Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Orlando da Silva Trindade

Supporting high quality cellular and Internet services inside railway vehicles: a statistical evaluation of wireless repeater performance on Railjet and Cityjet trains

Avaliação estatística do desempenho de repetidores sem fio em veículos sobre trilhos de longo trajeto: estudo de caso da melhoria da qualidade dos serviços de celular e Internet em trens europeus

Campinas

2018

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Orlando da Silva Trindade

Supporting high quality cellular and Internet services inside railway vehicles: a statistical evaluation of wireless repeater performance on Railjet and Cityjet trains

Avaliação estatística do desempenho de repetidores sem fio em veículos sobre trilhos de longo trajeto: estudo de caso da melhoria da qualidade dos serviços de celular e Internet em trens europeus

Dissertation presented to the School of Electrical and Computer Engineering at the State University of Campinas (UNICAMP) as partial requirement to obtain Master′s degree in Electrical Engineering, area Telecommunications and Telematics; in the scope of collaboration between UNICAMP and the Technical University of Vienna (TU Wien).

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas (UNICAMP) como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica, na Área de Telecomunicações e Telemática; no âmbito da colaboração Interinstitucional da UNICAMP com a Universidade Técnica de Viena (TU Wien).

Supervisor: Prof. Dr. Luiz C. Kretly

Co-supervisor Prof. Dr. Christoph Mecklenbräuker (TU Wien)

Master′s dissertation of oral examination of Orlando da Silva Trindade, supervised by Prof. Dr. Luiz C. Kretly, co-supervised by Prof. Dr. Christoph Mecklenbräuker (TU Wien).

Campinas

2018

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Inclua aqui o pdf com a ficha catalográfica fornecida pela BAE.

COMISSÃO JULGADORA - DISSERTAÇÃO DE MESTRADO

Candidato: Orlando da Silva Trindade RA: 160958
Data da Defesa: 06 de Junho de 2018

Título da Tese - Avaliação estatística do desempenho de repetidores sem fio em veículos sobre trilhos de longo trajeto: estudo de caso da melhoria da qualidade dos serviços de celular e Internet em trens europeus

Título da Tese (Inglês) - Supporting high quality cellular and Internet services inside railway vehicles: a statistical evaluation of wireless repeater performance on Railjet and Cityjet trains

Prof. Dr. . Luiz Carlos Kretly (Presidente, FEEC/UNICAMP)
Prof. Dr. Christoph Mecklenbräuker (TU Wien)
Prof. Dr. Silvio Ernesto Barbin (Membro externo, EP/USP)
Prof. Dr. . Michel Zamboni Hached (FEEC/UNICAMP)

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no processo de vida acadêmica do aluno.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

To Ms. Maria Olga da Silva, my mother.

Acknowledgements

This work has required several supports without which it would have been impossible. By mentioning the supporters, I wish to share this achievement with them. I wish to thanks the Brasilian Nacional Council of Science and Technology *(in Portuguesse: Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq)*; the State University of Campinas *(in Portuguesse: Universidade Estadual de Campinas - UNICAMP)*; the Technical University of Vienna *(in German: Technische Universität Wien)*. My special thanks to the Austrian Federal Railways-Technical Services *(in German: Österreichische Bundesbahnen-ÖBB-Technische Services GmbH)*, in the person of Dip. -Ing. Josef Resch and Dr. Stephan Ojak, whom initiated this research in first place and, personally, took part in measurement campaigns. Special thanks to some individuals are also necessary. In this way, I want to express my gratitude to my supervisor, Prof. Dr. Luiz C. Kretly, from State University of Campinas (UNICAMP), his multidisciplinary approach in science have allowed to bridge some apparently disconnected sections. In the same way, I thank Dr. Prof. Christoph Mecklenbräuker, my co-supervisor from TU WIEN, for his patience with some of my most uneasy times; and his encouraging explanations, true readjustment to my path. I also thank Dr. Philipp Svoboda and M.Sc. Taulant Berisha, both from TU WIEN, for having saved many of my days with effective support during data extraction.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

"The only true wisdom is knowing that you know nothing."

Socrates

Abstract

The present work aims to support improvement on Internet and telephony services on-board railway vehicles by providing theoretical methodological apparatus to evaluate technological solutions in this context. The proposed method is used to benchmark the impacts of an Amplify-and-Forward (AF) wireless repeater on Radio Frequency signal quality (downlink) inside High Speed Train (HST)s, with focus on statistics of Radio Frequency (RF) received signal. To do so, mobile phones, under quasi-real usage conditions (voice and data services usages), are deployed inside an operational High Speed Train (HST) equipped with AF repeaters. Field measurements took place on railway tracks in Austria and covered five relevant bands: Global System for Mobile Communications (GSM) (bands 900MHz and 1800MHz), Universal Mobile Terrestrial Services (UMTS) (band 2100MHz) and Long Term Evolution (LTE) (bands 800MHz and 2600 MHz). The work combines bootstrap with parametric and non-parametric statistical hypothesis tests to characterize distributional behavior of Radio Frequency signal power level on thousand-point-large data samples. Comparison of the *Student's* t-test and Wilcoxon test has shown that parametric t-test does not correctly quantify the change, yielding an unreliable statistical significance, this test is efficient to point the direction of improvement; however. It was found that the combination of this test with bootstrapping and non-parametric test is efficient in measuring the stochastic change in signal levels inside HSTs resulting repeater influence. Data analysis have shown that repeater effect changes the stochastic features of signal levels on HSTs, it produces a left-tailed and right-bounded distribution with large departure from the normal distribution, an indication that a measure of performance cannot be given solely via change in the mean. Moreover, data investigation strongly support downlink improvement under repeater condition in rural and urban areas, except for cases in which signal strength inside the HST averages above -80dBm most of the time. In addition, analytical results also support that repeater produces different improvements under diverse Radio Access Technology (RAT).

Keywords: Cellular communication inside trains; Cellular network repeater on trains; vehicular communications; Inferential Statistics; Hypothesis test.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Resumo

O presente trabalho pretende apoiar avanços nos serviços de telefonia e Internet em trens, através da elaboração de aparato teórico-metodológico para avaliar soluções técnicas nesse contexto. Com foco na avaliação estatística do sinal recebido, o método proposto é usado para estimar a influência de repetidores de sinal de rede celular-do tipo *Amplify-and-Forward (AF)*- na melhoria da qualidade do sinal dentro de trens. Para isso, celulares *(smart phones)*, operando sob condições quase-realísticas de uso dos serviços de voz e dados, são postos dentro de um trem de alta velocidade *(em inglês: High Speed Train-HST)* equipado com um repetidor sem fio *(em inglês: wireless repeater)*. Medições de campo foram realizadas em vários trechos da malha ferroviária da Áustria e incluiu cinco bandas de sinal celular; a saber: GSM (bandas 900MHz e 1800MHz), UMTS (banda 2100MHz) e LTE (bandas 800MHz e 2600 MHz). O trabalho combina re-amostragem *(em inglês: bootstrap resampling)* com testes de hipóteses paramétricos e não-paramétricos para caracterizar o comportamento da distribuição estatística do sinal de Radiofrequência dentro do trem, usando amostras com milhares de pontos. O método proposto consiste na comparação do teste paramétrico, *Student's* t-teste, com não-paramétrico, o teste de soma de ranques *(em inglês: Wilcoxon rank-sum test)*. Tal comparação mostrou que o teste paramétrico não quantifica corretamente a mudança e produz uma significância estatística não confiável, embora seja eficiente para apontar a direção do efeito. A combinação do *bootstrapping* com os testes paramétrico e não-paramétrico mostrou-se eficiente na quantificação e qualificação da mudança estocástica do sinal, consequente do uso do repetidor. Análise dos dados mostra que o efeito do repetidor muda as características estocásticas do sinal dentro do trem, resultando numa distribuição assimétrica negativa e, portanto, desviada da distribuição Gaussiana. Isso constitui clara indicação de que apenas a mudança na tendência central -na média, por exemplo- é insuficiente como medida de desempenho do repetidor. A investigação dos dados confirma que o repetidor melhora a qualidade do sinal recebido (*downlink*) dentro de trens, tanto em áreas rurais quanto em urbanas. Contudo, essa melhoria não aparece nos casos em que o valor médio da potência do sinal dentro do trem permanece acima de -80dBm a maior parte do tempo. Os resultados analíticos também confirmam mudança no desempenho do repetidor entre tecnologias de rádio celular.

palavras chave Comunicação celular em trens; repetidor de sinal de rede celular em trens; Comunicação veicular; Estatística inferencial; Teste de hipóteses.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

List of figures

Figure 1 – Simplified representation of the downlink with repeater. 42
Figure 2 – Simplified representation of the uplink with repeater. 43
Figure 3 – Procedures structure. 46
Figure 4 – Measurement equipment entity, image A; entity setup aboard, image B, GPS active antenna image C-1, and GPS modem image C-2. 48
Figure 5 – Measurement ajustment process. 49
Figure 6 – Row data file 50
Figure 7 – Rural section (A-B) on railway track - Vienna to Salzburg. 56
Figure 8 – Non-linear behavior under repeater effect. 58
Figure 9 – Histograms with equal sample size, N=16645; comparing contexts. 59
Figure 10 – Empirical t-test and two-sided alternative. Technology UMTS, variable RSCP. 60
Figure 11 – Non-parametric efffect-size, proportion P(X>Y). 62
Figure 12 – Signal strength distributions for GSM (900MHz) 16094 scores and UMTS (2100MHz) 16755. 63
Figure 13 – Comparison of signal strength inside the HST: rural vs. urban areas - UMTS (RSSI). 65
Figure 14 – Histogram for no-repeater context. 92
Figure 15 – Histogram for repeater context. 93
Figure 16 – Proportion P(X>Y). Wicoxon test for LTE. 94
Figure 17 – Histogram for no-repeater context, sample size, N=15191. Cityjet train. . 96
Figure 18 – Histogram for repeater context, sample size, N=15191. Cityjet train. . . 97
Figure 19 – Proportion P(X>Y). Wicoxon test for Cityjet. 98
Figure 20 – Example of onboard-repeater. 99
Figure 21 – Example of repeater configureation. Source: OBB report 100
Figure 22 – Example of outdoor antenna for onbord solution. 101
Figure 23 – Example of power spliter for onbord solution. 102
Figure 24 – Repater deployment onboard a railjet train. (Source: ÖBB) 108
Figure 25 – Railjet train. (Source: ÖBB) 109
Figure 26 – Cityjet train. (Source: ÖBB) 109

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

List of tables

Table 1 – Elementary statistics for UMTS, variable RSCP. 57
Table 2 – Measures of spread and shape 59
Table 3 – Empirical t-test. Technology UMTS, variable RSCP 60
Table 4 – Results for Wilcoxon sum-rank test 61
Table 5 – T-test- table. Variable RSSI 64
Table 6 – Results for Wilcoxon sum-rank test. Variable RSSI. 64
Table 7 – T-test results for UMTS in rural and urban areas; variable RSCP. 66
Table 8 – Wilcoxon sum-rank test results for UMTS in rural and urban areas; variable RSCP. 67
Table 9 – Elementary statistics. Variable RSSI. 91
Table 10 – Measures of spread and shape. Variable RSSI 92
Table 11 – Elementary statistics for LTE. Variable RSSI 93
Table 12 – Measures of spread and shape for LTE. Variable RSSI. 93
Table 13 – Empirical t-test for LTE, variable RSSI. 94
Table 14 – Wilcoxon sum-rank test for LTE, variable RSSI. 95
Table 15 – Rural and urban comparison. Variable RSSI. 95
Table 16 – Spread and shape for rural and urban comparison. Variable RSSI 95
Table 17 – Elementary statistics for Cityjet train. Variable RSSI 96
Table 18 – Measures of spread and shape for Cityjet train. Variable RSSI 97
Table 19 – Empirical t-test for Cityjet train. Alternative $H1\_M\_2 \neq M\_1$ 98
Table 20 – Wilcoxon sum-rank test for Cityjet 98

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

List of Abbreviations

Abbreviations and descriptions

OECD Organization for Economic Cooperation and Development

GPS Global Positioning System

ÖBB Austrian Federal Railways (in German:Österreichische Bundesbahnen)

CDF Cumulative Distribution Function

p.d.f. Probability Density Function

ECDF Empirical Cumulative Distribution Function

LTE Long Term Evolution

RV Random Variable

CLT Central Limit Theorem

MLE Maximum Lakelihood Estimation

OSI Open System Interconnections

iid Idependent identically distributed

FDR False Discovery Rate

MWW Man-Whitney-Wilcoxon

MRN Moving Relay Node

RAT Radio Access Technology

UDP User datagram protocol

AF Amplify-and-Forward

BS Base Station

MS Mobile Station

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

RAT Radio Access Technology

HetNet Heterogeneous Network

LoS Line-of-sight

ME Measurement Entity

RF Radio Frequency

DF Decode and Forward

ECDF Empirical Cumulative Distribution Function

GSM Global System for Mobile Communications

HST High Speed Train

LTE Long Term Evolution

LTE-A Long Term Evolution - Advanced

MRN Moving Relay Node

RSCP Received Signal Code Power

RSRP Received Signal Reference Power

RSSI Received Signal Strength Indicator

UMTS Universal Mobile Terrestrial Services

UE User Equipment

VPL Vehicular Penetration Loss

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Contents

Contents . . . . . .

I Overview **17**

1 Introduction **18**

1.1 General objective 19

1.1.1 Specific objectives 19

1.2 Literature review 20

II Theoretical framework **23**

2 Fundamentals of mobile communication **24**

2.1 Wireless transmission 24

2.1.1 Theoretical models 25

2.2 Empirical models 25

2.3 Stochastic behavior of the channel 26

2.3.1 Lognormal and large-scale fading 26

2.3.2 Small-scale fading: Rayleigh and Rician distributions 27

3 Statistical foundations for data evaluation **29**

3.1 Graphical analysis 29

3.1.1 Quantile-quantile plot (qq-plot) 29

3.1.2 Empirical Cumulative Distribution Function (ECDF) 29

3.2 Bootstrap resampling 30

3.3 Hypothesis test statistics 31

3.3.1 Multiple tests and significance probability 32

3.3.2 Parametric test–the Student's t-test 33

3.3.2.1 Parameter estimation 33

3.3.2.2 F-ratio 34

3.3.2.3 The empirical t-test design 34

3.3.2.4 Considerations on repeater effect and t-test 35

3.3.3 Rank-based non-parametric test 36

3.3.4 Wilcoxon rank-sum test 37

3.3.4.1 Significance probability 39

4 Theoretical considerations **41**

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4.1 Predicted stochastic behavior of signal on HSTs . . . 41
4.2 Mobile network resources for users on HSTs . . . 43
4.2.1 Influence of handover . . . 44

III Methodologies **45**
5 Measurement methodology . . . **46**
5.1 Routes and sectors . . . 46
5.2 Assessed solution . . . 47
5.3 Equipment deployment and setup . . . 48
5.4 Row metadata features . . . 49
6 Statistical Procedures . . . **51**
6.1 Parametric test procedure . . . 51
6.1.1 Data conditioning and parameter estimation . . . 51
6.1.2 Empirical critical value and t-statistic . . . 52
6.1.3 Estimated p-value . . . 52
6.1.4 Sample size . . . 52
6.1.5 Estimated effect size . . . 53
6.2 Wilcoxon sum-rank test procedures . . . 53
6.2.1 The Null hypothesis . . . 53
6.2.2 Non-parametric effect size . . . 54
6.2.3 Sample size estimation . . . 54
6.2.4 Critical value . . . 55
6.3 Data selection . . . 55
6.3.1 Spacial analysis . . . 55
6.3.2 Diverse RAT . . . 56
7 Results and discussion . . . **57**
7.1 Repeater effects on signal levels on HSTs . . . 57
7.2 Diverse mobile technology . . . 62
7.3 Rural and urban contexts . . . 64
Conclusion . . . **68**
References . . . **70**

References (Not cited) . . . **75**

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Appendices 78
A SUMMARY AND FURTHER INFORMATION FROM SUBMITTED PAPER 79
B SUMMARY OF SLIDES USED DURING DISSERTATION DEFESE 82
C Complements for chapter 7 91
C.1 RAT comparison: statistics on scale and shape 91
C.2 Summary of results for LTE- Railjet train 92
C.2.1 Basic statistics 92
C.2.2 Hypothesis test results 93
C.3 Rural and urban comparison: statistics on scale and shape 94
C.4 Summary of results for Cityjet train 95
C.4.1 Basic statistics 95
C.4.2 Hypothesis test results 97
D TECHNICAL INFOMATION FOR ON-BOARD REPEATER SOLUTION . . 99
D.1 Repeater electrical specification 99
D.2 Example of configuration for onboard-repeater 100
D.3 On-train rooftop mounted antenna - specification 101
D.4 Example of specification for power spliter 102
E NEMO TOOL 103
E.1 Nemo Walker Air 103
F TRAINS AND ONBOARD REPEATER 108
F.1 Example of deployment onboard Raijet trains 108
F.2 Raijet and CityJet trains 109
G AWK SCRIPTS - KEY POINTS 110
G.1 AWK 110
H R SCIPTS - KEY POINTS 113
H.1 R program 113
H.1.1 Macro scripts sequence 113
H.2 PARAMETRIC SECTION 113
H.2.1 Data conditioning 113
H.2.2 Empirical t-test 115
H.3 SCIPTS-NON-PARAMETRIC SECTION 118
H.3.1 Data conditioning 118
H.3.2 Sample size estimation 118
H.3.3 Wilcoxon rank-sum test 119

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Part I

Overview

1 Introduction

Modern life requires a great variety of remote and virtual interactions under mobility with impact in business and wellbeing; hence the possibility to communicate remotely, at anytime and anywhere, represents a current demand. Indeed, that demand is expected to increase in certain context, seeing that the number of mobile phone and broadband has been growing, as well as the incentives on usage of traffic-free public transportations. Nevertheless, while actual system solutions do not provide high quality mobile communication in trains, literature lacks on-real-world studies accessing performance of solutions for wireless communications on trains, considering the user's perception. Naturally, improving mobile communication on trains requires system characterization based on user demands.

Numbers of mobile phones and wireless broadband have been growing in the last ten years. In fact, mobile-broadband subscriptions per hundred inhabitants increased by 42.3 in Europe and 50.1 in the Americas, also there are 119 and 111 mobile per hundred inhabitants in those regions, respectively [1].

In parallel, worldwide environmental concerns impact on improvement and usage of public transportation. In the case of railways, this translates into modernization in terms of speed, safety and comfort, therefore a time-saving, productive and comfortable mobility solution, onto which passengers can perform a variety of activities during the trip. For example a student could connect through an e-learning platform; tourists could be able to book a restaurant table or by a ticket on-line; or one could relax watching a movie or accessing social media. Not to mention the new demands related to the smart-things as result of the Internet of Things (IoT). Those activities are only possible with high-quality mobile communication services inside trains.

Despite of that, on-real-world assessment of current solutions to improve mobile communications on trains seems to remains sparse which can provide material for theoretical studies and future solutions, seems to remains sparse. Studies on that direction can provide material for theoretical studies which will both improve and stimulate the development of future solutions. Indeed, performing mobile communication assessments on HSTs require multi-collaboration and it is both expensive and time-consuming. On one side, lie interests regarding the railway operator - i.e. safety and customer satisfaction aboard the train, once the measurement equipment and the wireless solution have to be deployed on operational trains in compliance

[1] This includes Internet connections for smart phones, tablets, and portable laptops. See <''http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> for details.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

with railway standards. If the measurement requires local monitoring, personnel and tickets for the trips have to be conceded. On the other side, mobile operator's strategy and interests vary among geographic regions, which might discourage investing in the study. Because the measurement requires real User Equipment (UE) usage, it consumes a significant number of hours for phone calls and data services. As a result, the assessment requires limitless data and voice services, which translates in a considerable investment. The manufacture of the on-train wireless solution–the manufacturer of the repeater in this case- appears as another entity to be considered. This partially justifies the lack of studies based on real mobile usage, which this study further contributes to change.

Combining the rise in mobile phones and broadband data subscriber with preferences for railway transportations, it is resonable to expect an increase in passengers using mobile devices on railway vehicles. Therefore providing satisfactory wireless communication in trains will impact positively on both the Economy and well-being, given that activities under mobility range from socializing to leisure, from working to learning, from on-line shopping to e-sales. Hence, a further understanding of the problem through assessment on real usage contexts consist in a valuable research topic.

1.1 General objective

The present work intend to support high quality cellular and Internet services inside railways vehicles, by providing methodological framework to assess improvements of those services on HSTs.

1.1.1 Specific objectives

In this study we undertake a statistical assessment of a wireless repeater performance onboard Railjet and Cityjet trains. From a research point of view, we question to what extent an active wireless repeater, installed inside a HST, influences on Internet and telephony services performance inside trains. More specifically, the present study begin answering this questions by performing a physical layer evaluation to answer the following primary questions:

- how the repeater effect changes the stochastic behavior of the received signal on board?
- what are the impacts on the received signal on board when considering different Radio Access Technology (RAT)?
- what are the impacts of geography on the repeater performance and, consequently, on the received signal on the train?

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- do statistics based on measure of central tendency (mean and median) constitute appropriate method to quantify and attribute statistical significance to the repeater effect?

In order to challenge the *status quo*, it is questioned whether the observed measurements support the statement that analog repeaters enhances mobile voice and data services on HSTs. The suggested research hypothesis states that deploying repeater inside trains improve mobile communication performance under some external conditions: terrain topology, distribution of base station, variety of RAT along the railway corridor and coverage profile, and repeater configuration. Moreover, a wireless repeater would even degrade the network performance in some of these cases.
The study comprised the following main activities:

- measure indicators from mobile phone Open System Interconnections (OSI) layer while mobile phones perform pre-configured voice and data tasks –e.g. call, web access, download;
- perform preliminary data analysis based on descriptive statistics;
- perform measurements on RF link quality using Received Signal Strength Indicator (RSSI), Received Signal Code Power (RSCP) and RSRP under context with and without repeater;
- provide general understanding of service performance with and without repeater – preliminary results;
- mathematically formalize hypotheses and choose the testing;
- select appropriate samples sets and and extract data from the data sets;
- choose appropriate tests for data analysis;
- apply parametric and non-parametric hypothesis testing and to analyze results to draw specific conclusions.

1.2 Literature review

Providing mobile communication services to vehicular users on public transportations appear to be a complex challenge dependent on vehicular structure, mobility state, mobile receiver and network infrastructure. Authors in (WU; FAN, 2016) analyze wireless communication for high mobility modems by presenting summary of main challenges on providing

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

wireless services under high mobility conditions - e.g. high speed trains. This comprises various entities of a communication system, such as channel model for high mobility, transceiver features and signal processing approach. The review points the main constraints on wireless mobile with high mobility, including those cited earlier in this chapter, and it further indicates fast time-varying fading, channel estimation errors, Doppler diversity, carrier frequency offset and inter-carrier interference as additional challenges. To that general approach, we could add also RAT handover. Preliminary analysis based on *Descriptive Statistics* indicates improvements on signal quality inside HST equipped with wireless repeater (TRINDADE *et al.*, 2016).

PROSPECTIVE SOLUTIONS – Despite of that, available technical solutions and theoretical studies seem to partially address the problem. Indeed, one approach is deployment on board the train, where the repeater interfaces outdoor and indoor links. For instance, a system with cooperative Moving Relay Node (MRN) can provide significant improvement on throughput compared to direct transmission. The system consists of deploying one Decode and Forward (DF) repeater – one on each carriage – configured in a master-slave architecture (SCOTT *et al.*, 2013). In contrast, an alternative consists in UE on board the train connecting directly to collaborative Remote unit (RU) deployed along the railway (MULLER *et al.*, 2015). This context is represented with a mathematical model, and evaluated using Long Term Evolution - Advanced (LTE-A) System Level Simulator. The optimum performance, in terms of spectral efficiency in [bit/s/Hz], occurs when remote units work in a cooperative manner, avoiding mutual interference and reducing handover occurrences to one every 50km along the railway for train speed 200km/h and inter RU distance of 1km.

SOLUTION ASSESSMENT – Assessing service performances of solution for serving vehicular receivers under operational condition provides means to better visualize constraints related to various sections of the communication protocol stuck, hence leading to improvements. For example, (LUTU *et al.*, 2016) presented a machine learning methodology to classify broadband service features perceived by passengers on a HSTs in different regions along a railway. The five-month-long field measurement was performed with a dedicated computer (anboard) communicating with a server through mobile networks of two mobile system operators. Focusing on (RAT) – no service, 2G, 3G and 4G - the study applies pattern identification and clustering approach to assess the metadata. Results indicated that insufficient service coverage coincides with inter-city and rural regions, and frequent transitions among no-service, 2G, and 3G. The work also points out that package loss higher than 1% occurred in more than 30% of sample in areas with sufficient service coverage, which is believed to result from frequent RAT handover.

In complement, an investigation found in (BALTRUNAS *et al.*, 2016) quantifies the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

relationship between RAT – i.e. no service, 2G, 3G or LTE - and packet loss inside trains, following the measurement method found in (LUTU *et al.*, 2016), that dedicated measurement computer one User datagram protocol (UDP) packed per second and two mobile broadband operators. The findings point out to packet loss occurrence of 2% to 12% in stationary case, and 30% to 50% under mobility. It highlights packet loss in diverse RAT: 3G has higher packet loss than LTE. With regards to handover and packet loss relationship, 30% of packet loss happened under single RAT, and 70% happened when at least one RAT change occurred. These figures were pointed as indication that handover maintains strong relation with packet loss, however cell changes – horizontal handover – was mentioned as the main cause of packet loss.

THE PRESENT WORK - Studies on performance assessment based on real measurement remains scarce and its data is hardly available to other researchers. The scarcity of those works results mainly from cost and complexity, as explained in earlier in this chapter. Moreover, studies covering voice and data services and including current RATs seem to be unpublished, or at least few, to the best knowledge of the author. Furthermore, further increment in methodology is also needed to widely support performance assessment of solutions to improve mobile communications inside HSTs. The proposed research performs a mobile communication assessment inside a train equipped with an AF wireless repeater, under current multi-RAT contexts, and quasi-real usage conditions. The study provides further contribution by applying a novel data analysis based on bootstrap re-sampling combined with parametric and non-parametric hypothesis test.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Part II

Theoretical framework

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

2 Fundamentals of mobile communication

The received signal amplitude and phase are elements of primary importance on detection processes. Once distortions of those elements are inherent in wireless signal transmissions, measures to counteract signal degradation are essential to overall system performance. In this chapter we consider classic theoretical and empirical models to support expectations on the subject and understanding of empirical results.

2.1 Wireless transmission

In wireless mobile communication, a Base Station (BS) on a fix position $S_{BS}(x, y, z)$ communicates with a Mobile Station (MS) with random position $S_{MS}(X, Y, Z)$ and velocity $V_{MS}(t)$ depict a typical context where the random nature of geometric path between transmitter and receiver models stochastic characteristics of the received signal. Consider a electrical field, with direction $\hat{x}$, at a distance very far away from the source, a possible mathematical description of this physical phenomenon is

$$\vec{E} = E_0 \exp[-(\alpha + j\beta)]\hat{x}, \tag{2.1}$$

where attenuation constant α, detailed in Equation 2.2, describes the decay of the field amplitude compared to the initial value E_0 in Equation 2.1; wheres the phase constant β, detailed in Equation 2.3, describes changes on the phase of the field in Equation 2.1.

$$\alpha^2 = \frac{\omega^2 \mu \epsilon}{2}\left[\sqrt{1 + \left(\frac{\sigma}{\omega\epsilon}\right)^2} - 1\right] \tag{2.2}$$

$$\beta^2 = \frac{\omega^2 \mu \epsilon}{2}\left[\sqrt{1 + \left(\frac{\sigma}{\omega\epsilon}\right)^2} + 1\right] \tag{2.3}$$

Classical wave phenomena - reflexion, refraction, transmission and diffraction- results from the spatial geometry of the transmitting media, as well as the media's constructive parameters: conductivity σ, magnetic permeability μ and electric permittivity ϵ. This provides further characterization of electrical field at a given point in terms of amplitude, phase and frequency.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.1.1 Theoretical models

Wireless signal propagates in space containing diverse material and geometries. For example, urban and rural topologies at macro level, and the respective physical properties of surfaces that constitutes those topologies. The non-deterministic combination of constructive parameters and forms at macro and micro levels interacts with the propagating wave, and the resulting physical phenomena alter the transmitted signal in terms of statistical distribution. The path loss is an estimate measure that relate the transmitted power P_t with the received power P_r as $l = P_r/P_t$ (in Watts) and the equivalent in dB,

$$L = -10\log(P_r) + 10\log(P_t). \tag{2.4}$$

Low values of L relates the absence of obstructions on the path, existence of Line-of-sight (LoS). This condition is described via *Friis free-space* model,

$$\frac{P_r}{P_t} = G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2, \tag{2.5}$$

in which, given constant antenna gains G_t and G_r, the loss increases with distance d between transmitter and receiver.

In the worst case scenario, obstructions block LoS between the communicating points, creating shadowed regions with large and positive *Fresnel diffraction parameter* ν in $|F(\nu)|$. This results in loss from the incident field E_i to the diffracted field E_d, given by Equation 2.6, and can be as much as 24dB (SAUNDERS; ZAVALA, 2007).

$$L = -10\log(|E_d/E_i|^2) = -20\log|F(\nu)| \tag{2.6}$$

2.2 Empirical models

Empirical models for path loss adjust theoretical models via parameter from measurement which includes elements from natural and artificial features. Okumura-Hata and Cost 231 are example of empirical model in current literature in the topic. The *Simplified path loss model* in Equation 2.7 pictures the presence of *power law* as part of the model for received signal in the term $d^{-\alpha}$.

$$\frac{P_r}{P_t} = K\frac{G_t G_r h_t^2 h_r^x}{d^\alpha f^y}, \tag{2.7}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

In this equation K is constant for environment (urban and rural characteristics); parameter for terrain topography, path loss slope, $2 \leq \alpha \geq 4$; y include the frequency dependency of the loss; also $h_t < 3$ gives for $x = 1$ and $3 \leq h_t \geq 10$ gives $x = 2$ includes the influence of the receiver antenna height on the path loss (YACOUB, 1993).

The path loss varies according to receiver position and its mobility state related to the serving BS. For instance, a vehicular mobile moving outward the base station in a region with the same path slop experiences a power decay from a point i to a point j give by

$$\frac{W_{rj}}{W_{ri}} = \left(\frac{d_j}{d_i}\right)^{-\alpha}. \tag{2.8}$$

In case of changes of environment -i.e. varying path slope- and change of distance from the BS, the loss compared to a given point d_0 is then

$$\frac{W_{rn}}{W_{r0}} = \left(\frac{d_n}{d_0}\right)^{-\alpha} C_1, \tag{2.9}$$

and the path loss correction constant resulting from variation of environmental the parameter (YACOUB, 1993) is given by

$$C_1 = d_1^{-\alpha_1} d_n^{\alpha_1} \prod_{i=2}^{n} \left(\frac{d_i}{d_{i-1}}\right)^{-\alpha_i}. \tag{2.10}$$

2.3 Stochastic behavior of the channel

Fading processes describes distortion on wireless signal. In this section we consider channels described with *large-scale fading* and *small-scale fading*.

2.3.1 Lognormal and large-scale fading

Sudden introduction of blocking or large diffractive obstacles (tall building, tunnel, hills) in a regions near (fraction of km) the mobile station causes *shadowing* or *slow fading*. In such a case, the decay in signal strength increases in inverse proportion to the distance between the interferer and the mobile station. Considering the maximum MS antenna height of 3m, the consequent large positive values of *diffraction parameter* places the MS in the deep *shadow region*.

Under blockage condition, scattering and transmission processes occurs at different degrees, depending on the physical characteristics of the obstacle and the form of the incident signal - i.e. multiple components or single strong component. In this condition, received

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

signal with multiple components of small amplitudes, occur when scattered incident signal penetrates the blocking surface–transmission phenomenon. Incident signal with single intense component can cause diffraction and/or scattering on the obstacle. This originates a multipath effect and the signal, at the shadow region, presents multiple components with significant phase shift. This result in destructive interactions and decay on amplitude (SEYBOLD, 2005). In this context, the received signal component E_n, transmitted through the nth obstruction with thickness Δr_i is

$$E_n = E_0 \exp\left(-\sum_{i=1}^{n} \alpha_i \Delta r_i \right), \tag{2.11}$$

and declaring $x = -\sum_{i=1}^{n} \alpha_i \Delta r_i$, we rewrite Equation 2.11 as $E_n = E_o \exp(x)$ which has normal distribution given by 2.12 for $n \to \infty$.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x}} \exp\left[\frac{1}{2} \left(\frac{x - \mu_x}{\sigma_x} \right)^2 \right] \tag{2.12}$$

2.3.2 Small-scale fading: Rayleigh and Rician distributions

In absence of large obstructions scattering and reflection effects are dominants on signal path and the received signal envelope consist of multiple components with random amplitude and phase. The predominant multipath propagation produces delay spread (time domain effect), while the vehicle speed originates Doppler spread (frequency domain effect). Both phenomena yield frequency-selective and non-frequency selective fading. In the case of frequency-selective fading, or fast fading, characteristic on wide-band channels, occur within a symbol period. Non-frequency selective fading or flat fading is characteristic on narrow band channels, where signal is considered constant within a symbol period (SEYBOLD, 2005). Those contexts describe a small-scale fading channel.

When scattering predominates and no strong signal component is observed, the received signal distribution with n-scattered and multipath components with different phase θ_i is

$$S_n = a_i \sum_{i=1}^{n} \exp\left[j(\omega_0 t + \theta_i) \right]. \tag{2.13}$$

The stochastic behavior is described via Rayleigh distribution with Probability Density Function (p.d.f.) given by

$$p(r) = \frac{r}{\sigma_r^2} \exp\left(-\frac{r}{2\sigma_r^2} \right), \text{ for } \; r \geq 0 \tag{2.14}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

where r=$|a\exp(j\theta)|$, and mean(r)=$\sigma_r\sqrt{\frac{\pi}{2}}$, var(r)=$\sigma_r^2\frac{4-\pi}{2}$ $r_y = a_i \sin\theta_i$ and r=$r_x^2 + r_y^2$.

In some contexts, LoS or intense reflexion produces signal with predominant strong component $a\exp(\omega_0 t)$ among the small scattered components $r\exp(j\omega_0 t + \theta)$ (YACOUB, 1993) and the received signal has the form

$$S = r\exp(j\omega_0 t + \theta) + a\exp(\omega_0 t), \tag{2.15}$$

which is then described via Rician p.d.f.

$$p(r) = \frac{r}{\sigma_r^2}\exp\left(-\frac{r^2+a^2}{2\sigma_r^2}\right)I_0\left(\frac{ar}{\sigma_r^2}\right). \tag{2.16}$$

The author understands that there might be other distribution that could provide more precise description of signal behavior. Those proposition are speculative and serve as starting point for further investigation interested in describing statistic distributions.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3 Statistical foundations for data evaluation

Descriptive statistics is applied to acquire an overview of the problem; whereas the inferential statistics, i.e. hypothesis test, is used to obtain more precise conclusions. This section contains theoretical elements for both topics.

3.1 Graphical analysis

Graphical analysis relies on non-parametric graph: quantile-quantile plot (qq-plot) and Empirical Cumulative Distribution Function (ECDF). Those elements are further explained in the following subsections.

3.1.1 Quantile-quantile plot (qq-plot)

A quantile-quantile compares two samples, or a sample against a theoretical distribution, using the samples' quantiles. The resulting graph is a scatter plot that reveals differences in location, dispersion and shape. Given two samples of Random Variable (RV) X and Y, it helps to answer whether these samples belong to the same population. This is, if they have a common distribution function; if they present a common location (mean, media, and mode) and spread (e.g. standard deviation, variance, range); if they have similar distributional shapes, or a similar tail behavior. When the samples have common distributional behavior, the points in the graph roughly fall onto a referential straight line through the origin with relationship $Y \approx X$ and an approximate unitary slope. Otherwise, points departure from that referential line, according to the distributional differences between the samples.

3.1.2 Empirical Cumulative Distribution Function (ECDF)

A Cumulative Distribution Function (CDF) represents the probability of the random variable (RV) assume a certain value less than or equal to a given value X, that is:

$$F_x(x) = P\Big\{X \leq x\Big\}. \tag{3.1}$$

The Empirical Cumulative Distribution Function (ECDF) is a CDF that represents collected sample–discrete values. Hence it constitutes an estimate of the CDF for samples of an experiment (NIST, 2012) and (KVAM; VIDAKOVIC, 2007). Considering an experiment in which n observations of a RV X give a sample $(x_1, x_2, \ldots, x_n)$. Often, the samples are

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

sorted in increasing order, giving the sorted sample
$(x_{(1)}, \ldots, x_{(n)})$ with $x_{(i)} \leq x_{(i+1)}$ for all i. Then, attributing equal probability $1/n$ to each value in the sample, the ECDF is the discrete step function – $F_n[x]$ - that assigns each ordered sample point in abscissas to a step in Y-axis. These steps are increased by $1/n$ at each point in X-axis (PAPOULIS, 1991). This function is represented by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\mathcal{X}_i}(x), \tag{3.2}$$

where $1_{\mathcal{A}}(x)$ denotes the indicator function,

$$1_{\mathcal{A}}(x) = \begin{cases} 0, & \text{for } x \notin \mathcal{A} \\ 1, & \text{for } x \in \mathcal{A} \end{cases} \tag{3.3}$$

and the events $\mathcal{X}_i$ are defined as

$$\mathcal{X}_i = \{X \geq x_i\}. \tag{3.4}$$

3.2 Bootstrap resampling

Bootstrap consists in a mechanism to exploit sample's distributions via resampling. The method does not require the Idependent identically distributed (iid) condition for the RV, hence it is applied to cases involving dependency (DASGUPTA, 2008). It consists in creating new sets of data samples from the main sample. In this way, it is possible to observe the variability of certain sample statistics –e.g. mean, median- among resampled sets of data. In addition, this method provides support for multiple hypothesis testing.
Consider a measurement and its respective sample

$$X_1, X_2, \ldots, X_n$$

from a population $G_x(x)$. Resampling B-times from the original data with replacement (non-parametric bootstrap) it generates bootstrapped samples–X^*– with size n, randomly drawn from the ECDF $G_n[x]$, and represented as

$$({X_{11}}^*, {X_{12}}^*, \ldots, {X_{1n}}^*); ({X_{21}}^*, {X_{22}}^*, \ldots, {X_{2n}}^*); \ldots, ({X_{B1}}^*, {X_{B2}}^*, \ldots, {X_{Bn}}^*). \tag{3.5}$$

Given a large sample size, the bootstrapped samples are considered to originate from the population $G_x(x)$ itself (DASGUPTA, 2008).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Considering original sample and its sample size N, an ideal resampling with replacement gives a total N^N number of boot strap from the original sample. That is impractical for large sample size, which raises the question on how many bootstrap resample are necessary to estimate a parameter; for instance. Literature on this topic provides no agreement on how many replicates should be used. A wide review of this matter, provided on (CHERNICK; LABUDDE, 2011), comments that while classical literature on the topic recommends from 100 to 1000 replicates; those numbers loose meaning in face of the currently available computational power, the author reinforces. Whereas, more recent work (CHERNICK; LABUDDE, 2011) uses 5000 replicates. The concept in (PATTENGALE *et al.*, 2009), supports that the ideal number of bootstraps can vary among different datasets; therefore, an ideal approach would be a stopping criteria where a pre-defined threshold interrupts the replication when enough replicates have been made.

3.3 Hypothesis test statistics

A statistical test is a mathematical rule that allows deciding whether or not the null hypothesis (H_0) should be rejected on the basis of the samples. The result of a hypothesis testing process is said to be statistically significant when it has very low likelihood to reject the (H_0) given that the null is true (GRAVETTER; WALLNAU, 2013). The process of rejecting and failing to reject the null hypothesis can yield two errors: type I and type II errors.

A basic hypothesis test design consists in pairs of hypothesis and both sustain statements about the system or condition under analysis. The null hypothesis (noted as H_0) represents the *status quo*, that is what it is assumed to be true or real, related to a system. In opposition, the alternative or research hypothesis (noted as H_1) predicts a change in the system under considered circumstances.

Outcomes from the test can lead to reject or fail to reject the null hypothesis, which may incur errors. Thus, a type I (type one) error consists in rejecting the null hypothesis when it is actually true according to the real tested system. The significance level α of a test represents an upper bound probability of such error. A type II error is the failure to reject the null hypothesis, when H_1 is true according to the actual system. The likelihood of that error is represented by β, and hence (1- β) represents the probability of a correct rejection of the null or the power of a test. An illustrative explanation of hypothesis testing can be found in Appendix B.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3.1 Multiple tests and significance probability

A context involving multiple independent hypotheses, that is various null hypotheses

$$H_I, H_{II}, \ldots, H_n, \tag{3.6}$$

result in equal number of p-values referred to each of the null and noted as

$$P_I, P_{II}, \ldots, P_n. \tag{3.7}$$

Assume a test with specified $\alpha = 5\%$, hence, as long as p-value$< \alpha$ the design is satisfactory. As the number of tests k increase, so does the likelihood of occurring the rare event $\alpha = 5\%$, that is type I error. Given the probability of not committing a type I error

$$1 - \alpha, \tag{3.8}$$

it is possible to write the probability of not commit such error in k-tests as

$$(1 - \alpha)^k, \tag{3.9}$$

from where the probability of committing such error at least once is

$$1 - (1 - \alpha)^k. \tag{3.10}$$

As a result, it is expected to find incorrect rejections, false positives, among the correct ones. The proportion of false positives in the entire result is named False Discovery Rate (FDR) or Type I error rate and discussed in (BENJAMINI; HOCHBERG, 1995). Literature provides a number of approaches to deal with this classical problem. A conservative approach, *Bonferroni correction*, for k-tests adjusts a new α/k which require a $p - value < 0.0005$ for $k = 100$ and $\alpha = 0.05$. An alternative consists in controlling the FDR via p-value adjustment. In this sense, *Simes method* (SARKAR; CHANG, 1997) gives the rejection condition as

$$P_{(i)} < i\alpha/k, \tag{3.11}$$

where $P_{(i)}$ is an element from p-values vector organized in crescent order. The main null hypothesis is rejected if at least one p-value satisfies the rejection condition in 3.11.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3.2 Parametric test–the Student's t-test

Parametric tests rely on assumptions regarding the sample's distribution, which appear as a necessary parameter or condition to apply the test. For instance, resemblance to normal and equal sample sizes are some of these parameters. In this sense, the *Student's* t-test allows verifying the significance of a linear shift between two distributions. In absence of referential parameters, parameter estimation can be used.

3.3.2.1 Parameter estimation

Given a RV X_i and its respective samples $\{x_1, x_2, \ldots, x_n\}$, we define the Maximum Lakelihood Estimation (MLE) for sample' mean

$$M_x = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.12}$$

and variances

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - M_x)^2. \tag{3.13}$$

When the sample size is large, the population estimator for mean $\hat{\mu}_x$ and variance $\hat{\sigma}_x^2$ can be estimated from the ECDF $F_x[x]$ when the these statistics are calculated via bootstrap resample from X. Define $B-th$ bootstrapped sample

$$X_B^* = (x_{B1}^*, x_{B1}^*, \ldots, x_{Bm}^*), \tag{3.14}$$

where X_B^* is a subsample from X_i. The estimated population parameters are calculated as in Equations 3.15 and 3.16. Repeating this process B-times yield $M_X = (M_{X_1^*}, M_{X_2^*}, \ldots, M_{X_B^*})$ and $S_{X_i}^2 = (S_{X_1^*}^2, S_{X_2^*}^2, \ldots, S_{X_B^*}^2)$. Since $E[\bar{X}] = \mu$ and $E[{S_i}^2] = \sigma^2$ we write

$$\hat{\mu}_X = \frac{1}{B}\sum_{i=1}^{B} M_{X_i^*} \tag{3.15}$$

and

$$\hat{\sigma}_X^2 = \frac{1}{B}\sum_{i=1}^{B} S_{X_i^*}^2, \tag{3.16}$$

where $\hat{\mu}_X$ and $\hat{\sigma}_X^2$ are the estimated population mean and variance, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3.2.2 F-ratio

Further variance behavior is explored variances ratio, F-ratio or, equally, F-value. In this measure, given variances S_x^2 and S_y^2, the F-value is

$$F - \text{value} = \begin{cases} S_x^2/S_y^2, & \text{for } S_x^2 > S_y^2 \\ S_y^2/S_x^2, & \text{for } S_x^2 < S_y^2 \\ 1, & \text{for } S_x^2 = S_y^2 \end{cases} \tag{3.17}$$

Variance homogeneity describes the case in which the max. F-value is about one.

3.3.2.3 The empirical t-test design

Given two samples from random variables, $X_i = (x_1, x_2, \ldots, x_{(m-1)}, x_m)^T$ and $Y_j = (y_1, y_2, \ldots, y_{(n-1)}, y_n)^T$, and a requirement to detect a shift in mean $\Delta M = M_y - M_x$, the t-test verify such a shift for a given significance α and degree of freedom $d.f. = m + n - 2$. This statistics, tests the null assumption of equality of expectations, that is

$$H_0 : [E(X)] = [E(Y)]. \tag{3.18}$$

On a trivial case, it is assumed that both samples are normal and the mean difference, under the null distribution, have distribution given by $t * (m + n - 2)$, and t-statistics is

$$t = \frac{(M_y - M_x) - (\mu_y - \mu_x)}{\tau} \tag{3.19}$$

where τ is the standard error for $(M_y - M_x)$ and can be computed in different ways depending on sample size and assumptions on the populations' variances - i.e. variance homogeneity[1]. Gaussian resemblance for both sample constitute a basic assumptions of this test.

This study focus here on the case of two independent samples X and Y with the following assumptions:

- The RV X_i and Y_i comes from unknown distributions;
- The variances of X_i and Y_i are unequal and unknown;
- The ratio of the variances is unknown;
- Independence exists between samples and among scores of a samples.

[1] When variances are nearly equal.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Under those conditions the t-statistics no longer have distribution in the form $t*(m+n-2)$ and the referential t-tables, from where the critical value $t_{(m,n,\alpha/2)}$ is no longer useful (DEKKING *et al.*, 2005). An alternative solution consist in resampling from the samples ECDF. Therefore, we subsample B-times, sample size with m and n from X and Y to obtain the MLE mean and variances, as in Equations 3.12 and 3.13, respectively. Then use the statistics

$$t_b^* = \frac{(M_{Y_b^*} - M_{X_b^*}) - (\hat{\mu}_y - \hat{\mu}_x)}{\tau_b^*}, \tag{3.20}$$

with

$$\tau_b^* = \left(\frac{S^2_{X_b^*}}{m} + \frac{S^2_{Y_b^*}}{n} \right)^{0.5}, \tag{3.21}$$

where the index b indicate the b-th bootstrapped sample. This processes yields a set of t-values,

$$T^* = (t_1^*, t_2^*, \ldots, t_B^*), \tag{3.22}$$

where T^* is a sample of the estimated null distribution and the ECDF of T^* provides the critical t-value for a given α at the quantile $q(\alpha)$ (DEKKING *et al.*, 2005) . This value is then compared with the statistics

$$T = \frac{(M_{Y^*} - M_{X^*})}{\tau}, \tag{3.23}$$

with τ defined as,

$$\tau^* = \left(\frac{S^2_{X^*}}{m} + \frac{S^2_{Y^*}}{n} \right)^{0.5}. \tag{3.24}$$

The null hypothesis is rejected at a level α when $T^* < T$.

3.3.2.4 Consideratios on repeater effect and t-test

Data evaluation aims to detect repeater influences and relies on hypothesis test to attribute significance to that change. For this purpose, it was decided to use measures of central tendency: mean, with mean shift detected via t-statistics; and median shift measured via Wilcoxon rank sum test. Concerning the t-test, some points needs careful attention. First, the test is applicable to linear models and *Student's* t-test does not confer reliability when variance differ between samples. In this condition, Type I error tend to increase with difference

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

in variance (RUXTON, 2006). Further consideration relates to the *Decibel transformation* or similarly log-transformation. A discussion on the inadequacy of the classical t-test to perform hypothesis on log-transformed data with unequal variance is presented in (FENG *et al.*, 2014), where the author suggest using the media as parameter to measure of central tendency. Alternatively, (ZHOU *et al.*, 1997) presents an adjusted parametric z-score test that can be applied without assumption related to the variances.

3.3.3 Rank-based non-parametric test

Statistical tests based on rank define a family of tests in which the statistics depend solely on position or location of the ordered sample's scores. The fundamental characteristic of those tests consists in transforming score into ranks from which the test statistics is calculated. Therefore, there is dissociation between the test statistics and the sample values or scores.

To obtain a generic notation for ranking, define samples $X = (X_1, X_2, \ldots, X_{(n-1)}, X_n)^T$ from a continuous Random Variable (RV) X, and the respective ordered sample set

$$(X_{(1)}, \ldots, X_{(n)}), \tag{3.25}$$

with $x_{(i)} < x_{(i+1)}$; the rank of $X_{(i)}$ is $R_{(i)} = j|X_i = X_{(j)}$. Therefore, the vector R with ranks of X is writen as:

$$R = (R_{(1)}, R_{(2)}, \ldots, R_{(n)})^T = (1, 2, \ldots, n-1, n), \tag{3.26}$$

from where we conclude that R, $i \in \mathbb{N}_+{}^*$.

There are cases when RV, $X = (x_1, x_2, ..., x_{(n-1)}, x_n)^T$, contains repeated scores (tied scores) and its ordered versions become

$$(X_{(1)} \leq X_{(2)}, \ldots, X_{(n-1)} \leq X_{(n)}), \tag{3.27}$$

in such a case the a *midranks* are assigned to the repeated scores with ranks i $X_{(i)} = X_{(i+1)} = \ldots = X_{(i+k-1)}$ which results in a *midrank*

$$\mathcal{R}^* = \frac{1}{k} \sum_{j=i}^{(i+k-1)} j, \tag{3.28}$$

that average rank substitute the ranks of all elements from i to $(i + k - 1)$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3.4 Wilcoxon rank-sum test

The Wilcoxon rank-sum test belong to a group of non-parametric tests that uses statistics based on ranks of samples to make inferences on the behavior of the sampling distribution. The statistics from this test maintain a close relationship with the one of Man-Whitney U test and literature showing their equivalence refers to both test as the Man-Whitney-Wilcoxon (MWW) test. Furthermore, there is recommendation to compute the Man-Whitney statistics to complement Wilcoxon rank-sum test in (LEHMANN, 1975), a work that is largely used to compose this section. This topic focuses on Wilcoxon rank-sum test and uses the equivalent U-test whenever it is convenient.

The present study adopts the treatment effect model which points whether the treatment have significant effect on the treated group. The test compares the shift in median or equivalent shift in the mean in cases where the treatment have a linear effect and the populations before and after the treatment are equal. The Wilcoxon rank-sum test allows to detect weather there is a shift between two populations represented by two dependent samples X and Y, based on the presumed homogeneity contained in the null hypothesis. A mathematical description of the null hypothesis follows.
Let F_X and G_X denote the CDF of X and Y respectively; the null statement on equality of $F_X(x)$ and $G_Y(y)$ is

$$H_0 : F_X(x) = P(X \leq x) \equiv P(Y \leq y) = G_Y(Y \leq y) \tag{3.29}$$

the two-tailed alternative is

$$H_1 : F_X(x) \neq G_Y(y) \tag{3.30}$$

and the one-tiled alternatives are $F_X(x) > G_Y(y)$ and $F_X(x) < G_Y(y)$.

Let $X = (X_1, X_2, \ldots, X_{(m-1)}, X_m)^T$ and $Y = (Y_1, Y_2, \ldots, Y_{(n-1)}, Y_n)^T$ be i.i.d. RV and the respective ordered vectors defined by Equation 3.25, the respective vector rank given by Equation 3.26. Also, conside the vector Z, containing all elements of X and Y in ordered version and represented by

$$Z(X;Y) = (Y_1, X_1, \ldots, X_{(m-1)}, ..., X_m, Y_{(m+1)}, \ldots, Y_{(m+n-1)}, Y_{(m+n)}). \tag{3.31}$$

Then, let R_z be the vector with the ranks of Z given by Equation 3.26, the Wilcoxon rank-sum test is based on sum of ranks of X or Y from Z. It is possible to chose any of the

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

two sets; and if Y is chosen, then rank sum W_y is

$$W_y = \sum_{j=1}^{n} j, \tag{3.32}$$

and for X, W_x is

$$W_x = \sum_{i=1}^{m} j. \tag{3.33}$$

Consider $m + n = N$ in Equation 3.31, we define the respective expectations of the distribution of W_x and W_y as

$$E(W_x) = \frac{m(N+1)}{2}, \tag{3.34}$$

$$E(W_y) = \frac{n(N+1)}{2}, \tag{3.35}$$

and the variances

$$Var(W_x) = Var(W_y) = \frac{mn(N+1)}{12} \tag{3.36}$$

In cases where the vector $Z(X;Y)$ contains repeated measures-tied score, the distribution of the ranks of X and Y containing *midranks*, as defined in Equation 3.28, not only dependent on m and n but also in the number of repeated measures (LEHMANN, 1975). The rank sum including *midranks* is noted as ${W_y}^*$and ${W_x}^*$ are the same as in Equations 3.33 and 3.32, respectively; as well,

$E(W_x) = E({W_x}^*)$
and $E(W_y) = E({W_y}^*)$.

Consider the number of random classes with repeated score b in Equation 3.31 and t_i the number of repetition in a given class i; then

$Var(W_x*) = Var(W_y*)$ is

$$Var({W_x}^*) = \frac{mn(N+1)}{12} - \frac{mn\sum_{i=1}^{b}({t_i}^3 - t_i)}{12N(N-1)}. \tag{3.37}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

3.3.4.1 Significance probability

Using the definition on Equations 3.25 and 3.26, we define a possible vector R_Z containing unequal ranks calculated in the combined vector $Z(X;Y)$

$$R_Z = (R_1, S_2, \ldots, S_m, \ldots, R_{(m+1)}, \ldots, R_{(m+n-1)}, R_N), \tag{3.38}$$

were R and S are the ranks of Y and X, respectively.
Consider m and n the respective sample size of Y and X, then, there is a number of combinations to randomly choose a set of n out-of N ranks in R_Z and the total number of those combinations is given by

$$C = \begin{pmatrix} N \\ n \end{pmatrix} = \frac{N!}{n!(N-n)!}. \tag{3.39}$$

Under the null hypothesis in Equation 3.29, the difference in the scores and consequently in the sum of ranks from X and Y are not statistically dissimilar. Therefore, all possible combination in Equation 3.39 have the same probability $1/C$; and hence no specific combination, yielding higher ranks sum in Y or X, will systematically occur. Considering, a model were Y represent the sample of the group under a given treatment and X the non-treatment group, the null hypothesis in Equation 3.29 is rejected when

$(W_y \geq c) \rightarrow W_x < c$ or when $(W_y \leq c) \rightarrow W_x > c$;

where c is a critical value (LEHMANN, 1975).

To postulate Y larger than X at a given significance level, we say that the probability of finding a sum of rank from Y smaller than the critical value c is as small as to the likelihood of finding a sum of ranks from X that is greater than the same critical value. Hence

$$P(W_y \leq c) = P(W_x > c) = \hat{\alpha}, \tag{3.40}$$

where $\hat{\alpha}$ is the significance probability or p-value, which compared to the specified significance level α, it is expected that $\hat{\alpha} \leq \alpha$.

Situations with N small– e.g. $N < 30$, there are statistical tables providing C from Equation 3.39 and critical value for a given m and n. Those details can be found in (LEHMANN, 1975). For large values of m, n and consequently N, the process is supported by asymptotic normality where the significance probability is given via z-statistics (LEHMANN, 1975). When m and n are very large, the significance probability for an one-sided alternative,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

H_1:$Y > X$,

is approximated as

$$P\left[\frac{W_y - E(W_y)}{\sqrt{Var_H(W_y)}} \leq z\right] \rightarrow \Phi(z), \tag{3.41}$$

and z is

$$z = \frac{W_y - E(W_y)}{\sqrt{Var_H(W_y)}}, \tag{3.42}$$

the significance probability calculated as

$$1 - \Phi(z) = \hat{\alpha}; \tag{3.43}$$

were $\Phi(z)$ represents the left side of the area under the normal curve.

Furthermore, because the $W_y - E(W_y)$ points difference between the observed value W_y and expectation $E(W_y)$, the possibles values in Equation 3.42 are $0 > z \leq 1$, which is consistent with the nature of the two directions of two-tiled alternatives. The symmetry of the normal distribution gives $\Phi(-z) = 1 - \Phi(z)$.

The two sided alternative is then,

$$\hat{\alpha} = 2\left\{1 - \Phi\left[\frac{|w_y - 0.5n(N+1)|}{\sqrt{Var(W_y)}}\right]\right\}. \tag{3.44}$$

and the one-tailed alternative is

$$\hat{\alpha} = 1 - \Phi\left[\frac{|w_y - 0.5n(N+1)|}{\sqrt{Var(W_y)}}\right]. \tag{3.45}$$

In presence of tied score -i.e *midranks*- we use the variance presented in Equation 3.37, W^* and the expectation $E(W^*)$ to calculate the significance probability.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

4 Theoretical considerations

A train can be conceived as quasi-metallic cylindrical structure with *state-of-the-art* thermally insulating windows, whose electromagnetic behavior approximates a Faraday cage. As a result, the train's structure imposes high signal penetration loss, naming Vehicular Penetration Loss (VPL), which constitute an immediate challenge on servicing users on board HSTs. There are various propositions for serving railway vehicular users (SCOTT *et al.*, 2013), among which solutions based on mobile cellular network appear as potential candidates. According to that approach, the basic idea consists in deploying a moving relay node (MRN) on the railway vehicle to relay signals from outside or to form a cell on the train(SUI *et al.*, 2013).

The simplified block diagrams in Figure 1 and Figure 2 represent the downlink and uplink inside the HST, respectively; from where contexts with and without on-board repeater can be analyzed. In this work, only downlink will be further explored once considerations on uplink is beyond the scope of this study.

4.1 Predicted stochastic behavior of signal on HSTs

The train carriage frame can be constructed of various materials such as metalized-glass windows, diverse metals and polymer-based materials, often placed in multi-layers. Combinations of those materials yield constructive parameters of the blocking surfaces and, as consequence, the level of penetration loss. In RF transmission phenomena, the decay in amplitude of the incident signal $X_0(\omega)$ is described by the attenuation constant, Equation 2.2, while the phase shift is described Equation 2.3; and both -phase and amplitude phenomena- yield $K_1(\omega)$. Those effects depend on the frequency band of the incident signal, therefore transmission loss is expected to appear differently among services provided via diverse frequency bands, e.g. diverse RAT and sub-bands in the same RAT. Furthermore, the train construction varies, and so does its constructive parameter, which justify the occurrence of different VPL - parameter $K_1(\omega)$- among train types.

Considering the transmission effect and context without repeater in Figure 1, the signal envelope on the HST, via channel $C_1(\omega)$, could be described via log-normal distribution. Equation 2.12. It seem reasonable to consider multipath effect and consequent Rayleigh behavior -Equation 2.14, which could allow it to expect a bimodal distribution. The repeater is equipped with an omni-directional antenna mounted on the train's rooftop, Chapter 5,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

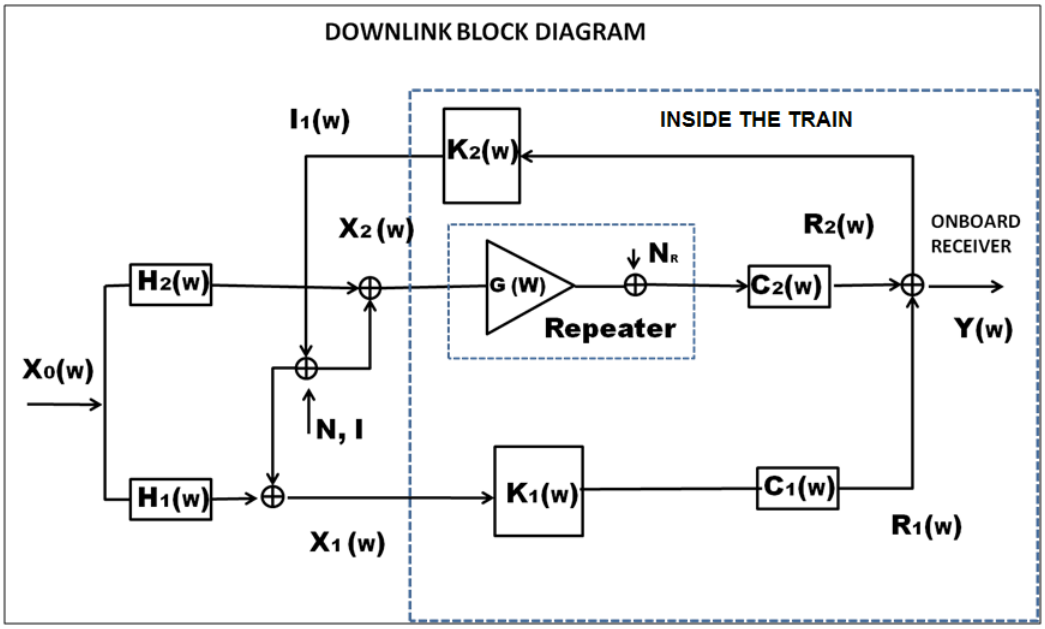


Figure 1 – Simplified representation of the downlink with repeater.

and VPL cancellation and more frequent LoS are immediate improvements. However, large scale fading also characterize the channel $H_2(\omega)$ once large blocking objects remain present. Under repeater situation, amplified signals propagate to the indoor environment by means of a leaky-feeder that maintain LoS with some UEs aboard. Signal interaction with geometric space on the HST results in scattered, diffracted and direct components composing the received signal. Hence, considering no transmission through the body of the train, a reasonable representation of such signal could be given by Equation 2.15 and the envelope of the received signal would be described via Rician distribution, Equation 2.16, or with a Rayligth distribution. Furthermore, in regions with poor coverage, the energy of the incoming signal $X_1(\omega)$ is completely dissipated on the RF transmission process through the train's frame. On the opposite case, when the field strength outside is enough to overcome the VPL, multipath components appear as multiple components of small amplitude. Eventual LoS between UE inside the HST and possible BS along the railway track, yields signals with dominant component for a receiver inside the train. Therefore, received signal $Y(\omega)$ comprises direct components from transmission processes $R_1(\omega)$ and repeated component $R_2(\omega)$. As result, a multi-modal distribution–Rayleigh, Rician and log-normal– is expected.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

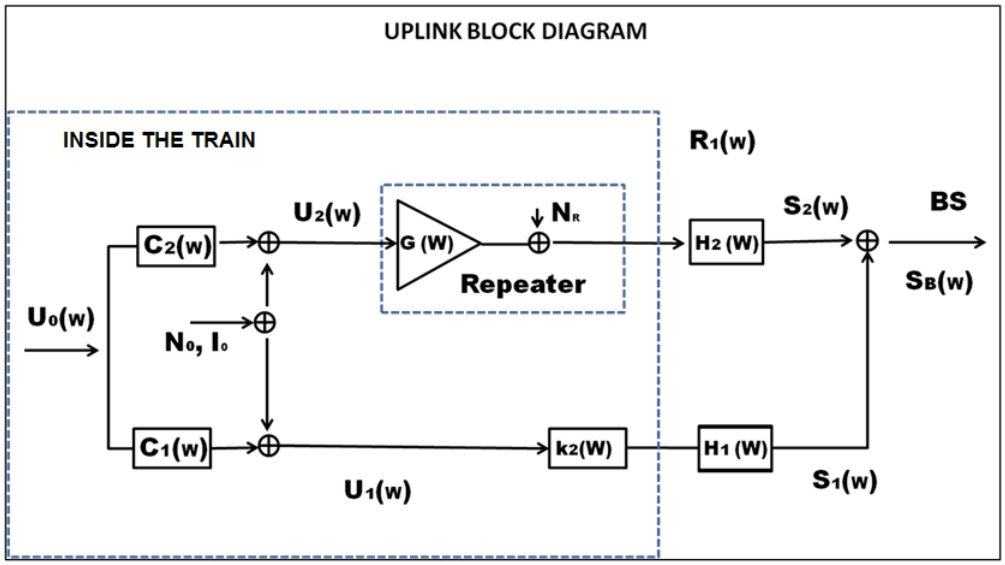


Figure 2 – Simplified representation of the uplink with repeater.

4.2 Mobile network resources for users on HSTs

Mobile cellular network services and demand are random variable in space and time. For instance, a given spatial user distribution points the direction of demand and hence traffic, generally defined in Erlang peer unity of time (minute or hour). The distribution of daily demand concentrates in certain parts of the day namely busy hours, which gives support point to estimate the number of channels per cell for a given user load (YACOUB, 2002).

For a given number of channels and a given area D, a large D implies a larger reuse distance D/R and fewer channel per cell, and more cell per cluster, which means lower traffic density per area. This causes large co-channel distance, leading to high C/I_c (carrier-to-co-channel interference ration) and resulting in better service quality. While, small values of D/R, means more channels per cell, higher spatial traffic density and consequent lower C/I_c. In current Heterogeneous Network (HetNet) data traffic demands predominates and mobile networks are designed to provide high traffic in dense urban areas, while this feature decreases on residential or rural regions. For railway context, this means lower data rate and lower handover frequency in rural regions.

Mobile network resource management entities have to cope with the dynamics of user distribution and consequent traffic (demand). Nonetheless, the commonly used assumption of Gaussian user distribution fails to represent the user behavior in a number of realistic contexts. This include railway corridors, where mobility and terrain characteristics (hilly and mountainous regions) deviate the center of the user distribution from the cell center (KHAN; JAMALIPOUR, 2014).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Despite the usage of modern algorithms to perform network resources management, certain mobility states, such as railway vehicles, can cause sudden changes on user distribution, resulting on overload in a cell region or cluster. For instance, consider hundreds, or even thousands, of passengers inside a HST, the number of active users might suddenly require services in the same coverage area beyond available network resources, resulting in handover failure as commented in (WU; FAN, 2016) and (PHAN *et al.*, 2010).

4.2.1 Influence of handover

Handover performance relies on overlapped coverage area between cells, which is defined by the mobile network radio frequency planning. Network infrastructure also affects the total time needed to conclude a handover process (handover delay). This includes interruption time whose duration depends on network latency and can further increase the handover delay. User experience is affected by this phenomenon during connected state, that is when performing a call, downloading a file or watching a video. Handover events in such conditions increases the probability of data connection failure, given that multimedia and voice season are highly sensitive to handover delay (WU; FAN, 2016)

Some handover processes are more prone to failure. This is the case of inter RAT handover, which requires termination of current connection with the serving BS before establishing a new connection with the target cell. In contrast, in soft handover, the mobile maintain simultaneous connection with previous cell, or cells, until the SNR threshold is achieved. Large scale-fading affects handover measurements and signal strength hysteresis curve between current and target cell, with consequent degradation of hysteresis trigger point. This can cause the mobile to move back to earlier cell shortly after the conclusion of a handover process, which characterize a *ping-poing* phenomenon with consequent increase in handover frequency (TOLSTRUP, 2015).

Those characteristics can affect users on railway vehicles. Intercity HSTs can achieve 100m/s or higher. A number of problems occur if mobile network deployment does not take into account characteristics regarding railway corridors. For instance, given certain speed, a train could cross an overlapping area between two cells very fast, and the short time-window might not be enough to complete the handover process, leading to handover failure. Further, the distribution of RAT along the railway in this case would not only increase the probability of service interruption (LUTU *et al.*, 2016), but also it might increase connection failure rate.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Part III

Methodologies

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

5 Measurement methodology

Reproducing a context of quasi-real usage of mobile communications services inside HSTs constitute the main idea of the measurement design. Therefore, the researches deployed groups of smart-phones inside operational and non-operational trains to perform several measurements. As illustrated in procedures structure in Figure 3, measurements were performed in trains with and without on-board repeater, resulting in two large groups of data sets from where samples from variables were selected. Statistical analysis based on descriptive and hypothetical-inferential methods supports data evaluation. The following subsections present further description of measurement setup and procedures, to apply the statistical methodology presented in Chapter 3.

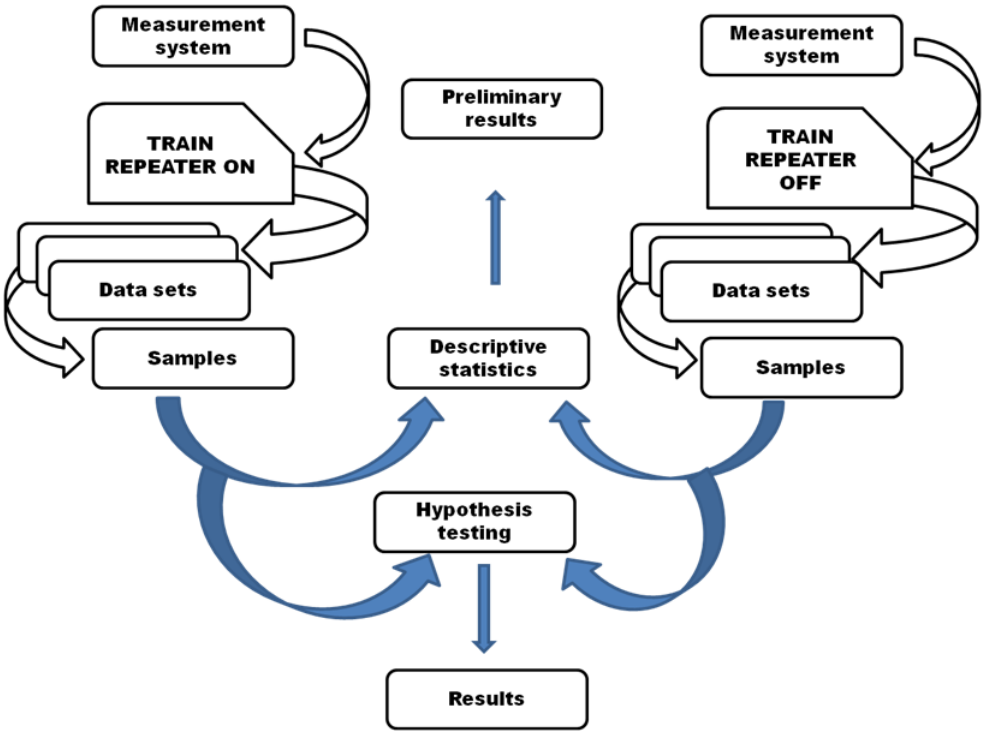


Figure 3 – Procedures structure.

5.1 Routes and sectors

Measurement campaigns were performed in operational trains from Austrian Federal Railways (in German:Österreichische Bundesbahnen) (ÖBB) in Austria from December-2015

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

to May-2016 and concentrated in the following two-way routs: Vienna <>Innsbruck; Vienna<>Salzburg; Vienna<>Graz; Vienna<>Bruck an der Mur. Those routes combine a variety of urban, sub-urban and rural areas, where measurements were performed in contexts with and without on-board repeater. The definition of rural and urban perimeters rely on the *Degree of Urbanization* presented in (European Commission, 2016) and on the list of urban areas in Austria from Organization for Economic Cooperation and Development (OECD)[1]. The resulting geographic limits were defined according following urban and metropolitan regions:

- Wien Large - metropolitan;
- Graz - Metropolitan area;
- Linz - Metropolitan area;
- Salzburg - Medium-sized urban.

Details on the limits rural and urban perimeters, including geographic coordinates, are provided in Section 6.3.1 and Appendix G.

5.2 Assessed solution

The assessed solution to enhance mobile communication on HSTs consists of a wireless wide-band AF-repeater that performs Frequency Division Duplex (FDD) and amplifies the selected signal bands in both directions (uplink and downlink). The repeater has a single omnidirectinal donor antenna, mounted on the rooftop of the train carriage. Service antennas are coaxial leaky cables interconnected, and crossing the carriage roofs longitudinally[2] Repeater systems were installed in two different trains. One repeater system was installed in a Railjet train (Austrian Federal Railways,): a High Speed Train (HST), composed of seven carriages totalizing 210m; it comports up to 400 passengers and achieves a maximum of 230 km/h. Another repeater system was installed in a CityJet train(Austrian Federal Railways,): a regional train comporting up to 259 passengers and achieving a maximum of 160 km/h. The Cityjet train has fixed carriages, without separated locomotive, and with standard length of 75,15m.

[1] For details, visit OECD web page at: <www.oecd.org/gov/regional/measuringurban>
[2] Find more on train types and deployment of the wireless repeater in Appendix F.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.3 Equipment deployment and setup

Data collection equipment – or Measurement Entity (ME) - consists in 18 smart phones, divided in three entities. Each entity consists of six smart phones - Figure 4 image A. Additionally, a single mobile connected to an outdoor antenna provided reasonable reference on the state of the mobile network outside of the train. The mobiles were equipped with dedicated measurement software–Anite's Nemo Walker Air (Keysight Technologies,). To guarantee space-stamped data points, the setup included at least one Global Positioning System (GPS) modem (Figure 4, image C-2) connected to one mobile via Bluetooth, and GPS active antenna (Figure 4, image C-1) boosted GPS signal for mobiles without dedicated GPS modem. These GPS data points has been used as a reference, which can be merged with data of the same measurement run, specially for spacial segmentation as detailed in Section 6.3.1.

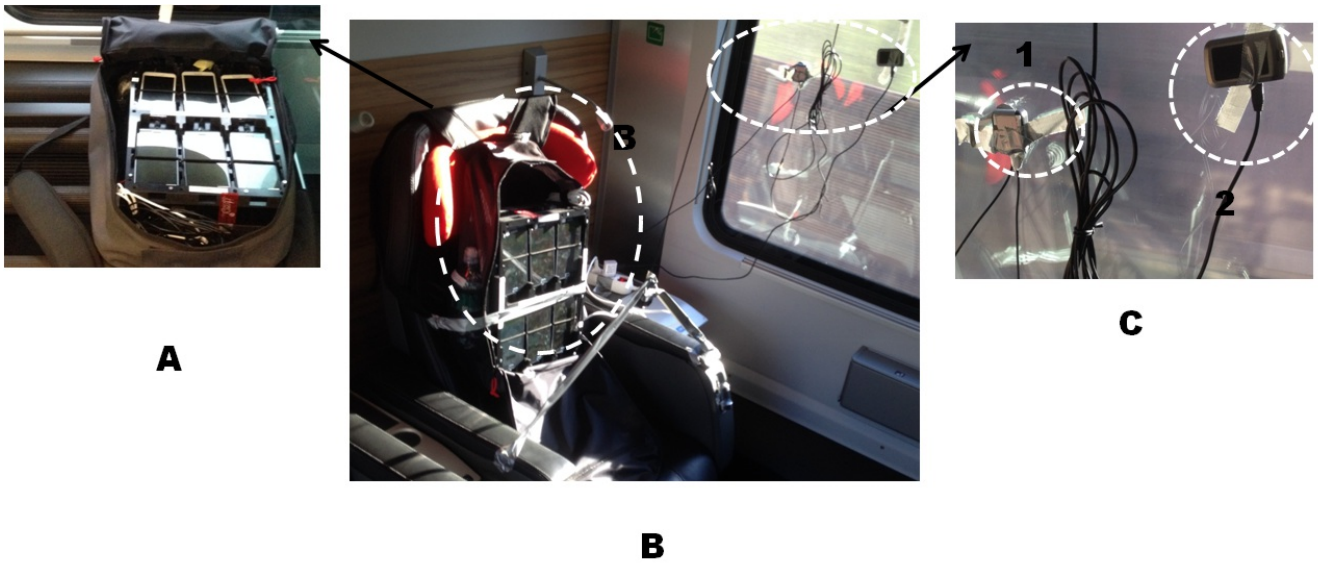


Figure 4 – Measurement equipment entity, image A; entity setup aboard, image B, GPS active antenna image C-1, and GPS modem image C-2.

In a ME, mobiles bear limitless voice and data engineering SIM CARDs from operators A, B, and C. Those devices were divided into three subgroups- of two mobile each, and those subgroups were configured with RAT locking. That is: subgroup one 2G only, subgroup two 2G and 3G, subgroup three no looking- i.e. free RAT choice according to availability. Voice and data session were concomitant events carefully configured to approach a true usage context. Telephone calls, started at a fixed time, were designed to phone a dedicated fixed automatic central telephone system. During data seasons, UEs with 3G and 4G transferred a 10MB-data packet and downloaded 10MB file from a web server in every script run.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Finding satisfactory values for service-related parameters constitute an extra challenge. For instance, parameters such as phone call duration or wait-before retry, attempt time-out, maximum duration, data packet size impact on event granularity and service performance. Figure 5 presents the method applied to perform parameter adjustment which combines preliminary field measurements and laboratory tests.

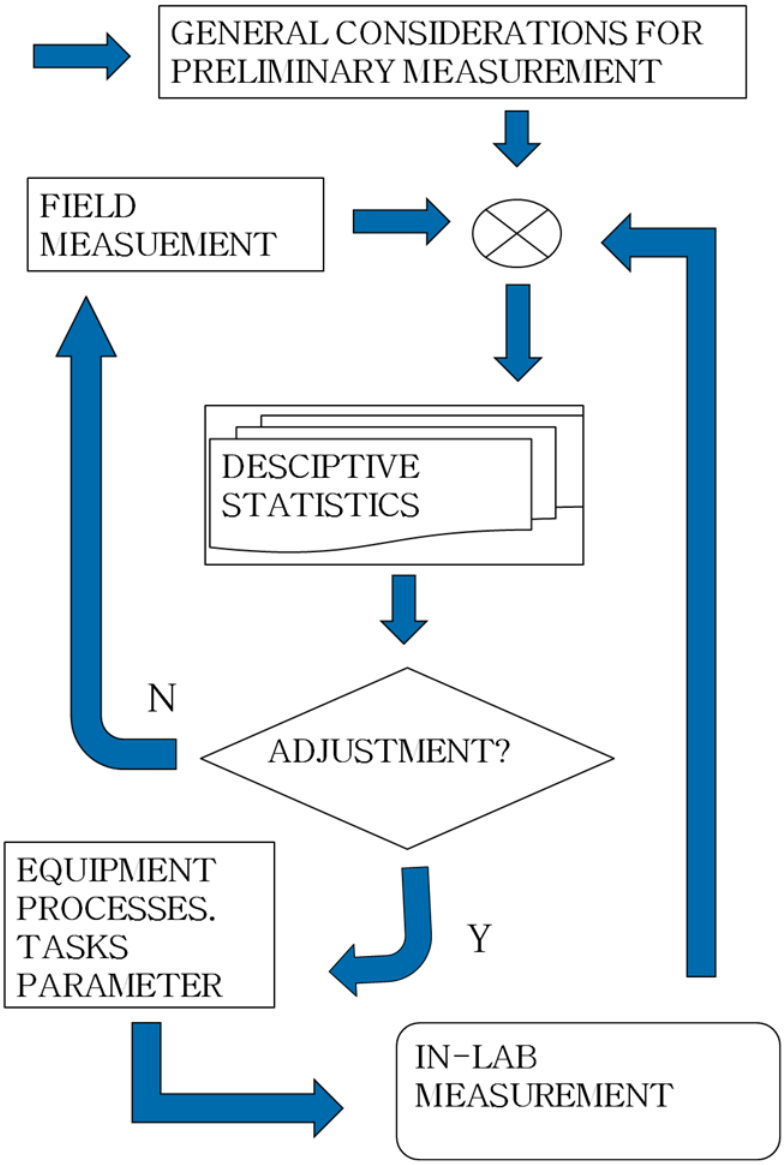


Figure 5 – Measurement ajustment process.

5.4 Row metadata features

Data log files consist in metadata that include various metrics and events from ISO layers. Some of these metrics are from:

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- physical layer - Received Signal Strength Indicator (RSSI), SNR (signal to noise ratio), Received Signal Quality (RSRQ), and Received Symbol Code Power (RSCP);
- network layer - Handover/hand-off attempt;
- service layer - Call attempt (CAA), call connect success, Data connection attempt (DAA), Data connection failed (DAF).

Those log files are generated as text files and can have as much as 200000 lines for a 4-hour measurement, for instance. A row-data file, Figure 6, contains and event ID on the first column, followed a time stamp, systems indicator and variables. Data analysis combines a script-based text processor *AWK*, [3] used to extract data from row-data files, and *RSoftware*[4] for statistical analysis.

```
MACRATE,15:07:10.743,,7,9195992,277,9.75,9.03,0.72,0.00,0.00,249,
MACRATEU,15:07:10.743,,7,281276,179,0.00,0.00,0.00,0.00,0.00,173
RLCRATE,15:07:10.772,,7,3,7364849,716,0.0,3,4,1,7363937,713,0.0,3
PDCPRATED,15:07:10.787,,7,2,8605093,360,3,3,1,8603455,358,33,1638
PDCPRATEU,15:07:10.837,,7,2,160848,187,3,3,1,160704,186,33,144,1,
CELLMEAS,15:07:11.245,,7,0,10,9,0,70003,1575,141,-74.6,-104.6,-11
MIMOMEAS,15:07:11.245,,7,0,2,9,70003,1575,141,0,0,-74.6,-11.3,-10
CI,15:07:11.245,,7,2,11.3,0,2,2,11.3,0,7.4,1
TXPC,15:07:11.245,,7,17.9,18.8,3.8,-6.0,1.0,2,2,137,0.0,1,1.0,2,2
PHRATE,15:07:11.245,,7,1,0,0.0,0
PHRATE,15:07:11.245,,7,2,6102982,0,245,10.2,363928,6102982,0.4,0,
PHRATE,15:07:11.245,,7,3,171307,0
PLAID,15:07:11.245,,7,5,912,22.4,22718,32880,0,7,6,73.1,0,,,,,0.2
CQI,15:07:11.245,,7,8,912,27554765,11,,,,,0,0,2,0,3
PLAIU,15:07:11.245,,7,4,912,1.1,858,3112,14,4,80.0,0,,,0.1,1,1,6,
MACRATE,15:07:11.245,,7,5529456,244,9.84,9.84,0.00,0.00,0.45,221,
```

Figure 6 – Row data file

[3] AWK scripts at Appendix G.
[4] RSoftware on scripts at Appendix H.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6 Statistical Procedures

In this topic it is presented procedures for parametric and a non-parametric tests. Here, not-treated population refers to sample from context without repeater and treated population refers to sample from context with repeater. In addition, all bootstrap re-sample are with replacement unless indicated otherwise.

6.1 Parametric test procedure

The null hypothesis states: **there are no changes between the contexts with and without repeater**. Hence it supports: that samples without repeater X and with repeater Y have no significant stochastic difference; meaning that $M_Y - M_X = 0$. In opposition, the alternative hypothesis states: **there are changes between the contexts with and without repeater**; supporting: $M_Y - M_X \neq 0$. Those statements is mathematically expressed as

$$H_0 : E[X] = E[Y], \tag{6.1}$$

while the two-sided alternative

$$H_1 : E[X] \neq E[Y]. \tag{6.2}$$

This constitutes the pair of hypothesis to be tested.

6.1.1 Data conditioning and parameter estimation

An one-way measurement sample typically varies from eight to fifteen thousand scores and the first step consist in subsampling from the main dataset to obtain equal sample sizes. That is, the smaller sample size between samples $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_m)$, named z, is used as the new sample size for both sets. Sub-sampling from both data sets yields a new pair of $(X_1^*, \ldots, X_z^*)$ and $(Y_1^*, \ldots, Y_z^*)$. Data conditioning for non-parametric test includes *TUKEY's outlier labeling* based on inter-quantile range (IQR).

To perform parameter estimation: first it is bootstrapped 5000 subsamples, each containing 100 scores; then the mean and variance are estimated as described in Section 3.3.2.1. Those values are used to estimate confidence intervals for the mean and F-value–variance ratio.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.1.2 Empirical critical value and t-statistic

In order to estimate the critical value, 10000 bootstrap subsamples, each with 100 scores are taken from the pair X_z^* and Y_z^*, then t-values are calculated using Equation 3.20 from Section 3.3.2.3, repeated here for convenience

$$t_b^* = \frac{(M_{Y_b^*} - M_{X_b^*}) - (\hat{\mu}_y - \hat{\mu}_x)}{\tau_b^*}. \tag{6.3}$$

This process yield a vector with ten-thousands t-values whose ECDF exibit the critical value at the quantile $q(0.25)$ for $\alpha = 0.25$, as described in Section 3.3.2.3. To calculate t-statics, bootstrapped subsamples are taken from the pair $(X_1^*, \ldots, X_z^*)$ and $(Y_1^*, \ldots, Y_z^*)$ and Equation 6.4 provides the t-statistic:

$$T = \frac{(M_{Y^*} - M_{X^*})}{\tau}. \tag{6.4}$$

6.1.3 Estimated p-value

The p-value, noted as $Pr > |t|$, define as the proportion of the null distribution that lies outside the region defined by the module of the critical value - i.e. left and right tails on the curve defined by the null distribution. As a consequence, the t-value, in Equation 6.4, has non-zero probability of being smaller than the critical value in this region if, and only if, the null and alternative distributions have an overlapped region.
To estimate such a region, 10000 subsample are taken from the pair X_z^* and Y_z^* via bootstrap to produce t-values via Equation 6.4. Those t-values are compared with the critical value as

$$p - value = 2 * [true(|t_{critic}^*| > t_{calc}^*)]/B, \tag{6.5}$$

where *true* denotes the number of times that $|t_{critic}^*| > t_{calc}^*$ and B is the number of iterations which coincide with the number of bootstraps $B = 10000$. The multiplication by two is required for the two-sided alternative[1].

6.1.4 Sample size

The samples size is fixed in 100 scores unless declared otherwise, once sample with this size satisfy the Central Limit Theorem (CLT).

[1] A one-sided p-value do not require the absolute value of t_{critic}^* which gives greater or less for each direction of the two-tailed hypothesis.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.1.5 Estimated effect size

It is used the *Glass-delta effect size*, Equation 6.6, estimated with samples from context without repeater influence–i.e. the pre-treatment standard deviation, SD_{OFF}, as recommended in (LAKENS, 2013). Hence, the effect size is a measure of the treatment effect compared to the standard deviation of the not-treated population.

$$\Delta_{Glass} = \frac{|M_Y - M_X|}{SD_{OFF}} \tag{6.6}$$

6.2 Wilcoxon sum-rank test procedures

In the present procedure, samples from context without repeater X are called the control group, represented by index w_x, W_X or U_x. The samples from context with repeater Y are named treatment group w_y, W_Y or U_y. The pre-defined power and significance level are $\beta - 1 = 0.98$ and the two-tailed $\alpha = 0.05$, respectively.

6.2.1 The Null hypothesis

The null postulates that there is no significant difference between the non-treatment group X and treatment group Y, from contexts with and without repeater, respectively. Considering the ECDF of X and Y as
$F_X(x)$ and $F_Y(y)$, respectively; the null hypothesis is mathematically denoted as

$$H_0 : F_X(x) = F_Y(y). \tag{6.7}$$

On the opposite side, the alternative hypothesis supports that the distributions differ, that is

$$H_1 : F_X \neq F_Y. \tag{6.8}$$

Considering that the null is true, subsamples, m and n, of equal sizes are taken from data set X and Y, respectively. Those subsamples are combined to calculate the rank vector as defined in Equations 3.31 and 3.38, respectively. Then, Wilcoxon rank sum test is computed and the significance probability is calculated via asymptotic normality as in Equation 3.43. Given the size of the main sample set–around fifteen thousand scores, the minimum estimated sample size between the pair of samples, X and Y, is taken to calculate the test. This process is repeated a hundred times in a repeated hypotheses approach.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.2.2 Non-parametric effect size

Treatment effect or non-parametric effect size, also named treatment effect, is usually defined as part of the experimental design and is connected with the desired power or significance level to be achieved. Once the experiment have already been performed, the effect size is calculated using the proportion U_y/mn (WALTERS, 2004), where

$$U_y/mn = Pr(X < Y)$$

which implies

$$U_x/mn = Pr(X > Y).$$

The plot of non-parametric effect size presents

$$Pr(X > Y) = 1 - Pr(X < Y).$$

The U_y/mn value is calculated in a hundred iterations, yielding a vector with size B=100 from which the confidence interval for non-parametric effect size is estimated.

6.2.3 Sample size estimation

The sample size is calculated using the non-parametric effect size $P_{Noether} = U_x/nm$ and formula to calculate sample size non-normal test Equation 6.9 provide by (WALTERS, 2004) and (NOETHER, 1987). First, an exploratory two-tailed U-test is performed on a vector containing a hundred of pairs of bootstrapped subsamples with $m = 100$ for set X and $n = 100$ for set Y; significance level 0.05. Then, from each value of the statistics U_x it is calculated the proportion U_x/nm and the mean of the 100-points long vector is used as the estimated proportion $P_{Noether}$ in Equation 6.9.

$$N = (z_{1-\alpha/2} + z_{1-\beta})^2/[6 * (P_{Noether} - 0.5)^2] \tag{6.9}$$

The sample size is dynamically estimated for each analysis section. When the treatment effect is really small -i.e. near 0.5- the required sample size is around few thousands which is impractical to calculate the test. Hence, when the estimated $P_{Noether} < 0.6$, the test is performed with $m = n = 100$.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.2.4 Critical value

Significance probability in Section 3.3.4.1 relies on asymptotic normality and the p-value, Equation 3.43, is the lower or upper portion of the area under the normal distribution. For the present analysis, the significance probability is measured with treatment rank sum W_Y, hence the null hypothesis is rejected based on $\alpha > \hat{\alpha}$ with $\hat{\alpha}$
given by Equation 3.44. The multiple tests approach produces a vector with 100 p-values, referring to a hundred of equal hypothesis and p-value adjustment is performed according to Section 3.3.1.

6.3 Data selection

Measurements produces large amount of sample groups, from which data is selected in diverse manner to allow specific analyses. The following lines contain detailed descriptions of methods used on those processes.

6.3.1 Spacial analysis

To perform spatial evaluation, urban and rural areas on railway corridors in from Vienna to Salzburg are classified with support of *Google maps*, and merging geographic coordinates from mobile with external antenna, Sec. 5.3 into the desired data file. First, it is fined the coordinates from referential point in Vienna Main Station (Wien Hbf) and then the coordinates for next train station on the track, that is *Sankt Pölten* main train statin (St.Pölten Hbf). This Station is defined as the limit of Vienna's urban area, once *Sankt Pölten* is not on the list of urban areas in Section 6.3.1.

The rural region that begins at (St.Pölten Hbf) finishes at the main train station in the city *Sankt Valentin* (St.Valentin Bahnhof), which also constitute the limit for the second urban area that finishes at Salzburg main train station (Salzburg Hbf). Rural section (A-B) appear in Figure 7. A number of 120 data-ponits (60 seconds) is used as transition region between urban and rural areas which is equivalent to approximately 2km for a average speed of 120km/h.

Having the coordinates for the limits of urban region, coordinates from the GPS file that approximate those points are found. Assuming that ME where roughly at the same point in space at a given the time, it possible to combine the time-stamp from GPS coordinate from one file with time in target file. This allows transferring coordinate to the target file, resulting in coordinates (Latitude., Longitude.) that defines the urban and rural perimeters as follows:

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

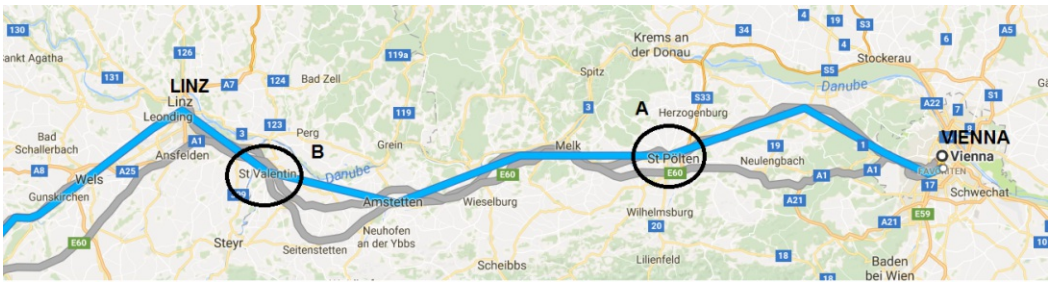


Figure 7 – Rural section (A-B) on railway track - Vienna to Salzburg.

- Vienna Main Station - (48.1666901, 16.407881);
- Vienna urban limit (St.Pölten Hbf) - (48.207881, 15.624819);
- Rural limit (Sankt Valentin) -(48.178786, 14.522511);
- End of track (Salzburg Hbf) - (47.813017, 13.045047).

6.3.2 Diverse RAT

In order to select data for RAT analyses, band filter is included, that is a single band in one RAT compared with single band in another. It is extracted and compared data from GSM band 900MHz and UMTS band 2100MHz.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

7 Results and discussion

In the following subsections it is presented results to evaluate improvements on RF signal level–i.e. physical layer of OSI model. Improvements on this level imply consequences on higher layers. Analyses are based on downlink signal strength– i.e. received signal- inside a HST using the variables RSSI, RSCP and Received Signal Reference Power (RSRP) (Third Generation Partnership-3GPP, 2005) whenever convenient. Those variable are presented in dBm in the row data file. In the present statistical analysis, they are treat as dimensionless number unless indicated otherwise.

7.1 Repeater effects on signal levels on HSTs

Radio frequency planning determines coverage characteristics along railway corridors, which controls the statistical behavior of signal inside trains. In case of repeated signal, the repeater's donor antenna appears as critical limit in case where signal from outdoors cannot overcome VPL. In Figure 9, histogram A, for instance, scores below -110dBm and above -90dBm represent situations with low and sufficient RF signal coverage; respectively. Those differences impact the distribution of signal quality received by an UE on board in contexts with and without repeater.

Table 1 – Elementary statistics for UMTS, variable RSCP.

Repeater	Min.	1st Q.[1]	Median	Mean	SE_M[2]	3rd Q.[1]	Max.	N
OFF	-131.50	-94.40	-83.20	-84.37	0.11	-73.80	-49.80	16645
ON	-129.50	-79.50	-73.70	-76.03	0.08	-69.80	-43.10	16645

[1] Q stands for quantile.
[2] SE_M -standar error from the mean.

There are areas with weak and strong signal along a railway track, which impacts on signal distribution under repeater condition. Despite the benefits from the donor antenna position, high rate of variation on the geometric path between the BS and repeater's donor antenna, causes shadowing as consequence of blockages. This occurs even in regions that allows LoS most of the time, as described in Sec. 4.1. In some portions of railway tracks, mobile network deployment is such that even the repeater antenna experiences low signal levels. Those conditions partially explain lower scores on the signal statistical distribution under repeater condition in Figure 9-histogram B.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

On the opposite side, there are occasions in which there is LoS between the repeater's donor antenna and the BS, as well as between mobile phones on board and the BSs deployed near the railway tracks. While the former occurs even when the railway line crosses a region with weak signal, the later is much likely in places with signal strong enough to overcome VPL–e.g. typically in urban regions. This partly explains high signal scores that populate the upper side of the signal strength distribution in Figure 9-histogram B.

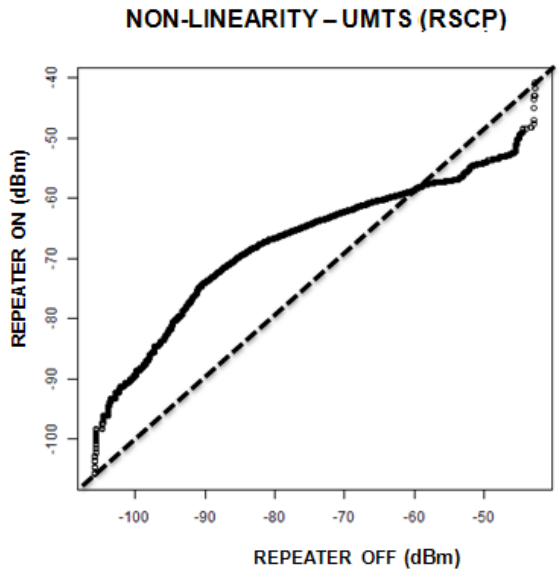


Figure 8 – Non-linear behavior under repeater effect.

Comparing histograms A - no repeater- with B-repeater- in Figure 9, it is seen a stronger amplification effect between the minimum value and the first quantile - i.e. signal strength well below -90dBm, compared to that between the second and third quantiles - signal strength above -70dBm; for example. Furthermore, the non-linearity of the shift between repeated and non-repeated signals also appear when the shift calculated as ($quatile_{ON} - quantile_{OFF}$). In Table 1, those values are, in approximate figures, 15dB, 9dB and 4dB for first, second and third quantiles; respectively. This non linearity appear more clearly on quantile-plot in Figure 8.

In terms of repeater performance, weaker signal components receive maximum available power (energy) while amplification of stronger signal components is bounded by the repeater's maximum output power. This fact, combined with repeater non-linear behavior, provides a partial explanation to the non-linear shift. Therefore, a strong repeater effect in areas with high quality signal yields signal with left-skewed and right-bounded distribution, displaying a large non-linear shift in comparison with non-repeated signal distribution.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

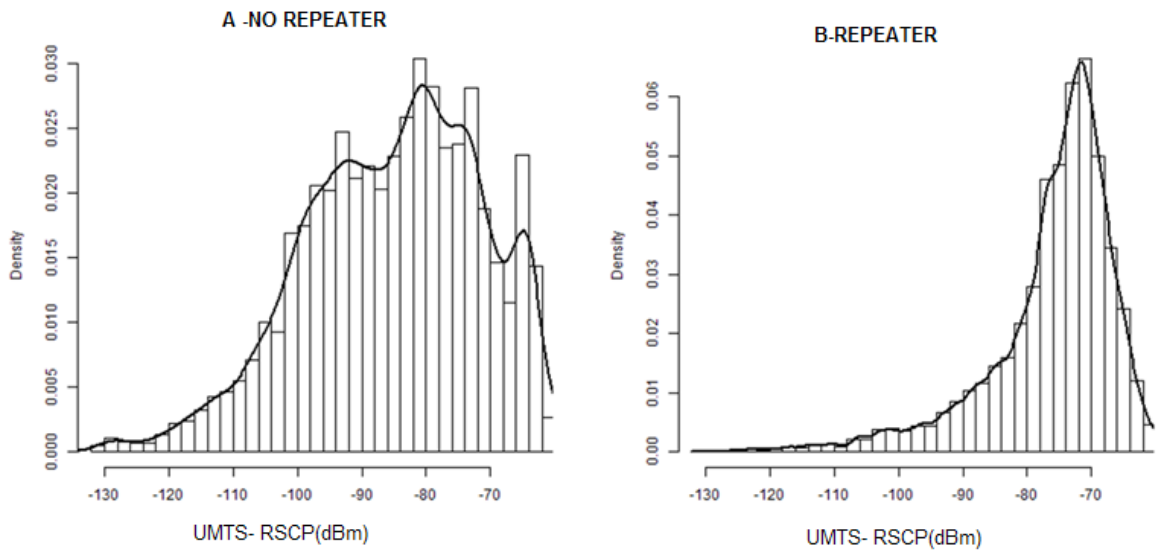


Figure 9 – Histograms with equal sample size, N=16645; comparing contexts.

Further effects on repeated signal levels appear in Table 2, where the amplification and power output limit cause the estimated variance to change from about 206 on the direct link to 98 on repeated link. This indicates a more stable signal level under repeater effect. In addition, the sample from the repeated link has significant departure from the normal distribution compared to the direct link. This becomes clear observing the respective change between contexts, which appear in form of decrease in symmetry with skewness measuring -0.28 and -1.51; and thicker tail with kurtosis measuring 2.80 and 6.17; Table 2.

Table 2 – Measures of spread and shape

Repeater	SD	Var.	Skewness[2]	Kurtosis[2]	Outliers/N [1]	N
OFF	14.35	206.00	-0.28	2.80	0.50%	16645
ON	9.91	98.22	-1.51	6.17	6.16%	16645

[1] Measured in percentages: Outliers/N.
[2] Values for normal distribution: kurtosis=3, skewness=0.

Describing the characteristics of the change under repeater condition raise an important question on how this shift is quantified and what degree of statistical significance would be attributed to it. A classical parametric *Student's* t-test is unsuitable in that case. It was seen that the resulting distribution under repeater condition strongly deviates from the non-repeater case and also from normal distribution as consequence. Furthermore, there are cases in which the repeater causes larger change in variance which not necessarily reflects in the measure $\Delta M = M_y - M_x$ [1]. This opens the question on whether the mean is the most

[1] M_y and M_x are the estimated mean for context with and without repeater, respectively; Sec. 3.3.2

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

appropriate statistical measure to quantify the observed distributional change in this case.

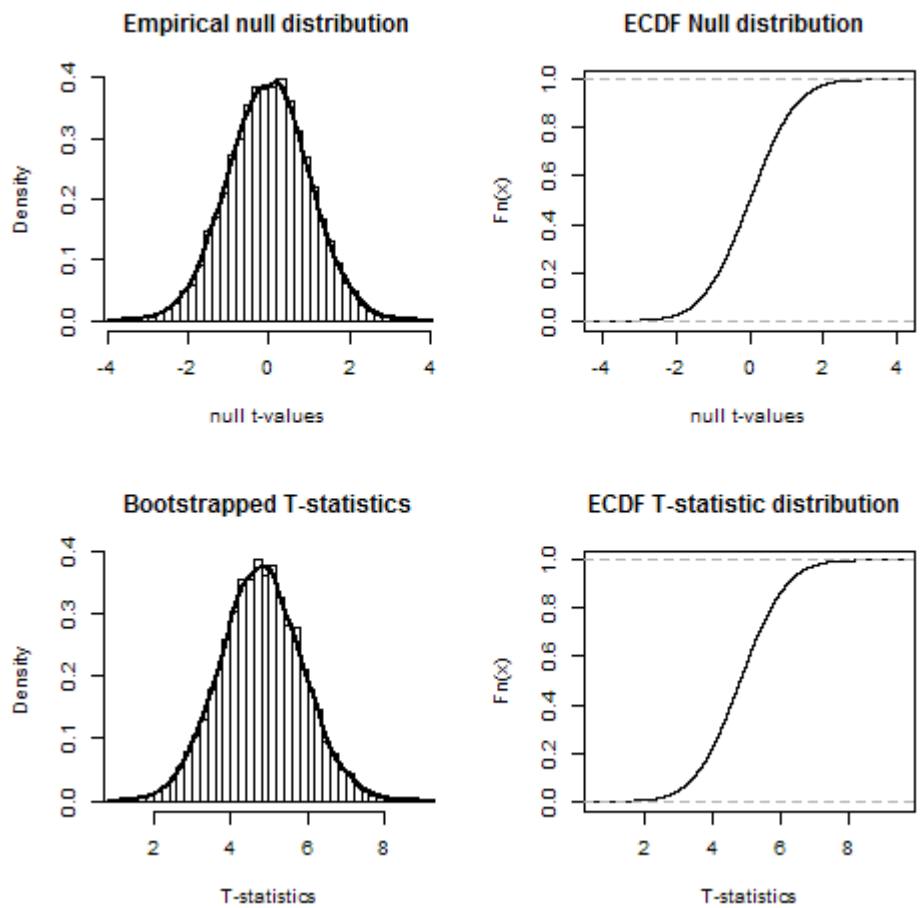


Figure 10 – Empirical t-test and two-sided alternative. Technology UMTS, variable RSCP.

Table 3 – Empirical t-test. Technology UMTS, variable RSCP

α	t.statistic	t-critic	$Pr > \|t\|$	Effect size	F-values	N	B
0.05	5.64	1.65	0.001	$[0.34, 0.78]^1$	$[1.28, 3.36]^1$	100	10000

[1] Quantiles .025 and .975, respectively.

A partial solution to this limitation is the empirical t-test, presented in Sec. 3.3.2.3, which is applied to the data using the procedure in Sec. 6.1. In Table 3, the t-statistics 5.64 quantifies the stochastic change between the empirical null distribution and the alternative distribution (bootstrapped t-statistics) both presented in Figure 10.

In addition, only a small portion from the empirical null distribution, a thin tail from t-critic 2—4, overlaps the alternative represented by the bootstrapped t-statistic. This means that the difference in signal strength between contexts with and without repeater is

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

very unlikely to be by chance. Therefore, this constitutes a strong evidence that supports the alternative hypothesis. In other words, outage in signal coverage is much less likely to happen under repeater condition.

Nonetheless, those statements require further assessment. First, outliers treatment causes the removal of 6.16%, or 1025 scores, from repeater's sample which decrease the tail in Figure 9 histogram A. Outliers removal also diminish the overlapped region between distributions, increases significance of the t-test. Second, the F-value, in Table 3, shows that variances can differ as much as 3.36 times. Furthermore, the value $(M_y - M_x)$ given via MLE in Equation 3.12, does not represent the true mean difference between distributions of direct link and repeated signal. This is because the scores result from the *decibel transformation* (i.e. are given in dB) and the distributions are unknown. Those two points undermine the accuracy of a t-test and, as result, the t-test does not provide a reliable measure of the shift; as considered in Sec. 3.3.2.4.

In order to overcome those limitations related to the t-test, the non-parametric Wilcoxon rank-sum test described in Sec. 3.3.4 is applied to the same pair of samples. The null and alternative hypotheses $H_0 : F_X = F_Y$ and $H_1 : F_X \neq F_Y$, respectively, as described in Sec. 6.2. The results obtained with the rank-sum test agree with those of the t-test once both support the alternative. They differ in their significances and measures of the effect, however. In Table 4, the confidence interval for non-parametric effect size U_y/mn has minimum value of 0.61, an illustration of that measure appear in Figure 11 represented by $P(X > Y) < 0.40$. Those values not only reinforce the alternative $H_1 : F_X \neq F_Y$, but also further support the one-sided alternative $H_1 : F_Y > F_X$.

In terms of significance, the value $Pr > |t| = 0.001$ in Table 3 indicates that it is unlikely that a context without repeater outranks the one with repeater in practical terms. By contrast, the *p-value max* of 0.04 consist on a comparable value from the non-parametric test and it indicates that there are cases in which the no-repeater context equals or surpasses the one with repeater as visualized in Figure 8. In other words, the non-parametric test reveals that there is a likelihood of 4% that a random sample from F_Y do not reject the null hypothesis. That agrees with theoretical expectations and preliminary findings.

Table 4 – Results for Wilcoxon sum-rank test

Rejection(%)[1]	p-value max.	U_x max.	U_y max.	(U_y/mn)	N	B
1.00	0.04	1839	2785	[0.61, 0.78][2]	68	100

[1] Percentage at which the null is rejected.
[2] Quantiles 0.025 and .975, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

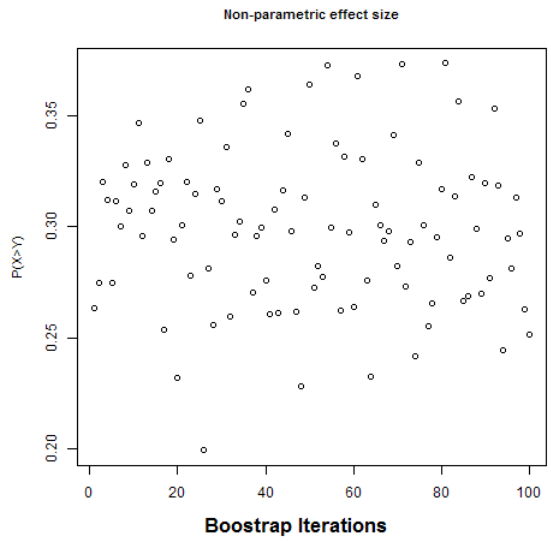


Figure 11 – Non-parametric efffect-size, proportion P(X>Y).

7.2 Diverse mobile technology

Diverse frequency bands imply diverse path loss and VPL characteristics, resulting in unequal repeater effect perceived by receivers on HSTs using different RAT, as commented in Chapter 4. In order to estimate those differences under repeater condition, we compare data from GSM (band 900MHz) with UMTS (band 2100MHz)[2].

In Figure 12, histograms display comparison of repeater and no-repeater contexts which shows the dissimilar distributional behavior of signal strength between mobile technologies. Direct link UMTS signal presents more variation, with variance of 161.2 compared to 97.4 for GSM[3]. Whereas under repeated link, UMTS variance diminishes to 78.1 compared to 60.0 for GSM. This change is more evident in Table 5, where maximum figures for F-values are about 2.3 and 3 for GSM and UMTS, respectively. This is a clear indication that treatment effect yields different degree of improvements in terms of signal regional stability for both technologies. In terms of normality resemblance, GSM signal distribution have no significant departure from the normal, resulting from repeater effect, with approximate values of 0 and 3 for skewness and kurtosis; compared to about -1.3 and 5 for UMTS.

Parametric and the non-parametric tests give similar significances for shift in both RAT. The t-test, Table 5, shows t-statistics is around 7 for both cases. In contrast, the *Glass* effect size points to a noticeable difference once confidence intervals for this figure are about [0.7, 1.3] and [0.5, 1] for GSM and UMTS, respectively. An average distance of 0.25, roughly.

[2] Consult Appendix D Sec. D.2 for details.
[3] Consult Appendix C for complete statistics on scale and spread.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

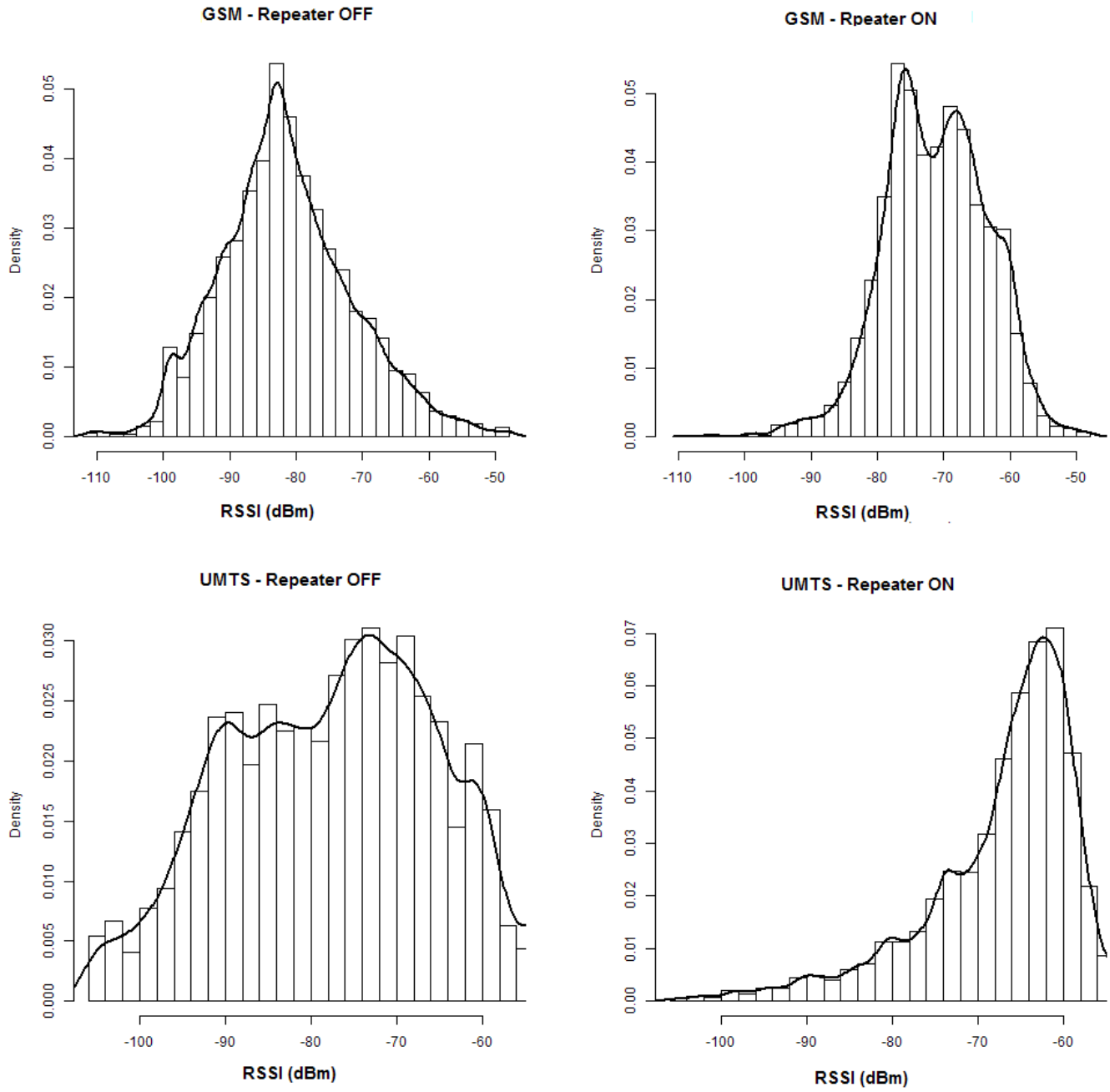


Figure 12 – Signal strength distributions for GSM (900MHz) 16094 scores and UMTS (2100MHz) 16755.

This result further demonstrates the limitations of t-test in this context.

Nonetheless, non-parametric test contrasts with t-test and favor GSM. Results in Table 6 indicate difference in performance under repeater condition between RATs; given that the maximum proportion $P(Y > X)$, represented by U_y/mn, measures about 0.9 for GSM and 0.8 for UMTS. The maximum p-value of 0.24 for UMTS, compared to 0.03 for GSM,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 5 – T-test- table. Variable RSSI

GSM							
α	t.statistic	t-critic	$Pr > \|t\|$	Effect size	F-values	N	B
0.05	6.99	1.61	0.001	[0.76, 1.34][1]	[1.08, 2.30][1]	100	10000
UMTS							
α	t.statistic	t-critic	$Pr > \|t\|$	Effect size	F-values	N	B
0.05	7.21	1.62	0.001	[0.53, 0.99][1]	[1.34, 3.10][1]	100	10000

[1] Quantiles .025 and .95, respectively.

expresses higher likelihood to commit type I error with samples from UMTS. Rejection rate of 98%, express no rejection of the null hypothesis in 2% of the cases, wheres result for GSM rejects the null in 100% of the cases. In practical terms, in some cases on board repeater provides no-improvement for UMTS or it might even degrade signal quality. This result supports that repeater's performance differ under diverse RAT. Larger improvement in form of signal strength stability for UMTS condition without repeater, which is partly justified given the difference in path loss and VPL for 900MHz and 2100MHz, as well as diverse strategy for RAT deployment.

Table 6 – Results for Wilcoxon sum-rank test. Variable RSSI.

GSM						
Rejection(%)[1]	p-value max.	U_x max.	U_y max.	(U_y/mn)	N	B
1.00	0.03	263.50	520.50	[0.69, 0.90][2]	28	100
UMTS						
Rejection(%)[1]	p-value max.	U_x max.	U_y max.	(U_y/mn)	N	B
0.98	0.24	1036.50	1364.50	[0.64, 0.81][2]	49	100

[1] Percentage at which the null is rejected.
[2] Quantiles 0.025 and .975, respectively.

7.3 Rural and urban contexts

Changes in geography between urban and rural areas raised the question, in Chapter 1, on how those changes impact the link quality inside HSTs under influence of an AF repeater. User concentration in urban regions encourages better RF planning for mobile networks and consequent lower service outage probability compared to rural areas. Therefore, characteristics of signal strength distribution on HSTs in those contexts are expected to differ. Aiming to verify that change, it is defined rural and urban regions, dividing the railway track and respective dataset as described in Sec. 6.3.1.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

As commented in Sec. 7.1, areas where signal strength remains much lower than maximum repeater's output power most of the time, generally characterize poor signal coverage. In a opposite situation–i.e regions with satisfactory signal coverage–context without repeater outperforms the one with repeater, decreasing repeater's performance. This is observed in quantile-plots (urban regions) in Figure 24 in which amplification is unnecessary for RSSI above -80dBm; as opposed to the rural region where repeater improves coverage onboard nearly at all time.

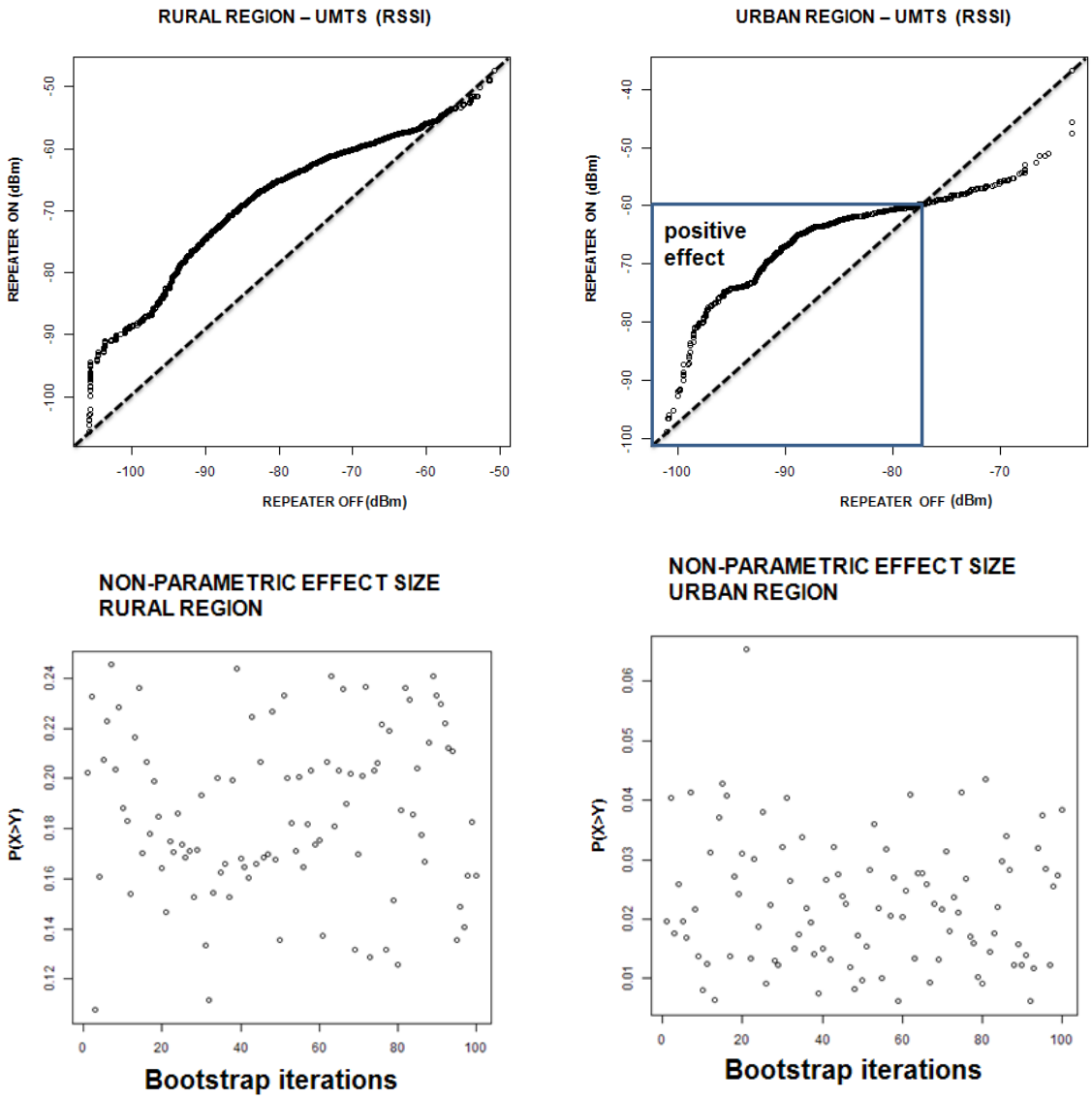


Figure 13 – Comparison of signal strength inside the HST: rural vs. urban areas - UMTS (RSSI).

Large difference between signal strength and maximum repeater's power output causes considerable change in variance. Indeed, estimated variance changes from 118.6 to 81.2 in

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

rural region and from of 45.8 to 39.5 in urban region as result of amplification [4].

Table 7 gives F-value interval of [1.02, 2.12], meaning a larger change in variance in rural region compared to [1.01, 1.74] for urban area. Difference in stability between those regions remain after repeater effect, which appear in form of dispersion of the non-parametric effect size in Figure 24; where upper limit of $P(X > Y) < 0.26$ and $P(X > Y) < 0.07$ for rural and urban regions, respectively. In addition, the effect size, Table 7, can be almost 4 in urban region compared to a maximum of 1.4 in rural areas. Those results point to a better performance on urban region, however a comparison via t-statistics cannot provide accurate measure relative performance in this context.

A more accurate comparison via non-parametric evaluation shows smaller differences. Comparing direct and repeated link, median changes from -88dBm to -72dBm in rural region and from -92 dBm to -70dBm in urban region. This reflects on maximum non-parametric effect size, Table 8, which can be as much as 0.9 in rural region compared to 1 in urban area. Also, both areas has 100% rejection rate, despite divergence in maximum significance, p-value<0.001, urban compared to p-value<0.03; meaning that samples from a rural a region is much likely to fail to support the alternative. In practical terms, this means a better signal strength onboard under repeater condition in urban area, given that coverage outdoor tends to be better in this region. In addition, non-parametric results suggest comparable performance in both areas rather then favoring a better repeater performance in a specific region.

Table 7 – T-test results for UMTS in rural and urban areas; variable RSCP.

RURAL							
α	t.statistic	t-critic	$Pr > \|t\|$	Effect size	F-values	N	B
0.05	9.54	1.60	0.00	[0.89, 1.43][1]	[1.03, 2.15][1]	100	10000
URBAN							
α	t.statistic	t-critic	$Pr > \|t\|$	Effect size	F-values	N	B
0.05	22.89	1.58	0.001	[2.58, 3.85][1]	[1.01, 1.74][1]	100	10000

[1] Quantiles .025 and .95, respectively.

[4] Table with more statistics in Appendix C.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 8 – Wilcoxon sum-rank test results for UMTS in rural and urban areas; variable RSCP.

RURAL						
Rejection(%)[1]	p-value max.	U_x max.	U_y max.	(U_y/mn)	N	B
1.00	0.03	256.50	527.50	[0.70, 0.91][2]	28	100
URBAN						
Rejection(%)[1]	p-value max.	U_x max.	U_y max.	(U_y/mn)	N	B
1.00	<0.001	21.00	148.00	[0.90 1.00][2]	13	100

[1] Percentage at which the null is rejected.
[2] Quantiles 0.025 and .975, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Conclusion

Mobile services inside railway transports have been in focus of many research groups in this decade. The majority of those works rely on simulations, as result of system complexities, as well as regulatory, financial and corporate constraints involved in measurements inside HSTs. Despite of its importance, measurement-based studies, such as (LUTU *et al.*, 2016), are rare. In the present study, it performed an assessment of RF downlink signal quality of GSM, UMTS and LTE on a HST. To do that, *ÖBB-Technische Services* deployed an AF RF repeater in Railjet and Cityjet trains and the *Institute of Telecommunications* from the TU Wien performed measurements under quasi-real usage conditions in various spatial contexts and investigate data via parametric and non-parametric hypothesis test. It has been found that the repeater effect yield a distribution with strong non-Gaussian behavior, therefore its statistical significance cannot be measured solely via mean shift and *Student* t-test. In addition, the on-board repeater yield signal improvements in both urban and rural areas, with change in performance under diverse RAT.

Data evaluation shows that the signal strength distribution inside the HSTs presents Gaussian behavior with small departure from the normal, which agrees with expectations in Sec. 4.1. The resulting distribution of signal strength under repeater condition has an uneven increase in scores with larger changes in lower quantiles, which does not totally appear in the mean before and after repeater influence. In addition, the repeater causes a diminution in variance which appears more significantly when the signal distribution without repeater has a large portion of scores much lower than repeater output power. Moreover, repeater effect yields a left-tailed and right-bounded distribution with departure from the normal distribution, an indication that a measure of performance cannot be given solely via change in the mean.

From a methodological perspective, a comparison of parametric and non-parametric tests demonstrates that empirical t-test indicates the direction of improvements, however it does not accurately quantify the change between samples using mean difference (FENG *et al.*, 2013), given the resulting non-Gaussian distribution under repeater condition. Furthermore, large difference in variances reduces parametric effect size and undermines the significance of t-test. Most importantly, the combination of this test with extensive bootstrap and non-parametric rank-sum test have been proved to be efficient in detecting and measuring stochastic changes in RF signal strength inside trains.

Results confirm improvements on RF (downlink) converge inside the train under re-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

peater condition, except for cases where signal strength remains above $-80dBm$ most of the time, more common in urban regions. A general reduction in variance observed under repeater condition represents improvement on signal stability, meaning signal in optimal condition most of the time. This allows saying that an user on a HST is much unlikely to experience signal outage in context with repeater.

Further, non-parametric evaluation supports the existence of diverse performance under repeater among RAT, which is confirmed via rank-test in which improvements in GSM (band 900MHz) surpasses the one in UMTS (band 2100MHz). Change in attenuation resulting from change in frequency is consistent with theory in Chapter 2 and expectations in Sec. 4.1. This means that the repeater produces unequal improvement for different RAT.

Although improvements on in-train mobile communications via repeater is proven statistically, some further consideration on the results appear to be necessary. First, repeater configuration influences improvements detected via measurement and having details of configuration for each measurement allows more precise measure of performance. Second, a non-linear shift under repeated link raises question on what would be the most appropriate statistics to measure improvement: central tendency -median or mean- or spread -variance. Finally, traceable relationship among row data sets and derived data files, its respective track and measurement condition would be an advantageous during data analysis. The concepts used for measurement and the proposed statistic methodology can be used into different transportation-e.g. buses, car, etc; diverse geographic areas-e.g. other countries. The statistic method can also be improved to convey higher layers of the communication network protocol stuck. Those points would be useful for future work.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

References

Third Generation Partnership (3GPP). *Technical specification TS25.215*. 2011. Avalable at: <http://www.3gpp.org/ftp/Specs/archive/>. Cited on page 57

AGGARWAL, C. C. An introduction to outlier analysis. In: . 2nd ed. 2017. ed. [S.l.]: Springer International Publishing, 2017, (Outlier Analysis). ISBN 978-3-319-47577-6. DOI: 10.1007/978-3-319-47578-3. Not cited in the text.

Austrian Federal Railways. *Railjet and Cityjet trains*. Avalable at: <www.oebb.at>. Cited on page 47

BAGDONAVICIUS, V.; KRUOPIS, J.; NIKULIN, M. *Nonparametric Tests for Complete Data*. 1. aufl.. ed. Wiley-ISTE, 2013. ISBN 978-1-84821-269-5. Avalable at: <http://ebooks.ciando.com/book/index.cfm/bok_id/499096>. Not cited in the text.

BALTRUNAS, D.; ELMOKASHFI, A.; KVALBEIN, A.; ALAY, O. Investigating packet loss in mobile broadband networks under mobility. In: . IFIP, 2016. p. 225–233. Avalable at: <http://ieeexplore.ieee.org/document/7497225>. Cited on page 21

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. v. 57, n. 1, p. 289–300, 1995. Avalable at: <http://www.jstor.org/stable/2346101>. Cited on page 32

CHERNICK, M. R.; LABUDDE, R. A. *An introduction to bootstrap methods with applications to R*. [S.l.]: Wiley, 2011. ISBN 978-0-470-46704-6. Cited on page 31

CORDER, G. W.; FOREMAN, D. *Nonparametric Statistics for Non-Statisticians*. Aufl. Wiley, 2011. ISBN 0-470-45461-X. Avalable at: <http://ebooks.ciando.com/book/index.cfm/bok_id/488684>. Not cited in the text.

COVER, T. M.; THOMAS, J. A. *Elements of Information Theory*. 2. ed. Hoboken, USA: Wiley-Interscience, 2006. v. 2nd. Not cited in the text.

DASGUPTA, A. *Asymptotic theory of statistics and probability*. [S.l.]: Springer, 2008. ISBN 978-0-387-75971-5. Cited on page 30

DEKKING, F.; KRAAIKAMP, C.; LOPUHAA, H.; MEESTER, L. *A Modern Introduction to Probability and Statistics*. [S.l.]: Springer London, 2005. DOI: 10.1007/1-84628-168-7. ISBN 978-1-85233-896-1. Cited on page 35

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

DIVINE, G.; KAPKE, A.; HAVSTAD, S.; JOSEPH, C. L. M. Exemplary data set sample size calculation for wilcoxon-mann-whitney tests. v. 29, n. 1, p. n/a, 2010. Avalable at: <http://www.ncbi.nlm.nih.gov/pubmed/19890884>. Not cited in the text.

European Commission. *Eurostat regional yearbook*. 2016. Avalable at: <http://ec.europa.eu/eurostat/statistics-explained/index.php/Eurostat_regional_yearbook>. Cited on page 47

FENG, C.; WANG, H.; LU, N.; TU, X. M. Log transformation: application and interpretation in biomedical research. *Statistics in Medicine*, v. 32, n. 2, p. 230–239, 2013. Avalable at: <http://onlinelibrary.wiley.com/doi/10.1002/sim.5486/abstract>. Cited on page 68

FENG, C.; WANG, H.; LU, N.; CHEN, T.; HE, H.; LU, Y.; TU, X. M. Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, v. 26, n. 2, p. 105, 2014. Avalable at: <http://www.ncbi.nlm.nih.gov/pubmed/25092958>. Cited on page 36

GRAVETTER, F. J.; WALLNAU, L. B. *Statistics for the behavioral sciences*. 9. rev. ed., internat. ed.. ed. [S.l.]: Thomson Wadsworth, 2013. ISBN 978-1-111-83955-0. Cited on page 31

HALL, P.; WILSON, S. R. Two guidelines for bootstrap hypothesis testing. *Biometrics*, v. 47, n. 2, p. 757–762, 1991–06–01. Avalable at: <http://www.jstor.org/stable/2532163>. Not cited in the text.

Keysight Technologies. *Nemo*. Avalable at: <www.keysight.com>. Cited on page 48

KHAN, A.; JAMALIPOUR, A. Downlink coverage performance of a relay cellular network considering non-uniform user distribution. In: . IEEE, 2014. p. 2260–2264. Avalable at: <http://ieeexplore.ieee.org/document/6952681>. Cited on page 43

KVAM, P. H.; VIDAKOVIC, B. *Nonparametric statistics with applications to science and engineering*. [S.l.]: Wiley-Interscience, 2007. ISBN 0-470-08147-3. Cited on page 29

LAKENS, D. D. Calculating and reporting effect sizes to facilitate cumulative science : a practical primer for t-tests and ANOVAs. v. 4, p. 1, 2013. Avalable at: <http://www.narcis.nl/publication/RecordID/oai:library.tue.nl:762429>. Cited on page 53

LEHMANN, E. L. *Nonparametrics: Statistical Methods Based on Ranks*. [S.l.]: Holden Day, 1975. ISBN 0-07-037073-7. Cited 3 times on pages 37, 38, and 39.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

LUTU, A.; ALAY, O.; BALTRUNAS, D.; ELMOKASHFI, A. Profiling mobile broadband coverage. In: . [s.n.], 2016. Published: IFIP International Workshop on Traffic Monitoring and Analysis (TMA). Avalable at: <http://tma.ifip.org/2016/papers/tma2016-final24.pdf>. Cited 4 times on pages 21, 22, 44, and 68.

MULLER, M. K.; TARANETZ, M.; RUPP, M. Performance of remote unit collaboration schemes in high speed train scenarios. In: . IEEE, 2015. p. 1–5. Avalable at: <http://ieeexplore.ieee.org/document/7391083>. Cited on page 21

NAUDTS, B.; SPRUYTTE, J.; OOTEGHEM, J. V.; VERBRUGGE, S.; COLLE, D.; PICKAVET, M. Internet on trains: A multi-criteria analysis of on-board deployment options for on-train cellular connectivity. In: . IEEE, 2014. p. 1–7. Avalable at: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6959256>. Not cited in the text.

NIST. *National Institute of Standards and Technology (NIST) /SEMATECH e-Handbook of Statistical Methods*. [s.n.], 2012. Avalable at: <http://www.itl.nist.gov/div898/handbook/>. Cited on page 29

NOETHER, G. E. Sample size determination for some common nonparametric tests. v. 82, n. 398, p. 645–647, 1987. Avalable at: <http://www.jstor.org/stable/2289477>. Cited on page 54

PAPOULIS, A. *Probability, random variables, and stochastic processes*. [s.n.], 1991. ISBN 978-0-07-048477-1. Avalable at: <http://catalog.hathitrust.org/Record/002455216>. Cited on page 30

PATTENGALE, N.; ALIPOUR, M.; BININDA-EMONDS, O.; STAMATAKIS, A. How many bootstrap replicates are necessary? In: . [S.l.]: Springer Berlin Heidelberg, 2009, (Research in Computational Molecular Biology, v. 5541). p. 184–200. ISBN 3-642-02007-0. DOI: 10.1007/978-3-642-02008-7_13. Cited on page 31

PHAN, V. V.; HORNEMAN, K.; YU, L.; VIHRIALA, J. Providing enhanced cellular coverage in public transportation with smart relay systems. In: . [s.n.], 2010. p. 301–308. ISBN 2157-9857. Avalable at: <http://ieeexplore.ieee.org/document/5698245>. Cited on page 44

POZAR, D. M. *Microwave engineering*. 4. ed.. ed. [S.l.]: Wiley, 2012. ISBN 0-470-63155-4. Not cited in the text.

RUXTON, G. D. The unequal variance t-test is an underused alternative to student's t-test and the mann-whitney u test. *Behavioral Ecology*, v. 17, n. 4, p. 688–690, 2006. Avalable at: <https://search.proquest.com/docview/204292670>. Cited on page 36

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

SAHA, C.; DHILLON, H. S. Downlink coverage probability of k-tier HetNets with general non-uniform user distributions. In: . IEEE, 2016. p. 1–6. Avalable at: <http://ieeexplore.ieee.org/document/7511509>. Not cited in the text.

SAHU, P. K. *Research methodology*. [S.l.]: Springer, 2013. Not cited in the text.

SARKAR, S. K.; CHANG, C.-K. The simes method for multiple hypothesis testing with positively dependent test statistics. v. 92, n. 440, p. 1601–1608, 1997. Avalable at: <http://www.jstor.org/stable/2965431>. Cited on page 32

SAUNDERS, S. R.; ZAVALA, A. A. *Antennas and propagation for wireless communication systems*. 2. ed.. ed. [S.l.]: Wiley, 2007. ISBN 0-470-84879-0. Cited on page 25

SCOTT, S.; LEINONEN, J.; PIRINEN, P.; VIHRIALA, J.; PHAN, V. V.; LATVA-AHO, M. A cooperative moving relay node system deployment in a high speed train. In: . IEEE, 2013. p. 1–5. ISBN 1550-2252. Avalable at: <http://ieeexplore.ieee.org/document/6691818>. Cited 2 times on pages 21 and 41.

SEO, S. *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*. phdthesis, 2006–08–09. Avalable at: <http://d-scholarship.pitt.edu/7948>. Not cited in the text.

SEYBOLD, J. S. *Introduction to RF propagation*. [S.l.]: Wiley, 2005. ISBN 978-0-471-65596-1. Cited on page 27

SHIVALDOVA, V.; PAULIN, T.; PAIER, A.; MECKLENBRAUKER, C. F. Performance measurements of multi-hop communications in vehicular ad hoc networks. In: . IEEE, 2012. p. 7162–7166. ISBN 1550-3607. Avalable at: <http://ieeexplore.ieee.org/document/6364791>. Not cited in the text.

SMITH, A. A. *Radio frequency principles and applications*. [S.l.]: IEEE Press, 1998. ISBN 978-0-7803-3431-1. Not cited in the text.

SUI, Y.; VIHRIALA, J.; PAPADOGIANNIS, A.; STERNAD, M.; YANG, W.; SVENSSON, T. Moving cells: a promising solution to boost performance for vehicular users. v. 51, n. 6, p. 62–68, 2013. Avalable at: <http://ieeexplore.ieee.org/document/6525596>. Cited on page 41

TAEGER, D.; KUHNT, S. *Statistical Hypothesis Testing with SAS and R*. 1. ed. John Wiley & Sons Ltd, 2014. ISBN 978-1-119-95021-9. Avalable at: <http://lib.myilibrary.com?ID=563515>. Not cited in the text.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

TOLSTRUP, M. *Indoor radio planning : a practical guide for 2G, 3G and 4G*. West Sussex, England : Wiley, 2015; Third edition, 2015. ISBN 978-1-118-91360-4. Cited on page 44

TRINDADE, O.; BERISHA, T.; SVOBODA, P.; MECKLENBRAUKER, C. F. Evaluation of wireless repeater performance on-board austrian railway trains by measurements. In: . Vienna Young Scientistis Symposium (VSS), 2016. p. 78–79. Available at: <https://publik.tuwien.ac.at/files/PubDat_251452.pdf>. Cited on page 21

VALLIANT, R.; DEVER, J. A.; KREUTER, F. *Practical tools for designing and weighting survey samples*. [S.l.]: Springer, 2013. ISBN 1-4614-6448-X. Not cited in the text.

VOLLANDT, R.; HORN, M. Evaluation of noether's method of sample size determination for the wilcoxon-mann-whitney test. v. 39, n. 7, p. 823–829, 1997. Not cited in the text.

WALTERS, S. J. Sample size and power estimation for studies with health related quality of life outcomes: a comparison of four methods using the SF-36. *Health and quality of life outcomes*, v. 2, n. 1, p. 26, 2004. Avalable at: <http://www.ncbi.nlm.nih.gov/pubmed/15161494>. Cited on page 54

WU, J.; FAN, P. A survey on high mobility wireless communications: Challenges, opportunities and solutions. v. 4, p. 450–476, 2016. Avalable at: <http://ieeexplore.ieee.org/document/7383229>. Cited 2 times on pages 20 and 44.

YACOUB, M. D. *Foundations of mobile radio engineering*. [S.l.: s.n.], 1993. ISBN 0-8493-8677-2. Cited 2 times on pages 26 and 28.

YACOUB, M. D. *Wireless technology: protocols, standards, and technique*. Boca Raton, FL: CRC Press, c2002, 2002. ISBN 0-8493-0969-7. Cited on page 43

ZHAO, H.; HUANG, R.; ZHANG, J.; FANG, Y. Handoff for wireless networks with mobile relay stations. In: . [s.n.], 2011. p. 826–831. ISBN 1525-3511. Avalable at: <http://ieeexplore.ieee.org/document/5779239>. Not cited in the text.

ZHAO, Y. D.; RAHARDJA, D.; QU, Y. Sample size calculation for the wilcoxon–mann–whitney test adjusting for ties. v. 27, n. 3, p. 462–468, 2008–02–10. Avalable at: <http://onlinelibrary.wiley.com/doi/10.1002/sim.2912/abstract>. Not cited in the text.

ZHOU, X.-H.; GAO, S.; HUI, S. L. Methods for comparing the means of two independent log-normal samples. v. 53, n. 3, p. 1129–1135, 1997. Avalable at: <http://www.jstor.org/stable/2533570>. Cited on page 36

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

References (Not cited)

AGGARWAL, C. C. An introduction to outlier analysis. In: . 2nd ed. 2017. ed. [S.l.]: Springer International Publishing, 2017, (Outlier Analysis). ISBN 978-3-319-47577-6. DOI: 10.1007/978-3-319-47578-3. Not cited in the text.

BAGDONAVICIUS, V.; KRUOPIS, J.; NIKULIN, M. *Nonparametric Tests for Complete Data*. 1. aufl.. ed. Wiley-ISTE, 2013. ISBN 978-1-84821-269-5. Available at: <http://ebooks.ciando.com/book/index.cfm/bok_id/499096>. Not cited in the text.

CORDER, G. W.; FOREMAN, D. *Nonparametric Statistics for Non-Statisticians*. Aufl. Wiley, 2011. ISBN 0-470-45461-X. Available at: <http://ebooks.ciando.com/book/index.cfm/bok_id/488684>. Not cited in the text.

COVER, T. M.; THOMAS, J. A. *Elements of Information Theory*. 2. ed. Hoboken, USA: Wiley-Interscience, 2006. v. 2nd. Not cited in the text.

DIVINE, G.; KAPKE, A.; HAVSTAD, S.; JOSEPH, C. L. M. Exemplary data set sample size calculation for wilcoxon-mann-whitney tests. v. 29, n. 1, p. n/a, 2010. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19890884>. Not cited in the text.

HALL, P.; WILSON, S. R. Two guidelines for bootstrap hypothesis testing. *Biometrics*, v. 47, n. 2, p. 757–762, 1991–06–01. Available at: <http://www.jstor.org/stable/2532163>. Not cited in the text.

HOLMA, H.; TOSKALA, A. *LTE for UMTS: Evolution to LTE-Advanced: Second Edition*. [s.n.], 2011. DOI: 10.1002/9781119992943. ISBN 978-1-119-99294-3. Not cited in the text.

NAUDTS, B.; SPRUYTTE, J.; OOTEGHEM, J. V.; VERBRUGGE, S.; COLLE, D.; PICKAVET, M. Internet on trains: A multi-criteria analysis of on-board deployment options for on-train cellular connectivity. In: . IEEE, 2014. p. 1–7. Available at: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6959256>. Not cited in the text.

PABST, R.; WALKE, B. H.; SCHULTZ, D. C.; HERHOLD, P.; YANIKOMEROGLU, H.; MUKHERJEE, S.; VISWANATHAN, H.; LOTT, M.; ZIRWAS, W.; DOHLER, M.; AGHVAMI, H.; FALCONER, D. D.; FETTWEIS, G. P. Relay-based deployment concepts for wireless and mobile broadband radio. v. 42, n. 9, p. 80–89, 2004. Available at: <http://ieeexplore.ieee.org/document/1336724>. Not cited in the text.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

POZAR, D. M. *Microwave engineering*. 4. ed.. ed. [S.l.]: Wiley, 2012. ISBN 0-470-63155-4. Not cited in the text.

SAHA, C.; DHILLON, H. S. Downlink coverage probability of k-tier HetNets with general non-uniform user distributions. In: . IEEE, 2016. p. 1–6. Available at: <http://ieeexplore.ieee.org/document/7511509>. Not cited in the text.

SAHU, P. K. *Research methodology*. [S.l.]: Springer, 2013. Not cited in the text.

SEO, S. *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*. phdthesis, 2006–08–09. Available at: <http://d-scholarship.pitt.edu/7948>. Not cited in the text.

SHIVALDOVA, V.; PAULIN, T.; PAIER, A.; MECKLENBRAUKER, C. F. Performance measurements of multi-hop communications in vehicular ad hoc networks. In: . IEEE, 2012. p. 7162–7166. ISBN 1550-3607. Available at: <http://ieeexplore.ieee.org/document/6364791>. Not cited in the text.

SMITH, A. A. *Radio frequency principles and applications*. [S.l.]: IEEE Press, 1998. ISBN 978-0-7803-3431-1. Not cited in the text.

TAEGER, D.; KUHNT, S. *Statistical Hypothesis Testing with SAS and R*. 1. ed. John Wiley & Sons Ltd, 2014. ISBN 978-1-119-95021-9. Available at: <http://lib.myilibrary.com?ID=563515>. Not cited in the text.

Third Generation Partnership-3GPP. *Technical specification TS36.133*. 2005. Available at: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2420>. Cited on page 57

VALLIANT, R.; DEVER, J. A.; KREUTER, F. *Practical tools for designing and weighting survey samples*. [S.l.]: Springer, 2013. ISBN 1-4614-6448-X. Not cited in the text.

VOLLANDT, R.; HORN, M. Evaluation of noether's method of sample size determination for the wilcoxon-mann-whitney test. v. 39, n. 7, p. 823–829, 1997. Not cited in the text.

ZHAO, H.; HUANG, R.; ZHANG, J.; FANG, Y. Handoff for wireless networks with mobile relay stations. In: . [s.n.], 2011. p. 826–831. ISBN 1525-3511. Available at: <http://ieeexplore.ieee.org/document/5779239>. Not cited in the text.

ZHAO, Y. D.; RAHARDJA, D.; QU, Y. Sample size calculation for the wilcoxon–mann–whitney test adjusting for ties. v. 27, n. 3, p. 462–468, 2008–02–10. Available at: <http://onlinelibrary.wiley.com/doi/10.1002/sim.2912/abstract>. Not cited in the text.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

–

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Appendices

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

A SUMMARY AND FURTHER INFORMATION FROM SUBMITTED PAPER

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

IEEE Transactions on Wireless Communications

Home

Author

Review

Submission Confirmation

Print

Thank you for your submission

Submitted to	IEEE Transactions on Wireless Communications
Manuscript ID	Paper-TW-Mar-18-0291
Title	Empirical study of signal penetration loss aboard high-speed trains with frequency selective windows
Authors	Trindade, Orlando Mecklenbräuker, Christoph Kretly, Luiz Carlos Berisha, Taulant Lerch, Martin Svoboda, Philipp Mayer, Lukas
Date Submitted	17-Mar-2018

Author Dashboard

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

IEEE Transactions on Wireless Communications

Empirical study of signal penetration loss aboard high-speed trains with frequency selective windows

Journal:	IEEE Transactions on Wireless Communications
Manuscript ID	Draft
Manuscript Type:	Original Transactions Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Trindade, Orlando; Universidade Estadual de Campinas, DECOM (FEEEC); Technische Universitat Wien, Telecommunications Mecklenbräuker, Christoph; Vienna University of Technology, Inst. f. Communications and Radio Frequency Engineering; ftw. Forschungszentrum Telekommunikation Wien, Kretly, Luiz Carlos; Universidade Estadual de Campinas, DECOM (FEEEC) Berisha, Taulant; Technische Universitat Wien, Telecommunications Lerch, Martin; Technische Universitat Wien, Telecommunications Svoboda, Philipp; Technische Universitat Wien, Telecommunications Mayer, Lukas; Siemens AG Osterreich
Keyword:	

SCHOLARONE™ Manuscripts

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

B SUMMARY OF SLIDES USED DURING DISSERTATION DEFESE

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

6/9/2018

MASTER'S DISSERTATION

ORLANDO DA SILVA TRINDADE **RA 160958 – FEEC UNICAMP**

TITLE
Supporting high quality cellular and Internet services inside railway vehicles: a statistical evaluation of wireless repeater performance on Railjet and Cityjet trains

TITLE (Portuguese)
Avaliação estatística do desempenho de repetidores sem fio em veículos sobre trilhos de longo trajeto: estudo de caso da melhoria da qualidade dos serviços de celular e Internet em trens europeus

June 18

MASTER'S DISSERTATION

Offcial examination Board:
Prof. Dr. Luiz Carlos Kretly ,-University of Campinas , **President**
Prof. Dr. Christoph Mecklenbräuker - Technische Universität Wien (TU WIEN) **Co-supervisor**
Prof. Dr. Silvio Ernesto Barbin – University of São Paulo (USP), **External member**
Prof. Dr. Michel Zamboni Hached – University of Campinas (UNICAMP) **Internal member**

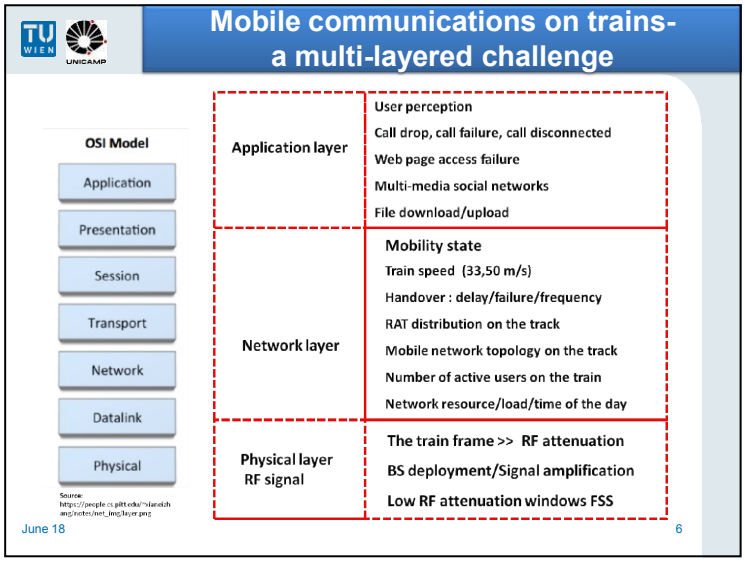
Substitute examination Board:
Prof. Dr. Gilmar Barreto – University of Campinas (UNICAMP)
Prof. Dr. Humberto Xavier de Araújo – Federal University of Tocantins (UFT)

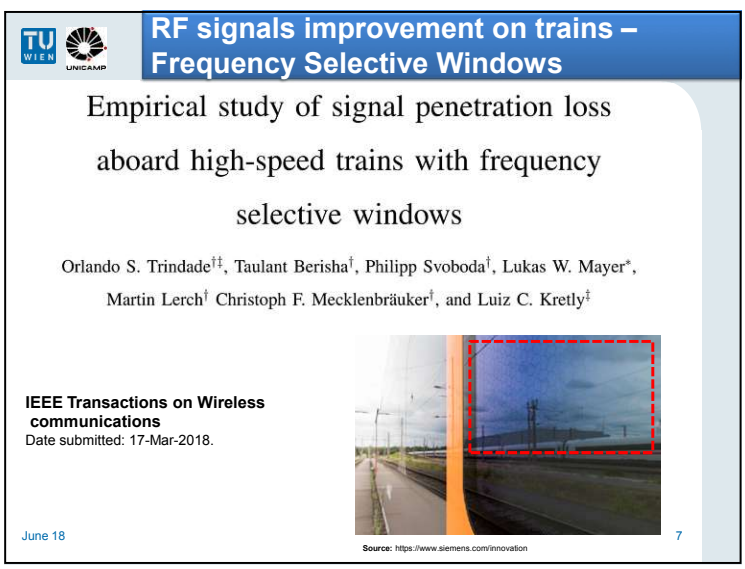
June 18

1

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

6/9/2018



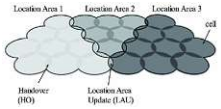


Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

6/9/2018

Challenges of measuring performance

Source: https://www.oebb.at



Source:http://www.kjit.bme.hu/images

- Railway size: 200km to 360km
- Train size: ~ 210m
- RAT deployment along the railway
- Cellular network topology
- Mimicking real services usage
- Absence of standard performance indicator (KPI) for such a context.

June 18 10

Sppliting data from in train measurements

ANDRA LUTU

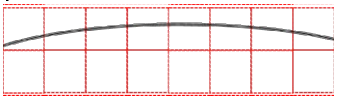
2km X 2km grid

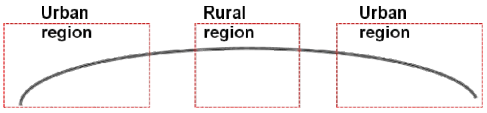
DŽIUGAS BALTRŪNAS

5-min bin and Speed >0

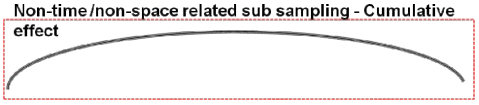
Location area codes (LAC)

Rout into time/or space grid – binning procedure





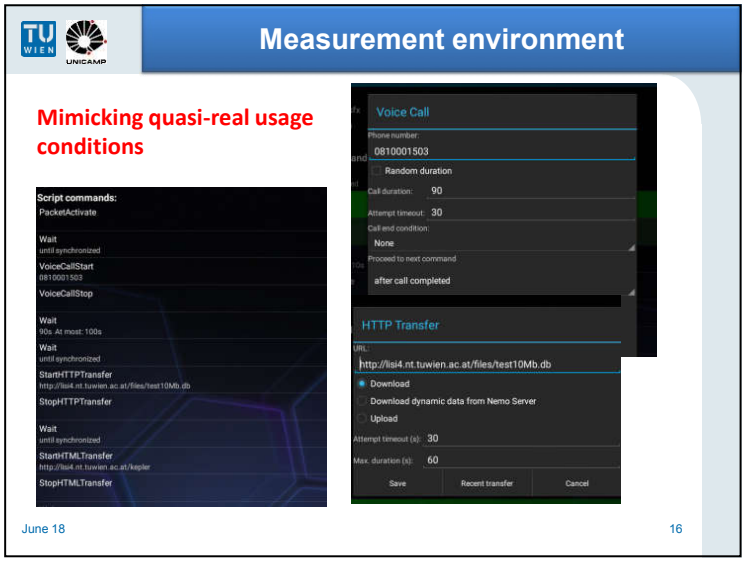
Non-time /non-space related sub sampling - Cumulative effect

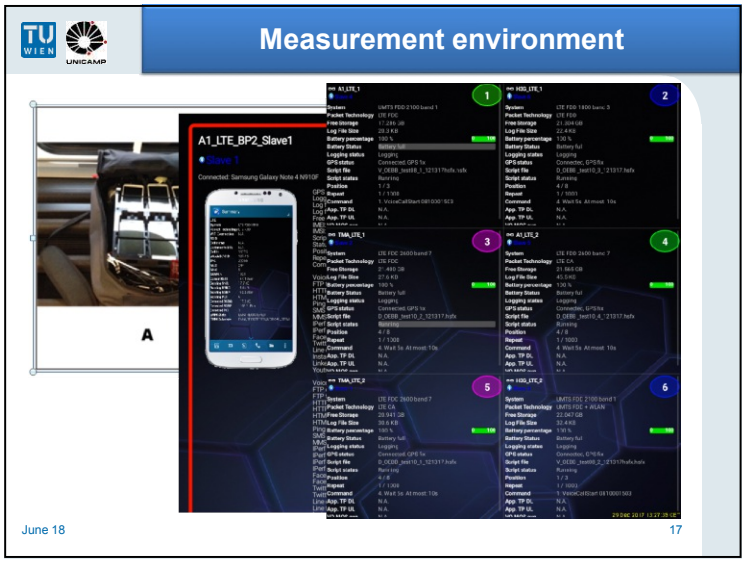


June 18 11

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

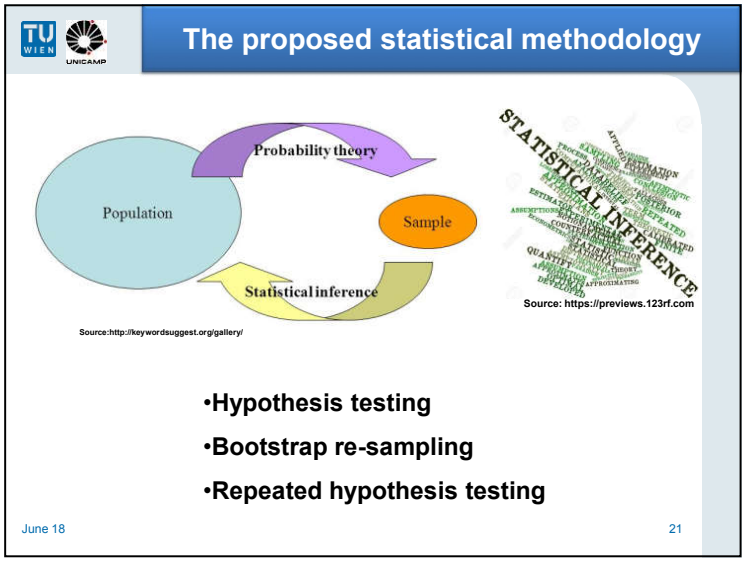
6/9/2018

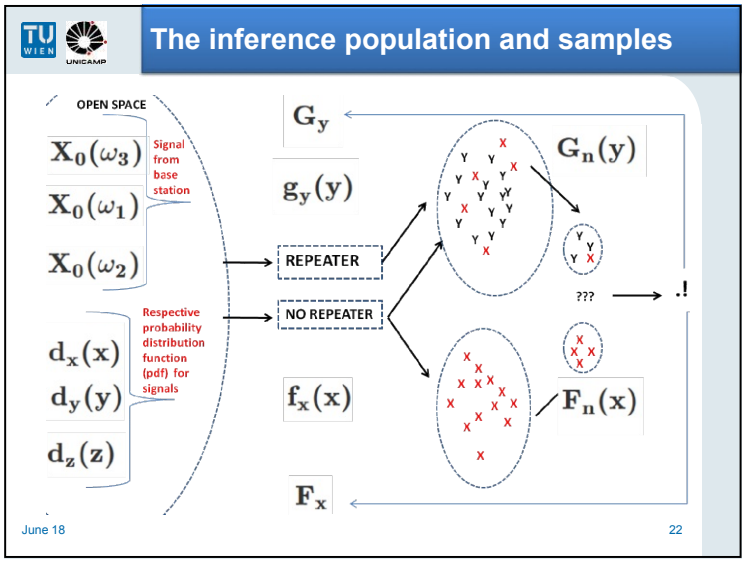




Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

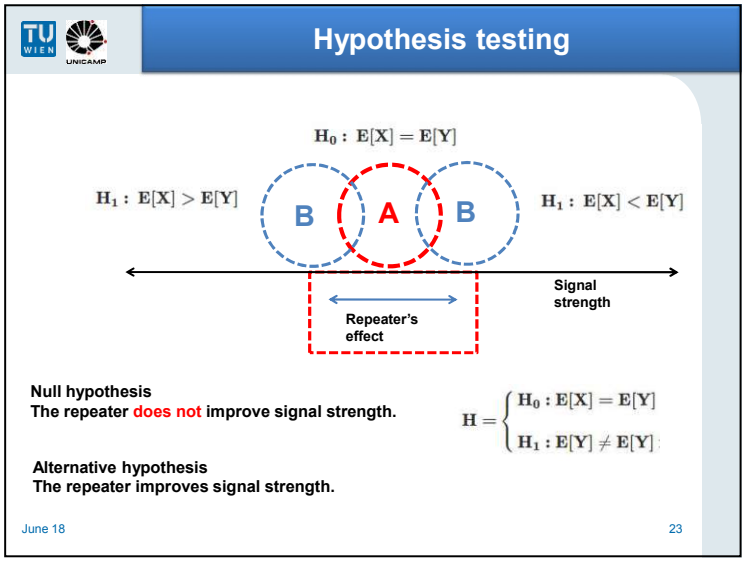
6/9/2018

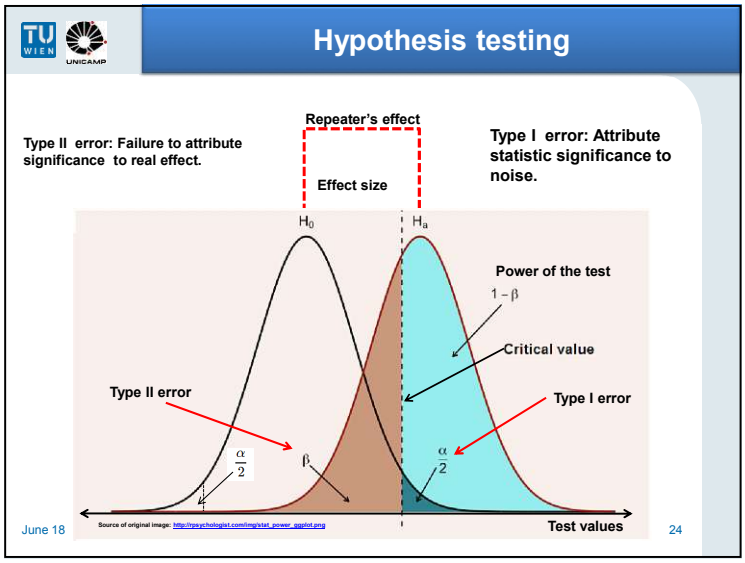




Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6/9/2018





Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6/9/2018

Repeated hypotheses; why?

What happens if we take one point to represent a 190000-score sample?

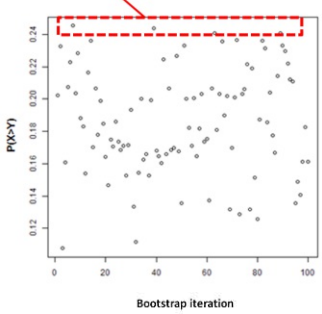
Compensates:

Sample variability and large sample size

Cumulative effect over 360km

User perception-coverage service outage

The bound used to attribute significance the effect.



June 18 25

Repeated hypotheses-consequences

SINGLE TEST
Type I error probability

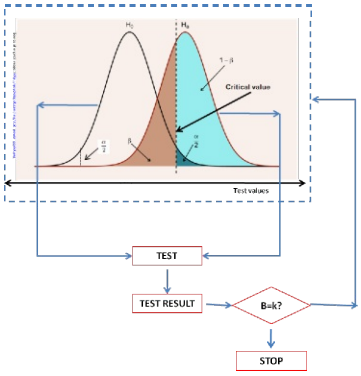
$$\alpha = 5\%$$

MULTIPLE TESTS
Type I error increases
Probability

$$1-(1-\alpha)^k = 99.40\% \quad \text{for } k = 100$$

ADJUSTMENT
Type I error Probability

$$P_{(i)} < i\alpha/k$$



June 18 26

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6/9/2018

Student´s t-test

Parametric test - model dependent

Student's t-test

- Gaussian model

Empirical - Student's t-test
- Quasi-Gaussian model

Assumptions under the empirical t-test
$f_x(x)$ and $g_y(y)$ are unknown;
$X \approx (\mu_x, \sigma_x{}^2)$ and $Y \approx (\mu_y, \sigma_y{}^2)$;
σ_x and σ_y are unequal and unknown.

Hypotheses pair

$$H = \begin{cases} H_0 : E[X] = E[Y], \text{ no repeater effect on signal (mean)} \\ H_1 : E[Y] \neq E[Y], \text{ repeater effect on signal (mean)} \end{cases}$$

June 18 27

Mann-Whitney-Wilcoxon (rank sum test)

Mann-Whitney-Wilcoxon (rank sum test)

Not based on model

samples of X and Y are mapped on $\mathbb{N}_+^*$

Hypotheses pair

$$H = \begin{cases} H_0 : F_x(x) = F_y(y), \text{ no repeater effect on signal (median)} \\ H_1 : F_y(y) \neq F_y(y), \text{ repeater effect on signal (median)} \end{cases}$$

June 18 28

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

C Complements for chapter 7

C.1 RAT comparison: statistics on scale and shape

Basic statistic refered to Section 7.2.

Table 9 – Elementary statistics. Variable RSSI.

Repeater	Min.	1st Q.[1]	Median	Mean	SE_M[2]	3rd Q.[1]	Max.	N
GSM								
OFF	-111.00	-88.00	-82.00	-81.46	0.08	-76.00	-48.00	16094
ON	-108.00	-76.00	-71.00	-71.06	0.06	-66.00	-48.00	16094
UMTS								
OFF	-105.80	-86.80	-76.10	-76.96	0.10	-67.70	-42.70	16755
ON	-105.80	-71.30	-64.90	-67.21	0.07	-61.20	-36.90	16755

[1] Q means quantile.
[2] SE_M -standar error from the mean.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 10 – Measures of spread and shape. Variable RSSI

GSM Repeater	SD	Var.	Skewness	Kurtosis	Outliers/N	N
OFF	9.87	97.45	0.35	3.24	2.31	
ON	7.75	60.00	-0.20	3.25	1.67	
UMTS Repeater	**SD**	**Var.**	**Skewness**	**Kurtosis**	**Outliers/N**	**N**
OFF	12.70	161.20	-0.04	2.39	0.00	
ON	8.84	78.12	-1.32	4.95	4.54	

[1] Measured in percentages: Outliers/N.

C.2 Summary of results for LTE- Railjet train

C.2.1 Basic statistics

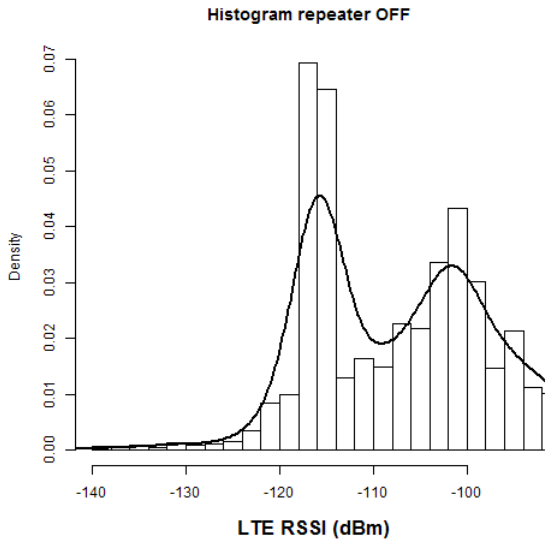


Figure 14 – Histogram for no-repeater context.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

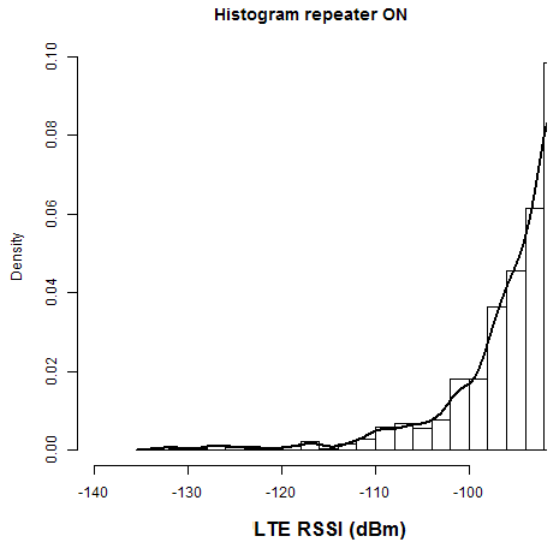


Figure 15 – Histogram for repeater context.

Table 11 – Elementary statistics for LTE. Variable RSSI

Repeater	Min.	1st Q.[1]	Median	Mean	SE_M[2]	3rd Q.[1]	Max.	N
OFF	-140.00	-115.30	-104.10	-103.50	0.28	-95.90	-75.90	2122
ON	-132.10	-95.30	-91.40	-92.06	0.16	-87.60	-73.60	2122

[1] Q means quantile.
[2] SE_M -standar error from the mean.

Table 12 – Measures of spread and shape for LTE. Variable RSSI.

Repeater	SD	Var.	Skewness[2]	Kurtosis[2]	Outliers/N [1]	N
OFF	13.08	171.01	0.55	2.59	0.00	
ON	7.43	55.27	-1.26	7.00	4.48	

[1] Measured in percentages: Outliers/N.
[2] Values for normal distribution: kurtosis=3, skewness=0.

C.2.2 Hypothesis test results

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 13 – Empirical t-test for LTE, variable RSSI.

α	t.statistic	t-critic	$Pr > \|t\|$	Effect size	F-values	N	B
0.05	5.72	1.59	<0.001	$[0.61, 1.14]^1$	$[1.87, 5.01]^1$	100	10000

[1] Quantiles .025 and .975, respectively.

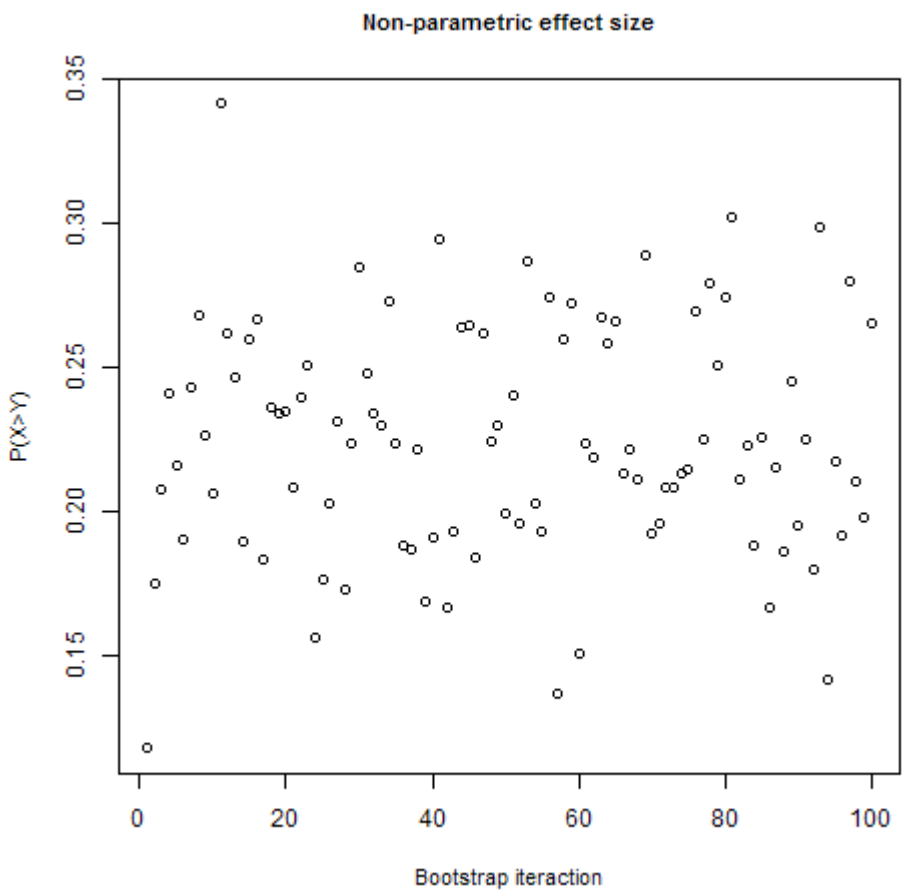


Figure 16 – Proportion P(X>Y). Wicoxon test for LTE.

C.3 Rural and urban comparison: statistics on scale and shape

Statistic refered to Section 7.3.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 14 – Wilcoxon sum-rank test for LTE, variable RSSI.

Rejection[1](%)	p-value max.	U_x max.	U_y max.	(Uy/mn)	N	B
0.96	0.10	504.00	792.00 [0.63, 0.87][2]	36	100	

[1] Percentage at which the null is rejected.
[2] Quantiles 0.025 and .975, respectively.

Table 15 – Rural and urban comparison. Variable RSSI.

Repeater	Min.	1st Q.[1]	Median	Mean	SE_M[2]	3rd Q.[1]	Max.	N
RURAL								
OFF	-105.80	-88.60	-80.60	-80.43	0.29	-72.40	-51.30	1435
ON	-104.60	-72.30	-65.60	-67.67	0.24	-61.30	-47.80	1435
URBAN								
OFF	-101.00	-92.30	-89.40	-87.94	0.18	-85.00	-58.90	1435
ON	-98.80	-70.50	-66.30	-66.86	0.17	-62.85	-36.90	1435

[1] Q means quantile.
[2] SE_M -standar error from the mean.

Table 16 – Spread and shape for rural and urban comparison. Variable RSSI

RURAL Repeater	SD	Var.	Skewness[1]	Kurtosis[1]	Outliers/N	N
OFF	10.89	118.60	0.01	2.55	0.00	
ON	9.01	81.21	-1.13	4.42	3.48	
URBAN Repeater	SD	Var.	Skewness	Kurtosis	Outliers/N	N
OFF	6.77	45.86	0.97	3.83	5.37	
ON	6.29	39.57	-0.79	5.96	2.09	

[1] Values for normal distribution: kurtosis=3, skewness=0.

C.4 Summary of results for Cityjet train

C.4.1 Basic statistics

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

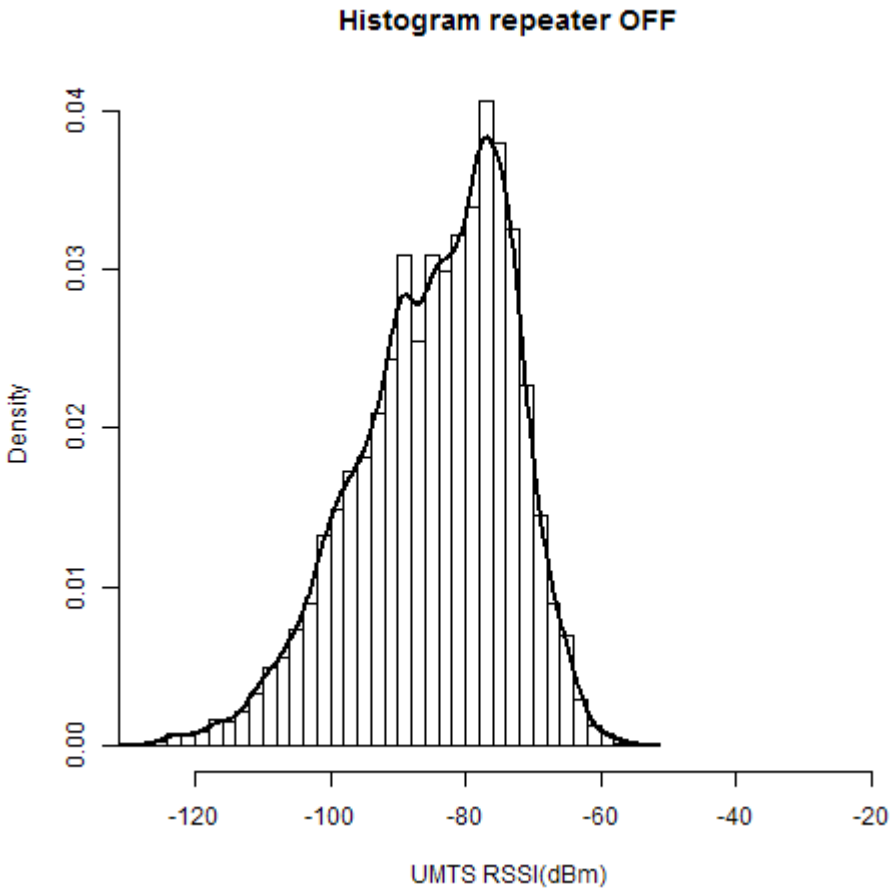


Figure 17 – Histogram for no-repeater context, sample size, N=15191. Cityjet train.

Table 17 – Elementary statistics for Cityjet train. Variable RSSI

Repeater	Min.	1st Q.[1]	Median	Mean	SE_M[2]	3rd Q.[1]	Max.	N
OFF	-127.10	-91.90	-83.10	-84.52	0.09	-75.80	-55.70	15191
ON	-118.60	-65.70	-58.80	-61.93	0.09	-55.20	0.00	15191

[1] Q means quantile.
[2] SE_M -standar error from the mean.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

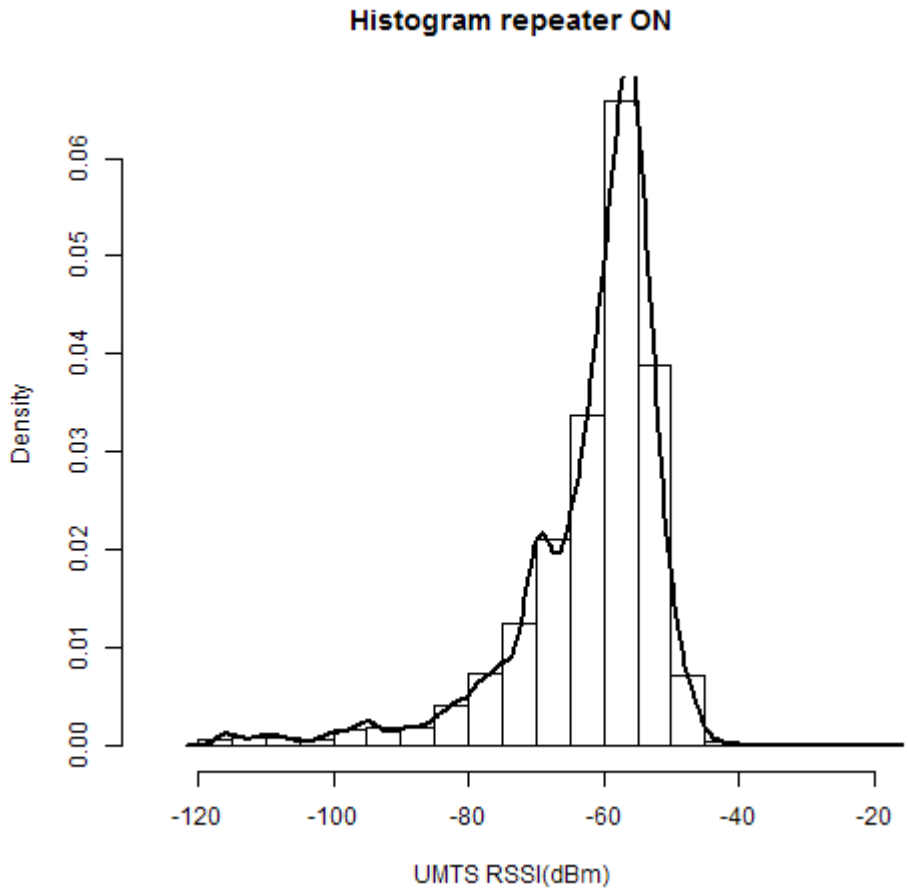


Figure 18 – Histogram for repeater context, sample size, N=15191. Cityjet train.

Table 18 – Measures of spread and shape for Cityjet train. Variable RSSI

Repeater	SD	Var.	Skewness[2]	Kurtosis[2]	Outliers/N [1]	N
OFF	11.40	129.90	-0.56	3.01	0.87	
ON	11.02	121.54	-1.95	8.56	5.75	

[1] Measured in percentages: Outliers/N.
[2] Values for normal distribution: kurtosis=3, skewness=0.

C.4.2 Hypothesis test results

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 19 – Empirical t-test for Cityjet train. Alternative $H1\_M\_2 \neq M\_1$

α	t.statistic	t-critic	$Pr > \|t\|$	Effect size	F-values	N	B
0.05	13.17	1.61	<0.001	$[1.67, 2.28]^1$	$[1.01, 1.90]^1$	100	10000

[1] Quantiles .025 and .975, respectively.

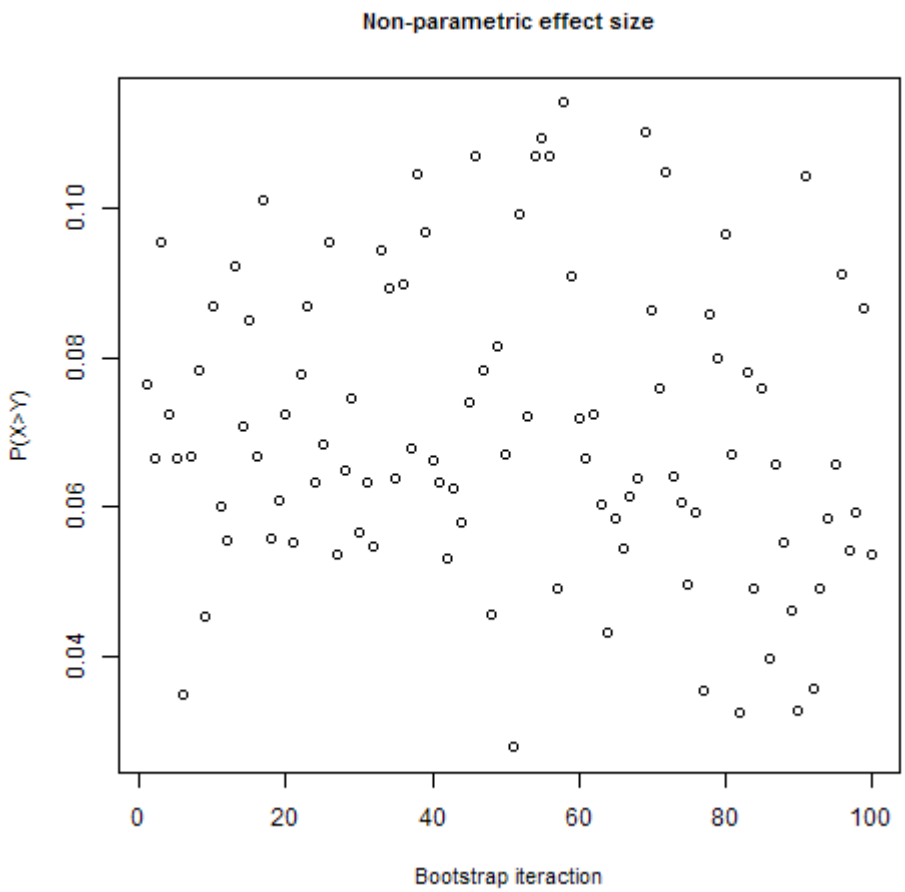


Figure 19 – Proportion P(X>Y). Wicoxon test for Cityjet.

Table 20 – Wilcoxon sum-rank test for Cityjet

Rejection[1](%)	p-value max.	U_x max.	U_y max.	(Uy/mn)	N	B
1.00	0.02	68.00	188	[0.80, 0.99][2]	16	100

[1] Percentage at which the null is rejected.
[2] Quantiles 0.025 and .975, respectively.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

D TECHNICAL INFOMATION FOR ON-BOARD REPEATER SOLUTION

D.1 Repeater electrical specification

Note: Power equally divide among operators.

The digital repeater units RUD-5 and RUD19-5 for up to 5 independent frequency bands have been designed to extend radio coverage in different radio networks. They are intended to be used for outdoor coverage extension, in-building and in-tunnel coverage. Due to their digital filtering they can operate as channel selective or band selective repeaters. The RUD-5 and RUD19-5 are available in different configurations:

- Repeater Unit FO-link (RUF)
- Repeater Unit Off-Air (RUOA)
- 19" or IP65 housing (RUD19 or RUD)

Features

- **5 bands** in one housing
- **50…2700MHz band support in one housing**
- **Digital** fibre optic and off air repeater
- **19" and outdoor housing**
- **40dBm output power**
- **16 channels** or sub bands
- Up to **14dB optical loss**

Options

- 24, 48 , 110VDC Power Supply
- COSweb Client Monitoring Software
- Remote access via wireless modem
- Extended Alarm Interface Module

Electrical Specification	
General RF Parameters DRM	
Delay: Band / Channel Selective Mode	1.9 us / 6.2 us (filter type dependent)
Return loss / Impedance	14 dB / 50 Ohm
Number sub bands / channels	up to 16
Max. input power: w/o destruction	10 dBm
Frequency ranges	TETRA390, TETRA410, LTE800, GSM/LTE-R900, EGSM/UMTS900, GSM/LTE1800, UMTS2100, LTE2600 (details see last sheet)
Downlink	
Gain RU with FO Link (RUF)	30–60dB ±1.5 dB in 0.5 dB steps
Gain Off Air low / high power	45-75 / 60-90dB±1.5 dB in 1dB step
Output Power RU:	
Composite output power (low/high)	25/40 dBm
@ 4 carriers	19/34 dBm / carrier
Noise figure (Off-Air)	typ. 5 dB
Uplink	
Gain RU with FO Link	15–35 dB ±1.5 dB in 0.5 dB steps
Gain Off Air low / high power	45-75 / 60-90dB ±1.5 dB in 1dB step
Composite output power (Off Air)	25dBm
@ 4 carriers	19/34 dBm / carrier
Noise figure	typ. 5 dB
Optical Interface FOM	
RF gain FO link	0 dB (+/-1dB)
Optical power	4mW (6dBm)
Optical wavelength	CWDM in 1310/1550 nm window
Max. optical loss	14dBo
IIP3	min. 19 dBm
Equivalent input noise power	
@ optical loss 0-8dBo	typ. -142 dBm/Hz
@ optical loss 8-14dBo	typ. -138 dBm/Hz

Figure 20 – Example of onboard-repeater.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

D.2 Example of configuration for onboard-repeater

Parameter	**Value**
LTE 800 MHz band	791 – 821 MHz, 832 – 862 MHz
LTE 2600 MHz band	2500 – 2570 MHz, 2620 – 2690 MHz
UMTS 2100 MHz band	1920 – 1980 MHz, 2110 – 2170 MHz
GSM 900 MHz band	880 – 915 MHz, 925 – 960 MHz
GSM 1800 MHz band	1710 – 1785 MHz, 1805 – 1880 MHz
Uplink power	25 dBm
Uplink gain	15 – 35 dB
Downlink power	25 dBm or 40 dBm
Downlink gain	30 – 60 dB
Delay of band/channel selective	1.9 μs/6.2 μs
Subbands/channel number	up to 16
Noise figure	5 dB

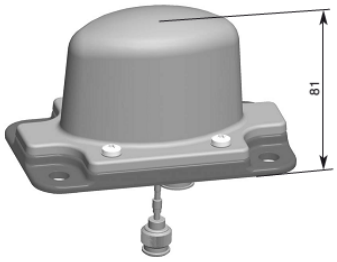
Figure 21 – Example of repeater configureation. Source: OBB report

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

D.3 On-train rooftop mounted antenna - specification

Train Antenna
790 – 2700 MHz and GPS 1575 MHz
870 10003

- Multi-band antenna: 800/900/1800/1900/UMTS/ UMTS II/W-LAN and GPS.
- The antenna can be operated in all frequency ranges simultaneously.
- Low profile antenna in fiberglass radome.
- The antenna fulfils the requirements according to EN 50155.



Type No.	870 10003
Antenna multi-band	
Input	N female
Frequency range	790 – 2700 MHz
VSWR	790 – 806 MHz: < 2.2 806 – 870 MHz: < 2.0 870 – 2550 MHz: < 1.5 2550 – 2700 MHz: < 2.0
Gain	0 dB (ref. to the quarter-wave antenna)
Impedance	50 Ω
Polarization	Vertical
Max. power	100 W (at 50° C ambient temperature)
Inner conductor	D.C. grounded
Antenna GPS	
Input	Cable RG 316/U of 225 mm length with TNC male connector
Frequency range	1575.42 ±1 MHz
VSWR	< 1.5
Polarization	Right hand circular
Gain (90° elevation)	2 dB (ref. to the circularly polarized isotropic antenna)
Impedance	50 Ω
Inner conductor	D.C. grounded
Weight	approx. 0.5 kg
Packing size	152 x 91 x 125 mm
Height	81 mm

Material: Radiator: Copper and brass. Flange: Aluminum. Radome: Fiberglass. All screws and nuts: Stainless steel. Colour: Grey.

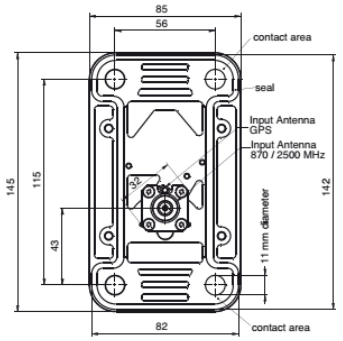
Mounting: On a conductive surface with a minimum size of 50 x 50 cm using 4 M10 bolts.

Grounding and high voltage protection: This antenna, tested by an independent institute and approved by the "Deutsche Bahn AG", is D.C. grounded to protect against lightning and high-tension lines.

Accessories: Low noise amplifier GPS 860 10142 (please order separately).

Warning: If the antenna is operated without the pre-amplifier type no. 860 10142, please note the following points.
- Due to the fact that the inner conductor of the antenna GPS is DC grounded, the input of the GPS receiver is loaded with a DC short circuit. If the GPS receiver provides a remote DC power supply, this could damage the GPS receiver.
- At the input of the antenna GPS a level of –25 dB below the signal applied at the input of the antenna multi-band appears. Depending on the level of the signal applied at the input of the antenna multi-band, the GPS receiver may be overloaded or damaged.

Mounting flange:

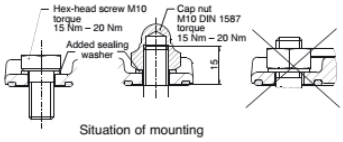


Mounting hole for the connector: 33 mm (max. 35 mm)

Note: Mounting surface must be free from paint for electrical contact.

Evenness of opposite surface 0.2 mm.

Use a cap nut or hex-head screw plus the enclosed sealing washer.



Situation of mounting

Figure 22 – Example of outdoor antenna for onbord solution.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

D.4 Example of specification for power spliter

Power Splitter, 2-Way, High Isolation, 790 - 2700 MHz, Nf

Features

- High isolation
- Low PIM
- Low insertion loss
- High power
- High stability and reliability
- Waterproof
- Applications: GSM, DCS,UMTS, LTE

Electrical Specifications	
Frequency range	790 - 2700 MHz
Insertion loss	≤ 3.4 dB
Return loss @ all ports	≥ 17 dB
Isolation	≥ 20 dB
Ripple	0.3 dB
Impedance	50 Ohm
Max. input power as a splitter	100 Watt
Max. input power as a combiner	1 Watt / Port
Passive Intermodulation (PIM)	≤-140 dBc @ 2 x +43 dBm

Mechanical Specifications	
Dimensions (L x W x H)	96 x 74 x 23 mm (incl. Connectors)
Weight	160 g
Connectors	N female
Material	Aluminum

Environmental Conditions	
Operating temperature range	-25 °C ~ +45°C

Figure 23 – Example of power spliter for onbord solution.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

E NEMO TOOL

E.1 Nemo Walker Air

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Keysight Technologies
Nemo Walker Air

Indoor Benchmarking On the Go

Nemo Walker Air is an ideal solution for performing indoor measurements and benchmarking on multiple technologies. It is lightweight, easy to carry and set up, and the complete system operates for up to 10 hours.

Anite is now part of **Keysight Technologies**

KEYSIGHT TECHNOLOGIES
Unlocking Measurement Insights

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Widest Selection of Android-Based Devices on the Market

Nemo Walker Air from Keysight Technologies, Inc. enables extensive synchronized measurements and script events with up to seven test terminals. One tablet acts as a master unit, conducting multi-measurements by sending commands to the test terminals via *Bluetooth*® connection. Log files can be uploaded to FTP/HTTP server from the master unit or collected manually from the slave units. The collected data can be directly analyzed with Keysight's Nemo post-processing tools. Nemo Walker Air supports the latest Samsung smartphones and tablets, for example, Samsung Galaxy S9, Samsung Galaxy S8, and Sony and Motorola terminals as slave units.

Nemo Walker Air Backpack

An optional Nemo Walker Air backpack provides a customized, lightweight, and compact carrying solution for Nemo Walker Air system. With high-quality USB batteries, Nemo Walker Air backpack can be used as a standalone system without an external power source.

Features

- Synchronize indoor benchmarking measurements and script events with up to seven test terminals
- Master conducts measurements, controls, and coordinates the slave units via *Bluetooth* connection
- Quick and easy set up
- Map navigation based on Google Maps
- Floorplans with markers and geodetic coordinates
- Enable geocoding indoor floorplans on top of Google Maps
- Remote management of drive testing with turn-by-turn navigation (optional)
- Audio quality and video quality streaming tests (POLQA, PESQ, PEVQ-S)
- DRT4311B and SeeGull IBflex scanner support
- Support for Qualcomm and Samsung Shannon chipsets
- Backpack option - a lightweight, compact carrying solution for the whole system
- Nemo Cloud option - a cloud-based platform for centralized remote control
- Nemo Instant Report option - a quick way to create reports in the field

KEYSIGHT TECHNOLOGIES
Unlocking Measurement Insights

Bluetooth and the Bluetooth logos are trademarks owned by Bluetooth SIG, Inc., U.S.A. and licensed to Keysight Technologies, Inc.

www.keysight.com/find/nemo

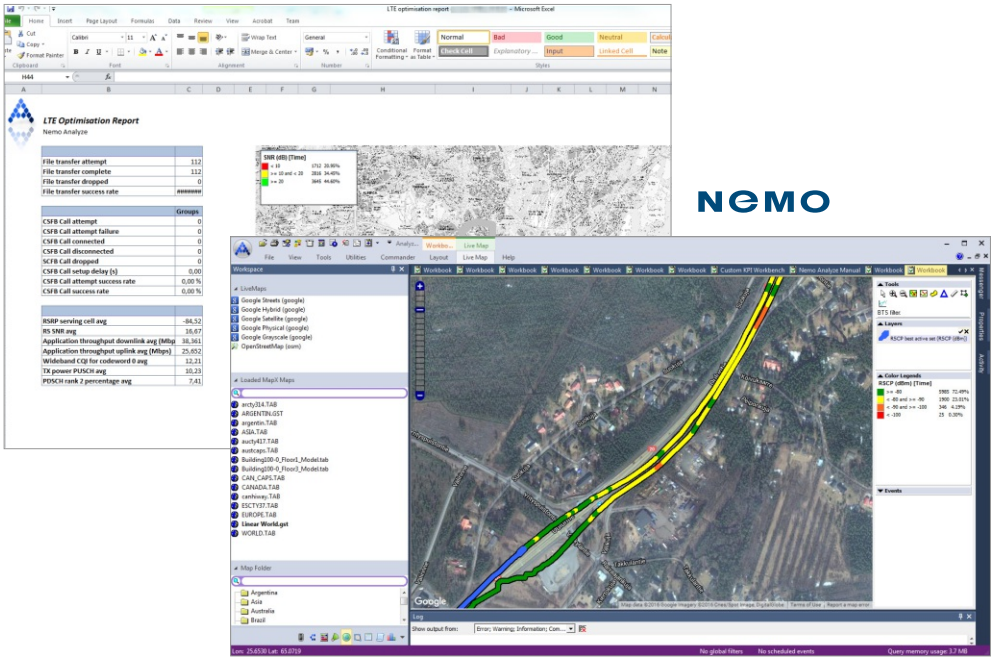
This information is subject to change without notice.
© Keysight Technologies, 2015 - 2018
Published in USA, April 24, 2018
5992-2051EN
www.keysight.com

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Keysight Technologies
Nemo Analyze

Professional Post-Processing of Drive Test Data

Nemo Analyze from Keysight Technologies, Inc. is a highly efficient and fully scalable analysis tool for benchmarking, automated troubleshooting, and statistical reporting based on drive test data. For benchmarking, its predetermined report templates help you compare key performance indicators (KPIs) from different operators, technologies, and time frames, and visualize the results in a single report. Nemo Analyze supports all 3GPP network technologies, including LTE-A CA, VoLTE/ViLTE, VoWiFi, NB-IoT, LTE-M, and 5G.



Anite is now part of **Keysight Technologies**

KEYSIGHT TECHNOLOGIES
Unlocking Measurement Insights

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Wide-Range Analytics and Reporting

Nemo Analyze with its full set of graphs, maps, and others data views for detailed analysis and playback is the perfect solution for post-processing and statistical reporting tasks for users with specific KPI reporting and analysis needs. Nemo Analyze enables the creation of custom KPIs and custom analysis tasks with KPI Workbench, a graphical flow chart-based scripting engine. Custom reports can be created with Microsoft Excel, Microsoft PowerPoint, and Microsoft Word. Nemo Analyze includes automated analysis tasks for most common GSM/UMTS/LTE network problems, such as call drop root cause analysis, BCCH clash analysis, and missing neighbor analysis. Drill-down from problem events on maps and grids enables efficient troubleshooting over large sets of data. Statistics and benchmarking based on operator, log file, time frame, and other attributes.

Best-In-Class Drive Test Data Visualization

The Nemo Analyze system from Keysight incorporates an innovative, low-maintenance database engine that has been designed and optimized specifically for high-performance post-processing of drive test data. The system is easy to set up and use and highly cost-effective to implement.

Nemo Analyze is the post-processing solution for data produced by Nemo tools. Nemo Analyze's support for character-separated value (CSV) format also lets you import ASCII data into the database, enabling for instance Wireshark/Ethereal and network counter data to be correlated and post-processed together with Nemo drive test data. Nemo Analyze supports Windows 7, Windows 8.1 and Windows 10, 64-bit. Integrating Nemo Analyze with other Nemo tools provides a complete automated data processing chain from raw measurement data to automatically generated results in workbook format.

Features

- Desktop/laptop drive test post-processing tool providing in-depth analysis, reporting, and optimization capabilities.
- Best-in-class data visualization with fully customizable, time-synchronized multi-page workbooks and a comprehensive set of data views, including maps, grids, line graphs, bar graphs, pie charts, surface grids, color grids, and spreadsheets.
- Support for web-based live maps, Google Maps, and OpenStreetMap.
- Diverse reporting options from instant reporting to Microsoft Office tools compatible reporting.
- Up-to-date support for all 3GPP network technologies, including LTE-A CA, Voice over LTE (VoLTE), Video over LTE (ViLTE), Voice over WiFi (VoWiFi), NB-IoT and LTE-M.
- Full automation of the data processing chain, from drive test data upload to full analysis results.
- A wide range of CDR-based off-the-shelf reports available for voice and data, with support for technologies from GSM to VoLTE, ViLTE, VoWiFi, and LTE 5CC.
- Seamless integration with Nemo products ensures 100% correct interpretation of the measurement data produced by your Nemo tools.

www.keysight.com/find/nemo

KEYSIGHT TECHNOLOGIES
Unlocking Measurement Insights

This information is subject to change without notice.
© Keysight Technologies, 2015 - 2018
Published in USA, May 17, 2018
5992-2047EN
www.keysight.com

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

F TRAINS AND ONBOARD REPEATER

F.1 Example of deployment onboard Raijet trains

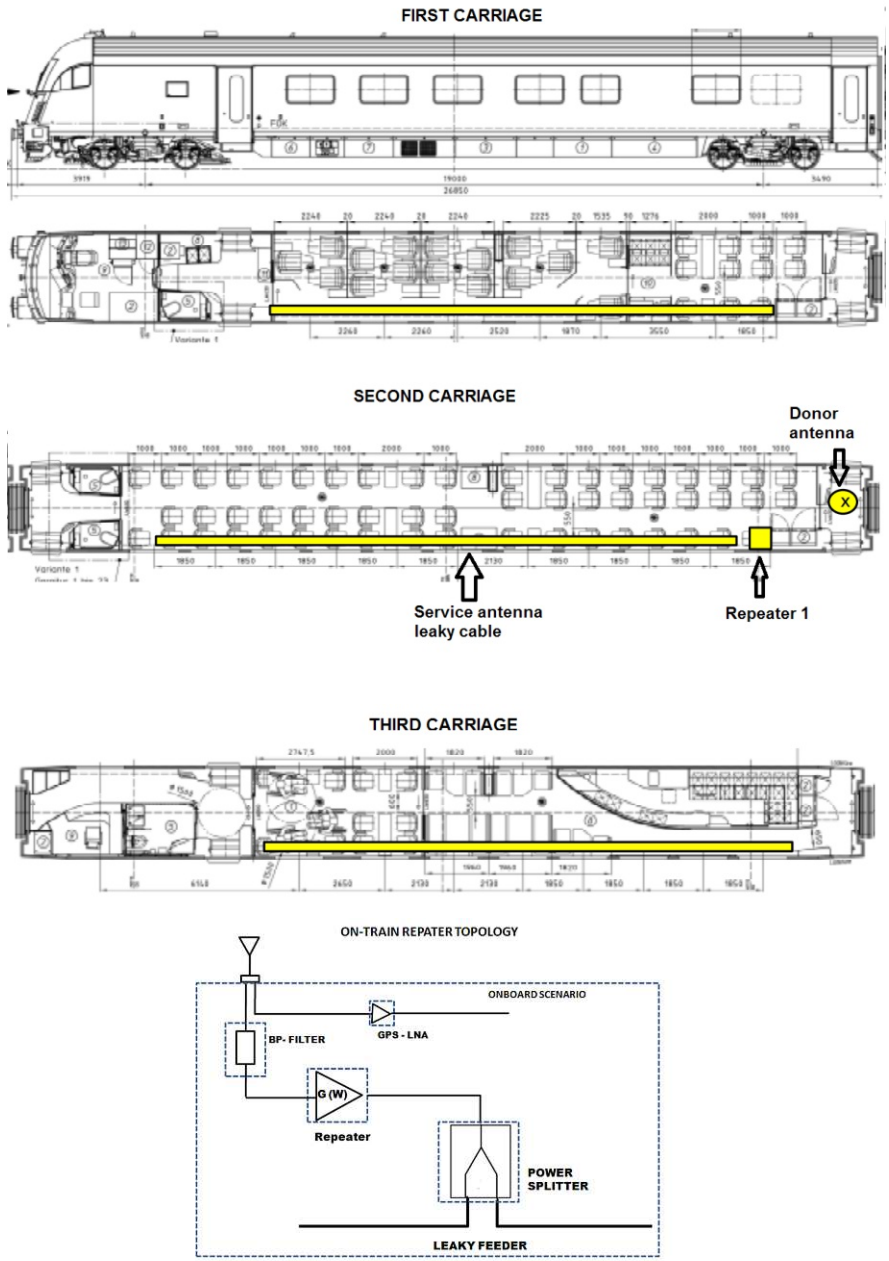


Figure 24 – Repater deployment onboard a railjet train. (Source: ÖBB)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

F.2 Raijet and CityJet trains

Figure 25 – Railjet train. (Source: ÖBB)

Figure 26 – Cityjet train. (Source: ÖBB)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

G AWK SCRIPTS - KEY POINTS

G.1 AWK

AWK program was used to extract data from text files. Key points for script are presented in the following lines.

```
# CREATING AND CLINNING ARRAYS

RSSI_GSM [gsm]=0
split("", RSSI_GSM)
SNR_GSM [gsm]=0
split("", SNR_GSM)


{        ####### MAIN SCRIPT OPENNING


if ($1 == "CELLMEAS")

{
cellmeas=++cellmeas;
#Number of occurences of CELLMEAS IN $1

################  FILTERS

#GSM

if ($4 == 1 && $8 == 1 && NF>=12) {
# active cell $8==1, neighbour $8==0 .
print $2, $10, $12, $9 > "RSSI_GSM.txt";

                        }

####     UMTS
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
if ($4 == 5 && $5 == 0) {
#Active cell:active $5==0, neighbour $5==1; $10==50001 to UMTS FDD 2100

umts=++umts;

        if (NF>=9) {
# NF line being processed.
It makes sure the minimum size of the line to
#avoid receiving wrong values

umts9=++umts9;

RSSI_UMTS [umts9]= $9;
print  $2, $8, $9, $10 > "RSSI_UMTS.txt";
                }

        if (NF>=19) {
# makes sure the minimum size of the line to
#avoid receiving wrong values
umts19=++umts19;

                if ($19 < 0) {
                        RSCP_UMTS [umts19]=$19;
                        }

                if ($19 > 0) {
                        RSCP_UMTS [umts19]=$22;
                        }

print $2, umts19 > "RSCP_UMTS.txt";
                }
                }

##### LTE ############

if ($4 == 7 && $8 == 0) {
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
lte=lte++;

        if (NF>12) {

lte12=++lte12;
RSSI_LTE [lte12]=$12;
print $2, $10, $12, $9 > "RSSI_LTE.txt";

                    }

        if (NF>13) {

lte=++lte13;
RSRP_LTE [lte13]= $13;

print $2, $10, $13  > "RSRP_LTE.txt";

                    }

                        } ##### MAIN LOOP CLOSING

}  # ##MAIN PROGRAM CLOSING


} # General structure closing
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

H R SCIPTS - KEY POINTS

H.1 R program

R program was used to process data from text files. User-friedly programming and variaty of packet-based functions make it all easy to implement. It is recommended to use RStudio to simplify scription processes. Implementation with RStudio should be straightforward given the backgrounds on statistics related to the topic. The following scripts lines intend to give gneral idea of implementation and can be used to implement via the more convenient language.

H.1.1 Macro scripts sequence

PARAMETRIC SECTION

Data conditioning (include outlier treatment) Non-parametric t-test

NONPARAMETRIC SECTION

Non-parametric data conditioning (no outlier treatment) Sample size estimation Non-parametric test

H.2 PARAMETRIC SECTION

H.2.1 Data conditioning

```
#_________Outliers with TUKEY's outlier labeling———


#IQR(rcpUoFF)

nQ1_off<-(1/4)*(length(rcpUoFF)+1)# first quantile
nQ3_off<-(3/4)*(length(rcpUoFF)+1)#third quantile
rcpUoFF_order_off<-sort(rcpUoFF,
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
 decreasing = FALSE, index.return = FALSE)
IQR_off<-rcpUoFF_order_off[nQ3_off]-
                                    rcpUoFF_order_off[nQ1_off]


#IQR(rcpUoN)
nQ1_oN<-(1/4)*(length(rcpUoN)+1)# first quantile
nQ3_oN<-(3/4)*(length(rcpUoN)+1)#third quantile
rcpUoN_order_oN<-sort(rcpUoN,
                  decreasing = FALSE, index.return = FALSE)
IQR_oN<-rcpUoN_order_oN[nQ3_oN]-rcpUoN_order_oN[nQ1_oN]

l_fence_off<-rcpUoFF_order_off[nQ1_off]-1.5*IQR_off
u_fence_off<-rcpUoFF_order_off[nQ3_off]+1.5*IQR_off

l_fence_oN<-rcpUoN_order_oN[nQ1_oN]-1.5*IQR_oN
u_fence_oN<-rcpUoN_order_oN[nQ3_oN]+1.5*IQR_oN


#Trimmed mean

trim_mean_off<-mean(rcpUoFF, trim = 0.5, na.rm = FALSE)
trim_mean_oN<-mean(rcpUoN, trim = 0.5, na.rm = FALSE)


#___IDENTIFYING

l_outl_off<-rcpUoFF[rcpUoFF<=l_fence_off]
#choose all values less than/equal than lower fence

if (length(l_outl_off)>0) { #runs only outliers exist
  for( i in length(rcpUoFF)) {#identifies upper-outlies

    if (rcpUoFF[i]<l_fence_off) {#identifies upper-outlies
      rcpUoFF[i]<-trim_mean_off}
  }
}
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
u_outl_off<-rcpUoFF[rcpUoFF>=u_fence_off]
#choose all values greater than/equal upper fence


if (length(u_outl_off)>0) { #runs only outliers exist
  for( i in length(rcpUoFF)) { #identifies upper-outlies

    if (rcpUoFF[i]>u_fence_off) { #identifies upper-outlies
      rcpUoFF[i]<-trim_mean_off}
  }
}
# END_OUTLIERS######


####BOOT ESTIMATES MEAN, SD, VAR———

for (i in 1:N_oN){
  R_rcpUoN<-sample(na.omit(rcpUoN), 100,
 replace = TRUE, prob = NULL) ## bootstrap resampling
  means_oN[i]<-mean(R_rcpUoN)
  sd_oN[i]<-sd(R_rcpUoN)
  var_oN[i]<-var(R_rcpUoN)
  means_M2_oN[i]<-means_oN[i]
  #means_M2_oN[i]<-(means_oN[i]+0.5*var_oN[i])
}

rcpUoN_M<-mean(means_oN)
rcpUoN_SD<-mean(sd_oN)
rcpUoN_var<-mean(var_oN)
```

H.2.2 Empirical t-test

```
#Receives two data vectors and calculates repeated test estatistics

# F-ratio
if (rcpUoN_var>rcpUoFF_var){Fmax=rcpUoN_var/rcpUoFF_var}
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
if (rcpUoN_var<=rcpUoFF_var){Fmax=rcpUoFF_var/rcpUoN_var}


#-------------------TEST PHASE ------------------------------

Tsamples_Size<-100

Td_rcpUoN<-sample(rcpUoN, Tsamples_Size,
                  replace = TRUE, prob = NULL) ## bootstrap resampling
Td_rcpUoFF<-sample(rcpUoFF, Tsamples_Size,
                  replace = TRUE, prob = NULL) ## bootstrap resampling

Sdxy<-sqrt((var(Td_rcpUoN)/length(Td_rcpUoN))+
                        (var(Td_rcpUoFF)/length(Td_rcpUoFF)))

#boot design
Td<-(mean(Td_rcpUoN)-mean(Td_rcpUoFF))/Sdxy

#normal
Td_table<-qnorm(1-.05/2) #one-sided


#________ #RESAMPLING WITH REPLACEMENT

T_boot<-10000


for ( i in 1:T_boot) {
  incem_T_bot[i]<-i

l_bot_rcpUoN<-sample(rcpUoN, loop_sz_rcpUoN,
                 replace = TRUE, prob = NULL) ## bootstrap resampling
l_bot_rcpUoFF<-sample(rcpUoFF, loop_sz_rcpUoFF,
          replace = TRUE, prob = NULL) ## bootstrap resampling

n1<-length(l_bot_rcpUoN)
n2<-length(l_bot_rcpUoFF)
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

```
#equal  samples size m=n
Sd_oNoFF<-sqrt((l_bot_rcpUoN_var/n2+
                l_bot_rcpUoFF_var/n2))
#deltm<-((l_bot_rcpUoN_M-
         l_bot_rcpUoFF_M)-(mean(Td_rcpUoN)-mean(Td_rcpUoFF)))
deltm<-((l_bot_rcpUoN_M-l_bot_rcpUoFF_M)-(rcpUoN_M-rcpUoFF_M))

deltm_TD<-((l_bot_rcpUoN_M-l_bot_rcpUoFF_M))


#From the empirical distribution
t_boot<-deltm/Sd_oNoFF
t_boot_vec[i]<-t_boot


#normal approach and repeated hypothesis
Td_boot<-deltm_TD/Sd_oNoFF
Td_boot_vec[i]<-Td_boot

# Glass' effect size distance
l_bot_effect_d<-(l_bot_diffMean)/sd(l_bot_rcpUoFF)

if (t_boot>0){

#null disign
if(t_boot<Td){count_tmore_score<-count_tmore_score+1}
inc_tmore<-inc_tmore+1
#normal design
    p_value_tmore[inc_tmore]<pnorm(t_boot, lower.tail = FALSE)

  }

#alternative

if (t_boot<0){ #
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
 #null disign
   if (abs(t_boot)<(Td)){count_tless_score<-count_tless_score+1}
   inc_tless<-inc_tless+1
#normal design
   p_value_tless[inc_tless]<-pnorm(t_boot, lower.tail = TRUE)

   }

#alternative
 if(abs(t_boot)<Td){count_t_score<-count_t_score+1}

} #end boot
```

H.3 SCIPTS-NON-PARAMETRIC SECTION

H.3.1 Data conditioning

The same used in paramtric section excluding outliers treatment.

H.3.2 Sample size estimation

```
for(i in 1:lop_w_size){

     U_incrmnt<-(U_incrmnt+1)
     U_incrmnt_vec[i]<-U_incrmnt

     Ux_oFF<-sample(rcpUoFF,U_size, replace = TRUE, prob = NULL)
     Uy_oN<-sample(rcpUoN,U_size, replace = TRUE, prob = NULL)

     testw_size<-wilcox.test(Ux_oFF, y = Uy_oN,
              alternative = c("two.sided"),
              mu = 0, paired = FALSE, exact = NULL, correct = FALSE,
              conf.int = FALSE, conf.level = 0.95)

     ux<-as.numeric(testw_size["statistic"])
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
    prop_yx[i]<-ux/(length(Ux_oFF)*length(Uy_oN))
}

P_noether<<-mean(prop_yx)
```

H.3.3 Wilcoxon rank-sum test

```
Sample size two-sided----

w_samplesize<-((z_crit_1side_w+z_crit_bet_w)^2)/(6*((P_noether-0.5)^2))
w_samplesize<-round(w_samplesize)+1

#Avids too large sample size
if((abs(P_noether-0.5))<0.1){w_samplesize<-100}

#y1_oN<-RSSI2_ON_WSW00[,3]
y1_oN<-na.omit(rcpUoN)

##END SAMPLE SIZE-----


  w_boot_incrmnt[j]<-j
  x_oFF<-sample(x1_oFF, w_samplesize, replace = TRUE, prob = NULL)
  y_oN<-sample(y1_oN, w_samplesize, replace = TRUE, prob = NULL)


#Creating raking the mixed vector (X;Y)

xy<-na.omit(c(x_oFF,y_oN))
xy<-sort(xy)
rank_xy<-vector()

#vectors with x-ranks and y-ranks

rank_x<-as.vector(rank_x)
rank_y<-as.vector(rank_y)
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
#————————— NORMALITY TEST AND U values———

sum_rank_x<-cumsum(rank_x)
# takes only the last value
wx_oFF<-sum_rank_x[length(sum_rank_x)]

sum_rank_y<-cumsum(rank_y)
# takes only the last value
wy_oN<-sum_rank_y[length(sum_rank_y)]

#U=w-n(n+1)/2

Uy_oN<-wy_oN-(length(y_oN))*((length(y_oN)+1)/2)
Ux_oFF<-wx_oFF-(length(x_oFF))*((length(x_oFF)+1)/2)

Uy_oN_vec[j]<-Uy_oN
Ux_oFF_vec[j]<-Ux_oFF

#Effect size: xy=x>y; yx=y<x
effect_xy<-Uy_oN/(length(y_oN)*length(x_oFF))
effect_yx<-Ux_oFF/(length(y_oN)*length(x_oFF))

#Adjustment for ties only calculated if there are ties
if(length(repeated_xy)>0){

for (i in 1:length(repeated_xy)) {
  count_repeat_xy<-1 # counts repetitions for Variance correction
#if classe di contain x(1)=y(5) it means 2 elements in the class di, hence

    for (d in 1:length(xy_1)){

    if (repeated_xy[i]==xy_1[d]) {
      count_repeat_xy<-count_repeat_xy+1
      }
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

```
        di_xy[i]<-count_repeat_xy
        di3_xy[i]<-(di_xy[i])^3-di_xy[i]
      }
}

}

# Maximum proportion of tied value

if(length(repeated_xy)>0)
{prop_ties_xy<-max(di_xy)/N_xy}

## NORMAL APPROXIMATION

mn_xy<-N_xy/2  #mn_xy = n=m sample size
var_ties_wy<-(mn_xy^2)*(N_xy+1)/12-(mn_xy^2)*(sum(di3_xy)/(12*N_xy*(N_xy-1)
sd_ties_wy<-sqrt(var_ties_wy)

#mean of wy same for wx
E_xy<-0.5*(mn_xy)*(N_xy+1)

# The Z value
Zvalue_tie_wy<-(wy_oN-E_xy)/sqrt(var_ties_wy)
```

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub