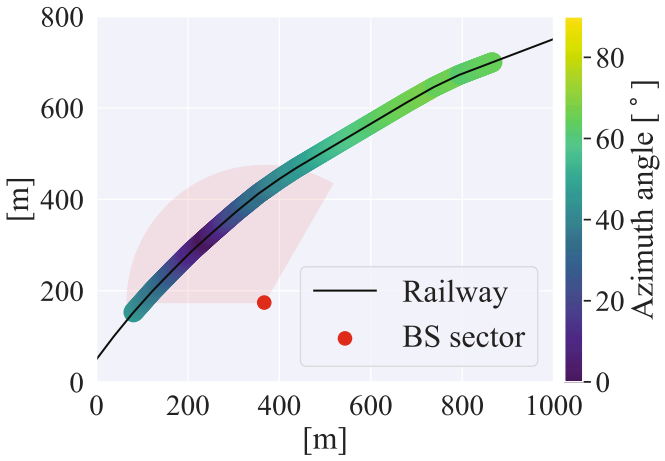
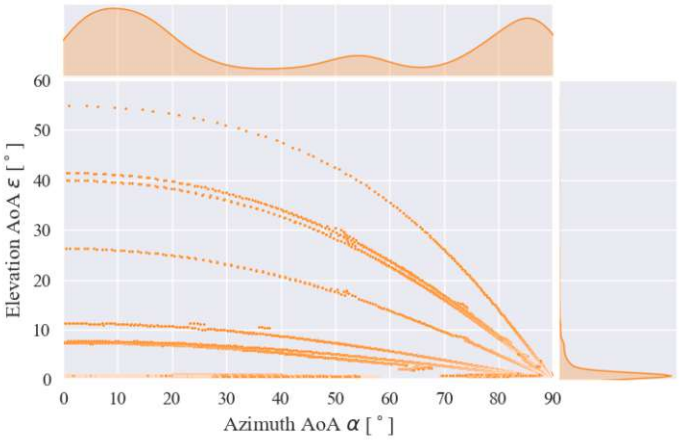


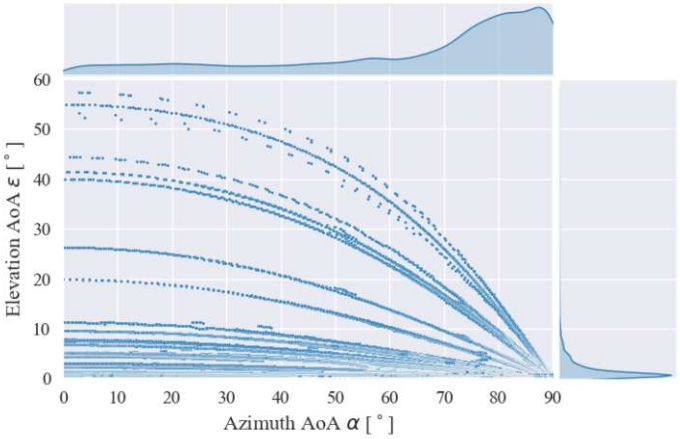
Figure 5.5: The color map represents the interpolated measurement locations along the track, depicting the corresponding computed azimuth AoA α. Additionally, it shows the position and orientation of the BS sector to which the UE was connected.

AoA value of 58°. However, it is essential to stress that this angle is only achieved when the train is in immediate proximity to the BS. As the train moves away from the BS, this angle rapidly decreases, with elevation AoA dropping below 10° for distances exceeding 150 m in most cases.

To investigate the individual impacts of elevation and azimuth AoA on VPL, we analyze their joint distributions in Figs. 5.6a and 5.6b. The scatter plot lines originate from distinct BSs and exhibit a consistent pattern across both frequency bands. We can notice that only very few lines (BSs) are falling into the higher elevation angle range, and most are concentrated within [0-10]° range, due to the terrain being relatively flat. As the train progresses along the track, the azimuth angle tends to be near 0° when it passes the BS, gradually increasing to its peak of 90° as the train moves away, as depicted in the example in Fig. 5.5. At the same time, the elevation angle reaches its maximum when the train is closest to the BS antenna, typically positioned atop a high mast or building. As the train travels away, the elevation angle decreases rapidly and soon falls within the $[0, 10]°$ range. This leads to a rarity of scenarios where both elevation and azimuth angles are simultaneously high. It is important to acknowledge that the distribution is significantly influenced by the terrain and track layout, which may vary across different scenarios.



(a) AoA joint distribution in 800 MHz band.



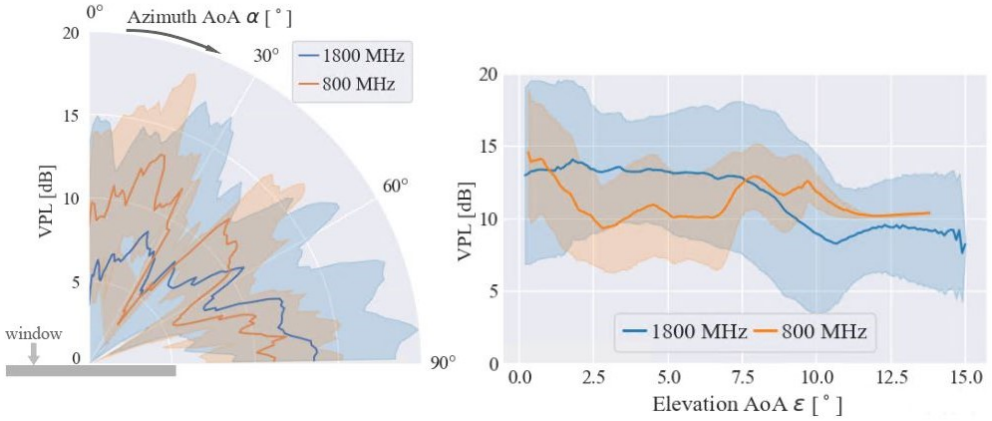
(b) AoA joint distribution in 1800 MHz band.

Figure 5.6: Elevation and azimuth AoA joint distributions for two frequency bands. Each line represents one PCI.

5.1.4 Impact of Angle of Arrival on Vehicle Penetration Loss

Our analysis is focused solely on attenuation, thereby excluding the influences of PL, BS antenna beam pattern, and sector orientation. This deliberate approach stems from the realization that these factors affect both outdoor and indoor signals equivalently, allowing us to effectively isolate and examine the attenuation dynamics in our investigation. In addition, a confidence threshold of 5 dB was established for both indoor and outdoor RSRP interpolated values, to ensure the exclusion of samples from our analysis that may lack robust reliability - primarily stemming from the limited number of samples available for interpolation.

Given that both elevation and azimuth AoA contribute to signal attenuation, it becomes necessary to fix one parameter while assessing the dependency of the VPL on



(a) (b)

Figure 5.7: VPL dependency on azimuth AoA (a) and elevation AoA (b), illustrated as moving averages with the standard deviation intervals ±.

the other. However, exclusively analyzing samples at a fixed value of either α or ε would yield an insufficient dataset for comprehensive analysis. As an optimal approximation, in the case of investigating the impact of α on the VPL, we restrict our sample set to instances within the ε range $[0, 5]°$. This 5° span in ε is considered as an approximation for $\varepsilon = 2.5°$. Conversely, for the assessment of ε influence on the VPL, our sample set is confined to the α range $[80, 90]°$, with this interval serving as an approximation for $\alpha = 85°$. The selection of both values was deliberate, aiming to align with the peaks of their respective AoA distributions (see Fig. 5.4b).

Upon analyzing Fig. 5.7a, it becomes evident that the minimal attenuation level is observed at a 0° azimuth angle, suggesting optimal signal penetration. Additionally, a discernible pattern emerges, with attenuation exhibiting a gradual increase as the azimuth angle approaches its 90° maximum value. Note that the ascending trend is much more prominent at 1800 MHz due to the higher number of samples in the given band. Overall the data shows $[3, 13]$ dB VPL-range over $[0, 90]°$ angle, emphasizing the significant impact of azimuth AoA on attenuation dynamics. We observe that positioning BSs exceedingly close to the track leads to higher average signal attenuation, given that azimuth AoA predominantly falls within the $[60, 90]°$ range along most of the track. However, placing the BSs excessively far from the track could compromise overall coverage, warranting a delicate balance to achieve optimal BS placement. The findings underscore the significance of identifying a strategic middle ground between these two considerations to optimize BS positioning for optimal communication performance.

Examining the influence of elevation AoA on attenuation as shown in Fig. 5.7b, our analysis is confined to samples within the $[80, 90]°$ α-range. Within this range, all corresponding ε samples remain below the maximum threshold of 15°, resulting in a noticeable yet slight reduction in attenuation. The decline in the curves beyond 10° can be attributed to the limited number of remaining samples in this region. Our finding underscores the limited influence of elevation angle on attenuation within our scenario,

where terrain flatness minimizes the elevation's role in signal degradation. Consequently, our results suggest that elevation angle's contribution can be predominantly disregarded, unless specific scenarios necessitate an increased consideration of ground elevation.

At this point, it's important to consider how the rooftop antenna's pattern might affect our measurements. Depending on the employed antenna, the signal strength of the outside reference alone would face an attenuation level that is dependent on the elevation AoA, as these antennas tend to be omnidirectional in the azimuth direction, but have a certain angle-dependent radiation pattern in lateral view, particularly when placed on top of the metal train, e.g. [114]. A detailed analysis of the rooftop antenna's radiation pattern and its potential influence on the results is beyond the scope of this work but remains relevant for future investigations.

5.1.5 Key Findings

We developed and implemented a comprehensive framework for determining the AoA of mobile communication signals in moving train environments. Our approach combines accurate train track geometry, detailed BS position and height data, and ground elevation information to enable precise calculation of both azimuth and elevation angles. By aligning and interpolating samples along the rail track at 1 m intervals, we effectively addressed GPS inaccuracies and captured the dynamic behavior of signal propagation with high spatial resolution. This interpolation not only reduced location uncertainty but also provided confidence intervals for each predicted value, allowing us to filter and analyze only the most reliable data points.

We applied this framework to real-world measurements from the controlled dataset described in Section 3.1, achieving robust results based on a large sample base. The full processed dataset has been made publicly available [66] to support further research into train-based signal attenuation and AoA-aware modeling.

Focusing on VPL allowed us to isolate the influence of the train cabin structure, avoiding effects from PL, BS antenna patterns, and sector orientation, which impact both indoor and outdoor signals equally. Our analysis revealed a clear trend: attenuation increases with azimuth AoA, ranging from approximately 3 dB to 13 dB across the [0-90]° range. In contrast, attenuation showed minimal dependence on elevation AoA, likely due to the flat terrain of the test environment.

Importantly, we uncovered a critical insight for future rail network planning. While BSs placed close to the track can temporarily provide low azimuth angles when the train passes directly beside them, the azimuth angle rapidly increases as the train moves along the track. In practice, most samples fall into the high azimuth range of [80-90]°, where attenuation is greatest. This presents a challenge for permeable window solutions, which are particularly sensitive to signal direction. Placing BSs farther from the track can reduce the azimuth angle and thus lower the VPL, but at the cost of reduced signal strength due to increased PL. This trade-off indicates that, in deployments relying solely on permeable windows, BS placement strategies must be revisited. Optimizing BS locations along rail corridors with AoA-driven design in mind could significantly enhance in-train connectivity performance - particularly when repeaters are no longer an option.

5.2 Impact of Window Size on Vehicle Penetration Loss

Assessing the effectiveness of wireless connectivity enhancement systems in real-world train environments, especially those involving FSS modifications to window design, presents a multitude of challenges, arising from variations in cabin layouts, window size and composition, orientation and angular dependency, environmental impact, and dynamic train conditions for trains traversing different routes.

Cabin Layouts and Window Modifications

Cabin layout and window modification differences among trains play a crucial role in shaping signal propagation within cabins. Window sizes and modifications, such as tinting or coatings, introduce varying degrees of influence on signal strength and quality across different train models. Additionally, the overall structural design of trains can differ significantly, including the number of windows present. Since for passive solutions, windows serve as the primary entry points for cellular signals into the cabin, these structural disparities can profoundly affect signal penetration and distribution within the train environment. To address these challenges, we conduct measurements in different cabin types of the same train, which feature varying window sizes and layouts. Since it is the same train, the materials used and window modifications remain consistent across all tested wagons, allowing us to isolate the impact of cabin layout variations on signal propagation while controlling for other factors such as materials and window modifications.

Diverse Track Routes Trains are assigned specific tracks, resulting in diverse cellular coverage and network setups along diverse routes, complicating the performance comparison across various train types. A reliable method to address this challenge is focusing on cabin attenuation. This metric, derived from indoor measurements and external rooftop references, quantifies signal attenuation within the cabin environment. By comparing the difference between indoor and rooftop signals, both equally influenced by BS positioning and orientation, we can accurately assess attenuation across different trains. This approach effectively minimizes the distinct track impact and network configurations on the evaluation of cellular performance.

When we evaluate solutions to reduce VPL, we need to consider that the level of signal loss changes depending on how the waves reach the antenna on the roof or the window. Building upon Chapter 2, where we introduced a benchmarking methodology for comparing solutions in an operational network under regular cabin operation, we utilize presented concepts to eliminate measurement bias, allowing for comparison of solutions across various trains. We demonstrated in Section 4.3 that, on average, the active repeater outperforms signal strength through the coated window, but is limited by the repeater system's maximum output power. In areas with excellent coverage, the passive solution of window modification surpassed the performance of the active repeater (see Fig. 4.10).

Here, we analyze measurements collected on the Nightjet train (see Section 3.3) to assess the impact of window size on VPL. This train was selected because it featured three wagons with different layouts and window sizes while maintaining the same window

pattern across all three, allowing for a focused evaluation of the influence of window dimensions on signal attenuation.

Additionally, since we have already measured two previous window types, MW1 and MW2, along with the Repeater solution and the SW baseline, we can now compare all of them, including the newest window type, MW3. This comparison enables us to assess which configuration achieves the lowest VPL and to evaluate the relative performance of different window modifications and repeater-based solutions.

5.2.1 Workflow for Vehicle Penetration Loss Calculation

For analyzing the influence of window size and cabin layout on VPL in mobile communications on moving trains, we utilize Nightjet dataset from Section 3.3.

This dataset includes LTE measurements collected by 6 UEs on three different wagons (see Fig. 3.5) all with MW3 solution (see Fig. 3.6f), over two consecutive days. The DT framework from Fig. 2.8 is employed to assess VPL by integrating indoor and outdoor measurements, as illustrated in Fig. 5.8. To compare VPL across different Nightjet wagons we conduct measurements both on the train rooftop, which serves as the outdoor coverage reference, and indoors to assess the signal quality experienced by passengers. We align $\mathrm{RSRP}_{\mathrm{out}}$ and $\mathrm{RSRP}_{\mathrm{in}}$ measurements based on frequency band, PCI, and the 100 m segments of the track to calculate the VPL as $\mathrm{VPL} = \mathrm{RSRP}_{\mathrm{out}} - \mathrm{RSRP}_{\mathrm{in}}$. This ensures a fair comparison of signal strength originating from the same BS[1].

The alignment process depicted in Fig. 5.1 involves the following:

1. For each segment s in each frequency band f and PCI p, take the minimum $\mathrm{RSRP}_{\mathrm{out}}$ value across all samples:

$$\mathrm{RSRP}_{\mathrm{out}}^{(f,p,s)} = \min_{i} \mathrm{RSRP}_{\mathrm{out}}^{(f,p,s,i)}. \tag{5.5}$$

   For each segment s in each band f, PCI p, and UE u, take the minimum $\mathrm{RSRP}_{\mathrm{in}}$ value across all samples:

$$\mathrm{RSRP}_{\mathrm{in}}^{(f,p,s,u)} = \min_{i} \mathrm{RSRP}_{\mathrm{in}}^{(f,p,s,u,i)}. \tag{5.6}$$

   The reported RSRP values are already averaged over 1 s intervals, which smooths out SSF and ensures more stable readings. We select the minimum RSRP per PCI and 100 m segment to avoid overestimating outdoor and underestimating indoor signal strength in cases where the train passes directly by the BS. At such points, the rooftop antenna often has a clear LOS, while the indoor signal - passing through coated windows - typically loses LOS and experiences significantly higher attenuation. Focusing on the minimum RSRP also helps keep the analysis focused on conditions where the repeater is, for the most part, not operating in saturation [13].

[1] Comparing this workflow with the one in the previous section, interpolation part can be omitted as we are focusing on the evaluation on a segment basis, and it is not required to have data samples indoor and outdoor at exact same positions, as it was the case for AoA calculation.

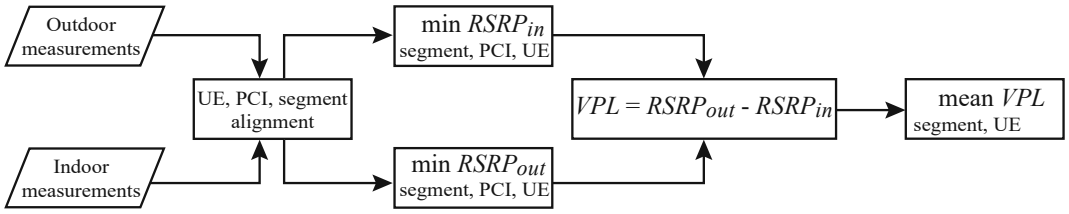


Figure 5.8: VPL calculation flowchart.

2. We calculate the difference between the outdoor and indoor RSRP to determine the VPL for a particular band f, PCI p, and UE u in each segment s:

$$\mathrm{VPL}^{(f,p,s,u)} = \mathrm{RSRP}_{\mathrm{out}}^{(f,p,s)} - \mathrm{RSRP}_{\mathrm{in}}^{(f,p,s,u)}. \tag{5.7}$$

3. We compute mean VPL over all PCIs in one band observed in the segment to obtain the expected value of the VPL for each UE along the entire track:

$$\mathrm{VPL}^{(f,s,u)} = \frac{1}{N_s} \sum_{p=1}^{N} \mathrm{VPL}^{(f,p,s,u)}, \tag{5.8}$$

where N_s is the number of observed PCIs in band f in segment s.

In the following sections, we present the results obtained by calculating the mean VPL using this approach.

5.2.2 Attenuation Level Relative to Measurement Position

In Section 2.2.2, we analyzed the RSRP variations across different measurement positions using repeated measurements conducted in a controlled, stationary environment. Our findings indicated that, after aggregation and interpolation, differences of up to 3 dB could be observed among various measurement locations. Building on this analysis, we now extend our investigation to examine the VPL, which is derived from RSRP, in a dynamic scenario. Specifically, we assess VPL while the train is in motion, continuously connecting to multiple BSs throughout the measurement process, rather than a single BS as in the static case. For this we utilize the two-sample KS test, a nonparametric statistical method that compares the empirical distributions of two datasets [105], to test the hypothesis that the attenuation levels of two VPL solutions arise from identical distributions, with and without mean shift. The KS test reports the maximum difference between the two cumulative distributions, and gives a numeric value related to that difference. Our results are presented in Fig. 5.9, where we illustrate the p_{value} corresponding to each position combination from the Nightjet dataset introduced in Section 3.3. It is noteworthy that the depicted matrix is symmetric. The p_{value} serves as a measure of the significance level of the test. Significance level of $p_{\mathrm{value}} = 1$, indicates that we cannot reject the proposed hypothesis in favor of the alternative [106]. In simpler terms, it suggests that the distributions of attenuation levels from the two

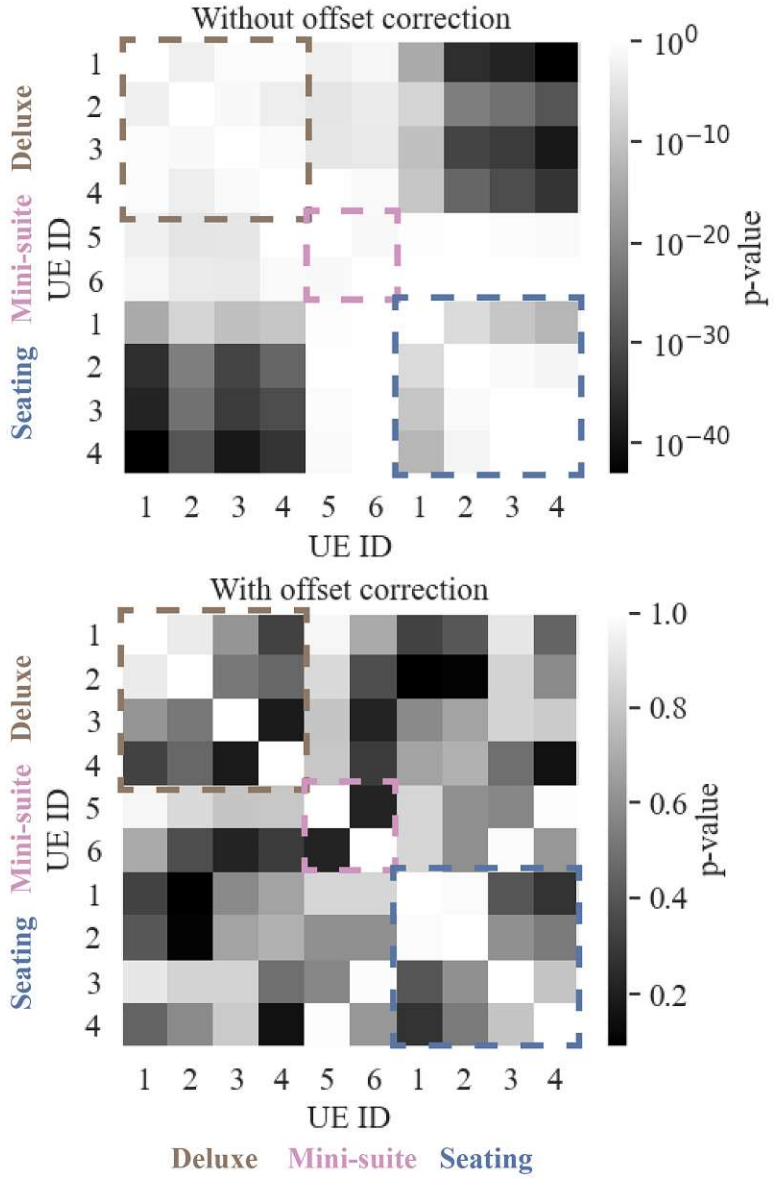


Figure 5.9: p_{values} obtained from a two-sided KS test in 1800 MHz band. Note that the scaling differs between the two plots.

configurations are statistically indistinguishable, which is why diagonal elements of the matrix in Fig. 5.9 are always equal to one (white). A lower p_{value} indicates a higher probability that two datasets stem from different distributions. Typically, $p_{\text{value}} = 0.1$ is chosen as the threshold for accepting or rejecting the hypothesis.

In Fig. 5.9, the colored blocks represent combinations of all positions within a single cabin type. We compare the mean VPL once without applying the offset correction and once with applying the offset correction, in order to determine if the mean shift is the

most relevant, or are the actual distribution differences caused by different curvatures, and therefore different fading properties within the cabin.

The brown block in Fig. 5.9 corresponds to the Deluxe cabins. Upon examining the upper plot, it becomes apparent that, according to the p_{value} level, the distribution for all four UEs positioned in that cabin aligns with the hypothesis of having the same distribution. The same conclusion can be drawn for the Mini-suite block in lilac. However, for the Seating coach block, depicted in blue, a slight discrepancy is observed in the distribution of UEs 1 compared to the other three UEs in the same cabin. Possible reasons for this discrepancy include measurement inaccuracies or potential issues with the UEs 1 phone itself. Furthermore, it is noteworthy that almost all elements not belonging to these three blocks have a p_{value} below 0.1, indicating that the attenuation distributions between the UEs in different cabin types are significantly different, at least when the bias in the distributions is not removed.

After applying the offset correction, by removing the mean at each position, differences among different positions level out, such that p_{value} for most UE combinations falls above 0.1 range, suggesting that we cannot reject the hypothesis for most UE combinations, that is, their ECDFs are very similar after the bias removal, independent of the cabin type and positions inside the cabin.

Similarly, this method can be employed on conditional distributions. For instance, we can compare VPL levels conditioned on observing only segments and PCIs where the BS was positioned very close to the track or, similarly, very far from the track. This allows us to differentiate between various cases in the analysis., but requires a larger sample size for conclusive analysis compared to a single drive test.

5.2.3 Attenuation Differences and Window Size Impact

Fig. 5.10a presents the mean ECDF curves derived from the ECDFs of various UEs within the same Nightjet cabin type. These curves represent the average values across multiple UEs, with shaded regions indicating the minimum and maximum ECDF values for each specific cabin type. The comparison of different Nightjet cabins illustrates the impact of window size and interior layout on signal attenuation levels.

Table 3.4 presents the window size and corresponding average VPL values for three different ÖBB Nightjet cabin types. The data show a clear inverse relationship between total window size and signal attenuation: cabins with larger window areas exhibit lower mean VPL. The Seating Coach and Deluxe Cabin have nearly identical total window areas per wagon ($\approx 22\,\text{m}^2$), yet differ in mean attenuation by over 3 dB. This deviation underscores the importance of internal cabin geometry. In Seating Coach, the open-plan design facilitates more direct propagation paths and reduces internal obstruction losses. In contrast, Deluxe and Mini-suite Cabins incorporate partition walls and doors, introducing additional barriers to signal propagation. The smallest windows and most enclosed layout of the Mini-suite configuration correspond to the highest observed VPL, at 18.4 dB. These results suggest that both window size and interior structure are critical contributors to in-cabin signal attenuation and must be jointly considered in future train design and connectivity optimization efforts.

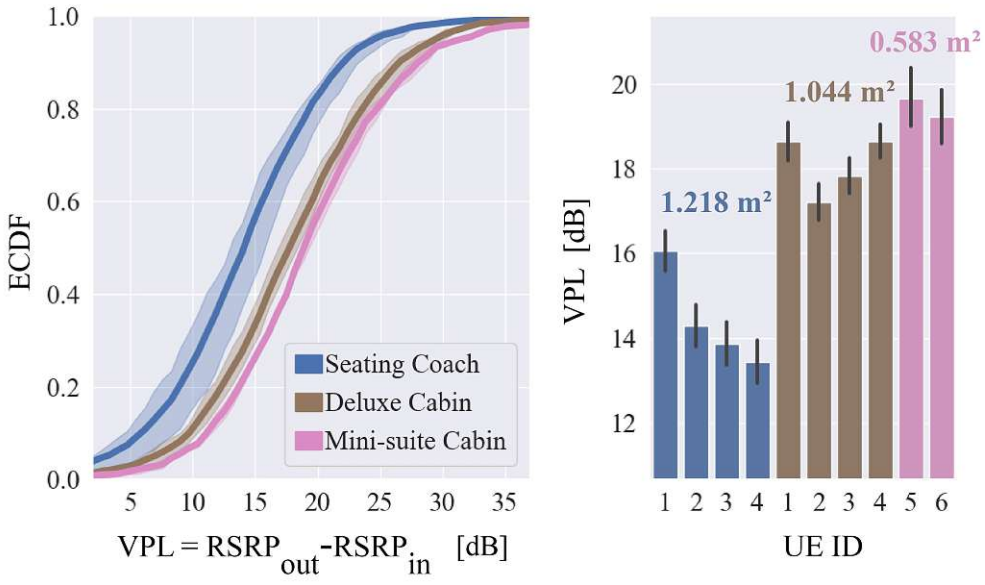


Figure 5.10: (a) Mean ECDF - derived from the ECDFs of various UEs within the same Nightjet cabin type. (b) Bar plot of the average attenuation per UE.

To further illustrate the variations in attenuation across different UEs, Fig. 5.10b presents bar plot of the average attenuation per UE. The results show that the attenuation levels in deluxe and mini-cabins remain relatively consistent, while the seating coach experiences significantly lower attenuation. This can be attributed to the larger number of windows and the absence of walls or door barriers that could obstruct signal propagation.

Additionally, the layout of UEs in the deluxe cabin further emphasizes the impact of direct window paths. UEs 2 and 3, positioned in the lower rack, benefit by occasionally having the LOS path through the window and thus experience lower attenuation levels. In contrast, UEs 1 and 4, located in the upper rack, are always in NLOS, leading to higher attenuation. This observation reinforces the importance of window placement and internal cabin structure in determining signal penetration efficiency.

5.2.4 Solution Comparison Across Different Trains

Fig. 5.11 compares attenuation levels across different train configurations, focusing on various window modification strategies (MW1-MW3), the SW configuration, and the use of the Repeater. These results highlight how different solutions impact outdoor-to-indoor signal penetration across Railjet, Nightjet, and Talent trains.

As expected, the highest attenuation levels are observed in the SW (Railjet) configuration, where outdoor-to-indoor attenuation averages around 22 dB. Implementing window modification MW2 on Railjet reduces attenuation by 5 dB, improving indoor

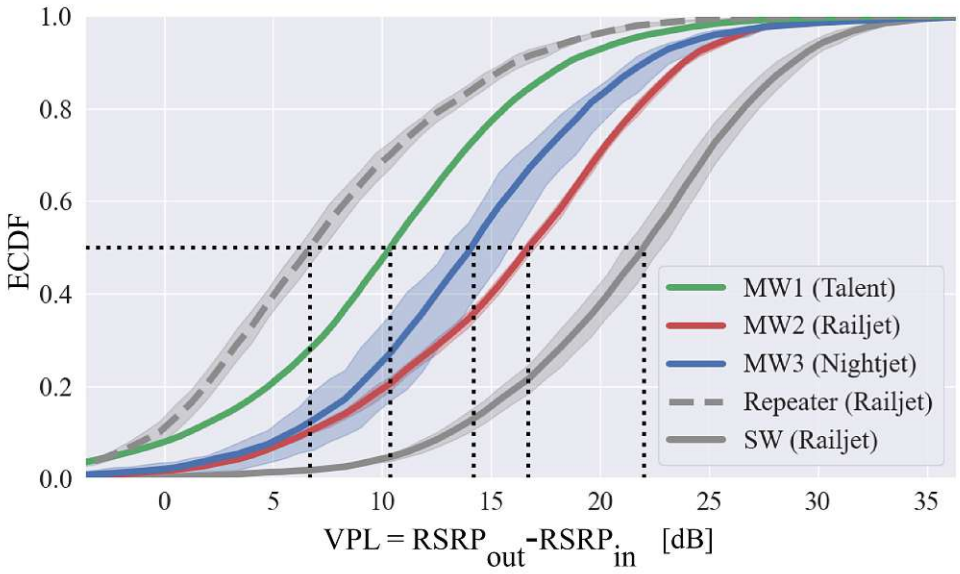


Figure 5.11: Attenuation[2] levels across different trains and train configurations.

Figure 5.12: Tracks used for comparative measurements across the three train types.

connectivity. In the Nightjet train, the MW3 seating cabin configuration provides an additional 2.5 dB improvement, attributed to the newer generation of permeable windows with a larger hexagonal grid.

The Talent train, which uses a regular MW1 window grid, shows a further 4 dB reduction in attenuation compared to the other trains. This improvement is likely driven by two factors. First, the Talent train features a larger window area than the Railjet and Nightjet trains. Second, its lighter construction and less tightly sealed design provide additional signal entry points, including windows and train doors. These results suggest

[2] Although attenuation, by definition, cannot be negative, the ECDF curves in the plot may start slightly below 0 dB. This occurs due to the way attenuation is computed as $\mathrm{RSRP}_{\mathrm{out}} - \mathrm{RSRP}_{\mathrm{in}}$ on a per-segment basis. In rare instances, especially with repeaters, the indoor signal can temporarily exceed the outdoor reference - caused by repeater gain, measurement mismatch, or local fading - which results in a negative computed VPL. These artifacts are rare but explain why the ECDF curve may extend into the negative range.

that window modifications alone are insufficient to significantly improve indoor coverage. A more effective approach combines larger window areas with optimized coating designs.

The repeater curve in Fig. 5.11 illustrates the optimal use case for repeaters, particularly in segments with poor coverage. Since the analysis focuses on minimum RSRP values per segment, the repeater effectively enhances indoor coverage in low-signal areas. Importantly, the repeater has not yet reached saturation levels, providing an additional 4 dB improvement over the best-performing window modification (MW3). This highlights the advantage of active signal enhancement solutions in scenarios where passive modifications are insufficient.

5.2.5 Key Findings

We conducted a targeted measurement campaign on the ÖBB Nightjet, specifically designed to evaluate how window size and cabin layout affect signal penetration. This setup enabled a controlled comparison across three distinct cabin types - Seating Coach, Deluxe Cabin, and Mini-suite Cabin - within the same train and using identical window coating pattern. To support further research and validation of our findings, we processed and published the Nightjet dataset at [68].

We observed that interior cabin layout plays a measurable role in signal attenuation. Despite having comparable total window areas, Seating Coach exhibited approximately 3 dB lower attenuation than more compartmentalized Deluxe cabin. This difference is attributed to the presence of internal partitions - such as walls and doors - that introduce additional propagation loss.

Beyond the controlled study within the Nightjet, we developed a generalizable benchmarking framework centered on VPL, enabling a fair comparison of different onboard connectivity solutions - such as repeater systems and modified window coatings - even when tested on different train types, tracks, and network deployments. By isolating the cabin-induced attenuation component, the framework allows us to decouple results from varying network deployments and propagation conditions.

Applying this framework, we quantified the attenuation characteristics across three ÖBB train types. Standard metal-coated windows consistently introduced high VPL around 22 dB. This was reduced to approximately 14 dB when modified FSS-based coatings were used. In trains with larger window areas and lighter construction - such as local commuter railcars - attenuation was further reduced to approximately 10 dB. This suggests that simply improving the window coating is not sufficient, and that the overall cabin design - particularly larger window areas - also plays a significant role in reducing attenuation.

Active repeater systems achieved the lowest observed attenuation, with average values as low as 7 dB. While they remain the most effective in absolute terms, window-based solutions significantly narrow the performance gap. Moreover, advanced window coatings offer practical benefits: they operate passively, support a broad frequency spectrum, and are inherently compatible with modern MIMO technologies. This makes them an attractive alternative in deployment scenarios where active systems are impractical or undesirable.

Chapter 6

Optimizing Measurement Strategies

So far, we focused on benchmarking in-train connectivity solutions in railway scenarios - such as repeaters and window modifications - that are independent of specific MNOs. These evaluations were conducted in a way that abstracts from specific BS deployments, making results applicable across different operators.

In this chapter, we shift our attention from evaluating in-train connectivity technologies to optimizing measurement strategies for performance monitoring. Our goal is to determine how many samples - and from which geographical areas - are needed to accurately estimate network KPIs. This reduces the overall number of measurements required while maintaining accuracy. Performance monitoring is essential for ensuring optimal service quality and efficient network operation, especially in modern network generations such as LTE, 5G, and beyond.

Real-time performance metrics are essential for network optimization and for enabling features such as network slicing and orchestration, as outlined in [115]. MNOs typically estimate network performance through large-scale data collection, using methods like drive tests or crowdsourcing. Commonly sampled KPIs include call drop rates [116] and RSRP, the latter being widely used by regulators to assess coverage obligations in both LTE and 5G networks [117]. The 3GPP standards [118, 119] define the relevant KPIs and measurement methodologies across the PHY, MAC, and service layers.

Obtaining reliable KPI estimates is challenging due to time- and location-dependent variability, user mobility, and network heterogeneity. While regulatory authorities like Austria's Regulatory Authority for Broadcasting and Telecommunications (RTR) enforce minimum performance standards, MNOs must demonstrate compliance efficiently.

Despite the importance of performance benchmarking, the literature lacks robust sampling methodologies tailored for cellular data. Few studies address how to minimize sample collection, handle confounding factors, or validate existing data, such as crowdsourced data. Although there are various experimental and exploratory studies on network monitoring [120, 121], they often lack comprehensive strategies for real-world network performance assessments.

To address this gap, we apply sampling theory - specifically, stratified sampling - to large-scale MDT data. The methodology and findings are drawn from our previously published work [16], demonstrating how stratified sampling can mitigate key benchmarking limitations such as user distribution imbalance and sampling bias.

Although demonstrated using urban and national-scale crowdsourced data, the same sampling principles can be applied to railway environments. In rail scenarios, static geographic tiles (e.g., 100 m × 100 m) can be replaced with fixed-length track segments (e.g., 100 m) representing GUs. Stratification can then be applied to account for tunnels, open tracks, stations, or high-speed segments. While large-scale rail monitoring data remains limited, we rely on simulations and MNO-provided crowdsourced data to illustrate the effectiveness of our approach. These strategies lay the groundwork for scalable, efficient, and operator-independent benchmarking on railway tracks.

In Section 6.1, we review existing research on benchmarking in cellular networks and motivate the need for more effective sampling strategies using a simulated example. Section 6.2 introduces the fundamentals of sampling theory in the context of benchmarking, explaining potential biases in Simple Random Sampling (SRS) and how SS can help address them. The theoretical foundations of both SRS and SS are detailed in Section 6.3. In Section 6.4, we compare their performance in controlled conditions, illustrating the benefits of stratification in specific scenarios. Section 6.5 demonstrates the application of our methodology to real-world MDT data - representing, to our knowledge, one of the first large-scale uses of SS in this context. Finally, Section 6.6 summarizes the main findings.

6.1 State of the Art in Mobile Network Operator Benchmarking

Benchmarking cellular network coverage is a crucial task in ensuring that mobile networks provide reliable and consistent service to users, particularly as demand for high-quality connectivity increases [116, 122]. This process helps operators assess network performance across various geographic areas, identify service gaps, and optimize network deployment. Sampling methodologies are essential for this task, allowing operators to gather accurate and representative data without the need for exhaustive measurements. Traditional methods, such as drive tests, as well as newer approaches like crowdsourcing, have been used to collect this data, each offering unique advantages and challenges [60].

Recent advancements in sampling techniques have been driven by the growing availability of high-quality geospatial data and the development of adaptive sampling mechanisms. These innovations allow for more efficient and targeted data collection, enabling network operators to focus on areas with specific performance issues or network conditions [123, 124]. Adaptive sampling, in particular, helps improve the accuracy of coverage assessments by adjusting in real-time to reflect changing network conditions, further enhancing the benchmarking process.

6.1.1 Drive Tests

Drive testing is one of the most commonly used methods for benchmarking mobile networks. In this approach, a vehicle equipped with specialized measurement equipment travels through a designated geographic area to collect data on network performance [125].

This method provides highly detailed and accurate data, offering direct measurements of signal strength and quality at specific locations. However, drive testing has several limitations. It is expensive and time-consuming, particularly when scaled to cover large geographic areas. Additionally, challenges such as speed variability, road conditions, and the repeatability of measurements introduce uncertainties, which can affect the consistency and reliability of the data [126].

According to 3GPP specifications, network performance measurements are critical for ensuring compliance with KPIs and service level agreements. For instance, 3GPP defines methodologies for measuring radio performance, including those related to coverage and signal strength, which are central to the drive test process [118, 119]. However, most drive test campaigns focus on a subset of major streets, often neglecting other important outdoor spaces such as minor streets, parks, and squares, where network quality is equally important. As a result, the measurements are highly correlated due to the environmental similarities along major streets, which limits the randomness of the sampling method. This lack of randomization makes it challenging to compute reliable and representative mean KPI values based on the restricted data, as it may not fully reflect the variability in network performance across other areas of interest.

In railroads, these challenges are even more pronounced. Tracks pass through a wide range of environments - from open rural areas to tunnels and complex station infrastructure - each with distinct propagation characteristics. It is therefore essential that drive test campaigns do not focus solely on routes with known good coverage, as this can lead to overly optimistic assessments. Different tracks often have different infrastructure deployments and coverage quality, and neglecting those with poorer conditions risks overlooking areas where connectivity issues are most severe. Comprehensive measurement must include stations, tunnels, and less-optimized segments to ensure realistic and representative performance evaluations.

6.1.2 Crowdsourcing

On the other hand, mobile crowdsourcing is increasingly recognized as an effective method for gathering network performance data from end-users through specialized applications installed on mobile devices. This technique allows for the collection of performance and location data across a broad geographic area, providing a more representative picture of network conditions, particularly in areas that drive testing may overlook. Unlike traditional drive testing, which is confined to predefined routes, crowdsourcing leverages a distributed network of devices to sample a wider array of environments, including residential and commercial areas that might not be adequately represented in targeted measurement campaigns.

In many cases, mobile users can trigger a network measurement at the press of a button through applications such as OpenSignal, Ookla, or Alladin [127–129]. This enables a high volume of measurements to be collected in real-time, allowing for more frequent and extensive data capture. The concept of mobile crowdsourcing was first introduced in [130], which not only described the principle but also provided examples of its application in various fields, including network performance monitoring, environmental sensing, and traffic management. In [131], this concept was further developed to incorporate mobile

computing, expanding the role of mobile devices from mere data collection points to active processors of data, allowing for preliminary analysis directly on the device before transmitting the results.

Despite its growing adoption, crowdsourcing presents certain challenges, particularly in ensuring a sufficient density of measurements in the desired locations and times. The success of a crowdsourcing initiative depends heavily on user participation, which can be uneven across different geographic areas or time periods. As a result, researchers have proposed various incentive models to motivate users to perform measurements in locations where data is needed. Studies [132–138] have developed cost functions and reward optimization strategies that encourage higher user engagement and improve spatial and temporal coverage of measurements.

Tutela [139], a prominent crowdsourcing platform, incentivizes app developers to integrate its Software Development Kit (SDK) into their applications. This SDK then collects anonymized network data from end-users, enabling network operators and researchers to monitor and evaluate real-world performance across diverse usage scenarios. While crowdsourced data provides valuable information on network conditions, it is important to recognize that this data can be subject to several biases. The quality of crowdsourced data may vary based on factors such as device type, network configuration, and user behavior. Additionally, crowdsourced data tends to have gaps in certain locations, particularly in areas where users are less likely to engage with network measurement apps.

The reliance on devices already in use, such as smartphones or IoT devices, can further introduce noise into the data, which may be amplified by factors like signal interference or variability in device performance. As discussed in [140], while crowdsourced data offers valuable granularity, it may not always align with the high precision typically achieved by dedicated measurement equipment in drive tests. Additionally, privacy concerns arise when collecting large-scale data from mobile users, making it essential to address data security and anonymization, as highlighted by [141].

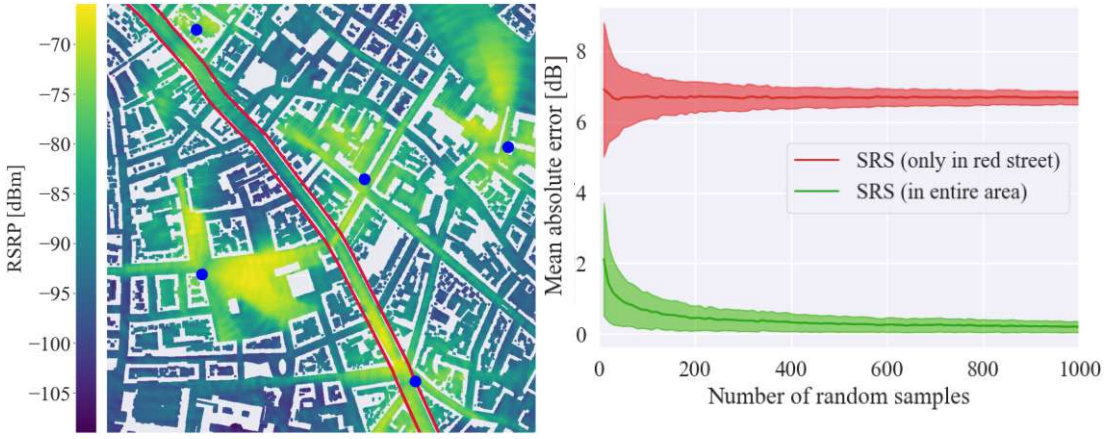
In Section 7.1 we discuss the use of crowdsourced data for benchmarking in railway scenarios as part of future work, outlining both its potential benefits and current limitations.

6.1.3 Crowdsourcing vs. Drive Testing for Mobile Network Operator Benchmarks

To illustrate the importance of random sampling and evaluate its effectiveness in both crowdsourced and drive test scenarios, we analyze a simulated map (shown in Fig. 6.1a), which depicts the received signal strength in a realistic urban network layout. The map was generated using the Deep Learning Network Planner (DLNP), as described in [63], with additional simulation details provided in Section A3. The goal of the analysis is to estimate the mean signal strength over the area represented on the map. In order to achieve this, we compare two distinct sampling approaches: one based on a targeted drive test along a single street and the other using crowdsourcing to collect data over the entire region. Specifically, we consider the following two sampling scenarios:

- *Drive test*: SRS of n RSRP samples only from the marked red street in Fig. 6.1a (red curve in Fig. 6.1b),
- *Crowdsourcing*: SRS of n samples from the entire RSRP dataset (green curve in Fig. 6.1b)

In both cases, the population mean of all simulated RSRP values (1m × 1m tile = one sample) acts as the ground truth. The number of random samples n (or 1m × 1m tiles of the simulated map) is depicted on the x-axis in Fig. 6.1b.



(a) (b)

Figure 6.1: (a) Simulated outdoor RSRP map in the third Vienna district. Red outlined street is the Landstrasser Hauptstrasse. Blue scatter points are the BS locations. (b) MAE of the RSRP mean estimation based on SRS.

For each n, the sampling process was repeated 500 times independently to compute the error bounds. The shaded areas in the curves represent the interval $[\mu_e - \sigma_e, \mu_e + \sigma_e]$, where μ_e denotes the MAE and σ_e refers to the Standard Deviation (SD) of the absolute error for each sample size n.

Among the total of 567 209 available samples[1] from the population, a random sample of 200 measurements from the entire region is sufficient to estimate the mean with an error of less than 1 dB. In contrast, focusing solely on a single major street, as might be the case in a typical drive test campaign, results in a bias of nearly 7 dB in the mean estimate. This bias persists even when the sample size is increased to cover all measurements within that street. This analysis highlights that crowdsourced data is more effective for producing accurate mean estimates compared to drive test data, which is often biased due to its concentration on specific areas, such as street canyons. In practical scenarios, the advantage of crowdsourcing is even more pronounced, as it can incorporate both outdoor and indoor measurements, unlike the simulation presented here, which omits indoor data.

[1]The number of samples is not rounded due to the exclusion of indoor building areas in the simulation (only outdoor measurements were considered).

Current state-of-the-art sampling methodologies for benchmarking cellular network coverage offer a variety of approaches that can be tailored to the specific needs and constraints of MNOs. The choice of methodology depends on several factors, including the desired level of detail, the geographical scope of the study, and the available resources for conducting measurements. Consequently, optimizing the sampling strategy to match the unique characteristics of each propagation environment is crucial for achieving the most reliable results. As new technologies and analytical techniques continue to evolve, it is expected that the methodologies for cellular network benchmarking will similarly advance and improve.

6.2 Benchmarking with Sampling Theory

In cellular network performance evaluation, it is essential to collect representative measurements from large areas. Sampling theory provides a framework for selecting subsets of data that allow for accurate population estimates. However, due to factors like geographic variation and different network deployments, SRS may lead to biased results. To address this, stratification is used to divide areas into homogeneous strata, improving accuracy and efficiency in the sampling process.

This section explains the role of sampling theory in benchmarking, potential biases in SRS, and how stratification helps mitigate these biases.

6.2.1 Sampling Theory and Bias in Cellular Network Benchmarking

Sampling theory is a crucial statistical framework for selecting a representative subset of samples from a larger population to estimate parameters of the whole population. The goal is to use a smaller sample of size n from a population of size N to make reliable inferences about the entire population. This is particularly important in network performance benchmarking, where acquiring measurements across a wide area is essential but often impractical due to resource constraints.

In scenarios like cellular network measurements, where conditions vary significantly (e.g., rural vs. urban, mobile vs. stationary, or close vs. far from the BS), the distribution of measurements can be skewed. A common approach in such cases is SRS, where samples are selected randomly from the population, assuming that each sample has an equal probability of being chosen. However, this method can introduce biases if the sample distribution does not accurately represent the diversity of the entire population. For instance, one MNO may have more measurements in rural areas, while another may have an over-representation in urban locations. Additionally, variations in network deployments among different operators, such as distances from the BS, can further distort the results.

To mitigate such biases, one approach is to limit the analysis to regions where all MNOs have a sufficient number of users or measurements. For example, focusing on urban areas, highways, or railways, where there is a uniform distribution of users across different operators, can reduce the risk of bias. Another mitigation strategy involves SS, which can improve the representativeness of the sample by addressing the heterogeneity within the population.

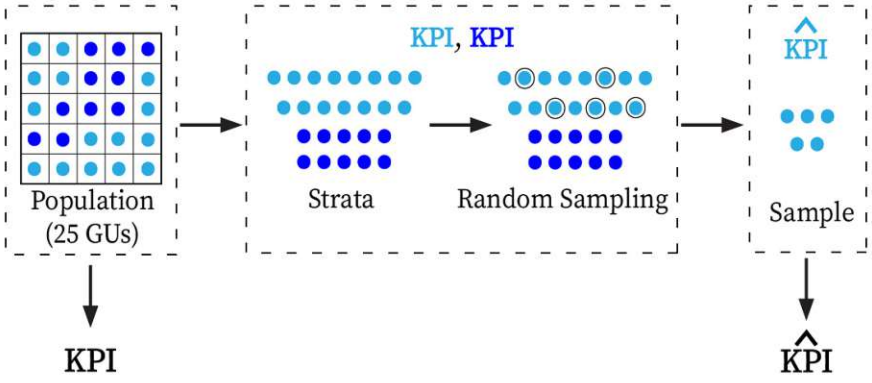


Figure 6.2: SRS of GUs in area of interest can lead to biased sample.

6.2.2 Stratified Sampling: Reduce Bias and Improve Efficiency

Stratification is a statistical technique that partitions the population into distinct, non-overlapping groups or strata based on specific characteristics, such as geographic location (urban vs. rural), traffic demand, or propagation conditions. These strata should be homogeneous within themselves, meaning that the variable of interest should exhibit low variance within each stratum. By dividing the measurement area into strata, it is possible to achieve more precise estimates with fewer samples compared to simple random sampling. The rationale is that, since the variance within each stratum is lower, fewer samples are needed to achieve the same level of confidence in the mean estimate.

In the context of benchmarking different MNOs across a large-scale area, stratification can help address distribution bias and improve accuracy. For example, instead of randomly sampling across the entire country, the area can be divided into smaller regions or GUs based on factors like population density, BS density, or signal quality. Each GU is then assigned to a specific stratum, ensuring that each region is represented proportionally to its importance in the overall analysis. Stratified sampling allows for more targeted data collection, ensuring that areas with varying propagation conditions are adequately represented.

Once the population is divided into strata, the next step is to determine how many GUs should be sampled from each stratum. This can be done using either proportional allocation or optimal allocation methods. In proportional allocation, the number of samples taken from each stratum is proportional to the size of the stratum, while optimal allocation adjusts the sample size based on factors such as the variance within each stratum and the cost of measurement.

For instance, in the case of a drive test campaign, the area of interest is split into tiles or GUs, and each tile is assigned to a specific stratum based on predefined criteria, such as BS density or traffic demand. The measurements taken from each GU are then averaged to compute a representative value for the entire stratum. This allows for a more efficient use of resources by ensuring that the number of samples required to achieve a reliable estimate is minimized.

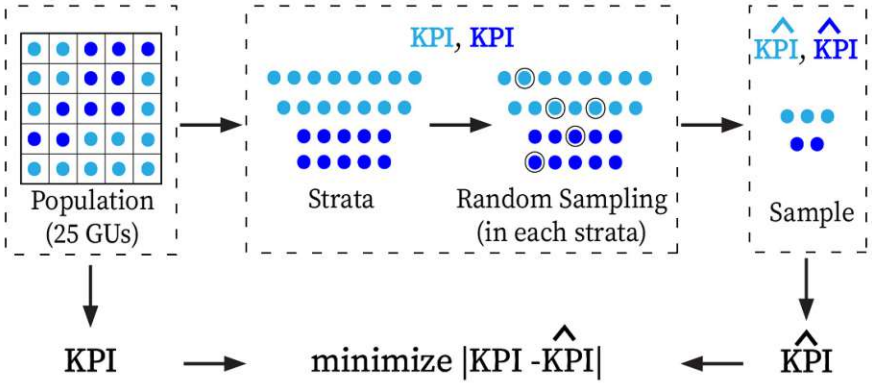


Figure 6.3: SS of GUs in area of interest.

Stratified sampling is also valuable for crowdsourced measurement campaigns, where data is collected by users with mobile devices. In these cases, the measurement points may be unevenly distributed, leading to biases if not properly accounted for. By grouping and averaging the data based on the strata, the impact of uneven measurement distribution can be removed, providing a more accurate map of the performance indicator.

6.3 Sampling Methodologies

This section provides a theoretical overview of two key sampling techniques - SRS and SS - in the context of cellular network performance benchmarking, including sample size requirements to achieve a specified level of estimation accuracy. Depending on the use case in Section 6.5, the population samples may refer to either individual KPI measurements or their aggregated values at the GU level.

6.3.1 Simple Random Sampling

Consider a population of size N, where each sample from the population is associated with a value of the variable of interest y. The sample mean of that population μ and the sample variance σ^2 are calculated as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i, \tag{6.1}$$

and

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu)^2. \tag{6.2}$$

By drawing n samples at random following a uniform distribution from the given population, we can define the sample mean as:

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \tag{6.3}$$

which is an unbiased estimator of the population mean μ [142]. Equivalently, we can also estimate the population variance σ^2, by computing the sample variance s^2:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2 . \tag{6.4}$$

It can again be shown, that s^2 is an unbiased estimator of σ^2 [142].

Further, we derive expressions for the expected variance of the estimators themselves. The variance of the population mean estimator $\overline{y}$ is given by:

$$\mathrm{var}(\overline{y}) = \left(\frac{N-n}{N}\right) \frac{\sigma^2}{n}, \tag{6.5}$$

and can be estimated using the following expression:

$$\widehat{\mathrm{var}}(\overline{y}) = \left(\frac{N-n}{N}\right) \frac{s^2}{n}. \tag{6.6}$$

To obtain a confidence interval I for a given estimate, we select a small number α, which denotes the probability of our population mean being outside of the confidence interval I. Therefore, for the estimation of the sample mean $\overline{y}$, the population mean μ should lie in the interval I with probability $1 - \alpha$, i.e., we require that

$$P(\mu \in I) = 1 - \alpha, \tag{6.7}$$

where we consider all possible samples of size n.

Under the assumption that the population mean estimates are normally distributed under random sampling, we obtain the interval as:

$$I = \left[\overline{y} - t\sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}}, \quad \overline{y} + t\sqrt{\left(\frac{N-n}{N}\right) \frac{s^2}{n}} \right], \tag{6.8}$$

where t denotes the upper $\alpha/2$ point of the Student-t distribution with $n-1$ degrees of freedom. In nearly all practical scenarios, Eq. (6.8) holds due to the central limit theorem. As a rule of thumb, we can replace the Student-t distribution with the standard normal distribution whenever $n > 50$. For a more detailed discussion, we refer the interested reader to [142–144].

6.3.2 Requirements on Sample Size in Simple Random Sampling

Let θ denote a population parameter of interest, such as the population mean, and let $\hat{\theta}$ be an estimator of θ. Given a basic understanding of SRS, a natural question arises: *What sample size n is required to achieve a specified level of estimation accuracy?*

We define a maximum allowable error d, either in absolute or relative terms, between the estimator and the true parameter value. Furthermore, we allow for a small probability α that the estimation error exceeds this threshold. Specifically, the probabilities that the absolute or relative error surpasses d are constrained to be less than α, formally expressed as:

$$\alpha > \begin{cases} P\left(|\hat{\theta} - \theta| > d\right) & \text{for absolute error,} \\ P\left(|\hat{\theta} - \theta| > d|\theta|\right) & \text{for relative error.} \end{cases} \tag{6.9}$$

Under the assumption of an unbiased normally distributed estimator, the distribution of error normalized by the square root of estimator variance:

$$\frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} \tag{6.10}$$

approaches a standard normal distribution for large n. This allows us to reformulate the inequalities from Eq. (6.9) into Eq. (6.11), by introducing z as the upper $\alpha/2$ point of the standard normal distribution:

$$P\left(\frac{|\hat{\theta} - \theta|}{\sqrt{\text{var}(\hat{\theta})}} > z\right) = P\left(|\hat{\theta} - \theta| > z\sqrt{\text{var}(\hat{\theta})}\right) = \alpha. \tag{6.11}$$

For the case of the population mean estimator under SRS, we have $\theta = \mu$ and $\hat{\theta} = \overline{y}$, we thus need to solve:

$$z\sqrt{\text{var}\left(\overline{y}\right)} = \begin{cases} d & \text{for absolute error,} \\ d|\mu| & \text{for relative error,} \end{cases} \tag{6.12}$$

with the variance of our population mean estimator specified in Eq. (6.5). Solving Eq. (6.12) for n gives us:

$$n = \frac{1}{1/n_0 + 1/N} := n_{\text{srs}}, \tag{6.13}$$

with

$$n_0 = \begin{cases} \frac{z^2\sigma^2}{d^2} & \text{for absolute error,} \\ \frac{z^2\sigma^2}{d^2\mu^2} & \text{for relative error,} \end{cases} \tag{6.14}$$

for absolute and relative error respectively. To distinguish n in SRS from n in Section 6.3.4, we denote it as n_{srs}.

Note, that the main challenge in this setup is the selection of the population variance σ^2, which has to be done beforehand.

6.3.3 Stratified Sampling

Now we consider a heterogeneous population, which we split into a total of L groups or strata, such that the samples in each individual stratum are as similar as possible. We denote the number of population samples per stratum h as N_h, such that the total number of samples in the population is given by:

$$N = \sum_{h=1}^{L} N_h. \tag{6.15}$$

Equivalently, we define n_h, as the number of randomly drawn samples from stratum h. Then the total number of drawn samples is given by:

$$n = \sum_{h=1}^{L} n_h. \tag{6.16}$$

Further, we introduce the mean in a stratum h as μ_h:

$$\mu_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_i. \tag{6.17}$$

Depending on our use case, y_i represents either the KPI of interest in the i^{th} GU or the i^{th} KPI measurement sample.

By summarizing the estimates for the individual strata we can derive an unbiased estimator of the population mean μ. We denote this estimate as the stratified sample mean $\overline{y}_{\text{st}}$, that is given by:

$$\overline{y}_{\text{st}} = \frac{1}{N} \sum_{h=1}^{L} N_h \overline{y}_h. \tag{6.18}$$

Note, that Eq. (6.18) assumes SRS estimates $\overline{y}_h$ for each stratum.

The variance of the estimator in Eq. (6.18) is given by:

$$\text{var}\left(\overline{y}_{\text{st}}\right) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{\sigma_h^2}{n_h}. \tag{6.19}$$

The corresponding estimator of this variance can be derived by replacing the population variance σ^2 with the sample variance s^2 for each stratum:

$$\widehat{\text{var}}\left(\overline{y}_{\text{st}}\right) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{s_h^2}{n_h}. \tag{6.20}$$

Given the total number of samples to be drawn n, we need to specify how to allocate these samples across different strata. In the following, we distinguish between two different types of allocation:

1. **Proportional Allocation**
   If the strata differ in size, the proportional allocation could be used to maintain a steady sampling fraction throughout the population. Here, we simply select the number of samples in accordance with the overall units N_h per stratum such that:

$$n_h = \frac{nN_h}{N}. \tag{6.21}$$

2. **Optimal Allocation**
   Optimal allocation results in the population mean estimate with the lowest variance for a fixed total number of samples n. Here, we have for the number of samples per stratum h:

$$n_h = \frac{nN_h\sigma_h}{\sum_{k=1}^{L} N_k\sigma_k}. \tag{6.22}$$

   For this, we again have to estimate the stratum variances σ_h^2 in advance. Typically, this is done with past data.

6.3.4 Requirements on Sample Size in Stratified Sampling

As in SRS, we begin by specifying a maximum allowable error d between the estimator and the true value, while permitting a small probability α that this error exceeds the threshold. When estimating the stratified sample mean $\overline{\mu}_{\text{st}}$, this corresponds to requiring that the population mean μ lies within a confidence interval I with probability $1-\alpha$, i.e., we require that:

$$\mathrm{P}(\mu \in I) = 1 - \alpha, \tag{6.23}$$

where we consider all possible samples of size n.
Assuming $\overline{y}_{\text{st}}$ to be normally distributed, using the central limit theorem the confidence interval can be constructed as follows:

$$I = \left[\overline{y}_{\text{st}} - z\sqrt{\text{var}\left(\overline{y}_{\text{st}}\right)}, \overline{y}_{\text{st}} + z\sqrt{\text{var}\left(\overline{y}_{\text{st}}\right)}\right], \tag{6.24}$$

where z is the upper $\alpha/2$ point of the standard normal distribution. For a more detailed discussion, we refer the interested reader to [143, 144].
Following the same approach as in Eqs. (6.9) and (6.10) we can derive the required sample size for fixed choices of d and α. W.l.o.g. we define $d_{\text{abs}} := d$ and $d_{\text{rel}} := d|\mu|$, and solve for $d_{\text{X}} = z \cdot \sqrt{\text{var}\left(\overline{y}_{\text{st}}\right)}$, where d_{X} denotes either d_{abs} or d_{rel}. Thereby, we insert the expression for variance from Eq. (6.19) and corresponding allocation according to Eqs. (6.21) and (6.22), to get for:

1. **Proportional Allocation**

$$\begin{aligned}\frac{d_{\mathrm{X}}^2}{z^2}N^2 + \sum_{h=1}^{L} N_h\sigma_h^2 &= \sum_{h=1}^{L} \frac{N_h^2\sigma_h^2}{n_h} \\ &= \frac{1}{n}\cdot\sum_{h=1}^{L} N_h\sigma_h^2 N.\end{aligned} \tag{6.25}$$

Hence, the required number n to achieve the error bound d with probability $1-\alpha$ using proportional allocation is given by:

$$n = \frac{\sum_{h=1}^{L} N_h\sigma_h^2 N}{\frac{d_{\mathrm{X}}^2}{z^2}N^2 + \sum_{h=1}^{L} N_h\sigma_h^2} := n_{\mathrm{prop}}. \tag{6.26}$$

2. **Optimal Allocation**

$$\begin{aligned}\frac{d_{\mathrm{X}}^2}{z^2}N^2 &= \sum_{h=1}^{L} N_h^2\left(\frac{N_h - n_h}{N_h}\right)\frac{\sigma_h^2}{n_h} \\ &= \sum_{h=1}^{L} \frac{N_h^2\sigma_h^2}{n_h} - \sum_{h=1}^{L} N_h\sigma_h^2, \\ \frac{d_{\mathrm{X}}^2}{z^2}N^2 + \sum_{h=1}^{L} N_h\sigma_h^2 &= \sum_{h=1}^{L} \frac{N_h^2\sigma_h^2}{n_h} \\ &= \frac{1}{n}\cdot\sum_{h=1}^{L} N_h\sigma_h\cdot\left(\sum_{k=1}^{L} N_k\sigma_k\right) \\ &= \frac{1}{n}\cdot\left(\sum_{h=1}^{L} N_h\sigma_h\right)^2.\end{aligned} \tag{6.27}$$

Hence, the required number n to achieve the error bound d with probability $1-\alpha$ using optimal allocation is given by:

$$n = \frac{\left(\sum_{h=1}^{L} N_h\sigma_h\right)^2}{\frac{d_{\mathrm{X}}^2}{z^2}N^2 + \sum_{h=1}^{L} N_h\sigma_h^2} := n_{\mathrm{opt}}. \tag{6.28}$$

Finally, by substituting d_{X} with d_{abs} or d_{rel} in Eqs. (6.26) and (6.28), depending on the error measure in question, the number of samples (n_{prop}, n_{opt}) required to remain below the given error bound is calculated. As was the case in SRS, the main challenge in stratification setup is the selection of the strata variances σ_h^2, which has to be done beforehand. The main advantage of SS is that it typically requires a smaller sample size n to achieve the same error bound compared to SRS. This allows MNOs to obtain accurate estimates of overall network quality using fewer GUs or measurement samples.

In the following section, we analyze how the strata means, variances, and sizes affect the performance of SRS, SS with proportional allocation, and SS with optimal allocation.

6.4 Sampling of Simulated Data Set

Three parameters influence the sampling performance with respect to the given error bound d:

- strata means μ_h,
- strata variances σ_h^2, and
- strata sizes N_h,

where $h = 1, 2, \ldots, L$.

To determine the required number of samples in regions that have not yet been measured, we must estimate variances based on similar previously measured areas. This represents a key challenge, as calculating the required sample size in advance depends on information we typically only obtain after measurements are conducted. While Section 6.5 distinguishes between the use cases where each sample corresponds either to a single measurement or to the measurement average within a single GU, the general simulation results presented in this section are applicable to both.

Despite this challenge, we can investigate how sampling parameters influence estimation accuracy by simulating artificial datasets under controlled conditions. Each artificial sample may represent either a single measurement or a GU average, depending on the later application. The simulation-based approach allows us to quantify how strata means, variances, and sizes affect the required number of samples in different sampling schemes.

In Sections 6.4.1 and 6.4.2, we examine the impact of these parameters on sampling performance by analyzing a general use case and two boundary special cases to gain a comprehensive understanding of their influence.

6.4.1 Sampling Performance: General Use Case

Using a set of parameters $S = \{\mu_h, \sigma_h, N_h\}$ we generate an artificial stratified dataset, by drawing samples from a normal distribution with mean μ_h and standard deviation σ_h of size N_h for each stratum, finally combining these strata into a full artificial dataset. Various statistical distributions can be employed in generating the artificial dataset, e.g., Student-T, Gumball, and even multi-modal Gaussian, provided that the resulting sample mean is sufficiently close to a normal distribution to justify the use of standard statistical inference techniques.

W.l.o.g., we show the results simulated using normal distribution. The simulation parameters we used for generating an artificial RSRP dataset consisting of two strata are provided in Table 6.1 (General Use Case). Next, we choose $\alpha = 0.05$ and generate a range for acceptable sampling error d, which is represented on the x-axis of Fig. 6.4. For each of the d-values, we use Eqs. (6.13), (6.26) and (6.28) to calculate the required number of samples for each sampling method (solid curves in Fig. 6.4). Thereby we assume perfect knowledge of N_h and σ_h for both strata - something we will need to

Table 6.1: Simulation parameters for the three use cases shown in Fig. 6.4, listed in the same order. σ denotes the SD, i.e., the square root of the variance.

	Stratum 1 (S1)	Stratum 2 (S2)
General Use Case	$\mu_1 = -110\,\text{dBm}$ $\sigma_1 = 15\,\text{dBm}$ $N_1 = 8\,000$	$\mu_2 = -85\,\text{dBm}$ $\sigma_2 = 5\,\text{dBm}$ $N_2 = 5\,000$
Equal Strata Variances	$\mu_1 = -110\,\text{dBm}$ $\sigma_1 = 10\,\text{dBm}$ $N_1 = 8\,000$	$\mu_2 = -85\,\text{dBm}$ $\sigma_2 = 10\,\text{dBm}$ $N_2 = 5\,000$
Equal Strata Means and Sizes	$\mu_1 = -95\,\text{dBm}$ $\sigma_1 = 15\,\text{dBm}$ $N_1 = 6\,500$	$\mu_2 = -95\,\text{dBm}$ $\sigma_2 = 5\,\text{dBm}$ $N_2 = 6\,500$

estimate for real-world data. For proportional and optimal SS we further use Eqs. (6.21) and (6.22) respectively, to calculate the number of samples required in each stratum (dashed curves in Fig. 6.4). Horizontal dashed lines illustrate N_h - the population total for each stratum. Considering Fig. 6.4a, assume we want to predict the mean value of the simulated RSRP dataset. If we are willing to accept that the absolute error between the mean estimate (calculated from the sample) and the population mean is in 95% of the cases $(1-\alpha)$ below $d = 10^{-1}$, then we require $n_{\text{srs}} \approx 11\,800$ samples ($\approx 90\%$) using the SRS method. On the other hand, using the proportional stratification method, this number reduces to $n_{\text{prop}} \approx 10\,600$ ($\approx 80\%$), while the optimal stratification method requires only $n_{\text{opt}} \approx 9\,000$ (less than 70%) samples in total (solid lines in Fig. 6.4a). While in proportional stratification the number of samples in each stratum proportionally rises until reaching its boundary at respective N_h (red dashed curves), the optimal scheme exploits the knowledge of strata variances.

For lower strata variance, fewer samples are required for accurate stratum mean estimation (*S2 SS opt* curve in Fig. 6.4a). However, if the variance in a stratum is too high, the algorithm would require more samples than we have available (*S1 SS opt* curve overshoots the N_1 level). This might represent a problem if we are, for instance, trying to determine how many tiles need to be measured in the area of interest, as we cannot simply introduce more tiles in that area. However, if we are determining the number of required measurements, we merely would have to measure more in those areas where the variance is high. The dotted part of the n_{opt} (*total SS opt* curve in Fig. 6.4a) curve symbolizes the overshot of one of the stratum curves.

For verifying that under SRS of n_h samples in each stratum, we truly remain under set estimate error level d in $(1-\alpha)100\%$ of the cases, we apply the verification algorithm provided in Section A5.

6.4.2 Special Cases

To assess under which conditions the added complexity of SS is justified, and when SRS remains preferable, we consider the following specific scenarios.

Scenario I: Equal Variances Let us assume all strata variances are equal, i.e., the SD are equal:

$$\sigma_1 = \sigma_2 = \cdots = \sigma_L := \sigma_s, \tag{6.29}$$

then we can simplify Eqs. (6.26) and (6.28) to:

$$\begin{aligned} n_{\text{prop}} &= \frac{\sum_{h=1}^{L} N_h \sigma_h^2 N}{\frac{d_X^2}{z^2} N^2 + \sum_{h=1}^{L} N_h \sigma_h^2} = \frac{N \sigma_s^2 \sum_{h=1}^{L} N_h}{\frac{d_X^2}{z^2} N^2 + \sigma_s^2 \sum_{h=1}^{L} N_h} \\ &= \frac{\sigma_s^2 N^2}{\frac{d_X^2}{z^2} N^2 + \sigma_s^2 N}, \end{aligned} \tag{6.30}$$

$$\begin{aligned} n_{\text{opt}} &= \frac{\left(\sum_{h=1}^{L} N_h \sigma_h\right)^2}{\frac{d_X^2}{z^2} N^2 + \sum_{h=1}^{L} N_h \sigma_h^2} = \frac{\sigma_s^2 \left(\sum_{h=1}^{L} N_h\right)^2}{\frac{d_X^2}{z^2} N^2 + \sigma_s^2 \sum_{h=1}^{L} N_h} \\ &= \frac{\sigma_s^2 N^2}{\frac{d_X^2}{z^2} N^2 + \sigma_s^2 N}. \end{aligned} \tag{6.31}$$

From Eqs. (6.30) and (6.31), we notice that n_{prop} and n_{opt} are equal as long as there is no difference in variance among strata. This is also supported by the results of our simulations, as illustrated in Fig. 6.4b and based on the simulation parameters outlined in Table 6.1 (equal strata variances). In this simulation, two strata were used, each characterized by different means and sizes, but identical variances. The maximum number of available samples per stratum - i.e., the upper sampling limit - is indicated by horizontal dashed lines. The plotted curves show the number of samples required to remain within a specified absolute error bound d. In situations where the variances of the strata are identical, the n_{prop} and n_{opt} curves coincide, and both outperform the SRS approach. Thus, when the variances of all strata are equal, the use of an optimal allocation scheme - which is computationally more demanding than proportional allocation - is not beneficial. However, employing stratification in cases where the strata means and sizes differ surpasses the simple SRS approach.

Scenario II: Equal Means and Sizes
Let us now assume that strata variances are not equal, but strata means and strata sizes are, i.e.,

$$\mu_1 = \mu_2 = \cdots = \mu_L, \tag{6.32}$$

$$N_1 = N_2 = \cdots = N_L = \frac{N}{L}. \tag{6.33}$$

By substituting Eq. (6.33) into Eq. (6.26) we get:

$$\begin{aligned} n_{\text{prop}} &= \frac{\sum_{h=1}^{L} N_h \sigma_h^2 N}{\frac{d_X^2}{z^2} N^2 + \sum_{h=1}^{L} N_h \sigma_h^2} \\ &= \frac{\sum_{h=1}^{L} \frac{N}{L} N \sigma_h^2}{\frac{d_X^2}{z^2} N^2 + \frac{N}{L} \sum_{h=1}^{L} \sigma_h^2} = \frac{1}{\frac{d_X^2}{z^2} \frac{L}{\sum_{h=1}^{L} \sigma_h^2} + \frac{1}{N}}. \end{aligned} \tag{6.34}$$

Since the means μ_h and strata sizes N_h are equal, we can prove that under the assumption of large N_h, the averaged variances of different strata will amount to the variance of the entire combined sample (see Appendix A4), i.e.:

$$\sigma^2 = \frac{\sum_{h=1}^{L} \sigma_h^2}{L}. \tag{6.35}$$

By inserting Eq. (6.35) into Eq. (6.34) the SRS result from Eq. (6.13) is again obtained:

$$n_{\text{prop}} = \frac{1}{\frac{d_X^2}{z^2 \sigma^2} + \frac{1}{N}} = n_{\text{srs}}. \tag{6.36}$$

It follows that the proportional stratification provides no benefit to SRS in cases where the means and sizes of strata are identical. Hence, in such a scenario, proportional stratification is unnecessary as it does not improve estimate accuracy while only bringing higher computational costs. The optimal allocation remains the only viable option that can bring further improvement, as it still depends on different strata variances.

This behavior is also noticed in our simulation results in Fig. 6.4c with two strata S1 and S2. Simulation parameters are given in Table 6.1 (Equal Strata Means and Sizes). Here, the n_{srs} and n_{prop} curves overlap. However, n_{opt} outperforms them both as it requires fewer samples in total to achieve the same estimate. The dotted part of n_{opt}, where $d < 10^{-1}$, results from high variance in strata S2. The optimal scheme here requires more sampled segments than we have available in that strata, making such sampling unfeasible. However, for $d > 10^{-1}$ the performance of the optimal scheme is better than the proportional one. Thus, depending on the mean estimate error d we are willing to accept, we can use simulations to choose the sampling scheme for a specific use case at hand.

If we additionally assume equal variances across all strata (Table 6.1 (Equal Variances), the sampling methods - SRS, SS with optimal allocation, and SS with proportional allocation - can be reduced to a common formulation. In this special case, all three approaches yield equivalent performance in terms of error bounds. However, they still differ in terms of implementation complexity. Given this, it is generally advisable to select the simplest method, namely SRS.

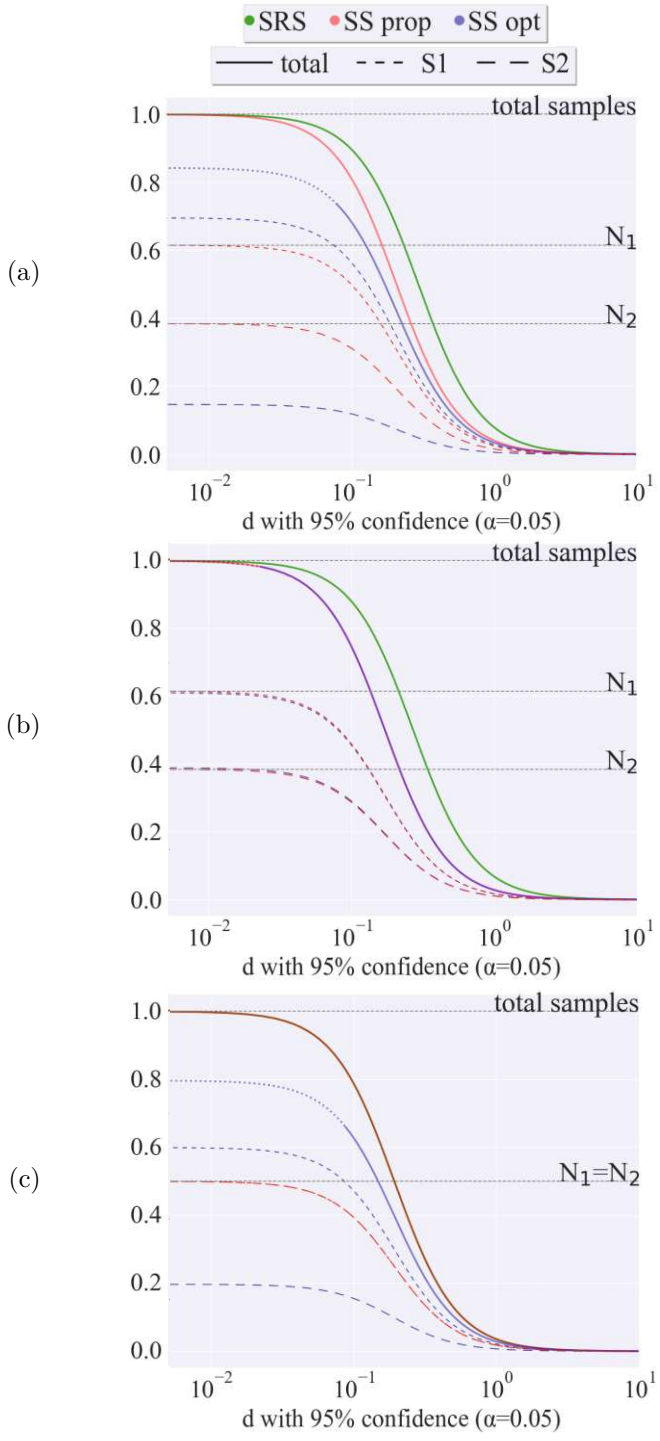


Figure 6.4: Simulation results for two strata (S1 and S2): (a) General Use Case, (b) Equal Variances, and (c) Equal Means and Sizes. Simulation parameters are given in Table 6.1. N_1 and N_2 are the total samples in each stratum, with $N_1 + N_2 = N$ as the total. Curves and vertical lines are normalized by N to show n/N as a fraction.

6.5 Sampling the Crowdsourced Data Set

This section demonstrates the practical applicability of the proposed sampling methodologies by applying them to a real-world crowdsourced MDT dataset. Section 6.5.1 provides an overview of the dataset and its characteristics. In Sections 6.5.2 and 6.5.3, we evaluate and compare the sampling methods at both the individual measurement level and the GU level, respectively. Our results demonstrate how the methods perform under real network conditions and validate their effectiveness beyond synthetic simulations.

6.5.1 Crowdsourced Data Set

We evaluated the sampling methods on real-world data using MDT datasets from a live LTE network provided by a major Austrian MNO, offering extensive coverage and spatial diversity. The sampling algorithms were applied directly to the data in a secure, on-premise environment, ensuring that all data remained within the operator's infrastructure in compliance with data protection requirements. The first dataset contains only measurements in the city of Vienna, Austria, consisting of a total of 10 000 000 samples. The second dataset contains 10 000 000 samples in the entire country of Austria in the same network. Features of both datasets are summarized in Tables 6.2 to 6.4.

Both datasets contain only measurements from the UEs connected to LTE macro BSs in the area of interest. Before applying sampling algorithms, we first clean up the data, by removing all measurements with erroneous or inaccurate attributes. To this end, we apply the following data processing:

- Using the reported GPS location, we filter out points outside the area of interest, even if they are connected to a BS within it. Since our goal is to estimate the mean KPI within a specific region, we apply geographic filtering based on city and country boundaries, respectively.
- Next, we filter out all points where the in meters reported Timing Advance (TA) is smaller than the calculated LOS Euclidean distance to the BS. TA corresponds to the time a signal takes to reach the BS from a mobile phone. The BS can use precise arrival time to determine the distance to the UE [145]. Since it is physically impossible for the TA values to be smaller than the minimum LOS distance, we exclude such measurement points from our datasets. For calculating the distance to the serving BS, we use Cellular Network DT.
- Since the operator's indoor/outdoor classification relies on GPS accuracy, we exclude measurements with high positional uncertainty by filtering out those with a reported GPS uncertainty greater than 45 m.
- Finally, we split the datasets by frequency band to account for the differing propagation characteristics across bands.

To test and compare different sampling methodologies on each of the MDT datasets, we have to split the data into statistically meaningful strata. Our goal is to obtain an accurate mean RSRP estimate with a minimum number of measurements and/or a minimum number of GUs used for the mean estimate calculation in each method. We investigate both alternatives in the following sections.

Table 6.2: Overview of Vienna and Austria MDT datasets.

	Vienna	Austria
Frequency [MHz]	800/1800/2600	800/1800/2600
Duration [days]	8	3
Area [km^2]	414	83 871
Total Samples	10 000 000	10 000 000
Filtered Samples	5 771 426	4 592 394
BS type	Macro cells	Macro cells
Number of Cells	693	3 468

Table 6.3: Austria, RSRP statistics per frequency band.

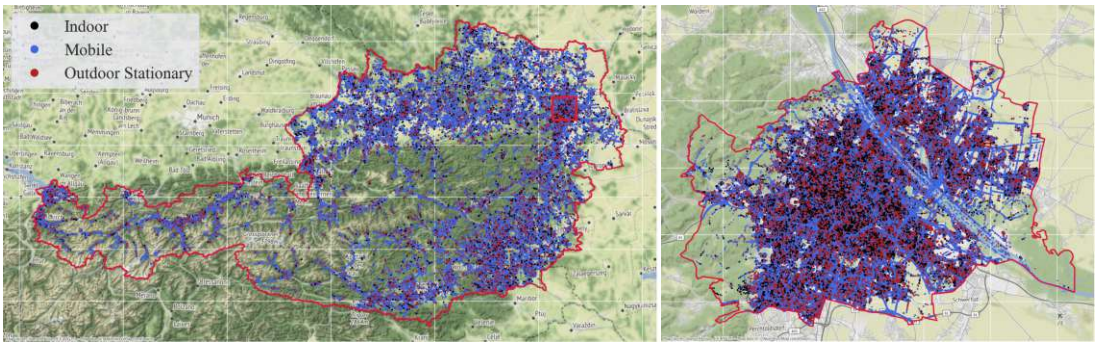
	800 MHz	1800 MHz	2600 MHz
RSRP Mean	−98.66 dBm	−94.96 dBm	−95.48 dBm
RSRP SD	12.33 dBm	10.95 dBm	11.29 dBm
Measurement Count	1 476 179	2 560 008	556 207

Table 6.4: Vienna, RSRP statistics per frequency band.

	800 MHz	1800 MHz	2600 MHz
RSRP mean	−96.77 dBm	−93.88 dBm	−96.21 dBm
RSRP SD	11.85 dBm	10.87 dBm	10.73 dBm
Measurement Count	1 100 826	3 546 656	1 123 944

6.5.2 Required Number of Measurements

In a real-world context, effectively leveraging the benefits of stratification necessitates identifying the specific conditions that strongly influence the reported KPI values. To achieve a suitable separation of strata, it is important to base the split on the statistical characteristics of the KPI in question. For the case of RSRP, one possible approach is splitting the data into groups based on factors such as radio conditions, position, and motion of the UE, since the reported value correlates with these parameters. The resulting groups exhibit distinctive RSRP statistics with different means and smaller standard variations when compared to the full dataset. On the other hand, for other KPIs such as reported T value, the relevant parameters for the stratification split decision are expected to be user tariffs, available bandwidth, and CL. If there is no prior knowledge regarding the statistics and dependencies of the KPI, clustering methods can be employed to identify underlying correlations in the dataset and determine the groups or strata. In a practical realization, the split can be based on the various regions of interest (e.g. tunnels, stadiums, stations, industry sites, rural and urban areas) that share common propagation conditions.



(a) Austria (b) Vienna, Austria

Figure 6.5: Filtered MDT measurement datasets of a large MNO in Austria. In the map on the left, the red square marks the city of Vienna. Different colors represent TrueCall Netscouts' radio condition-dependent *environment* classification, with the following categories: *Indoor*, *Mobile*, and *Outdoor Stationary*. We use these classes as three separate strata in the following.

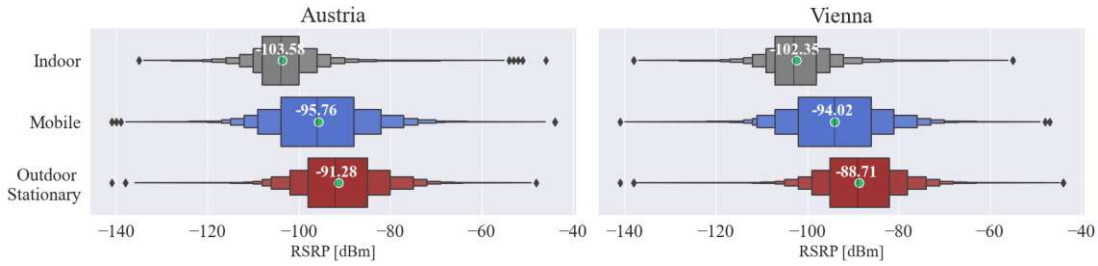


Figure 6.6: RSRP distribution per environment and dataset in 1800 MHz frequency band. Marked green points represent the mean RSRP values. Strata tendencies observed in the 800 and 2600 MHz bands are similar.

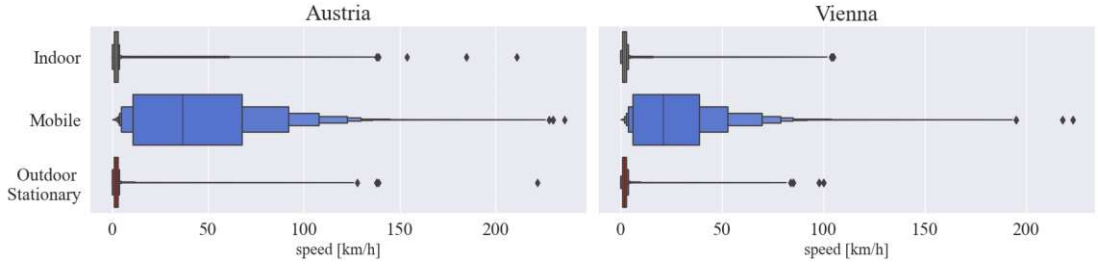


Figure 6.7: Speed profile per environment strata and dataset for all three frequency bands combined.

Fig. 6.5 depicts the filtered measurement datasets. Classification to *Indoor*, *Mobile*, and *Outdoor Stationary* measurements is provided by the operator and is based on radio conditions at the time of the measurement, as well as the GPS details. We use this classification to separate our measurement datasets into three corresponding groups

or strata. Figs. 6.6 and 6.7 provide more detailed characteristics of the measurement datasets in the form of boxenplots [146]. Since we are only considering UEs connected to LTE macro BS, the lowest mean RSRP value is observed in the *Indoor* scenario, due to high BPL [147, 148]. In the *Mobile* scenario, UEs frequently undergo BS handovers, along with numerous PCI changes, especially near cell edges where signal quality is lower. Conversely, the *Outdoor Stationary* case provides the best signal quality, as it avoids handovers and BPL. This trend is visible in Fig. 6.6 for both Austria and Vienna datasets, where indicated values represent the mean in the particular stratum. The speed profiles in Fig. 6.7 further illustrate that only *Mobile* stratum has average speeds higher than 5 km/h, with few outliers in both static strata. We can also notice the speed difference between the country and city dataset, with the city (right) having lower average speeds compared to the country (left) as expected.

We use these three strata, to calculate how many measurements from each group and in total are required to achieve a mean RSRP estimate with an error below level d in over 95% of the cases ($\alpha = 0.05$). The statistics of each stratum are provided in Tables 6.5 and 6.6.

Fig. 6.8 shows the comparison between different sampling methods. The x-axis represents the acceptable error d, while the y-axis indicates how many measurements are required per strata (dashed) and in total (solid curves) to achieve the mean RSRP estimate under the error bound d with 95% accuracy. Notice that in Tables 6.5 and 6.6 for both datasets, the *Mobile* stratum has the highest SD. Therefore, to achieve the estimation error of less than 10^{-2} dB we require more samples than we have available in that stratum in our crowdsourced datasets, which is why we have an overshoot over N_2 level in both cases for *optimal* allocation scheme. On the other hand, the *Indoor* stratum has the lowest SD in both datasets. Therefore, the *optimal* scheme does not require all available *Indoor* measurements for any requested estimation error level and remains well below the N_1 level. On the total (solid lines), *proportional* (red) and *optimal* (blue) allocation schemes performed almost identically since the difference among the SD levels in all three strata was much smaller compared to our simulated dataset in Section 6.4. However, both stratification schemes outperform the SRS (green). Considering that the performances of proportional SS and SRS would overlap for equal strata means and sizes, we notice that this difference originates in the discrepancy among different strata means. However, since their discrepancy is not as high as in the simulated scenario, we see only a minor advantage to SRS, which is valid for both city and country levels. Note, however, by finding an even better-suited strata split, than the one the operator is providing, the strata mean differences may increase, such that stratification schemes show a higher advantage to SRS than the one we are currently seeing. The findings demonstrate that, in the present configuration, obtaining a mean RSRP estimate with a mean absolute estimation error lower than 10^{-2} dB requires the utilization of 50-65% of the available measurements in both datasets. The specified threshold for mean absolute estimation error of 10^{-2} dB was chosen as an indication of a substantially high degree of accuracy in the RSRP estimation.

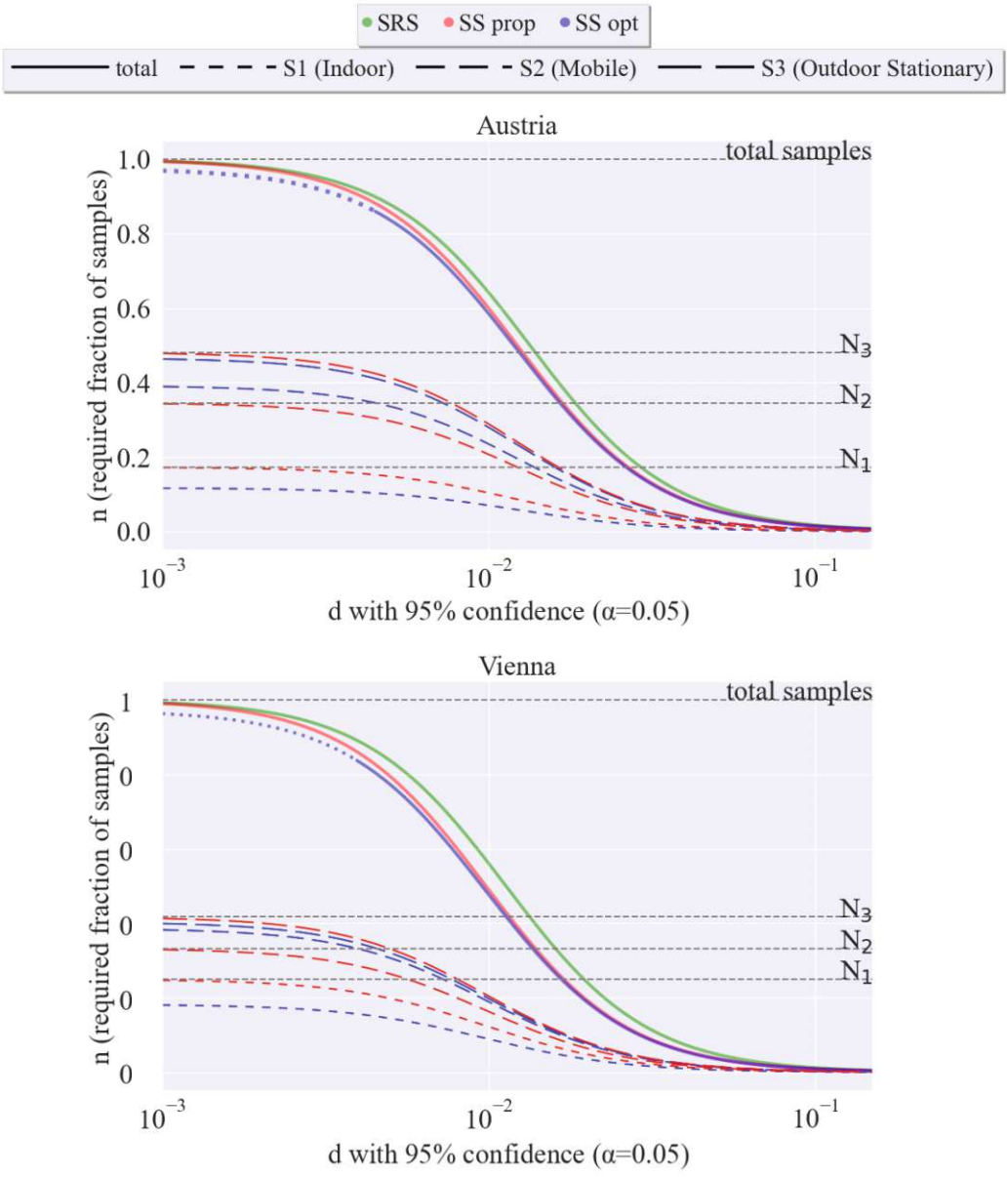


Figure 6.8: Required fraction of samples n plotted over acceptable estimation error d for different sampling schemes in Austria (top) and Vienna (bottom) dataset (1800 MHz), based on radio-condition and speed-dependent strata split. All curves are normalized by N.

Table 6.5: Strata RSRP statistics for Austria dataset in the 1800 MHz band.

Austria Data Set in 1800 MHz Frequency Band		
Indoor	Mobile	Outdoor Stationary
$\mu_1 = -103.5$ dBm	$\mu_2 = -95.76$ dBm	$\mu_3 = -91.28$ dBm
$\sigma_1 = 6.87$ dBm	$\sigma_2 = 11.49$ dBm	$\sigma_3 = 9.82$ dBm
$N_1 = 442\,864$	$N_2 = 884\,781$	$N_3 = 1\,232\,363$

Table 6.6: Strata RSRP statistics for Vienna dataset in the 1800 MHz band.

Vienna Data Set in 1800 MHz Frequency Band		
Indoor	Mobile	Outdoor Stationary
$\mu_1 = -102.35$ dBm	$\mu_2 = -94.02$ dBm	$\mu_3 = -88.71$ dBm
$\sigma_1 = 6.99$ dBm	$\sigma_2 = 11.12$ dBm	$\sigma_3 = 9.25$ dBm
$N_1 = 885\,728$	$N_2 = 1\,180\,224$	$N_3 = 1\,480\,704$

6.5.3 Required Number of Geographical Units

Given a measurement dataset, whether it is crowdsourced, drive- or train-test data, it often happens that measurements get accumulated in certain areas, and are infrequent in others. For illustration, a driving train in an ongoing measurement campaign makes longer stops at train stations, while on some parts of the tracks, it reaches speeds of 250 km/h. Under such conditions measurement dataset gets very confounded in time [126]. To solve this problem, we can bin the measurement data into GUs, take the average, and get representative KPI values per train length, independent of the train speed [13]. To remove measurement bias created by having more measurements in very crowded areas, compared to very few measurements in suburban/rural regions, we can apply the same binning strategy with crowdsourced data and then work with GU averages instead of the measurement samples.

Stratification sampling can also be used to determine which environment types and in what amount should be covered in a measurement campaign. For instance, rural and urban regions have different BS deployments, in terms of BS density, propagation loss, and LOS connectivity. This fact is particularly relevant in developing countries, where the population is more concentrated in rural areas despite poorer BS deployment [149]. To have a better understanding of the overall network quality in such conditions, and to be able to plan a measurement campaign more efficiently, we can split the area of interest, e.g. country, into fixed-sized GUs, and assign rural or urban property (strata) to each. With an estimate of the KPI SD in each stratum, we can determine how many GUs of each stratum we should cover to gain an accurate KPI benchmark.

We test sampling techniques on a GU-level, and we only take a look at a single dataset - Austria. To reduce the influence of the BPL we split the data into *Indoor* and *Outdoor* datasets and look at these scenarios separately. By splitting the entire country area

Table 6.7: Strata RSRP statistics for Austria Outdoor (left) and Austria Indoor (right) dataset in the 1800 MHz frequency band.

Austria Outdoor (1800 MHz)		Austria Indoor (1800 MHz)	
S1(*BS-sparse* GUs)	S2(*BS-dense* GUs)	S1(*BS-sparse* GUs)	S2(*BS-dense* GUs)
$\mu_1 = -102.75$ dBm	$\mu_2 = -96.49$ dBm	$\mu_1 = -109.29$ dBm	$\mu_2 = -105.11$ dBm
$\sigma_1 = 9.41$ dBm	$\sigma_2 = 7.79$ dBm	$\sigma_1 = 6.62$ dBm	$\sigma_2 = 5.10$ dBm
$N_1 = 23\,102$	$N_2 = 7\,778$	$N_1 = 5\,791$	$N_2 = 5\,177$

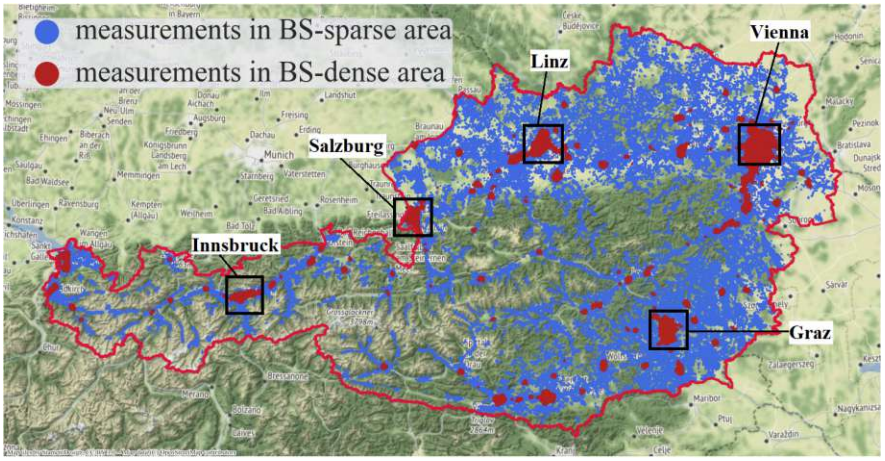


Figure 6.9: Filtered MDT measurement dataset of a large MNO in Austria. Different colors represent measurement classification based on the BS-density in a 2 km radius of the measurement GU. *BS-dense* deployments in red overlap with larger cities in Austria (denoted by black squares), while the blue corresponds to more suburban and rural areas.

into, e.g. 500 m × 500 m large, GUs, we can determine how many of them need to be taken into account for an accurate mean RSRP estimation.

Since we do not have a clear rural/urban area split in Austria, we use the macro BS locations in Austria (provided by the operator), to determine how many BSs are in a 2 km radius of each GU. If more than 50 BSs are found in the radius, then the GU is classified as urban or *BS-dense*, otherwise, we classify it as rural or *BS-sparse*. Measurements are then mapped to their belonging GU and thus to their corresponding stratum - Fig. 6.9 depicts the mapping of the outdoor measurement dataset. Notice that the red *BS-dense* areas overlap with larger cities in Austria, e.g. Vienna, Graz, Linz, Salzburg, Innsbruck.

All crowdsourced data are then binned by their GUs and the mean for each GU is calculated. These GU averages represent our population ground truth in the following sampling schemes. Table 6.7 presents the RSRP GU statistics in the 1800 MHz band. The statistics show as expected, that indoor and outdoor measurements have ≈ 7 dB discrepancy, while the strata alone in each dataset, show a difference of 4 to 6 dB.

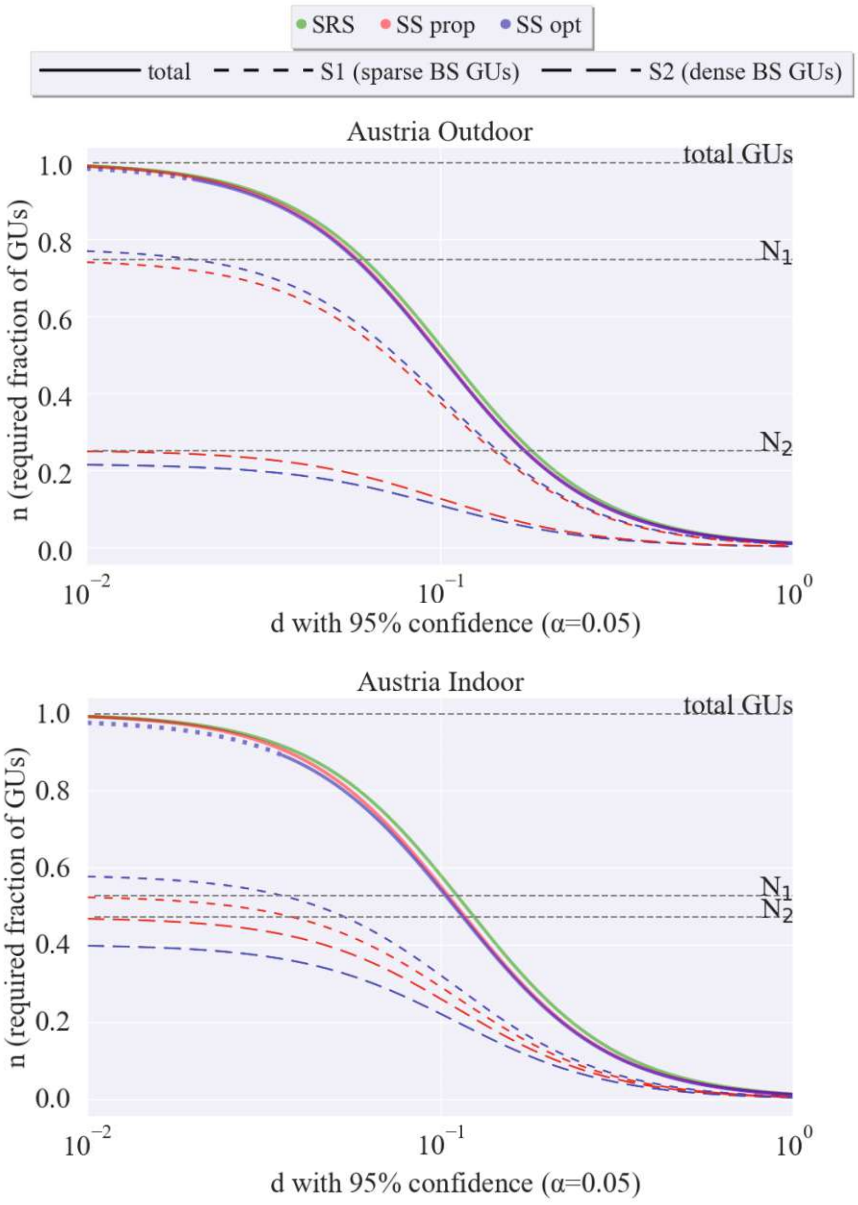


Figure 6.10: Required fraction of GUs n plotted over acceptable estimation error d for different sampling schemes in Austria Outdoor (top) and Indoor (bottom) dataset (1800 MHz), based on BS-density strata split. All curves are normalized by N.

In the Outdoor scenario, we have predominantly *BS-sparse* GUs, while in the Indoor scenario number of *BS-sparse* and *BS-dense* GUs is in the same order of magnitude. This indicates, that many GUs in Austria are missing Indoor measurements in the crowdsourced data. If we look at the performance comparison between applied sampling strategies in Fig. 6.10, we can hardly notice a difference in the performance of the three

sampling methods. This is the case due to the very small SD difference among the strata, as well as insufficient mean discrepancies. This testifies to the fact that the coverage of the operator in question is as good in rural areas as it is in urban, as separating the dataset into strata brings almost no advantage. Due to its higher variance, the *BS-sparse* stratum (S1) experiences an overshoot of the N_1 level when approaching the 10^{-2} dB error bound in both datasets.

To achieve a mean estimation error below 10^{-1},dB in 95% of cases - regardless of the sampling method - approximately 50% (15 440 GUs) of the currently measured GUs in Austria need to be covered in the Outdoor scenario. Similarly, in the Indoor scenario, we would need to cover around 55% (6 032 GUs) of the currently measurement-covered GUs to achieve the same level of accuracy.

6.6 Conclusion

Accurate estimation of KPIs in mobile networks is critical for improving network performance and customer satisfaction. Sampling methods can be used to estimate KPIs with an acceptable error level while minimizing the number of measurements required. In this chapter, we investigated the behavior of three sampling methods for accurate KPI mean estimation in mobile networks: SRS, SS with proportional allocation, and SS with optimal allocation.

To understand the conditions under which stratification is most beneficial, we evaluated the performance of different sampling methods on both synthetic and real-world data, focusing on scenarios relevant to cellular network measurements. We then analyzed the same sampling methods on two MDT crowdsourced datasets from a major Austrian operator. To test and compare different sampling methods on both MDT datasets, we stratified the data into meaningful strata to obtain an accurate mean RSRP estimate with a minimum number of measurements and/or a minimum number of GUs used for mean estimate calculations in each method. The analysis revealed that the first strata split, based on the GPS position, speed, and radio channel conditions, offered a subtle advantage of the SS methods over the SRS method. All three approaches resulted in between 50 and 65% of the total measurements being required to remain below a mean absolute estimation error of 0.01 dB in both datasets.

We further binned the data into equally sized GUs, removing the confounding by determining a single representative KPI value for each GU. Using BS-density-based stratification, we determined how many rural and urban GUs are required for accurate mean prediction. Again, we compared three sampling techniques while using calculated mean GU KPI values as our population ground truth. The analysis revealed that we would have to cover around 50% of the GUs to remain below a mean absolute estimation error of 0.1 dB. Stratification provided a minimal advantage to SRS due to the comparable coverage of this operator in rural and urban regions in Austria, with minor differences in mean and SD among these two strata.

While stratification provided only limited gains in real-world datasets for RSRP mean estimation, we can utilize these methods to determine how many samples or areas are required for determining the mean of any KPI in the network. For instance, considering

T and CL, possibly more distinct strata can be found to utilize stratification to its full benefit. The results presented in this chapter provide practical guidelines for network operators seeking to balance measurement effort with accuracy in KPI estimation. By leveraging SS, operators can determine the required number of measurements and measurement areas needed for reliable benchmarking, while minimizing costs and time. Across both theoretical derivations and simulation results, we observed that the required sample size is primarily driven by the variance - either of the full population (in the case of SRS) or the individual stratum variances (in SS). Additionally, SS offers clear advantages over SRS when there are substantial differences in the means across strata. However, in scenarios where strata are similar in both variance and mean, this advantage diminishes, and SRS becomes a more practical and equally effective approach. Moreover, the strata-specific variance values derived from real-world MDT data can serve as initialization parameters for future studies or optimization frameworks employing similar methodologies.

These measurement strategies can be directly transferred to railway scenarios, where measurement resources are often constrained, and coverage variability is high due to terrain, tunnels, and heterogeneous antenna deployments. In such use cases, fixed-length track segments (e.g., 100 m) would serve as the fundamental GUs, and stratification can naturally be applied across environmental classes such as rural open tracks, urban areas, tunnels, or stations. Although large-scale MDT-like data from railway operators was not available during this study, the simulation framework and theoretical derivations in Sections 6.3 - 6.4 remain entirely valid and can guide railway-specific campaign design. Thus, this chapter not only evaluates sampling approaches on existing crowdsourced data, but also provides a scalable strategy for future railway connectivity benchmarking efforts.

Chapter 7

Future Research Directions

Modern trains are no longer just vehicles - they are digital platforms on wheels, hosting a multitude of onboard services that go far beyond critical railway operations. Applications such as real-time video surveillance, passenger counting, infotainment, predictive maintenance, and energy monitoring all rely on continuous, high-quality wireless connectivity. These non-critical, yet essential services demand consistent network performance throughout the journey, particularly as passengers and operators increasingly expect seamless digital experiences on the move.

Extending on findings and methodologies established in this research, future work should focus on:

- **Crowdsourced Data Exploitation:** Enhancing the established benchmarking framework by incorporating crowdsourced MDT data to continuously validate and refine network performance predictions and optimization strategies.
- **Integration with FRMCS:** Further studies should evaluate the proposed methodologies and frameworks in the context of FRMCS deployments, leveraging advanced 5G and beyond technologies to fully digitalize railway operations and enhance connectivity robustness.
- **Holistic DT Expansion:** Further integrating various DT components - including predictive analytics, real-time monitoring, and extensive scenario simulations - to create an even more comprehensive, scalable tool for network performance management.

In the following sections, a brief discussion is provided on the implications of crowdsourced MDT data exploitation and the integration with FRMCS, as these are areas for future development and open challenges.

7.1 Railway Digital Twins Enabled by Crowdsourcing

Data-driven methods are becoming essential for cellular network optimization, enabling efficient performance evaluation. Crowdsourced data - particularly MDT - plays a central role in this shift, replacing traditional drive tests in several applications such as coverage validation, operator benchmarking, and spatiotemporal performance analysis [17, 150]. MDT provides structured access to user- and network-level measurements across idle and connected states [151], offering a cost-effective alternative to manual drive testing. However, it comes with challenges: reporting frequency varies greatly due to device and contextual factors, and GPS inaccuracies in typical UEs are especially pronounced on

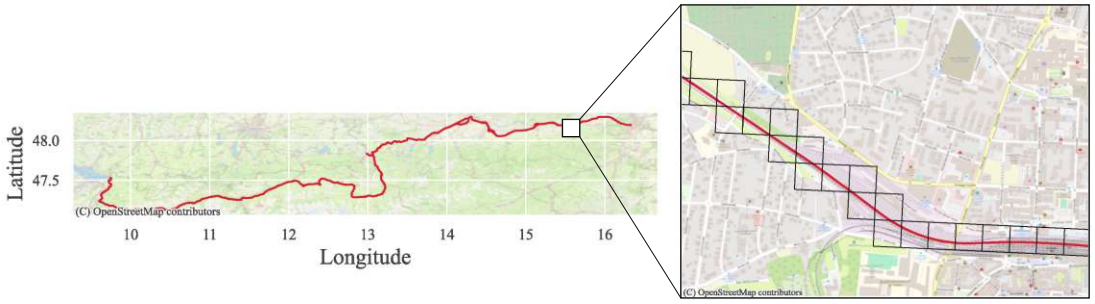


Figure 7.1: Visualization of the Vienna-Bregenz measurement track, showing the full route and a detailed view of intersecting Statistics Austria GUs [61]

trains due to movement, tunnels, train structure, and visibility constraints. This leads to scattered data that hinders track-aligned performance evaluations.

We propose a methodology to classify MDT data as either *in-train* or *out-of-train* using user-level speed and spatial movement across Statistics Austria's 100 m × 100 m GUs. Each measurement m includes a timestamp, anonymized UE ID, coordinates, speed, and network KPIs. A sample is labeled as *in-train* if the user has a median daily speed above 10 km/h and appears in at least five unique GUs along the track:

$$C(m) = \begin{cases} \textit{in-train}, & \text{if } \bar{v}_u > 10 \text{ km/h and } \mathrm{GU}_u \geq 5, \\ \textit{out-of-train}, & \text{otherwise.} \end{cases} \tag{7.1}$$

We applied this filtering to MDT data collected over a month along the Vienna-Bregenz train track. Although the raw dataset typically showed less than five GUs per user across the month, the classified *in-train* subset achieved around 14 GUs per user, offering improved coverage. However, this is still too sparse to track individual train journeys or evaluate performance per train type - highlighting the current limitation in MDT data granularity for benchmarking specific onboard configurations. Nevertheless, such crowdsourced datasets lay the groundwork for future train-specific DT models, where statistical movement, connectivity patterns, and onboard experience could be inferred without active measurement campaigns.

Fig. 7.2 shows a side-by-side comparison between MDT and controlled drive-test data collected on the same route. The drive test, sampled every 500,ms, captures detailed spatial variations, whereas MDT sampling - affected by device type, network conditions, and battery status - occurs at irregular intervals with a median of 60 s, during which the train may travel several kilometers. This results in sparse measurements and lower spatial resolution. Despite this, signal strength distributions (Fig. 7.3) remain comparable, and aggregated GU-level averages show a low MAE of 6.25 dB over segments where both datasets report measurements on 1800 MHz.

These findings indicate that with targeted filtering - potentially enhanced by Artificial Intelligence (AI)-driven classification - MDT data can serve as a viable proxy for drive tests in large-scale, long-term coverage evaluation.

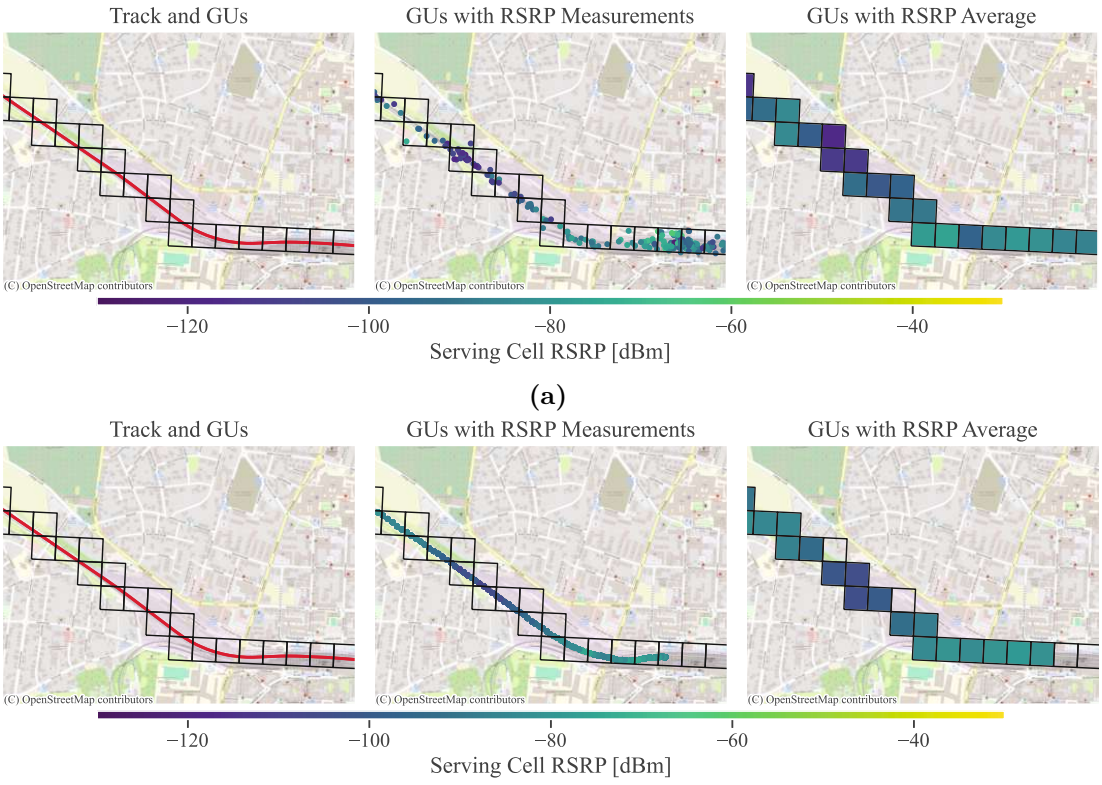


Figure 7.2: Granularity comparison: (a) MDT data and (b) drive test data. The red line indicates the track, scatter points are RSRP samples, and right-most maps show averaged GU-level RSRP.

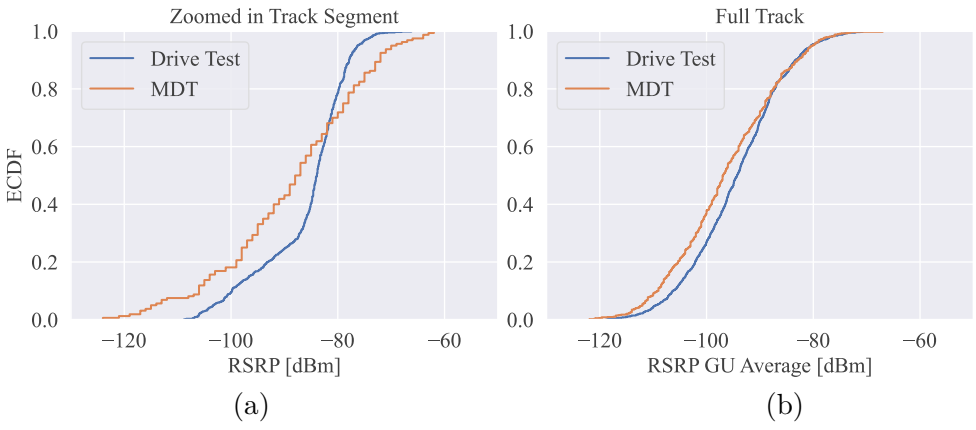


Figure 7.3: ECDF comparison of GU-averaged RSRP for MDT and drive test data: (a) zoomed-in segment from Fig. 7.1, (b) full Vienna-Bregenz track.

Over time, with access to contextual metadata (e.g., train schedules, seating positions, or equipment tags), these data streams could inform DT models that emulate train characteristics such as antenna placement, cabin attenuation, and user density - enabling passive, data-driven modeling of wireless environments onboard. The long-term vision is to enable real-time online learning of railway DTs from end-user observations, eliminating the need for repeated measurement campaigns. Future work should focus on integrating such data streams with AI-based tools to enable intelligent, real-time optimization of railway network performance.

7.2 Digital Twins for Enabling FRMCS Deployment

The FRMCS is envisioned as the successor to GSM-R, addressing its limitations and enabling the digital transformation of railway communications. Building on 5G principles, FRMCS aims to deliver enhanced data rates, Ultra-Reliable Low-Latency Communication (URLLC), and advanced features such as network slicing to support a wide variety of use cases - from mission-critical applications like European Train Control System (ETCS) and ATO to non-critical services such as passenger infotainment.

While this dissertation has focused on benchmarking current commercial LTE and early 5G solutions in railway environments, the presented methods are directly applicable to the evaluation of future FRMCS deployments. In particular, the use of drive-test models and crowd-sourced MDT-based performance estimation could be extended to the FRMCS context, supporting early deployment validation and optimization. For example, measurement-assisted RT could be a valuable tool for assessing interference risks and propagation characteristics in the FRMCS band prior to commercial roll out.

One of the key concerns surrounding FRMCS is its spectrum coexistence with existing Mobile/Fixed Communication Networks (MFCN). The European Commission has allocated the [1900-1910] MHz unpaired band for Railway Mobile Radio (RMR) applications, which is to be used by FRMCS. This Time Division Duplex (TDD) band lies adjacent to the DL of 3GPP Band 3 (1800 MHz) and the UL of Band 1 (2100 MHz)- two of the most commonly used frequency bands in commercial LTE and 5G NR networks. As shown in Figure 7.4, this spectral adjacency introduces the potential for harmful interference, particularly between the high-powered FRMCS DL and the comparatively weaker MFCN UL signals.

This poses significant challenges, as FRMCS networks are expected to serve safety-critical functions. The risk of interference from or to adjacent commercial networks - particularly MFCN BS emissions into the FRMCS UL, and FRMCS DL leakage into MFCN UL-must be carefully assessed. DTs provide a promising framework for this task, enabling virtual coexistence testing of FRMCS and commercial networks under various interference conditions, antenna configurations, and deployment topologies. Future work should explore these interference mechanisms using simulation tools such as RT (e.g., Sionna RT) and validate them through targeted field measurements. While FRMCS hardware is still in development, the propagation characteristics of nearby commercial bands (e.g., Band 3) can serve as a useful proxy for modeling expected FRMCS signal behavior.

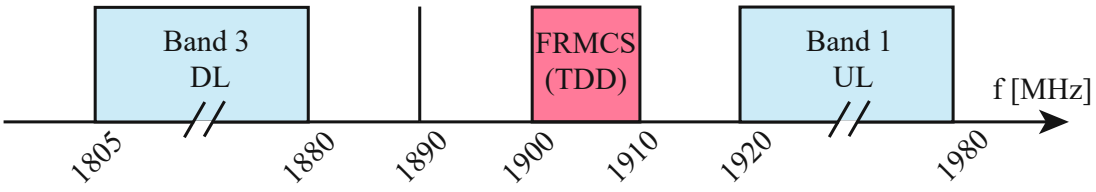


Figure 7.4: Radio spectrum planned for FRMCS frequency band and neighboring MFCN bands.

Additionally, future studies should investigate coordinated deployment strategies to mitigate interference, including:

- spatial separation and alignment of FRMCS and MFCN BSs (e.g., minimum separation distances),
- the use of notch filters to suppress out-of-band emissions,
- adaptive antenna configurations and sector alignment, and
- potential reductions in transmit power in sensitive areas.

Current European Conference of Postal and Telecommunications Administrations (CEPT) and Electronic Communications Committee (ECC) reports (e.g., [152–154]) underscore the absence of real-world deployment data for future railway communication systems. To address this, future work should explore simulation-assisted, measurement-informed approaches that can guide planning and evaluation in the absence of standardized FRMCS equipment and live deployments.

There is also considerable potential in integrating FRMCS into existing railway infrastructure. Reusing legacy GSM-R sites or mounting equipment on catenary masts offers practical paths forward. These integration scenarios call for further investigation into deployment planning, infrastructure reuse, and the impact of train-borne system characteristics such as handover behavior and penetration loss.

Looking ahead, advancing tools for scalable benchmarking - both through simulations and real-world data - will be essential for evaluating hybrid commercial and mission-critical networks coexisting along the same tracks. In particular, research into AI-assisted planning, adaptive network configuration, and DT-based optimization will be key to ensuring reliable, interference-aware deployment of next-generation railway communication systems like FRMCS.

Chapter 8

Conclusion

This dissertation has explored the benchmarking and optimization of cellular connectivity in railway environments, addressing both theoretical and practical challenges of ensuring reliable communication on moving trains.

Our investigation began with high-resolution RT simulations aimed at modeling wireless performance under realistic railway conditions. While these simulations provided an initial view of coverage patterns and propagation behavior, they exhibited significant discrepancies compared to real-world measurements. This was not due to a lack of metadata or deployment information, but rather the need to calibrate simulation parameters using ground-truth data. Even after calibration, deviations persisted - highlighting that RT, while powerful, must be complemented by real-world measurements to capture the full complexity of live network conditions.

To address this gap, we designed and executed extensive measurement campaigns aboard moving trains, capturing both indoor and outdoor RSRP using scanners and user equipment. These campaigns generated large, heterogeneous datasets that required robust methodologies for processing and interpretation. In response, we developed a scalable measurement framework that integrated real-world measurements with DT representations of the environment and network. This enabled spatial alignment of samples, correction on confounding effects, and extraction of environment-aware metrics - ensuring that benchmarking results were both accurate and reproducible. As a further contribution to the research community, we published multiple processed datasets from these campaigns, enabling transparent validation and supporting future studies on railway connectivity.

Since trains operate on fixed routes and hardware-level DTs remain largely unavailable, direct comparison under identical conditions is infeasible. To fairly evaluate in-train connectivity solutions across different train types and routes, we focused on generalizable performance indicators. We used VPL as a key metric to isolate the cabin's contribution to signal degradation. This approach removes the influence of PL and BS deployment differences around individual measurement tracks, as both indoor and outdoor signals are equally affected by these external conditions. By examining outdoor-indoor RSRP differences, we could robustly evaluate and compare different train configurations and connectivity solutions.

Motivated by the industry's increasing shift toward passive window solutions, We then investigated the influence of the signal's AoA on VPL. While window coatings offer advantages such as broad frequency transparency, low maintenance, and compatibility

with modern MIMO systems, their effectiveness is sensitive to the direction of incoming signals. To support this analysis, we developed a robust framework for estimating AoA in motion using GPS-tagged UE and scanner measurements in combination with Environment and Rail DT and Cellular DT - eliminating the need for controlled lab setups. This framework is applicable not only to controlled datasets but also to future use cases involving crowdsourced data.

Our AoA-aware analysis revealed that high azimuth angles - common when BSs are placed close to the tracks - lead to significantly higher VPL. These placements, historically suitable for repeater-equipped trains, pose challenges for permeable window designs. As a result, network planning strategies may need to be re-evaluated: for instance, increasing the lateral distance of BSs from the track could reduce azimuth angles and improve penetration, but would also increase PL. This trade-off highlights a promising optimization space in which BS positioning is tailored to onboard connectivity architecture.

To further explore how train design affects mobile signal propagation, we conducted a targeted measurement campaign on the ÖBB Nightjet, comparing three cabin types within the same train using identical window coatings. This controlled study showed that, despite having comparable total window areas, cabins with more internal partitions (e.g., walls and doors) introduced up to 3 dB higher attenuation. Applying our benchmarking framework, we also quantified VPL across various ÖBB train types: standard metal-coated windows resulted in VPL of approximately 22 dB, while modified FSS-coated windows reduced this to around 14 dB. In lightweight commuter trains with larger windows, attenuation dropped further to approximately 10 dB - highlighting the combined impact of coating technology and overall cabin design.

Recognizing the resource-intensive nature of train measurements, we developed a SS methodology tailored to mobile network benchmarking. This approach enables reliable estimation of key performance indicators (KPIs) while significantly reducing the number of required measurements. By allocating sampling efforts based on strata such as environmental type, BS density, and user speed, our method ensures that measurements are concentrated where they are most informative.

Although railway-specific large-scale datasets were unavailable during this study, we validated our methodology using two real-world MDT datasets: one city-scale (Vienna) and one national-scale (Austria). Across both, we demonstrated that SS can achieve mean RSRP estimates with minimal absolute error using only [50-65]% of the original data. This demonstrates the method's efficiency and its applicability to large-scale mobile measurement campaigns. Stratification proved especially useful in heterogeneous environments, and its design is highly transferable to railway settings - with fixed-length track segments as natural GUs and strata based on tunnel presence, urbanization level, or topography.

Looking ahead, we recognize the untapped potential of crowdsourced datasets for railway connectivity evaluation. As access to onboard-generated data - such as MDT reports or logs from railway operator routers - increases, these sources may offer a scalable, low-cost alternative to traditional measurement campaigns. Integrating such

data into our SS framework could enable benchmarking of railway connectivity across train types, regions, and time periods.

In parallel, the ongoing FRMCS roll-out presents new challenges and opportunities. As this mission-critical standard replaces GSM-R, further research will be essential to evaluate its coexistence with commercial networks, its propagation behavior in diverse railway environments, and its infrastructure requirements. While the tools developed in this dissertation were not designed specifically for FRMCS, they can support related investigations: the DT-based environment modeling provides a solid foundation for RT simulations of FRMCS deployments, while the measurement datasets and crowdsourced frameworks introduced here can assist in calibrating these simulations against real-world conditions.

In summary, this dissertation delivers a comprehensive framework for assessing cellular connectivity in railway environments, grounded in real-world measurements and supported by scalable analytical tools. The developed methods address current benchmarking challenges and lay the groundwork for future network deployments, contributing both to academic research and practical improvements in rail communication systems.

Appendix

A1 Gaussian Process Regression

This appendix provides a concise overview of GPR as presented in Rasmussen and Williams' *Gaussian Processes for Machine Learning* [155]. GPR is a non-parametric Bayesian method for regression tasks that defines a prior over functions and uses observed data to update this prior into a posterior distribution.

Gaussian Process (GP) Definition

A Gaussian Process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution. A GP is fully specified by a mean function $m(x)$ and a covariance function $k(x, x')$:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')). \tag{A1.1}$$

In practice, the mean function is often assumed to be zero: $m(x) = 0$.

Prior and Posterior Distributions

Given a training dataset $\mathcal{D} = \{X, \mathbf{y}\}$, where $X \in \mathbb{R}^{n \times d}$ and $\mathbf{y} \in \mathbb{R}^n$, and a testing data $\mathbf{X}_*$, the joint distribution of the observed targets and the function value at $\mathbf{X}_*$ is:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right). \tag{A1.2}$$

The posterior predictive distribution at $\mathbf{X}_*$ is then Gaussian, with predictive mean and variance given as:

$$\mathbb{E}[\mathbf{f}_*] = K(X_*, X)[K(X,X) + \sigma_n^2 I]^{-1}\mathbf{y}, \tag{A1.3}$$

$$\mathbb{V}[\mathbf{f}_*] = K(X_*, X_*) - K(X_*, X)[K(X,X) + \sigma_n^2 I]^{-1}K(X, X_*), \tag{A1.4}$$

where

$$K(X,X) \in \mathbb{R}^{n \times n}, \quad [K(X,X)]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) \tag{A1.5}$$

is the kernel matrix for training data,

$$K(X, X^*) \in \mathbb{R}^{n \times n^*}, \quad [K(X, X^*)]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j^*) \tag{A1.6}$$

is the kernel matrix between training and test points, and

$$K(X^*, X^*) \in \mathbb{R}^{n^* \times n^*}, \quad [K(X^*, X^*)]_{ij} = k(\mathbf{x}_i^*, \mathbf{x}_j^*) \tag{A1.7}$$

is the kernel matrix for test data.

Kernel Functions

The choice of kernel function $k(\cdot, \cdot)$ encodes assumptions about the function being modeled (e.g., smoothness, periodicity). Common kernel functions include: squared exponential (or RBF), Matern, polynomial, exponential, rational quadratic, linear, periodic, constant, etc. A new kernel can also be created by multiplication and/or addition of the existing kernels. For the purpose of RSRP interpolation, we chose RBF kernel multiplied with the constant kernel:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2}\|\mathbf{x}_i - \mathbf{x}_j\|^2\right). \tag{A1.8}$$

The hyper-parameter $l > 0$ of the RBF kernel is the characteristic length scale, encapsulating the distance at which function values do not correlate anymore. Small l indicated quickly changing function, large l slowly changing one. The $\sigma_f^2 \geq 0$ of the constant kernel accounts for the signal variance that scales the RBF kernel.

In the case we do not know the parameters of the kernel, we can learn them by maximizing the log marginal likelihood function [155].

Advantages of GPR

- Provides probabilistic predictions with uncertainty estimates.
- Highly flexible through kernel selection.
- No need for an explicit parametric form of the model.

GPR Limitations:

- Computational complexity is $\mathcal{O}(n^3)$ due to matrix inversion.
- Performance depends heavily on the choice of kernel and hyperparameters.

Application to RSRP Interpolation

GPR can be effectively applied to interpolate RSRP values in mobile network measurement campaigns, particularly in geospatial scenarios such as train tracks or urban drive tests. Given sparse RSRP measurements along GPS coordinates, GPR can infer the RSRP at unmeasured locations while also quantifying prediction uncertainty.

Let $\mathbf{x}_i = \begin{bmatrix} \text{lat}_i \\ \text{lon}_i \end{bmatrix} \in \mathbb{R}^2$ denote the GPS coordinates of a measurement point, and y_i the corresponding RSRP value. The full set of training locations is represented by a matrix $\mathbf{X} \in \mathbb{R}^{n \times 2}$, where each row corresponds to a geographic coordinate pair. Similarly, let $\mathbf{X}^* \in \mathbb{R}^{n^* \times 2}$ denote the set of test locations where we wish to estimate RSRP. With a trained GPR model using a spatial kernel (e.g., RBF kernel with geodesic or Euclidean distance), we can interpolate the RSRP at any location $\mathbf{x}_*$ along the route.

GPR can be effectively applied to interpolate RSRP measurements in mobile network datasets, especially in scenarios where multiple samples are recorded at or near the same geographic location. This typically occurs during repeated train runs or slow movement, resulting in clusters of RSRP measurements per position.

In wireless propagation, the received signal strength is influenced by both LSF - caused by distance-dependent PL and SF from obstacles - and SSF, which results from rapid constructive and destructive interference of multipath components. SSF leads to

strong fluctuations over short distances (on the order of a wavelength), whereas LSF evolves more smoothly over space.

By applying GPR to the spatial coordinates and corresponding RSRP values, the predictive mean function yields a smoothed representation of the RSRP field. This effectively suppresses the influence of SSF and reveals the underlying LSF component. As such, GPR serves as a powerful spatial averaging tool to extract the location-dependent mean signal strength from noisy, repeated measurements.

This is particularly useful when:

- **Handling measurement gaps:** In real-world data collection, there are often missing measurements due to equipment malfunctions, temporary signal loss, or filtering during preprocessing. GPR enables interpolation across these gaps by leveraging spatial correlation, providing smooth estimates of RSRP even in regions without direct observations.
- **Enhancing spatial resolution of coverage maps:** Raw measurements are typically collected at irregular intervals and limited spatial density. GPR produces a continuous estimate of the RSRP field, allowing for the generation of high-resolution coverage maps that support visualization, performance assessment, and network planning.
- **Supporting propagation modeling and analysis:** Propagation models rely on accurate spatial representations of signal strength. GPR facilitates this by reconstructing the RSRP field over an entire area of interest, serving as a basis for LSF analysis, PL modeling, or RT validation.
- **Smoothing SSF to extract LSF trends:** By applying GPR to the spatial coordinates and corresponding RSRP values, the predictive mean function yields a smoothed representation of the RSRP field. This effectively suppresses the influence of SSF and reveals the underlying LSF component. As such, GPR serves as a powerful spatial averaging tool to extract the location-dependent mean signal strength from noisy, repeated measurements.

Furthermore, the predictive variance can indicate regions with high uncertainty, guiding where additional measurements may be needed or where model confidence is low due to lack of data.

A2 Kolmogorov-Smirnov Test

The KS test is a classical nonparametric method used to compare empirical distribution functions. It can be applied in both the one-sample and two-sample settings. This section is based on the treatment in Wasserman's *All of Statistics* [156].

One-Sample KS Test

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample from an unknown distribution F. Suppose we want to test the null hypothesis $H_0 : F = F_0$, where F_0 is a fully specified, CDF.

The test statistic is based on the ECDF $\hat{F}_n(x)$, defined as:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}. \tag{A2.9}$$

The Kolmogorov-Smirnov test statistic is:

$$D_n = \sup_x |\hat{F}_n(x) - F_0(x)|, \tag{A2.10}$$

where supremum means we are computing the maximum vertical distance between the ECDF $\hat{F}_n(x)$ and the CDF $F_0(x)$ across all possible values of x.

Under H_0, the distribution of D_n does not depend on F_0, and its asymptotic distribution can be used to compute p-values or critical values.

Two-Sample KS Test

In the two-sample setting, we are given two independent samples:

$$X_1, \ldots, X_n \sim F, \qquad Y_1, \ldots, Y_m \sim G, \tag{A2.11}$$

and we wish to test the null hypothesis $H_0 : F = G$.

Let $\hat{F}_n(x)$ and $\hat{G}_m(x)$ denote the ECDFs of the two samples. The two-sample KS test statistic is:

$$D_{n,m} = \sup_x |\hat{F}_n(x) - \hat{G}_m(x)|. \tag{A2.12}$$

Under H_0, the test statistic $D_{n,m}$ has a known limiting distribution, and approximate p-values can be calculated.

KS Test Properties

- The KS test is distribution-free under the null hypothesis, meaning its critical values do not depend on the underlying distribution.
- The test is sensitive to differences in both location and shape between distributions.
- The one-sample test requires that F_0 is fully specified - it cannot be estimated from the data.

A3 Reference Signal Received Power Simulation

To illustrate the importance of random sampling and compare it in crowdsourced and drive test scenarios, we simulate an outdoor RSRP map for a part of Vienna's third district using the DLNP from [63], with the following simulation parameters: 15 sectors (three sectors at each of the five BS locations) with $P_{\mathrm{TX}} = 15\,\mathrm{W}$, $f = 1\,800\,\mathrm{MHz}$, sector down-tilt of 10° and BS height of 30 m. The simulated area has a dimension 1 000 m × 1 000 m, with an RSRP map resolution of 1 m. The DLNP utilizes the geospatial building model, obtained from the GSW and a realistic network layout. The serving RSRP map is obtained by computing the maximum RSRP value across all 15 sectors at each location of the map grid and is depicted in the final map shown in Fig. 6.1a. Blue scatter points represent five BS locations, while the red line outlines the Landstrasser Hauptstrasse street in Vienna, obtained from OSM.

A4 Variance Approximation

Assuming two separate sample sets or stata $S_1 = \{x_1^{(1)}, x_2^{(1)}, \ldots, x_{N_h}^{(1)}\}$, $S_2 = \{x_1^{(2)}, x_2^{(2)}, \ldots, x_{N_h}^{(2)}\}$ that have same mean μ and sample size N_h, we define their variances as:

$$\begin{aligned} \sigma_1^2 &= \frac{\sum_{i=1}^{N_h}(x_i^{(1)} - \mu)^2}{N_h - 1}, \\ \sigma_2^2 &= \frac{\sum_{i=1}^{N_h}(x_i^{(2)} - \mu)^2}{N_h - 1}. \end{aligned} \tag{A4.13}$$

Then the average variance of the two groups is given as

$$\begin{aligned} \frac{\sigma_1^2 + \sigma_2^2}{2} &= \frac{\sum_{i=1}^{N_h}\left[(x_i^{(1)} - \mu)^2 + (x_i^{(2)} - \mu)^2\right]}{2(N_h - 1)} \\ &= \frac{\sum_{i=1}^{2N_h}(x_i^{(1,2)} - \mu)^2}{2N_h - 2}. \end{aligned} \tag{A4.14}$$

In comparison, if we combine these two strata into one set

$$S_{1,2} = S_1 \cup S_2 = \{x_1^{(1)}, x_2^{(1)}, \ldots, x_{N_h}^{(1)}, x_1^{(2)}, x_2^{(2)}, \ldots, x_{N_h}^{(2)}\},$$

then the variance of the combined set is given by

$$\sigma = \frac{\sum_{i=1}^{2N_h}(x_i^{(1,2)} - \mu)^2}{2N_h - 1}. \tag{A4.15}$$

In the limit $N_h \longrightarrow \infty$, the denominator terms in Eq. (A4.14) and Eq. (A4.15) can be approximated with $2N_h$. This approximation results in $\sigma = \frac{\sigma_1^2 + \sigma_2^2}{2}$, the equation that can be easily generalized to account for arbitrary L strata:

$$\sigma = \frac{\sum_{h=1}^{L}\sigma_h^2}{L}. \tag{A4.16}$$

Hence, we can use the approximation from Eq. (A4.16) for sufficiently large strata sizes.

A5 Verification Algorithm

After calculating how many samples n_h are required in each stratum $h = 1, 2, \ldots, L$ for remaining under a certain error bound d with 95% accuracy, we can verify this result by using calculated n_h in each of R sampling iterations. In each iteration, we compute the strata sample means, the stratified mean estimate and the absolute error between the true population mean and the stratified mean estimate. Finally, we compute the α, representing the percentage for which the condition estimation error level is violated. If α indeed lies below 5% (100-95), then randomly sampling previously calculated n_h

samples from the corresponding h stratum, results in the stratified mean estimate under the estimation error bound d in 95% of the cases. Note that in the optimal allocation scheme, it may happen that the calculated n_h is higher than N_h. We address this in step 6. Similarly, step 4. corrects for the n_h values rounded to zero, which can happen in rare cases. Clearly, a sample size of zero or drawing more samples than are available in the population is unfeasible. The verification algorithm is summarized below.

Algorithm for Stratification Verification

Input: population samples d, R, y_i, n_h for $h = 1, 2, \ldots, L$
Output: α
  Compute strata mean $\mu_h = \frac{1}{N} \sum_{i=1}^{N_h} y_i$ for $h = 1, 2, \ldots, L$
  Compute population mean as $\mu = \frac{1}{N} \sum_{h=1}^{L} \mu_h N_h$
  for $h = 1$ to L **do**
    if $n_h = 0$ **then**
      $n_h = 1$
    end if
    if $n_h > N_h$ **then**
      $n_h = N_h$
    end if
  end for
  Initialize error list as an empty set
  for $i = 1$ to R **do**
    SRS of n_h from N_h in each strata $h = 1, 2, \ldots, L$
    Compute sample means $\hat{\mu}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{h_i}$
    Compute SS mean $\overline{y}_{st} = \frac{1}{N} \sum_{h=1}^{L} \hat{\mu}_h N_h$
    error$[i] = |\overline{y}_{st} - \mu|$
  end for
  return $\alpha = \frac{\sum_{i=1}^{R} \mathbf{1}(\text{error}[i] > d)}{R}$

Bibliography

[1] European Telecommunications Standards Institute (ETSI). *GSM-R; Requirements for EIRENE; Part 1: Core Requirements Specification*. Tech. rep. ETSI TS 102 610 V1.1.0. ETSI, Mar. 2003. URL: https://www.etsi.org/deliver/etsi_ts/102600_102699/102610/01.01.00_60/ts_102610v010100p.pdf.

[2] Global Railway Review. "Powering up FRMCS in Austria". In: *Global Railway Review* (2020). URL: https://www.globalrailwayreview.com/article/96344/powering-up-frmcs-austria/.

[3] Ericsson. *Integrating FRMCS: Enhancing Rail Communications with 5G Technology*. 2024. URL: https://www.ericsson.com/en/reports-and-papers/white-papers/5g-powered-frmcs.

[4] International Telecommunication Union, Radiocommunication Sector (ITU-R). *Assessment of the technical feasibility of IMT in the frequency bands above 100 GHz*. Report M.2442-1. ITU-R, 2025. URL: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2442-1-2025-PDF-E.pdf.

[5] X. Zhou, Y. Wang, Z. Li, M. Chen, and F. Liu. "A Conceptual Model-based Digital Twin Platform for Holistic Large-scale Railway Infrastructure Systems". In: *International Journal of Rail Transportation* 10.3 (2022), pp. 123–137.

[6] G. C. Doubell, A. H. Basson, K. Kruger, and P. D. F. Conradie. "A Digital Twin System for Railway Infrastructure". In: *R & D Journal of the South African Institution of Mechanical Engineering* 39 (2023), pp. 23–34.

[7] N. P. Kuruvatti, M. A. Habibi, S. Partani, B. Han, A. Fellan, and H. D. Schotten. "Empowering 6G communication systems with digital twin technology: A comprehensive survey". In: *IEEE access* 10 (2022), pp. 112158–112186.

[8] S. Ghaboura, R. Ferdousi, F. Laamarti, C. Yang, and A. El Saddik. "Digital Twin for Railway: A Comprehensive Survey". In: *IEEE Access* 11 (2023), pp. 120237–120257.

[9] S. Kaewunruen and N. Xu. "Digital twin for sustainability evaluation of railway station buildings". In: *Frontiers in Built Environment* 4 (2018), p. 430624.

[10] E. Bernal, Q. Wu, M. Spiryagin, and C. Cole. "Augmented Digital Twin for Railway Systems". In: *Vehicle System Dynamics* 62.1 (2024), pp. 67–83.

[11] I. Errandonea, J. Goya, U. Alvarado, S. Beltrán, and S. Arrizabalaga. "IoT approach for intelligent data acquisition for enabling digital twins in the railway sector". In: *2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC)*. IEEE. 2021, pp. 164–168.

[12] L. U. Khan, Z. Han, W. Saad, E. Hossain, M. Guizani, and C. S. Hong. "Digital twin of wireless systems: Overview, taxonomy, challenges, and opportunities". In: *IEEE Communications Surveys & Tutorials* 24.4 (2022), pp. 2230–2254.

[13] S. Tripkovic, P. Svoboda, and M. Rupp. "Benchmarking of Mobile Communications in High-Speed Scenarios: Active vs. Passive Modifications in High-Speed Trains". In: *2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring)*. IEEE. 2022, pp. 1–6. DOI: 10.1109/VTC2022-Spring54318.2022.9860953.

[14] S. Tripkovic, P. Svoboda, and M. Rupp. "Measuring the Effects of AoA on Vehicle Penetration Loss in Cellular Networks". In: *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*. IEEE. 2023, pp. 1–7. DOI: 10.1109/VTC2023-Fall60731.2023.10333537.

[15] S. Tripkovic, P. Svoboda, and M. Rupp. "Enhancing Mobile Communication on Railways: Impact of Train Window Size and Coating". In: *2024 19th International Symposium on Wireless Communication Systems (ISWCS)*. IEEE. 2024, pp. 1–6. DOI: 10.1109/ISWCS61526.2024.10639152.

[16] S. Tripkovic, L. Eller, P. Svoboda, and M. Rupp. "Unbiased Benchmarking in Mobile Networks: The Role of Sampling and Stratification". In: *IEEE Access* 11 (2023), pp. 53772–53787. DOI: 10.1109/ACCESS.2023.3280828.

[17] V. Raida. "Data-driven estimation of spatiotemporal performance maps in cellular networks". PhD Thesis. Technische Universität Wien, 2021. DOI: 10.34726/hss.2021.80644. URL: https://repositum.tuwien.at/handle/20.500.12708/18354.

[18] ETSI. *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements (3GPP TS 36.214 version 18.0.0 Release 18)*. Technical Specification. ETSI, May 2024. URL: https://www.etsi.org/deliver/etsi_ts/136200_136299/136214/18.00.00_60/ts_136214v180000p.pdf.

[19] E. Dahlman, S. Parkvall, and J. Skold. *5G NR: The next generation wireless access technology*. Academic Press, 2020.

[20] 3rd Generation Partnership Project (3GPP). *TS 36.321: Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification*. 3rd Generation Partnership Project Technical Specification 36.321. Version 17.7.0. Available at: https://www.3gpp.org/ftp/Specs/archive/36_series/36.321/36321-g70.zip. 3GPP, Mar. 2024.

[21] 3rd Generation Partnership Project (3GPP). *TS 36.322: Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) protocol specification (Release 18)*. Technical Specification 36.322. Version 18.0.0. ETSI, Mar. 2024. URL: https://www.etsi.org/deliver/etsi_ts/136300_136399/136322/18.00.00_60/ts_136322v180000p.pdf.

[22] 3rd Generation Partnership Project (3GPP). *TS 36.323: Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) specification (Release 18)*. Technical Specification 36.323. Version 18.0.0. ETSI, Mar. 2024. URL: https://www.etsi.org/deliver/etsi_ts/136300_136399/136323/18.00.00_60/ts_136323v180000p.pdf.

[23] 3rd Generation Partnership Project (3GPP). *TS 36.331: Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC) protocol specification (Release 18)*. Technical Specification 36.331. Version 18.0.0. ETSI, Mar. 2024. URL: https://www.etsi.org/deliver/etsi_ts/136300_136399/136331/18.00.00_60/ts_136331v180000p.pdf.

[24] 3rd Generation Partnership Project (3GPP). *TS 36.213: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 18)*. Technical Specification 36.213. Version 18.0.0. ETSI, Mar. 2024. URL: https://www.etsi.org/deliver/etsi_ts/136200_136299/136213/18.00.00_60/ts_136213v180000p.pdf.

[25] 3rd Generation Partnership Project (3GPP). *TS 23.203: Policy and Charging Control (PCC) architecture (Release 18)*. Technical Specification 23.203. Version 18.3.0. ETSI, Mar. 2024. URL: https://www.etsi.org/deliver/etsi_ts/123200_123299/123203/18.03.00_60/ts_123203v180300p.pdf.

[26] 3rd Generation Partnership Project (3GPP). *TS 36.423: Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP) (Release 18)*. Technical Specification 36.423. Version 18.0.0. ETSI, Mar. 2024. URL: https://www.etsi.org/deliver/etsi_ts/136400_136499/136423/18.00.00_60/ts_136423v180000p.pdf.

[27] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005. ISBN: 978-0521837163.

[28] T. S. Rappaport. *Wireless Communications: Principles and Practice*. 2nd. Prentice Hall, 2002. ISBN: 978-0130422323.

[29] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005. ISBN: 978-0521845274.

[30] L. Ahlin, J. Zander, and B. Slimane. *Principles of Wireless Communications*. Studentlitteratur AB, 2006. ISBN: 978-9144030807.

[31] European Telecommunications Standards Institute (ETSI). *TR 103 257-1 V1.1.1: Electromagnetic compatibility and Radio spectrum Matters (ERM); Methods for the assessment of mobile network coverage performance*. Technical Report TR 103 257-1. Version 1.1.1. ETSI, Oct. 2015. URL: https://www.etsi.org/deliver/etsi_tr/103200_103299/10325701/01.01.01_60/tr_10325701v010101p.pdf.

[32] W. Stallings. *Wireless Communications and Networks*. 2nd. Prentice Hall, 2005. ISBN: 978-0131918351.

[33] L. Eller, P. Svoboda, and M. Rupp. "A differentiable throughput model for load-aware cellular network optimization through gradient descent". In: *IEEE Access* 12 (2024), pp. 14547–14562.

[34] C. E. Shannon. "A Mathematical Theory of Communication". In: *The Bell System Technical Journal* 27.3 (Sept. 1948). Part I: July 1948, pp. 379–423; Part II: October 1948, pp. 623–656, pp. 379–423.

[35] H. T. Friis. "A Note on a Simple Transmission Formula". In: *Proceedings of the IRE* 34.5 (1946), pp. 254–256.

[36] C. A. 231. *Urban Transmission Loss Models for Mobile Radio in the 900 and 1800 MHz Bands*. Chapter 4 in the Final Report for COST Action 231. European Cooperation in Science and Technology (COST), 1999.

[37] M. Hata. "Empirical formula for propagation loss in land mobile radio services". In: *IEEE transactions on Vehicular Technology* 29.3 (2013), pp. 317–325.

[38] Y. Okumura. "Field strength and its variability in VHF and UHF land-mobile radio service". In: *Review of the Electrical communication Laboratory* 16.9 (1968).

[39] C. Wen, T. Tang, B. Ai, K. Guan, Z. Zhong, Y. Cui, R. He, Z. Wang, L. Wang, and M. Ma. "Measurement and modeling of LTE-railway channels in high-speed railway environment". In: *Radio Science* 55.5 (2020), e2019RS006933. DOI: 10.1029/2019RS006933.

[40] H. Wei, Z. Zhong, K. Guan, and B. Ai. "Path loss models in viaduct and plain scenarios of the High-speed Railway". In: *2010 5th International ICST Conference on Communications and Networking in China*. 2010, pp. 1–5.

[41] R. He, Z. Zhong, B. Ai, and J. Ding. "An Empirical Path Loss Model and Fading Analysis for High-Speed Railway Viaduct Scenarios". In: *IEEE Antennas and Wireless Propagation Letters* 10 (2011), pp. 808–812. DOI: 10.1109/LAWP.2011.2164389.

[42] PCTEL. *IBflex® Scanning Receiver*. https://www.pctel.com/products/test-measurement/scanning-receivers/ibflex-scanning-receiver/. Accessed: 2025-04-02.

[43] Z. Yun and M. F. Iskander. "Ray tracing for radio propagation modeling: Principles and applications". In: *IEEE access* 3 (2015), pp. 1089–1100.

[44] J. Hoydis, F. A. Aoudia, S. Cammerer, F. Euchner, M. Nimier-David, S. Ten Brink, and A. Keller. "Learning radio environments by differentiable ray tracing". In: *IEEE Transactions on Machine Learning in Communications and Networking* (2024).

[45] Beijing Jiaotong University State Key Laboratory of Advanced Rail Autonomous Operation (RAO). *CloudRT: Cloud-Based Ray Tracing Platform*. Accessed: 2024-11-20. URL: http://cn.raytracer.cloud:9090/.

[46] Remcom. *Wireless InSite: High-Fidelity Ray Tracing Software*. Accessed: 2024-11-20. 2024. URL: https://www.remcom.com/wireless-insite-em-propagation-software/high-fidelity-ray-tracing.

[47] MathWorks. *Matlab: Ray Tracing in Wireless Networks*. Accessed: 2024-11-20. URL: https://www.mathworks.com/help/antenna/ref/rfprop.raytracing.html.

[48] J. Hoydis, F. A. Aoudia, S. Cammerer, M. Nimier-David, N. Binder, G. Marcus, and A. Keller. "Sionna RT: Differentiable ray tracing for radio propagation modeling". In: *2023 IEEE Globecom Workshops (GC Wkshps)*. IEEE. 2023, pp. 317–321.

[49] C. of Vienna. *Boundary Cadastre Map (BKM)*. Last Accessed: 2024-11-29. 2024. URL: https://www.wien.gv.at/stadtentwicklung/stadtvermessung/geodaten/bkm/.

[50] Blender Online Community. *Blender – a 3D Modelling and Rendering Package*. Blender Foundation. Amsterdam, The Netherlands, 2025. URL: https://www.blender.org.

[51] u-blox AG. *High Precision GNSS Modules*. Accessed: March 19, 2025. 2024. URL: https://www.u-blox.com.

[52] Keysight Technologies. *Nemo Handy Handheld Measurement Solution*. (Accessed: 2025-03-20). URL: https://www.keysight.com/us/en/product/NTH50047B/nemo-handy-handheld-measurement-solution.html.

[53] I. PCTEL. *MXflex Scanning Receiver Brochure*. Accessed: March 19, 2025. 2020. URL: https://www.pctel.com/wp-content/uploads/2020/06/PCTEL-MXflex-Scanning-Receiver-Brochure-web.pdf.

[54] Keysight Technologies. *"Nemo file format specification" Manual, Version 2.37*. 2018.

[55] OpenStreetMap contributors. *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org. 2017.

[56] Geoland.at. *Digital Terrain Model (DTM) Austria*. Accessed: 2024-11-29. 2024. URL: https://www.data.gv.at/katalog/dataset/d88a1246-9684-480b-a480-ff63286b35b7.

[57] C. L. M. Service. *Corine Land Cover (CLC) 2018*. Accessed: 2024-11-29. 2018. URL: https://land.copernicus.eu/en/products/corine-land-cover/clc2018.

[58] E. Krakiwsky, C. Harris, and R. Wong. "A Kalman filter for integrating dead reckoning, map matching and GPS positioning". In: *IEEE PLANS '88.,Position Location and Navigation Symposium, Record. 'Navigation into the 21st Century'*. 1988, pp. 39–46. DOI: 10.1109/PLANS.1988.195464.

[59] K. Gao, B. Wang, L. Xiao, and G. Mei. "Incomplete road information imputation using parallel interpolation to enhance the safety of autonomous driving". In: *IEEE Access* 8 (2020), pp. 25420–25430.

[60] International Telecommunication Union. *Measurement campaigns, monitoring systems and sampling methodologies to monitor the quality of service in mobile networks*. Standard. ITU-T, June 2019. URL: https://www.itu.int/rec/T-REC-E.806/en.

[61] Statistics Austria. *Statistical Grid (100m x 100m) ETRS89-LAEA Austria*. https://www.statistik.at/gs-inspire/www/inspire2/download/daten/sugrid_100m_77679c2c-302c-11e3-beb4-0000c1ab0db6.zip. Last Accessed: 2024-11-12.

[62] A. Bonyar, A. Geczy, O. Krammer, H. Santha, B. Illes, J. Kaman, Z. Szalay, P. Hanak, and G. Harsanyi. "A Review on Current eCall Systems for Autonomous Car Accident Detection". In: *2017 40th International Spring Seminar on Electronics Technology (ISSE)*. IEEE. 2017, pp. 1–8.

[63] L. Eller, P. Svoboda, and M. Rupp. "A Deep Learning Network Planner: Propagation Modeling Using Real-World Measurements and a 3D City Model". In: *IEEE Access* 10 (2022), pp. 122182–122196. DOI: 10.1109/ACCESS.2022.3223097.

[64] D. He, B. Ai, K. Guan, L. Wang, Z. Zhong, and T. Kuerner. "The Design and Applications of a High-Performance Ray-Tracing Simulation Platform for 5G and Beyond Wireless Communications: A Tutorial". In: *IEEE communications surveys & tutorials* 21.1 (2018), pp. 10–27.

[65] J. Johansson, W. A. Hapsari, S. Kelley, and G. Bodog. "Minimization of Drive Tests in 3GPP Release 11". In: *IEEE Communications Magazine* 50.11 (2012), pp. 36–43.

[66] S. Tripkovic. *Cellular measurements (4G) onboard a moving train*. 2023. DOI: 10.21227/abdd-2y20. URL: https://dx.doi.org/10.21227/abdd-2y20.

[67] L. Mayer, A. Demmer, A. Hofmann, and M. Schiefer. *Metal-Coated Windowpane Particularly for Rail Vehicles*. Oct. 2014. URL: https://patents.google.com/patent/WO2014166869A1/en.

[68] S. Tripkovic. *4G measurements onboard a moving train - FSS windows*. 2024. DOI: 10.21227/b919-cq93. URL: https://dx.doi.org/10.21227/b919-cq93.

[69] BMEIA. "Austria's 2030 Mobility Master Plan". In: *Climate Action, Environment, Energy, Mobility, Innovation and Technology* (2021).

[70] S. Schwarz and M. Rupp. "Society in motion: Challenges for LTE and beyond mobile communications". In: *IEEE Communications Magazine* 54.5 (2016), pp. 76–83.

[71] Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2021: The Physical Science Basis*. Tech. rep. IPCC, 2021.

[72] N. Jamaly, S. Mauron, R. Merz, A. Schumacher, and D. Wenger. "Delivering Gigabit Capacities to Passenger Trains: Tales from an Operator on the Road to 5G". In: *IEEE Communications Magazine* 57.9 (Sept. 2019), pp. 18–23. ISSN: 1558-1896. DOI: 10.1109/MCOM.2019.1800949.

[73] M. Fadoul, M. B. Morsin, C. Y. Leow, and A. A. Eteng. "Using Amplify-and-Forward Relay for Coverage Extension in Indoor Environments". In: *J. Theor. Appl. Inf. Technol.* 91.2 (2016).

[74] P. Laehdekorpi, T. Isotalo, S. U. Khan, and J. Lempiaeinen. "Implementation Aspects of RF Repeaters in Cellular Networks". In: *21st Annual IEEE PIMRC*. 2010, pp. 2511–2516.

[75] P. Lähdekorpi, T. Isotalo, S. U. Khan, and J. Lempiäinen. "Implementation Aspects of RF-Repeaters in Cellular Networks". In: *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Istanbul, Turkey, Sept. 2010.

[76] M. Madani Fadoul, M. Morsin, C. Y. Leow, and A. Eteng. "Using Amplify-and-forward Relay for Coverage Extension in Indoor Environments". In: *Journal of Theoretical and Applied Information Technology* 91.2 (Sept. 2016), pp. 304–312.

[77] B. Ai, K. Guan, M. Rupp, T. Kurner, X. Cheng, X.-F. Yin, Q. Wang, G.-Y. Ma, Y. Li, L. Xiong, and J.-W. Ding. "Future Railway Services-Oriented Mobile Communications Network". In: *IEEE Communications Magazine* 53.10 (2015), pp. 78–85.

[78] M. Tolstrup. *Indoor Radio Planning: A Practical Guide for 2G, 3G and 4G*. 3rd ed. Wiley, Apr. 2015. ISBN: 978-1-118-91359-8.

[79] T. Zhou, C. Tao, S. Salous, L. Liu, and Z. Tan. "Implementation of an LTE-Based Channel Measurement Method for High-Speed Railway Scenarios". In: *IEEE Transactions on Instrumentation and Measurement* 65.1 (Jan. 2016), pp. 25–36. ISSN: 0018-9456. DOI: 10.1109/TIM.2015.2477166.

[80] F. Kaltenberger, A. Byiringiro, G. Arvanitakis, R. Ghaddab, D. Nussbaum, R. Knopp, M. Bernineau, Y. Cocheril, H. Philippe, and E. Simon. "Broadband Wireless Channel Measurements for High-Speed Trains". In: *IEEE International Conference on Communications (ICC)*. June 2015. DOI: 10.1109/ICC.2015.7248720.

[81] K. Guan, Z. Zhong, B. Ai, C. Briso-Rodriguez, and L. Zhang. "Large scale fading characteristics in rail traffic scenarios". In: *Antennas and Propagation & USNC/URSI National Radio Science Meeting, 2015 IEEE International Symposium on*. IEEE. 2015, pp. 83–84.

[82] T. Dominguez-Bolano, J. Rodriguez-Pineiro, J. A. Garcia-Naya, and L. Castedo. "Experimental Characterization of LTE Wireless Links in High-Speed Trains". In: *Wireless Communications and Mobile Computing* 2017 (Sept. 2017). DOI: 10.1155/2017/5079130.

[83] M. Lerch, P. Svoboda, V. Platzgummer, and M. Rupp. "Analysis of LTE in Two-Path Vehicular Repeater Channels". In: *IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. Honolulu, USA, Sept. 2019.

[84] M. Lerch, P. Svoboda, J. Resch, and M. Rupp. "Cellular Uplink Impairments in Vehicular Repeater Deployments". In: *IEEE Open Journal of Vehicular Technology* (2025).

[85] COMLAB AG. *Multiband InTrain Repeater RUD19-5*. https://www.railway-technology.com/products/comlab-remote-unit-digital-rud-5/. (Accessed: 2019-01-15).

[86] A. Karttunen, M. Moekkoenen, and K. Haneda. "Investigation of 5G radio frequency signal losses of glazing structures". In: *All Eyes on Smarter Glass Performance Day 2019*. 2019.

[87] L. Yunos, M. L. Jane, P. J. Murphy, and K. Zuber. "Frequency selective surface on low emissivity windows as a means of improving telecommunication signal transmission: A review". In: *Journal of Building Engineering* 70 (2023), p. 106416.

[88] J. Fleury, M. Lanini, C. Pose, L. Burnier, A. Salvadè, E. Zimmermann, C. Genoud, and A. Schüler. "Wide band-pass FSS with reduced periodicity for energy-efficient windows at higher frequencies". In: *Applied Physics A* 126.5 (2020), p. 417. DOI: 10.1007/s00339-020-03547-w.

[89] O. Bouvard, M. Lanini, L. Burnier, R. Witte, B. Cuttat, A. Salvade, and A. Schueler. "Mobile Communication Through Insulating Windows: A New Type of Low Emissivity Coating". In: *Energy Procedia* 122 (2017), pp. 781–786.

[90] M. Raspopoulos and S. Stavrou. "Frequency selective buildings through frequency selective surfaces". In: *IEEE Transactions on Antennas and Propagation* 59.8 (2011), pp. 2998–3005.

[91] C. Wang, A. Ghazal, B. Ai, Y. Liu, and P. Fan. "Channel Measurements and Models for High-Speed Train Communication Systems: A Survey". In: *IEEE Communications Surveys Tutorials* 18.2 (2016), pp. 974–987. ISSN: 1553-877X. DOI: 10.1109/COMST.2015.2508442.

[92] K. Guan, D. He, M. Rupp, M. Shahid, and Z. Zhong. "Challenges and Future Research Trends of Window Glass for Smart Rail Vehicles: From the Perspective of Wireless Propagation". In: *IEEE Access* (2022).

[93] H. Ostendarp. *Glass Pane with Structured Conductive Coating*. WO2019185345A1, World Intellectual Property Organization. Sept. 2019. URL: https://patents.google.com/patent/DE102018107559B4/en.

[94] R. Z. Jiang, Q. Ma, J. C. Liang, Q. Y. Zhou, J. Y. Dai, Q. Cheng, and T. J. Cui. "A Single-Layered Wideband and Wide-Angle Transparent Metasurface for Enhancing the EM-Wave Transmissions Through Glass". In: *IEEE Transactions on Antennas and Propagation* 71.8 (2023), pp. 6593–6605. DOI: 10.1109/TAP.2023.3281879.

[95] O. Bouvard, M. Lanini, L. Burnier, R. Witte, B. Cuttat, A. Salvade, and A. Schueler. "Structured Transparent Low Emissivity Coatings with High Microwave Transmission". In: *Applied Physics A* 123 (2017), pp. 1–10.

[96] R. Chueca, R. Alcain, C. Heras, and I. Salinas. "Modelling of Frequency Selective Surfaces on Multi-layered Glazing Windows". In: *2022 Global Conference on Wireless and Optical Technologies (GCWOT)*. IEEE. 2022, pp. 1–5.

[97] H. Xiao, M. Shi, and J. Chen. "Electromagnetic Function Textiles". In: *Electromagnetic Materials and Applications*. IntechOpen, 2019, pp. 1–25. DOI: 10.5772/intechopen.85586. URL: https://www.researchgate.net/publication/333401215_Electromagnetic_Function_Textiles.

[98] O. Trindade, T. Berisha, P. Svoboda, E. Bura, and C. F. Mecklenbrauuker. "Assessment of Treatment Influence in Mobile Network Coverage Onboard High-Speed Trains". In: *IEEE Access* 8 (2020), pp. 162945–162960.

[99] V. Raida, P. Svoboda, and M. Rupp. "Modified Dynamic Time Warping with a Reference Path for Alignment of Repeated Drive-Tests". In: *2020 IEEE 92nd VTC-Fall*. 2020, pp. 1–6.

[100] ITU. "Quality of Experience Metrics for Mobile Telephony Communication during Rail Travel". In: *Rec. ITU-T G.1034* (2020).

[101] L. Eller, V. Raida, P. Svoboda, and M. Rupp. "Localizing Base Stations from End-User Timing Advance Measurements". In: *IEEE Access* 10 (2022).

[102] M. Lerch, P. Svoboda, S. Ojak, M. Rupp, and C. Mecklenbraeuker. "Distributed Measurements of the Penetration Loss of Railroad Cars". In: *IEEE 86th Vehicular Technology Conference (VTC2017-Fall)*. Toronto, Canada, Sept. 2017. DOI: 10.1109/VTCFall.2017.8287899.

[103] V. Raida, P. Svoboda, M. Koglbauer, and M. Rupp. "On the Stability of RSRP and Variability of Other KPIs in LTE Downlink". In: *GLOBECOM 2020*. 2020, pp. 1–6.

[104] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator". In: *Ann. Math. Stat.* 27.3 (1956), pp. 642–669.

[105] J. L. Hodges. "The Significance Probability of the Smirnov Two-Sample Test". In: *Ark. Mat.* 3.5 (1958), pp. 469–486.

[106] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. "SciPy 1.0: Fundamental algorithms for scientific computing in python". In: *Nat. Methods* 17.3 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[107] T. Zhou, C. Tao, S. Salous, L. Liu, and Z. Tan. "Implementation of an LTE-based channel measurement method for high-speed railway scenarios". In: *IEEE Transactions on Instrumentation and Measurement* 65.1 (2015), pp. 25–36.

[108] K. Guan, B. Ai, Z. Zhong, C. F. López, L. Zhang, C. Briso-Rodríguez, A. Hrovat, B. Zhang, R. He, and T. Tang. "Measurements and Analysis of Large-Scale Fading Characteristics in Curved Subway Tunnels at 920 MHz, 2400 MHz, and 5705 MHz". In: *IEEE Transactions on Intelligent Transportation Systems* 16.5 (Oct. 2015), pp. 2393–2405. ISSN: 1524-9050. DOI: 10.1109/TITS.2015.2404851.

[109] ITU. "Compilation of Measurement Data Relating to Building Entry Loss". In: *Report ITU-R P.2346-4* (2021).

[110] P. Karlsson, C. Bergljung, E. Thomsen, and H. Borjeson. "Wideband Measurement and Analysis of Penetration Loss in the 5 GHz Band". In: *Gateway to 21st Century Communications Village. VTC 1999-Fall. IEEE VTS 50th Vehicular Technology Conference (Cat. No. 99CH36324)*. Vol. 4. IEEE. 1999, pp. 2323–2328.

[111] E. Tanghe, W. Joseph, L. Verloock, and L. Martens. "Evaluation of Vehicle Penetration Loss at Wireless Communication Frequencies". In: *IEEE Transactions on Vehicular Technology* 57.4 (2008), pp. 2036–2041.

[112] C. U. Bas, R. Wang, S. Sangodoyin, T. Choi, S. Hur, K. Whang, J. Park, C. J. Zhang, and A. F. Molisch. "Outdoor to Indoor Propagation Channel Measurements at 28 GHz". In: *IEEE Transactions on Wireless Communications* 18.3 (2019), pp. 1477–1489.

[113] T. Bertin-Mourot, N. F. Borrelli, S. Yurevich Ten, and Y. ZINGER. "Energy-Efficient, Microwave-Transparent Window Compatible with Present Design". In: *US Patent* (2019).

[114] H. und Suhner. *Huber+Suhner antenna pattern example*. `https://ecatalog.hubersuhner.com/media/documents/pattern.pdf/en/zip/84459410`.

[115] 3rd Generation Partnership Project (3GPP). *5G; Management and orchestration; 5G performance measurements*. Technical Specification (TS) 28.552. 3GPP, Oct. 2022.

[116] GSMA. *4G/5G Network Experience Evaluation Guideline*. `https://www.gsma.com/futurenetworks/wp-content/uploads/2020/02/4G5G-Network-Experience-Evaluation-Guideline-_GSMA.pdf`. (Accessed: March 2023).

[117] Rundfunk und Telekom Regulierungs-GmbH (RTR). *Ausschreibungsunterlage im Verfahren betreffend Frequenzzuteilungen in den Frequenzbereichen 800 MHz, 900 MHz und 1800 MHz*. Tech. rep. RTR, Mar. 2013.

[118] 3rd Generation Partnership Project (3GPP). *Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Telecommunication management; Performance Management (PM); Performance measurements; Core Network (CN) Circuit Switched (CS) domain; UMTS and combined UMTS/GSM*. Technical Specification (TS) 32.407. 3GPP, Apr. 2022.

[119] 3rd Generation Partnership Project (3GPP). *Universal Mobile Telecommunications System (UMTS); LTE; Telecommunication management; Key Performance Indicators (KPI) for Evolved Universal Terrestrial Radio Access Network (E-UTRAN): Definitions*. Technical Specification (TS) 32.450. 3GPP, Apr. 2022.

[120] R. V. Akhpashev and V. G. Drozdova. "Spatial interpolation of LTE measurements for minimization of drive tests". In: *2018 19th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM)*. IEEE. 2018, pp. 6403–6405.

[121] R. Enami, S. Gupta, D. Rajan, and J. Camp. "LAIK: Location-Specific Analysis to Infer Key Performance Indicators". In: *IEEE Transactions on Vehicular Technology* 70.5 (2021), pp. 4406–4418.

[122] International Telecommunication Union. *Crowdsourcing Approach for the Assessment of End-to-End Quality of Service in Fixed and Mobile Broadband Networks*. Standard E.812. ITU-T, May 2020. URL: https://www.itu.int/rec/T-REC-E.812/en.

[123] M. P. Armstrong, S. Wang, and Z. Zhang. "The Internet of Things and fast data streams: prospects for geospatial data science in emerging information ecosystems". In: *Cartography and Geographic Information Science* 46.1 (2019), pp. 39–56.

[124] P. Katsikouli, D. Madariaga, A. C. Viana, A. Tarable, and M. Fiore. "DuctiLoc: Energy-efficient Location Sampling with Configurable Accuracy". In: *IEEE Access* (2023).

[125] S. Farthofer, M. Herlich, C. Maier, S. Pochaba, J. Lackner, and P. Dorfinger. "An Open Mobile Communications Drive Test Data Set and Its Use for Machine Learning". In: *IEEE Open Journal of the Communications Society* 3 (2022), pp. 1688–1701.

[126] V. Raida, P. Svoboda, and M. Rupp. "On the Inappropriateness of Static Measurements for Benchmarking in Wireless Networks". In: *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. 2020, pp. 1–5. DOI: 10.1109/VTC2020-Spring48590.2020.9128867.

[127] Opensignal. *Apps | Opensignal*. https://opensignal.com/apps. (Accessed: 2019).

[128] Ookla LLC. *Speedtest by Ookla — The Global Broadband Speed Test*. http://www.speedtest.net/. (Accessed: 2019).

[129] Alladin IT. *The Alladin Nettest*. https://nettest.alladin.at/home. (Accessed: 2018).

[130] R. K. Ganti, F. Ye, and H. Lei. "Mobile crowdsensing: current state and future challenges". In: *IEEE communications Magazine* 49.11 (2011), pp. 32–39.

[131] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou. "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm". In: *ACM computing surveys (CSUR)* 48.1 (2015), pp. 1–31.

[132] F. Ma, X. Liu, A. Liu, M. Zhao, C. Huang, and T. Wang. "A time and location correlation incentive scheme for deep data gathering in crowdsourcing networks". In: *Wireless Communications and Mobile Computing* 2018 (2018).

[133] C.-K. Tham and T. Luo. "Quality of contributed service and market equilibrium for participatory sensing". In: *IEEE Transactions on Mobile Computing* 14.4 (2014), pp. 829–842.

[134] S. Reddy, D. Estrin, and M. Srivastava. "Recruitment framework for participatory sensing data collections". In: *Pervasive Computing: 8th International Conference, Pervasive 2010, Helsinki, Finland, May 17-20, 2010. Proceedings 8*. Springer. 2010, pp. 138–155.

[135] F. Campioni, S. Choudhury, K. Salomaa, and S. G. Akl. "Improved recruitment algorithms for vehicular crowdsensing networks". In: *IEEE Transactions on Vehicular Technology* 68.2 (2018), pp. 1198–1207.

[136] Y. Wang, Z. Cai, G. Yin, Y. Gao, X. Tong, and G. Wu. "An incentive mechanism with privacy protection in mobile crowdsourcing systems". In: *Computer Networks* 102 (2016), pp. 157–171.

[137] Z. Song, C. H. Liu, J. Wu, J. Ma, and W. Wang. "QoI-aware multitask-oriented dynamic participant selection with budget constraints". In: *IEEE Transactions on Vehicular Technology* 63.9 (2014), pp. 4618–4632.

[138] F. Zhang, B. Jin, H. Liu, Y.-W. Leung, and X. Chu. "Minimum-cost recruitment of mobile crowdsensing in cellular networks". In: *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE. 2016, pp. 1–7.

[139] Tutela. *Tutela*. `https://www.tutela.com`. (Accessed: 2023).

[140] F. Wamser, A. Seufert, A. Hall, S. Wunderer, and T. Hossfeld. "Valid Statements by the Crowd: Statistical Measures for Precision in Crowdsourced Mobile Measurements". In: *Network* 1.2 (2021), pp. 215–232.

[141] C. Midoglu and P. Svoboda. "Opportunities and challenges of using crowdsourced measurements for mobile network benchmarking a case study on RTR open data". In: *2016 SAI computing conference (SAI)*. IEEE. 2016, pp. 996–1005.

[142] W. G. Cochran. *Sampling Techniques*. John Wiley and Sons, 1977.

[143] S. K. Thompson. *Sampling*. Wiley, 2012. ISBN: 978-0-470-40231-3.

[144] Shalabh. *Sampling Theory*. Lecture Notes, IIT Kanpur.

[145] L. Eller, V. Raida, P. Svoboda, and M. Rupp. "Localizing Basestations From End-User Timing Advance Measurements". In: *IEEE Access* 10 (2022), pp. 5533–5544. DOI: `10.1109/ACCESS.2022.3140825`.

[146] H. Hofmann, H. Wickham, and K. Kafadar. "Value plots: Boxplots for large data". In: *Journal of Computational and Graphical Statistics* 26.3 (2017), pp. 469–477.

[147] R. Hoppe, G. Wolfle, and F. Landstorfer. "Measurement of building penetration loss and propagation models for radio transmission into buildings". In: *Gateway to 21st Century Communications Village. VTC 1999-Fall. IEEE VTS 50th Vehicular Technology Conference (Cat. No. 99CH36324)*. Vol. 4. IEEE. 1999, pp. 2298–2302.

[148] M. Rindler, S. Caban, M. Lerch, P. Svoboda, and M. Rupp. "Swift Indoor Benchmarking Methodology for Mobile Broadband Networks". In: *IEEE 86th Vehicular Technology Conference (VTC2017-Fall)*. Toronto, Canada, Sept. 2017. DOI: `10.1109/VTCFall.2017.8288190`.

[149] L. T. Gwaka, J. May, and W. Tucker. "Towards low-cost community networks in rural communities: The impact of context using the case study of Beitbridge, Zimbabwe". In: *The Electronic Journal of Information Systems in Developing Countries* 84.3 (2018), e12029.

[150] V. Raida, P. Svoboda, M. Lerch, and M. Rupp. "Crowdsensed Performance Benchmarking of Mobile Networks". In: *IEEE Access* 7 (2019), pp. 154899–154911. DOI: 10.1109/ACCESS.2019.2949051.

[151] 3rd Generation Partnership Project (3GPP). *Technical Specification Group Radio Access Network; Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Measurement Collection for Minimization of Drive Tests (MDT); Overall Description; Stage 2*. Technical Specification (TS) 37.320. 3GPP, June 2022.

[152] Electronic Communications Committee. *CEPT Report 076*. 076. CEPT ECC, Nov. 20, 2020. URL: https://docdb.cept.org/download/137 (Zugriff am 09/30/2024).

[153] Electronic Communications Committee. *ECC Report 314*. 314. CEPT ECC, May 21, 2020. URL: https://docdb.cept.org/download/1425 (Zugriff am 09/30/2024).

[154] Electronic Communications Committee. *ECC Report 318*. 318. CEPT ECC, July 3, 2020. URL: https://docdb.cept.org/download/1433 (Zugriff am 09/30/2024).

[155] C. K. Williams and C. E. Rasmussen. *Gaussian Processes for Machine Learning*. Vol. 2. 3. MIT press Cambridge, MA, 2006.

[156] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. New York: Springer, 2004. ISBN: 978-0-387-21736-9.