Assessment of Treatment Influence in Mobile Network Coverage on Board High-Speed Trains

ORLANDO S. TRINDADE1, Taulant Berisha2, Philipp Svoboda©1, (Senior Member, IEEE), Efstathia Bura1, and Christoph F. Mecklenbräuker©1, (Senior Member, IEEE)
1TU Wien, 1040 Vienna, Austria
2Dimetor GmbH, 1040 Vienna, Austria
Corresponding author: Philipp Svoboda (philipp.svoboda@tuwien.ac.at)
This work was supported by the TU Wien Bibliothek through its Open Access Funding Program.

ABSTRACT In this paper, we investigate the systematic change in the signal quality of mobile telecommunications systems inside a high-speed train. Using commercial mobile phones, we perform measurements inside a carriage of a high-speed train with an installed repeater and inside a carriage of a different high-speed train equipped with low radio frequency attenuation windows. We carry out additional reference measurements inside regular carriages of a high-speed train. We use the adjusted Harrell–Davis method and the nonparametric bootstrap to test multiple quantiles of samples from the key performance indicators produced by the commercial mobile phones. In addition, we propose a common framework to standardize future comparisons on such studies. The results confirm considerable improvements in signal quality resulting from both the repeater and low RF attenuation windows.

INDEX TERMS Vehicular channel characterization, inferential statistics, mobile network measurement, mobile network service quality, RF signal, high-speed trains, multiple hypotheses, multiple quantile testing.

I. INTRODUCTION
Over the last decades, mobile terminals have become an essential communication device. The current mobile network infrastructure supports large-scale economies in a variety of ways. Commuters and travelers depend on reliable wireless connections to perform a variety of tasks that used to be tied to fixed access points. This is especially true for professionals and commuters who use High-speed Trains (HSTs) to travel for long periods. In this sense, it is critical for mobile service providers to monitor their mobile network periodically to assure that consumers are provided with a dependable, secure, and high-quality service network [1]. Enabling mobile service providers to accurately access the improvements of their mobile services ultimately determines their ability to enhance user experience [2]. Key Performance Indicators (KPIs) are the metrics monitored in network infrastructures and Mobile Station (MS). The KPIs are measured through the physical and logical channels of the mobile communication protocol to enable providers to deploy, operate, and improve mobile telecommunication networks efficiently.

KPIs are metrics estimated by the network infrastructure and the MS based on the measurements derived from the physical and logical channels of the mobile communication protocol stack. Large-scale Key Performance Indicator (KPI) measurements and advanced statistical analyses are necessary to deploy, operate, and improve mobile telecommunication networks efficiently.

The KPI statistics of a mobile communication network vary over space and time; KPI statistics also change due to network improvements, varying loads, and disruptions in mobile networks.

However, a statistically sufficient characterization of channels under High-speed Train (HST) context is yet to be proposed [3]. It is crucial to detect even small changes in the statistical properties of KPIs that are noticeable by commercial users of mobile services in the network.

Channel characterization becomes challenging due to: 1) the variety of frequency bands of mobile communication systems; 2) the velocity of the HST under investigation peaks around 250 km/h, and forecasts indicate that this velocity would further increase in the future; and [4]; 3) the large number of mobile users on board HSTs. Accordingly, network and transceiver designs in the several layers of the protocol stack need to be improved to meet these challenges [5].

Some of the significant issues on the physical layer of the protocol stack include small-scale fading, large-scale fading, and the time and frequency selectivity [5]–[7]. On the higher layers of the protocol stack, the so-called signaling storm is caused by the large number of mobile users that simultaneously access the network, which overloads the available signaling resources [5]. The diverse nature of those challenges impede mobile network providers from delivering high-quality services to consumers while on board HSTs. As a result, this motivates researchers to conduct measurement campaigns in an attempt to address those challenges. Evaluations performed in [8] report degradations in the form of packet drops, network disconnections, and Round Trip Time (RTT) at velocity ranging 200–300 km/h. A similar study in [9] has found sufficient throughput up to 100 km/h; however, service quality significantly degenerated at velocities of 300 km/h and beyond.

Commercially available solutions to improve network on board HST vary in applicability, achievable gains, and complexity. We first distinguish between active and passive network improvements. Among the active improvements, there are the amplify-and-forward repeaters [10], which are transparent to the protocols, and the decode-and-forward repeaters [11]. Passive improvements constitute structural changes that reduce signal penetration loss [12], [13]. The work in [14] attempted to optimize handover. The studies in [15] and [16] propose an improved resource-scheduling scheme to reduce Quality of Service (QoS) problems. These studies evaluate the proposed scheme under static-, low-, and high-mobility contexts. In terms of statistical evaluation, the majority of the previously mentioned analyses have relied solely on descriptive statistics of the KPIs. An exception is the study in [17], where the authors suggest using nonparametric hypothesis testing and then divide the traveled trajectory into kilometer-long segments in order to evaluate the collected KPIs.

A. CHALLENGES

The KPIs in mobile networks vary at different time-space scale, and this poses the first challenge in KPI evaluation. The variations in a KPI due to a specific modification in a mobile network do not occur in isolation, but results from a combination of many different factors. Some of these factors are due to variations in space and time of the propagation channel, in the traffic load in the mobile network, and in the measurement sample [18], [19]. Secondly, the characteristics of the implemented modification need to be first evaluated before the KPIs of interest are selected. KPIs differ from one another in terms of the underlying statistical distribution and data type. This becomes a challenge for researchers when the KPIs need to be compared among diverse studies. Thirdly, the length of the trajectory, velocity of the vehicle, and the KPI’s recording rate produces large data sets with large variations in KPI samples. Some KPIs have sample sizes in the order of tens of thousands, whereas others have just a few hundreds, therefore making it difficult to perform statistical comparisons. Furthermore, the mobile networks contain several Radio Access Technologies (RATs) that have discontinuities along the sampled trajectories [20].

For operational reasons, we cannot control the factors mentioned above, and any statistical inference concerning a KPI needs to be robust against these challenges. For this reason, we do not assume any specific form of KPI distribution. This motivates the use of nonparametric methods [21] to detect and infer changes in the KPIs of mobile communication networks under operational constraints and high mobility. In addition, the lack of standard terminologies among studies that focus on detecting change in mobile networks is another constraint. This undermines the comparability of results from the current literature and those from new studies. These conditions make it difficult for researchers to develop a protocol-independent method that can detect changes in the QoS of the mobile network as perceived by the user.

B. CURRENT WORK

We rely on inferential statistics as it enables comparing probability distributions of the underlying data generating processes.

In addition, we introduce terminology to standardize comparisons among similar studies. Likewise, the present contribution proposes an empirical KPI assessment methodology based on distribution-free hypothesis testing combined with the bootstrap [22], [23]. More specifically, the proposed scheme uses the adjusted Harrell–Davis (HD) estimator [24] developed by Wilcox [25] which can compare multiple quantiles of two distributions simultaneously. This method helps us to determine more accurately whether the distributions differ and how they differ from each other.

In the experiment, the unmodified HST carriage is called the placebo carriage. In an experimental setup, KPI samples are collected from a modified carriage and from a placebo carriage. We performed measurement in two experimental setups: repeater and low Radio Frequency (RF) attenuation windows. We compared the KPI distribution under the influence of a repeater and the KPI distribution under the influence of low RF attenuation windows with that of the placebo. The statistical analysis indicates that signal quality significantly improved.

This paper is organized as follows: Sec. II defines the proposed terminology, formulates the problem statement, and describes the collected data. Section IIIA discusses our method for analyzing the data, whereas Sec. IV describes two case studies where the proposed method is applied. The results are discussed in Sec. IV-F, and we conclude with some remarks.

II. ASSESSING THE KEY PERFORMANCE INDICATORS OF MOBILE COMMUNICATION SCENARIOS

Conducting an effective empirical assessment of KPIs requires understanding the phenomena that induce the changes in these KPIs. In the following subsections, we introduce a terminology from inferential statistics, describe the mobile communication scenarios, and characterize the collected data.
A. DEFINITIONS AND TERMINOLOGY
The following definitions characterize various measurement contexts in a mobile network. Figure 1 illustrates two different contexts. Fig. 1(a) shows the controlled condition where there is no mobility. Under the controlled condition, it is possible to isolate variables to a large degree. The uncontrolled condition is illustrated in Fig. 1(b). In this scenario the mobile network infrastructure is static and the users are mobile. In context (b), the variable isolation is limited and the samples display non-stationary statistical properties.

We further define the velocities of the transceivers \(v(p_1)\) and \(v(p_2)\) and the relative velocity as \(v_r = v(p_1) - v(p_2)\). When the transceivers are fixed, \(v_r = 0\), we name it as static condition which is identical to the controlled conditions in Fig. 1(a).

The non-static condition depicts the case in which either \(p_1\) or \(p_2\) moves along the traveled trajectory and \(v_r \neq 0\) as illustrated in Fig. 1(b). In terms of space, we define \(V\) as a volume equivalent to an indoor environment. When \(V\) is large, it can be further divided into spatial sections to improve the representation of the context under evaluation.

Figure 2 illustrates the modifications done on the mobile networks or on a vehicle similar to those shown in Fig. 1. Any systematic modification is called a treatment and is denoted as \(Q\). The context without systematic modification is called placebo \(q\). The effect of a treatment is called treatment effect and its magnitude is the effect size. Assessing a treatment effect requires comparing it with at least one of the following: 1) reference values, 2) absence of treatment or placebo \(q\), 3) an alternative treatment e.g. \(Q_1\). In addition, we use \(y\) to indicate the observations from a treatment and \(x\) to denote placebo-related observations. The treatments, samples, and the comparison with reference values are shown in Fig. 2(d).

A treatment without active elements is called passive treatment. The effect of such treatment is related to the changes in the characteristics of volume \(V\), such as building materials and internal and external geometries [26], [27]. The treatment effect, which in this case is the Vehicular Penetration Loss (VPL), is estimated by the signal power levels \(P_{\text{inside}}\) and \(P_{\text{outside}}\) [28]. In controlled conditions, we control the transmitted signal power \(P_{TX}\) at the transmitting antenna. Thus, we get a good estimate of \(P_{TX}\), namely \(\hat{P}_{TX}\), instead of using the estimated \(\hat{P}_{\text{outside}}\) [29]. Since the volume’s walls represent the treatment \(Q\), we can estimate the treatment effect by

\[
\hat{L}_Q(\text{dB}) = -10 \log_{10}\left(\frac{\hat{P}_{\text{inside}}}{\hat{P}_{TX}}\right) (\text{dB}) \quad (1)
\]

where \(\hat{L}_Q\) represents the transmission loss \(L_Q\) due to a passive treatment.

A treatment composed of active systems that superimposes power gain on the signal levels is called active treatment. In the controlled condition, we assess the treatment effects via the signal power levels and express it as a gain \(\hat{G}_Q\).

\[
G_Q(\text{dB}) = 10 \log_{10}\left(\frac{\hat{P}_Q}{\hat{P}_{TX}}\right) (\text{dB}) \quad (2)
\]

The measure \(L_Q\) in (1) is similar to the signal transmission losses in buildings [30] or vehicles [13], [31]. Both \(L_Q\) and \(G_Q\) are compared to referential values, e.g., measurements inside an anechoic chamber [13].

The variables \(L\) and \(\hat{P}\) are empirically derived from measuring power levels via averaging or a computation of a given quantile. In this study, the passive treatment is a Frequency Selective Surface (FSS), and the active treatment is a repeater.

Measurements performed via commercial mobile phones allow mobile network operators to assess the KPIs of interest. The number of observations available for a given KPI depends on the recording rate \(r\) and on the duration of the measurement \(d\). The recording rate is the number of KPI points recorded per second. The product, \(r\times d\), represents the expected observations from a measurement. We refer to the observed data as recorded observations. The reliable observations result from data preprocessing. In measurements performed under uncontrolled conditions, given a KPI and its recording rate, we define spatiotemporal granularity as the number of KPI points spread along the traveled trajectory. A mobile phone is said to be in a connected state when it performs a task that actively generates user data, e.g., a download or a phone call. Further, a mobile phone is said to be in an idle state when it is registered and connected to the mobile network without any active user data transfer.

B. THE NEED FOR A NONPARAMETRIC APPROACH
Fig. 1 (a), shows the controlled condition in which the mobility state can be controlled through constant distance \(p_1 \rightarrow p_2\) and has no velocity. The signals can be controlled through transmitted power levels, angle of incidence, and polarization of the RF signal. Meanwhile, the propagation conditions can be predesigned and controlled through reflective and dispersive behavior. As a result, we can fully control the distribution of the received signal by limiting the physical conditions in the experiment. This will ultimately control the statistical properties of the observed KPIs. Then, we compare with the KPI probability distributions inferred for the no-treatment scenario to assess the treatment effect.

Figure 1 (b) depicts a context that is characterized by non static and uncontrolled conditions, which constitute a real-world operational situation of transceivers inside HSTs. In this operational context, the users on board HSTs move at trajectories \(s(x, y, z)\) while being served by the mobile network. We assume that the Base Station (BS) is at a fixed position \(p_1(x, y, z)\), and the user is at position \(p_2(x, y, z)\) on board the moving HST. The relative instantaneous velocity \(v_r = v(p_2(x, y, z))\) along the trajectory contains stops, accelerations, and intervals of steady velocity.

The velocity of HSTs causes fast and frequent transitions in propagation conditions, that is, hilly and flat terrains, combinations of high and low building, and different
mobile network deployments. In railway tracks with dedicated mobile network BSs, the distance between HSTs is within few hundreds meters; thus, large blockages cause rapid transitions in the Line of Sight (LOS), strong multipath components, and Non Line of Sight (NLOS). The conditional distribution of the signal received by MSs inside HSTs changes under LOS and NLOS, respectively. The position of the BSs relative to the HSTs is considered to be random as there is no dedicated deployment for railway coverage. Hence, multipath and LOS propagation conditions are dominant in urban and rural geographical regions, respectively.

The velocity of the HSTs cause fast transitions in the propagation condition and it can be reasonably assumed that the received signal is characterized by a mixed distribution. We assume that a shorter travelled trajectory would decrease the number of such combinations. The treatment effect is then superimposed on effects from the propagation conditions. In such experimental setting, the resulting data do not have an easily identifiable or tractable distribution. Hence, we rely on a nonparametric approach and consider the distributions as unknown.

Mobile network resources are not uniformly distributed in space. That is, the available resources vary along the travelled trajectory. This variation is primarily due to the differences in the characteristics of the network deployment between dense urban, suburban, and rural regions. In this case, the spatial variation in radio resources results in variations in the QoS. The user perceives this variation through conditions such as No Service (NoS), call drops, insufficient data throughput [19], or frequent Radio Access Technology (RAT) handover [20]. Network providers have little or no control over the variables being measured when the KPIs are being assessed under operational conditions. In addition, accessing some KPIs requires a special setup such as that in [32], which is not feasible in HSTs.

Mobile operators have full control of their mobile networks and can collect KPI samples from both the infrastructure and the user side. The combined analyses of both sides is of
strategic interest for a mobile operator. Therefore, the mobile operator wishes not to disclose KPI observations at the infrastructure side. Since we aim to assess the treatments regardless of any specific mobile operator, we are then forced to base the assessment solely on KPI samples at the user’s side. In this way, Fig. 2 (a) and (c) represent the user’s side. The mobile network side – e.g. BSs – is represented by Fig. 2 (b). In this study, we compare the KPI distribution under the treatment context with that under the no treatment, given that both distributions are unknown. Also, the distribution of $\Delta x_Q$ is unknown. In the following paragraphs, we discuss some possible changes in the KPI distributions resulting from treatment $Q$.

a) **Treatment effect $\Delta x_Q$ has no statistical significance.** At a given level of significance, we cannot distinguish between the two KPIs that are with and without treatments.

b) **Treatment effect $\Delta x_Q$ changes the mean of $x_q$, i.e., $\hat{M}_Q = \hat{M}_x + \Delta M_x$, where $\Delta M_x$ is a constant.** Here, we infer the KPI distribution of $y_Q$ is related to the KPI distribution of $x_q$ by a deterministic shift.

c) **Treatment effect $\Delta x_Q$ impacts the variance of $x_q$, i.e., $\hat{\sigma}^2_{\Delta Q} \neq \hat{\sigma}^2_{x_q}$.** Consequently, the $y_Q$ and $x_q$ differ.

### C. DATA DESCRIPTION

Figure 3 illustrates specific attributes of the KPI samples considered in this work: namely, the mobile network operator, radio access technology, frequency band, and physical channel.

We analyse KPI distributions for specific choices of the attributes. The required filtering according to Fig. 3 is parametrized in the measurement setup and performed during data preprocessing. In this manner, we obtain a data set for one chosen set of attributes, which is then used to infer the treatment effect. The data set used in this work is attributed by one network operator, the RAT 3G, frequency band 1, and the downlink direction, cf. Fig. 3.

The mobile phone, MS, is a system that recovers information from the application layer contained in $g(t, \omega, s)$, e.g., voice, text files, pictures. In order to accomplish this task, the MS exchange a large amount of information and many KPIs are computed from both sides. The *data array* $D$ in (3) is a metadata file that contains records of the computed KPIs and mobile network information, e.g., system information, user information.

$$
D = \begin{bmatrix}
A_{i1}, & a_{12}, & a_{13}, & \ldots, & a_{1n} \\
C_{i1}, & c_{22}, & c_{23}, & c_{24} \\
D_{i1}, & d_{32}, & d_{33} \\
A_{i11}, & a_{42}, & a_{43}, & \ldots, & a_{4n} \\
C_{i1}, & c_{52}, & c_{53}, & \ldots, & c_{5n} \\
F_{i1}, & f_{62}, & f_{63}, & \ldots, & f_{mn}
\end{bmatrix}
$$

We represent the data array from the contexts with and without treatment by $D_Q$ and $D_q$, respectively. A representation of the treatment effect through the data array is

$$D_Q = D_q + \Delta D_Q, \quad \text{(4)}$$

where $\Delta D_Q$ is the treatment effect in the data array.

The rows in (3) list $A, B, \ldots, E$ as *measurement events*. Both the measurement events and their content depend on various factors, such as the type of task performed on the application layer, e.g., file transfer, voice call, and position. Other factors are events from the mobile network during the measurement, e.g., absence of signal and RAT coverage. Under optimum conditions, the measurement events are recorded at a fixed interval. The outcome of a KPI is an element of a measurement event and appears in the same column of (3). For instance, a sample from a KPI $(a_{ij})$ in the measurement event $A$ is

$$A_{ij} = (a_{12}, a_{42}, \ldots, a_{m2})^T. \quad \text{(5)}$$

Here, the samples of a selected KPI from the contexts with and without treatment are $x_q$ and $y_Q$, respectively,

$$x_q = (x_1, \ldots, x_n)^T, \quad y_Q = (y_1, \ldots, y_m)^T. \quad \text{(6)}$$

Using (4), we write the expression of the KPI under the treatment as

$$y_Q = x_q + \Delta x_Q. \quad \text{(8)}$$

where $\Delta x_Q$ is the treatment effect observed via a selected KPI $x$.

As the duration of the measurement increases, so does the length of the KPI vector in (6). A mobile phone connects only to one RAT at a time, and the distribution of the RAT is not necessarily continuous along a travelled trajectory. Given a RAT dependent KPI, the sample size of this KPI decreases by the proportion of the RAT available along the travelled trajectory, as depicted in Fig. 8. Computing a KPI depends on the parameters measured by the MS and on the values reported by the BS. However, mobile signal coverage gaps along the traveled trajectory are not uncommon and result in missing reports from the BS at many positions, which reduces the KPI sample size. The samples from the vector in (6) and (7) contain information on the temporal and spatial dynamics of the stochastic process.

In addition, train stations have dedicated cellular network that have better QoS than that along the railway. The trains that approach or stop at a train station are at least partially served by this dedicated cellular network. In this case, the spatial variation of the signal $g(t, \omega, s)$ is considerably smaller than the variation inside an HST that is moving steadily away from a train station. Both signal levels and service quality are better under these conditions. Consequently, a portion of the KPI samples in $x$ and $y$ contain only temporal variations, which results in better signal and service levels.
III. DATA ANALYSIS

For the statistical evaluation, we use a nonparametric statistical approach based on multiple hypotheses testing combined with bootstrap to test for treatment effect and to estimate the effect size. This approach assesses whether the treatment $Q$ results in different KPI distribution based on the collected samples.

In order to design the data evaluation approach, we assume the following:

A1 The variations in the observed KPIs are considered random independent identically distributed within a given geographical region and a single travel direction.

A2 The KPI probability distribution only depends on treatment $Q$. Other factors can be neglected.

A3 The KPI distribution is unknown.

To satisfy assumption A1, the travel trajectory is divided systematically into geographical regions. This results in urban regions $S_{u1}, \ldots, S_{uk_u}$ and rural regions $S_{r1}, \ldots, S_{kr_r}$. The number of urban regions $k_u$ and rural regions $k_r$ depend on the length of the trajectory $s(x, y)$, among other features. Selecting a geographical region results in a pair of data sets $D_Q$ and $D_q$ which contain samples from $y_Q$ under treatment $Q$ and samples $\hat{y}_q$ under treatment $q$ (placebo).

A. ESTIMATION OF TREATMENT EFFECT

The adjusted HD method is a nonparametric approach for estimating the difference between sample quantiles of independent groups [24], [25]. This method uses the HD estimate [24] to compute the $p$th sample quantile $F^{-1}(p) = \theta_p$. Then, the difference between quantiles is computed via multiple comparisons.

Let a pair of quantiles from the samples with and without treatment be $\hat{\theta}_{pq}$ and $\hat{\theta}_{pQ}$, respectively. We estimate the difference between those quantiles as

$$\Delta \hat{\theta}_p = \hat{\theta}_{pq} - \hat{\theta}_{pQ}. \quad (9)$$

The null hypothesis of no treatment effect is

$$H_0 : \theta_{pq} = \theta_{pQ}. \quad (10)$$

where $\theta_{pq}$ and $\theta_{pQ}$ are the true quantiles of the with and without treatment distributions, respectively. The function $qcomhd$ in R package WRS2 [33] tests whether the corresponding single or multiple quantiles of two distributions are the same. That is, it computes (9), tests (10) and provides the $1 - \alpha$ confidence interval.

In order to specify the sample size for this estimation, we define a time window of 150 seconds in the measurement data. The samples from the desired KPI inside this time window form the sample size. The time window corresponds to the duration of the longest preprogrammed task and maintains partial agreement with recommendations in [34]. This results in 300 and 450 scores for the recording rates of 2 and 3 samples/sec, respectively. When the available number of scores in a KPI vector is smaller than 300, the entire sample is used. In order to perform the test, the function $qcomhd$ produces bootstrap replicates with size equal to the size of the input data. For example, given input data of 100 scores, the function produces 2000 bootstrap replicates with size 100. Our study also includes cases where the size of the input data is as large as 5000, and we exploit this large sample by extracting subsamples with smaller size. For this reason, we adjust the package to produce bootstrap replicates with prespecified sample size as the input in the function $qcomhd$.

B. STATISTICAL SIGNIFICANCE

The testing design in [25] is based on the null hypothesis significance testing [35]. A Type I error occurs when the null hypothesis is true, but is wrongly rejected. The prespecified statistical significance, $\alpha = 0.05$, represents the acceptable probability to commit Type I error. In [25], the empirical $p$-value, that is the empirical probability of type
I error or \( \alpha \), is defined as

\[
\hat{p} = \frac{A + 0.5C}{B},
\]

(11)

where \( A \) is the number of times that \( H_0 \) is rejected, \( C \) is the number of times that \( H_0 \) is not rejected, and \( B \) is the number of bootstrap replicates.

Using (11), we further compute the \( p \)-value as \( 2 \min(\hat{p}, 1 - \hat{p}) \). The HD method applies multiple comparisons, and the overall significance level is controlled via the classical Bonferroni method [36], in which the prespecified statistical significance is adjusted as

\[
a_n = \frac{\alpha}{n}, \quad \forall n = 1, \ldots, k,
\]

(12)

where \( k \) is the number of tested quantiles. The null hypothesis is rejected if, and only if, at least one \( p \)-value satisfies

\[
p\text{-value}_n < \alpha_n.
\]

(13)

IV. CASE STUDY

In this case study, we apply the aforementioned nonparametric statistical approach based on multiple hypotheses testing to investigate the advantages of two commercial solutions to improve mobile network services inside HSTs, see Fig. 4.

B1 A passive treatment based on modified windows equipped with a FSS. Fig. 5 depicts a FSS [30] that reduces the VPL of HSTs and hence improves mobile network services inside HSTs.

B2 An active treatment using rooftop antennas and Amplify-and-Forward (AF) repeaters. In Fig. 6 the AF repeater creates a path that bypass the VPL in HSTs. Due to its external antenna, the repeater also allows users inside HSTs to access mobile network infrastructures near and distant from the railway track.

The Device under Test (DUT) is the entire 210 -meter-long HST which consists of individual carriages of 26.5 -meters each. Each treatment is prototypically applied to carriages of two separate railjet trains [37]. The Measurement Entity (ME) shown in Fig. 7 is deployed on board railjet trains. Selecting and applying treatments \( Q \) and \( q \) to two different carriages yields \( V_Q \) and \( V_q \). In the following sub-sections we further explore those treatments and their respective effects.

A. PASSIVE TREATMENT: FREQUENCY SELECTIVE WINDOWS

Current HSTs are equipped with metal-coated windows to provide thermal and visual comfort to passengers. As a side effect, the metal coating attenuates the wireless signal transmissions of cellular phones and Global Positioning System (GPS) up to 40 dB. The untreated HST is fully equipped with lightly coated windows, which are referred to as RW [12]. In general, a FSS consists of a two-dimensional periodic geometric structure composed of metallic and dielectric materials [30], [38]. The geometric structure defines the effective electromagnetic properties, i.e. permittivity, conductivity, and magnetic permeability.

The transmission of electromagnetic waves through a FSS is highly linear over a large dynamic range of wave amplitudes. Its transfer function depends on the frequency, polarization, and angle of incidence of the waves [12], [39], [40]. An engineered FSS functions as a frequency selective absorber, passive repeater [41], filter, or reflector [42]. The passive treatment consists in equipping a section of a HST with a FSS window, named SW, that yields a wide-band bandpass filter within the frequency range 0.7—3.5 GHz [43]. This SW features substantially smaller RF attenuation than a RW does. The Austrian Federal Railways partially upgraded a single carriage of a HST with SWs to evaluate how this modification improves the QoS of the mobile network inside the HST.

Using (1), we estimate the transmission loss of a HST without treatment by

\[
\hat{L}_q = -10 \log_{10} \left( \frac{\hat{P}_q}{\hat{P}_{TX}} \right) \text{(dB)}
\]

(14)

and the loss of a HST with treatment by,

\[
\hat{L}_Q = -10 \log_{10} \left( \frac{\hat{P}_Q}{\hat{P}_{TX}} \right) \text{(dB)}.
\]

(15)

Here \( \hat{P}_q \) and \( \hat{P}_Q \) are the received signal power estimates inside the HST.

To arrive at an equivalent representation for (8), we subtract (14) and (15) to express the estimated difference in transmission loss \( \Delta \hat{L} \),

\[
\Delta \hat{L} = \hat{L}_q - \hat{L}_Q \text{(dB)}.\]

(16)

Taking the difference ensures that the value of \( \hat{P}_{TX} \) cancels, as well as the antenna characteristics, cable losses, and many other setup specific properties. In the context with treatment, the reduced transmission loss increases the received signal power levels. The condition that attributes improvement to the current treatment is expressed as

\[
\Delta \hat{L} > 0 \iff \hat{L}_q > \hat{L}_Q.
\]

(17)

The improved signal power level can be represented as

\[
\hat{P}_Q = \hat{P}_q + \Delta \hat{P},
\]

(18)

where \( \hat{P}_q, \hat{P}_Q \) are quantile estimates of the KPI samples and represent the received signal power levels.

B. ACTIVE TREATMENT: AF REPEATER

The active treatment consists of deploying Moving Relay Nodes (MRNs) on board the HST. A Moving Relay Node (MRN) is a system composed of rooftop mounted donor antennas, AF repeaters, and coaxial leaky cables [10]. A single omnidirectional antenna is used as donor antenna for a single AF repeater.

The AF repeater amplifies signal and noise at selected signal bands in both the up- and the downlink of the Frequency Division Duplex (FDD) mobile network. The bands
for amplification are selected by digitally programmable filters. The filter transfer functions have steep transitions between the pass- and stopbands which are associated with increased group delay [44].

The AF repeater is a device with non-linear behavior because there is a maximum output power level due to physical limits of the RF circuit and its power transmission limits. Furthermore, the AF repeater is a Single-Input Single-Output (SISO) device that uses a single antenna. This fundamentally influences the achievable data rates for high-speed mobile network services that employ Multiple Input Multiple Output (MIMO) techniques using multiple antennas at the BS and MS.

The active treatment creates an additional propagation path between the outside and the inside of the HST in parallel to the passive propagation paths through the windows. The performance of this active treatment strongly depends on the optimization of the AF repeater’s configuration as well as the scenario under test, e.g. base station deployment along the track [45].

Similarly to Sec. IV-A, the gain under active treatment is written as

$$\hat{G} = 10 \log \left( \frac{\hat{P}_Q}{\hat{P}_q} \right) \text{ (dB)}.$$  \hspace{1cm} (19)

The condition that attributes improvement to the current treatment is

$$\hat{G} > 0 \iff \hat{P}_q < \hat{P}_Q.$$  \hspace{1cm} (20)

$\hat{P}_Q = \hat{P}_q + \Delta \hat{G},$  \hspace{1cm} (21)

where $\hat{P}_q$, $\hat{P}_Q$ are quantile estimates of KPIs samples and represent the received signal power levels.

In the equations above, the notation $\hat{P}$ is a selected KPI that represents an estimate of the received signal power level.
C. MEASUREMENT SETUP

We carried out measurement campaigns in two railjet trains [37]. The train with a carriage equipped with SW traveled from Vienna to Graz (198 km, 3 hour duration), while the other with a carriage equipped with an AF repeater traveled from Vienna to Salzburg (305 km, 3 hour duration), as detailed in Fig. 8.

The ME in Fig. 7, consists of a group of six commercial off-the-shelf smart phones. For our measurement, we divided the smart phones in three groups of two. The first group was locked to 2G RAT only. The second group was locked to 2G and 3G RATs. The third group had access to 2G, 3G, and 4G. Each mobile was equipped with dedicated software (Anite Nemo Walker Air) [46] which was configured to repeat a fixed sequence of tasks. The tasks consisted of one voice call with total duration of 120s, one download and one upload both with 10 MB data volume. The tasks emulate user activities in quasi-real usage conditions. In both measurements, one ME was deployed inside section $V_Q$, where treatment $Q$ predominates, and another ME in section $V_q$ under the predominance of treatment $q$. The ME is placed on a passenger’s seating, as shown in Fig. 7.

D. VARIABLES OF INTEREST

We expect to detect a statistically significant treatment effect in any selected KPI under the direct influence of the treatment. Different KPIs show different sensitivities to the treatments. Table 1 displays the group of KPIs used to assess the treatment effect.

Both the AF repeater and the SW are expected to improve link quality at the physical layer. For this reason, the treatment effect is expected to be associated with the metrics from this layer. Based on this association, three KPIs report signal levels in the downlink: Received Signal Strength Indicator (RSSI), Received Signal Code Power (RSCP) and received Signal-to-Interference Ratio (SIR).

In Universal Mobile Terrestrial Services (UMTS), the User Equipment (UE) reports the RSSI and RSCP according to the Third Generation Partnership (3GPP) standards. The RSSI report measures the total received signal plus interference and noise power within the signal band of interest. It also serves as a rudimentary indicator of the power level of the signal of interest [47] and forms the basis for estimating more specific link quality-related KPIs [48]. The selected RSCP from the primary cell is such a more specific measure of link quality. The SIR indicates the signal quality and is estimated via the Dedicated Physical Control Channel (DPCCH) [47]. It is a metric that contains interference from co-channels and neighbor cells. Therefore, the variation in this KPI depends on the radio resource control and network topology (among others). In MRN contexts, the SIR is affected by the multipath components passing through the vehicle frame [49].

E. DATA PREPROCESSING

The size of the data sets depends on the length of the regions, the average velocity of the train, and the recording rate. Before the statistical evaluation of the KPIs discussed in Sec. IV-C, we filter the data sets with respect to mobile network operator and RAT of interest, cf. Sec. II-C. In addition, we select the geographical regions from the travelled trajectory. The 3G (UMTS) network offers continuous coverage along the travel trajectories and sufficient spatial granularity of the data sample.
We select the group of mobile phones that was configured to access 3G preferably and 2G alternatively (see Sec. IV-C) for a single operator. The mobile phones repeatedly execute the tasks described in Sec. IV-C, and hence they remain predominantly in a connected state (only 2.8% in idle state along the route Vienna to Graz). The route Vienna to Salzburg traverses a well-covered flat terrain, thus contribution of the samples from an idle state is neglected in the statistical evaluation from both tracks. The stops at train stations are part of the regular train commute, and we have not imposed any special treatment of those.

With support from the Eurostat [50], we classify segments of the travelled trajectories Vienna—Graz $s_1(x, y)$ and Vienna—Salzburg $s_2(x, y)$ into regions as described in Sec. III. Figure 8 shows this classification, where each region corresponds to data sets with treatment $Q$ and without treatment $q$.

**F. KPI HISTOGRAMS**

Figures 9 and 10 present histograms of KPIs comparing data with and without active treatment in urban and rural regions. In these histograms, we observe that the treatment effect imposes a considerable shift on the lower quantiles and imposes truncation due to the limited output power of the AF repeater on the upper quantiles. The KPI histograms with treatment indicate left-skewed distributions, upper-bounded, and show a large overlap with the histograms without treatment.

Figures 11 and 12 show histograms of KPIs comparing data with and without passive treatment in urban and rural regions. Here the histograms show that the treatment effect imposes a moderate shift on the lower quantiles and a small shift on the upper quantiles. The histograms indicate that the distributions with and without passive treatment share a large overlapped region.

**G. EVALUATION FOR TREATMENT EFFECT**

Tables 2 and 3 summarize the changes in the distributions from both active and passive treatments and attribute statistical significance to these.

Traditionally, the median is used, for assessing changes in KPIs from the physical layer of mobile communication systems from measurements.

The treatment effect of the active treatment assessed via the median is summarized in Table 2. Its magnitude for the RSCP varies from 12.38 dB to 15.12 dB and from 16.08 dB to 20.02 dB in the urban and rural regions, respectively. The treatment effect of the passive treatment assessed via the median is summarized in Table 2. Its magnitude for the RSCP varies 1.80 dB to 7.97 dB and from 3.53 dB to 10.67 dB in the urban and rural regions, respectively.

Due to the characteristics of the active and passive treatments, the effect size that we evaluated solely via the median prevents the treatment effects from being detected in lower and upper quantiles. Indeed, the treatment effects from the active and passive treatments are not equal among the sample quantiles, as seen in Figs. 9–12. This behavior of the treatment effect is presented in Table 3 displaying results of multiple hypotheses testing carried out by the HD method.

In the following, we discuss the effect size for the 0.25 and 0.75 quantiles as summarized in Table 3. The column “Effect size” gives the 95% confidence interval for the shift in the data due to the treatment.

In the active treatment, the magnitude of the effect size for RSCP varies from 9.93 dB to 14.91 dB and from 15.56 dB to 27.93 dB in the urban and rural regions, respectively.

In the passive treatment, the magnitude of the treatment effect for RSCP varies between 3.53 dB and 14.08 dB for urban and from 5.22 dB to 22.79 dB for rural regions, respectively. The treatment effect in rural regions is more varied, i.e., the samples exhibit a wider range. This is due to the poor signal coverage in those areas. As a result, both treatments lead to larger improvements in those regions.

The values presented in Table 3 show that the treatment effect among the quantiles is unequal. This implies that both treatments change the KPI distribution. The null hypothesis of equal distribution is rejected for all evaluated quantiles.

The active treatment effect is more considerable in the rural region (Fig. 10) then in urban (Fig. 9). This is explained by the low signal levels in rural regions at the input of the AF.
repeater. The received signals are amplified with the desired gain because the RF amplifiers of the AF repeater operate far below its output power limit.

The passive treatment shows no output power limit. However, the treatment effect depends on distribution of Angle of Arrival (AoA) and polarization. The distribution of AoA and polarization differs between the urban and rural regions, as a result of the differences in deployment of mobile network and characteristics of the terrain between those regions. Multipath propagation conditions are dominant in urban regions, whereas LOS is dominant in rural regions.

The FSS used in the passive treatment imposes higher attenuation on the weaker multipath components with highly
FIGURE 10. Histograms of KPIs without (a, b, c) and with (d, e, f) active treatment ("Repeater") in rural region.

FIGURE 11. Histograms of KPIs without (a, b, c) and with (d, e, f) passive treatment ("Structured Windows") in urban region.
The analysis shows that the effect size depends on the characteristics of the treatment (active or passive), as well as the network deployment (rural or urban). The mobile service coverage in the urban regions is better than in the rural regions where signal levels are much lower. Therefore, the overall benefits of a treatment depend on the mix of networks deployments along the travelled trajectory.

The chosen segmentation of the travelled trajectory in terms of length (km) and type (urban vs. rural) influences the validity of assumption A2 in Sec. III, see Fig. 8. Short segments, e.g., 5–10 km, result in small spatial variations in the KPIs with treatment effect. Long segments, e.g., hundreds of kilometers, hide significant variations in the treatment effect on the KPI. This effect can be observed when comparing the RSCP statistics of the whole data set with the RSCP statistics in rural region for the active treatment.

scattered AoA. Accordingly, the passive treatment shows benefits in urban regions.

| Repeater (Active treatment) - Vienna to Salzburg | Data group | sample size(mm) | Bootstrap replicates | Medians $\hat{\theta}_Q$ | Medians $\hat{\theta}_G$ | Effect size | p-value | Rej. $H_0$ |
|---|---|---|---|---|---|---|---|---|
| RSSI urban | 5003 | 300 | -66.34 | -53.20 | [-14.18, -11.93] | < 0.001 | YES |
| RSSI rural | 3727 | 300 | -64.59 | -49.44 | [-16.28, -13.94] | < 0.001 | YES |
| RSCP urban | 1988 | 300 | -75.75 | -61.89 | [-15.12, -12.38] | < 0.001 | YES |
| RSCP rural | 1625 | 300 | -75.32 | -57.68 | [-20.02, -16.08] | < 0.001 | YES |
| RSCP track | 7219 | 300 | -74.52 | -62.56 | [-13.68, -10.43] | < 0.001 | YES |
| Rx SIR urban | 2629 | 300 | 4.72 | 6.13 | [-0.86, -0.00] | 0.05 | NO |
| Rx SIR rural | 3111 | 300 | 4.91 | 6.79 | [-2.54, -1.28] | < 0.001 | YES |

| Structured windows (Passive treatment) - Vienna to Graz | Data group | sample size(mm) | Bootstrap replicates | Medians $\hat{\theta}_Q$ | Medians $\hat{\theta}_G$ | Effect size | p-value | Rej. $H_0$ |
|---|---|---|---|---|---|---|---|---|
| RSSI urban | 5157 | 300 | -81.25 | -69.66 | [-14.52, -8.62] | < 0.001 | YES |
| RSSI rural | 4773 | 300 | -83.67 | -78.92 | [-7.97, -1.80] | < 0.001 | YES |
| RSCP urban | 2133 | 300 | -87.86 | -81.14 | [-10.67, -3.53] | < 0.001 | YES |
| RSCP rural | 915 | 300 | -96.51 | -85.68 | [-14.11, -7.62] | < 0.001 | YES |
| Rx SIR urban | 73 | 73 | -13.67 | -16.70 | [0.39, 5.15] | 0.03 | NO |
| Rx SIR rural | 131 | 131 | -8.02 | -8.65 | [-1.29, 3.70] | 0.48 | NO |

1 Rural and urban geographical regions; track - entire travelled trajectory.
2 Estimated medians with treatment Q and without treatment $\tilde{Q}$.
3 95% confidence interval for the treatment effect.
4 Provided that p-value < adjusted p-value.
The results presented in Table 2 show this bias: the effect size along the 64-km track in the rural regions is larger at approximately 18 dB, compared to an effect size of approximately 12 dB along the entire 305-km trajectory.

The treatment effect differs among the set of investigated KPIs. Some KPIs may consistently display no effect, in contrast to others. In our case study, the goal of the treatment is an enhanced signal level on board the HST.

Specifically, we emphasize that the results in Table 2 show that the effect sizes in the Rx SIR under both active and passive treatments are much smaller than in the RSCP and RSSI. This is partially explainable by the closed-loop transmit power control in the network for voice calls. This control loop varies the transmit power to achieve an operator-specified target value for Rx SIR.

In regions, where the interference level is lower than the noise level plus VPL, the Rx SIR on board the HST is close to the Rx Signal-to-Noise Ratio (SNR). This is true for both treatments: active and passive. Here, any treatment reducing the VPL has the potential of improving the measured Rx SIR on board the HST.

Active treatments can completely compensate the VPL in rural areas where the received signal levels are low, but come at the expense of adding some noise. Passive treatments can only partially compensate the VPL and they do not add noise. Therefore, the active treatment shows more considerable KPI improvements in rural regions than the passive treatment. In urban regions, neither treatment shows much improvement.

### V. CONCLUSION

In this paper, we inferred significant benefits of treatments in mobile network coverage on board high-speed trains (HSTs) in urban and rural regions through measurements of KPIs in Austria. Two treatments are assessed: an active treatment based on AF repeaters and a passive one using structured windows with low RF attenuation.

We use the modified HD method and nonparametric bootstrap to test multiple sample quantiles from collected key performance indicators (KPIs) for significance. This approach handles well the differences in mobility, regional conditions, and from the varying sample sizes along the railway track.

The results support the hypothesis that signal quality improves with both treatments. The improvement in the mobile network coverage on board the HST with the AF repeater is higher than that due to the installed low RF attenuation windows. The enhancements from both treatments are larger when the HST travels in rural regions.

Regarding the effect of the difference in signal quality, prior research has shown that QoS enhancement with both treatments is similar along the trajectories travelled in Austria. The experiments in this study cannot consider the effect of the network deployment along the railroad.

### ACKNOWLEDGMENT

The authors thank J. Resch and S. Ojak at the Austrian Federal Railways (OBB) for providing access to prototypical railjets equipped with Structured Windows (SW) and Amplify-and-Forward (AF) repeaters. The measurement campaign was funded by OBB.
ORLANDO S. TRINDADE received the bachelor's degree in electrical engineering and the master's degree in telematics from the State University of Campinas (Unicamp), São Paulo, Brazil. He is currently undergoing research with the Institute of Telecommunications, Technische Universität Wien, Vienna, Austria. Some of his research interests are stochastic modeling, industrial networks, Macro Raman spectroscopy, mobile networks quality assessment, vehicular communications, the IoT, and IIoT.

TAULANT BERISHA received the B.S. and M.S. degrees in electrical engineering from the University of Prishtina, Kosovo, in 2011 and 2014, respectively, and the Ph.D. degree in telecommunications from TU Wien, Vienna, Austria, in 2019. From 2015 to 2019, he was a Project Assistant with TU Wien, and since 2019, he has been a Researcher with Dimetor GmbH, Vienna. His research interests are in vehicular communications, BVLoS UAV operations, benchmarking, statistical modeling, and service quality measurements.

PHILIPP SVOBODA (Senior Member, IEEE) received the Dr.Ing. degree in electrical engineering from TU Wien. He is currently a Senior Scientist with TU Wien, with a research focus on the performance aspects of mobile cellular technologies. He is currently examining the feasibility of using crowdsourcing to conduct performance measurements on 4G and 5G mobile networks. His research aims to establish a common framework for evaluating the performance of mobile networks, guaranteeing reliable, and fair connectivity for end-users.

EFSTATHIA BURA received the B.S. degree in mathematics from the University of Athens, Greece, the M.S. degree in mathematics from the University of Illinois at Chicago, and the M.S. and Ph.D. degrees in statistics from the University of Minnesota, USA, in 1996. She served as an Assistant, Associate (with tenure), and Full Professor with the Statistics Department, George Washington University, and currently holds the Chair of Applied Statistics at the Institute of Statistics and Mathematical Methods in Economics, Faculty of Mathematics, TU Wien. She is an elected member of the International Statistical Institute and a Fulbright scholar, has published widely in methodological and applied statistics and has been serving in several editorial boards.

CHRISTOPH F. MECKLENBRÄUKER (Senior Member, IEEE) received the Dipl.-Ing. degree (Hons.) in electrical engineering from Technische Universität Wien, Vienna, Austria, in 1992, and the Dr.-Ing. degree (Hons.) from Ruhr-Universität Bochum, Bochum, Germany, in 1998. From 1997 to 2000, he worked at Siemens AG Austria and engaged in the standardization of UMTS. From 2000 to 2006, he held a senior research position with the Telecommunications Research Center Vienna (FTW), Vienna. In 2006, he joined TU Wien as a Full Professor. From 2009 to 2016, he lead the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility. He has authored approximately 250 papers in international journals and conferences, for which he has also served as a reviewer, and was granted several patents in the field of mobile cellular networks. His current research interests include radio interfaces for peer-to-peer networks (vehicular connectivity, sensor networks), ultra-wideband radio, and MIMO transceivers.

Dr. Mecklenbräuker is a member of the Antennas and Propagation Society, the Intelligent Transportation Society, the Vehicular Technology society, the Signal Processing society, as well as VDE and EURASIP. His doctoral dissertation received the Gert-Massenberg Prize, in 1998. He is the councilor of the IEEE Student Branch Wien.