Repeatability for Spatiotemporal Throughput Measurements in LTE

Vaclav Raida, Philipp Svoboda, Martin Lerch, Markus Rupp
Institute of Telecommunications
Technische Universität Wien
Vienna, Austria
{firstname.lastname}@nt.tuwien.ac.at

***Abstract*—Evaluating mobile broadband in an operational cellular network is a challenging measurement task that depends on many parameters. Throughput is an essential performance indicator in any data network. Particularly in cellular mobile networks, the throughput is very sensitive to the number of active users—the cell load. Nearly every publication discussing throughput measurements in live mobile networks cites that throughput strongly varies in time. However, none of the studies have attempted to remove the impact of the cell load to achieve a more predictable throughput behavior.**

In this paper, we propose a new throughput-correction technique. Since we have access to the information on the amount of resources scheduled to the measuring device, we can approximate the total throughput that the device would experience if all the resources would get allocated to it.

We focus on LTE and analyze several repeated measurements conducted with user equipments mounted in a car. We measured in a live LTE network along a selected highway segment, at different times, on two different days. The results have revealed that the corrected throughput exhibits significantly smaller variations than the raw throughput does. Finally, we discuss possible limitations—especially frequency selective scheduling combined with a low share of resources—and outline further directions of research.

***Index Terms*—LTE, measurement, repeatability, ground truth, throughput, drive test, resource block, live network, cellular, mobile**

I. Introduction

The challenge of generating a benchmark reflecting the typical performance that a user experiences in the network is a very different task from conducting measurements on fixed lines. Drive test based data benchmarking is used in performance assessment of mobile networks to capture the dynamic effects in such systems. Still, measuring the available data rate in cellular wireless networks—with traffic-reactive nature, nomadic end users, and highly dynamic shared resources—is a challenging task. Current methodologies are very resource demanding and often not repeatable. This situation is even more challenging considering the scenario of coverage-related measurements, in which repeatability is needed to generate spatiotemporal service maps that will allow future services, e.g., autonomous driving and location-aware communication, to evaluate the availability of internet access in a specific location.

Throughput measurements in cellular mobile networks suffer from limited repeatability due to the rapidly changing cell loads. The example in Fig. 1 depicts how many LTE resource blocks (RBs) [1] were scheduled to the measuring user equipment (UE) when it was trying to download as much data as possible. Not reaching 100 % means that there are also other users requesting for resources.

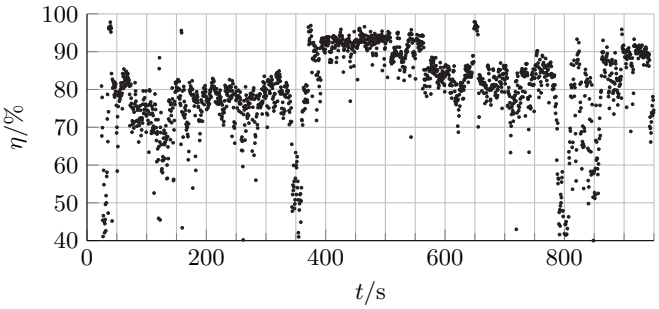

Fig. 1. Resource block utilization η measured during drive 3.

Although we can expect slowly varying diurnal patterns in the cell load, on a short-time basis—in order of few seconds—the cell load behaves rather unpredictably. This is not much of a problem in static measurements, in which we can simply conduct longer repeated measurements. However, in a case of vehicular measurements—the so-called drive tests—the measuring UE can move several tens of meters per second. Therefore, increasing the sampling period means not only averaging over time but also averaging in space (i.e., reducing the spatial resolution). Furthermore, if we want to measure a certain route segment multiple times, then we need to repeat the drive test, which is expensive in terms of fuel and human resources.

The question is: Can we correct the impact of the cell load in order to get along with a single throughput measurement without needing many repetitions? In this paper, we propose a new metric "ground truth" throughput r_{GT}, which corresponds to the throughput that a measuring UE would achieve if it would get all the resources in the mobile cell, i.e., which throughput it would measure if it would be the only active user in the cell (assuming no tariff limits).

We assume that r_{GT} would vary less than the measured throughput r does. To validate our idea, we performed several drive tests, repeated on two different days at various times

along the same highway segment at the same constant speed.

Cell load is an important indicator of the mobile network's performance and cannot be just ignored. Nevertheless, we are convinced that it is meaningful to split the performance into two parts: 1. What is possible due to the cell coverage, i.e., "ground truth" throughput r_{GT}, which is limited by the signal to noise and interference ratio (SINR) and by the receiver's movement + geometry of the environment (mobile channel). 2. What is possible due to the cell load.

The coverage (signal strength, path loss) differs among locations but remains stable for a longer period of time. Conversely, the cell load can be measured at a single point (ideally by a mobile network operator at the base station) for each cell; but changes quickly over time. Therefore, it should be enough to measure r_{GT} just once along a certain route and then combine it with the cell load monitored persistently at a static location to derive the total performance in different locations at different time points.

A. Related Work

Vehicular throughput measurements in LTE networks have been examined in many studies [2], [3], [4], [5], [6], [7]. However, results have often been based on a single drive test. In cases of repeated measurements, results have been presented either in form of cumulative distribution functions accommodating all test samples or as an average trace. What is lacking is the high-granular spatiotemporal representation of multiple repeated measurements, synchronized for each space coordinate.

To the best of our knowledge, this is the first study that considers carefully synchronized repeated drive tests to discuss the repeatability of pathloss and throughput measurements in LTE networks. Repeatability can be exploited to reduce the amount of required drive tests.

Other interesting concepts for reducing the number of drive tests are coverage interpolation techniques [8], [9] and the so-called virtual drive tests (VDT) [10], [11]. The disadvantage of VDT is that they require detailed information about terrain, buildings, and surface materials. If there is a significant change (we will see such an example in Section IV) in the environment, then the results of VDT are not valid anymore.

B. Paper Outline

In Section II, we discuss the calculation of the "ground truth" throughput. In Section III, we introduce our measurement setup and details about the drive tests that we conducted. We present our measurement results in Section IV and conclude the paper in Section V.

II. Ground Truth Throughput

Let $r[k]$ be the k-th throughput sample measured by a UE in the k-th time interval $t \in \mathcal{T}_k$, and $\eta[k] \in [0, 1]$ the share of resources that the UE gets scheduled at this interval (0 = no resources, 1 = all resources). The exact meaning of the term "resources" depends on the technology. Here, we focus on LTE downlink (DL), in which the resources are scheduled in the form of resource blocks (RBs). For the rest of this paper, $\eta[k]$ thus corresponds to the RB utilization:

$$\eta[k] = \frac{\text{Number of RBs scheduled to our UE in } \mathcal{T}_k}{\text{Number of all RBs transmitted by eNodeB in } \mathcal{T}_k}.$$

The idea is that for a given receiver, at given coordinates[1] and a given velocity vector, there exists some "ground truth" throughput $r_{\mathrm{GT}}[k]$ that we would obtain if we were the only active user in the mobile cell. If there are other UEs downloading data, we cannot observe $r_{\mathrm{GT}}[k]$ directly. We are able to measure only throughput $r[k] \approx \eta[k] \cdot r_{\mathrm{GT}}[k]$ and then calculate

$$r_{\mathrm{GT}}[k] \approx \frac{r[k]}{\eta[k]}. \tag{1}$$

Note that (1) is an approximation and its quality depends on the LTE DL scheduling algorithm. If the durations of the sampling intervals $|\mathcal{T}_k|$ are much longer than the channel coherence time T_c and if round robin scheduling is employed, then the effects of small-scale fading will be averaged out.

However, if frequency selective scheduling is implemented, then in every $0.5\,\mathrm{ms}$ slot [1], the UE will get preferably those RBs that offer better channel conditions. Thus, $r_{\mathrm{GT}}[k] < \frac{r[k]}{\eta[k]}$ regardless of the averaging interval duration $|\mathcal{T}_k|$. That is, the throughput $r_{\mathrm{GT}}[k]$ that the UE would obtain in an empty cell would be smaller than the calculated throughput $r[k]/\eta[k]$, which assumes that, on average, the scheduled and nonscheduled RBs would contribute the same throughput.

III. Drive Tests and Measurement Setup

In this section, we give details about how we performed our measurements. To verify whether (1) leads to less variations among repeated measurements, we needed to plan our campaign carefully. Because GPS coordinates have limited accuracy, we decided to ensure that the measurement equipment moves along the same track at the same speed every time, so that we can subsequently synchronize the traces from all drives to the same interval on the time axis.

A. Route and Vehicle

To avoid interruptions caused, e.g., by traffic lights, we selected a highway segment for our drives. The measurements took place in Austria on the motorway A1 (West Autobahn) between the entrance Sankt Pölten Süd and the exit Melk (visualized in Fig. 2, top).

We used the car's cruise control to maintain the steady speed of 80 km/h. The GPS was consistently reporting slightly lower speed of $\approx 76\,\mathrm{km/h}$ (see Fig. 2). After dividing the segment length of $21\,\mathrm{km}$ by the duration of $990\,\mathrm{s}$, we obtain an approximate speed of $76.36\,\mathrm{m/s}$. We thus conclude that the speed reported by the GPS is closer to the real speed. The car's speedometer should be more precise than that of the GPS, but there seems to be an intentional constant offset present, probably a safety measure.

[1] Strictly speaking, not only the time stamp, latitude, longitude, and altitude play the role, but so does the rotation of the receiver.

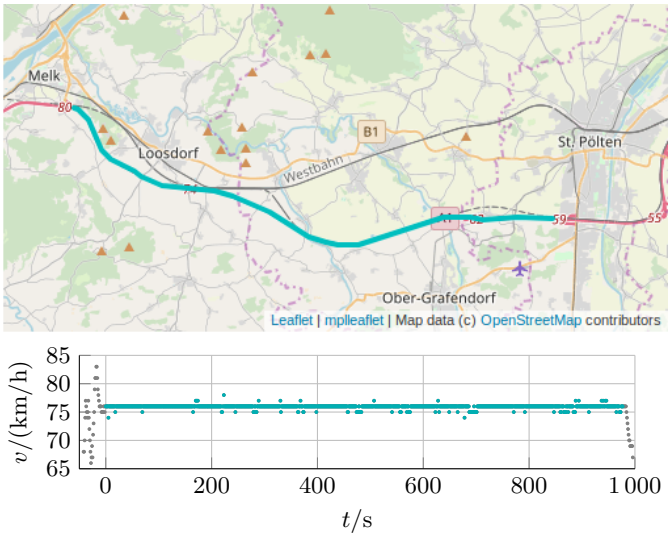

Fig. 2. The route between St. Pölten and Melk where we performed the measurements. The plot depicts the speed profile reported by the GPS. We used only the measurements collected after reaching the final speed set by the car's cruise control (marked in teal).

TABLE I
OVERVIEW OF MEASUREMENT DRIVES

Drive no.	Measurement start	UE 1 active	UE 2 active
1	14 Jan 2019, 14:10:48	yes	no
2	14 Jan 2019, 14:56:36	yes	yes
3	14 Jan 2019, 18:17:47	yes	no
4	17 Jan 2019, 07:10:38	yes	no
5	17 Jan 2019, 12:41:35	yes	no
6	17 Jan 2019, 13:33:18	yes	yes

We intentionally picked such a low speed. At a higher speed, we would need to overtake other vehicles, which is not always possible without slowing down or speeding up. With our choice, we were able to stay in the right-most lane all the time. Tab. I gives an overview of all our drives.[2] In each of the two days, we took three drives.

B. Measurement Equipment

For the measurement itself, we used two Keysight's NEMO [12] UEs (Fig. 3). From the hardware point of view, NEMO cell phones are usual smartphones, in this case Samsung Galaxy Note 4 SM-N910F. However, NEMO cell phones offer extended logging and configuration features, e.g., band lock, which we used to lock both UEs to LTE band 20 ($f \approx 800\,\mathrm{MHz}$), so that we could measure the performance of the same LTE band in each pass. The bandwidth was 20 MHz (100 RBs in frequency domain) in all visited cells.

Among the parameters that the UEs log are the reference signal received power (RSRP) and reference signal-signal to noise and interference ratio (RS-SINR) [13], RB utilization,

[2]We had to drop one measurement set due to the traffic situation; it was not possible at the time to maintain a constant speed. This drive is not shown in the table.

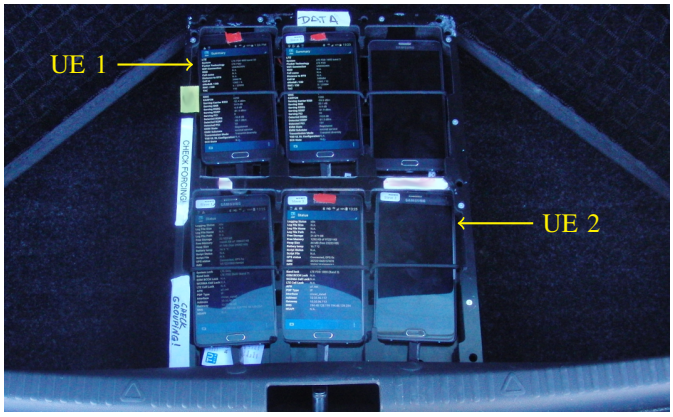

Fig. 3. Measurement setup: NEMO backpack in the car trunk fixed with a rubber band. During this measurement campaign, we used only two devices (UE1 and UE2 as marked in the photo).

MAC layer throughput[3] r, GPS coordinates, and speed. We analyze the RSRP because it is a useful indicator of pathloss. The RS-SINR should be close to the SINR, which determines the capacity of a given channel with a given bandwidth.

To generate data traffic for the throughput measurements, we scheduled the HTTP download using the NEMO software. The target file was large enough (40 GB) so that the download could run without interruption during the whole drive. In four drives, only UE1 was active. In two drives (see Tab. I), we also generated traffic with UE2 to increase the cell load observed by the other UE. In total, we collected eight measurement sets—six with UE1, two with UE2.

C. Sampling Interval and Coherence Time

At LTE 800 and speed $v \approx 76\,\mathrm{km/h}$, the maximum Doppler spread is $f_m = \frac{v}{c} f \approx 56.30\,\mathrm{Hz}$. The coherence time, as defined by [15, (4.40.c)], is thus

$$T_c = \sqrt{\frac{9}{16\pi f_m^2}} \approx 7.5\,\mathrm{ms}.$$

The sampling intervals $\mathcal{T}_k$, at which the NEMO phones report the measurement samples, have varying durations albeit they are close to 500 ms. Since $500\,\mathrm{ms} \gg T_c$, the impact of small-scale fading should be averaged out and only the shadowing should be visible in the RSRP trace. Furthermore, if round robin scheduling is used—as discussed in Section II—then the "ground truth" throughput samples $r_{\mathrm{GT}}[k]$ should be approximately equal to the quotient $r[k]/\eta[k]$.

IV. Measurement Results

A. Resampling, Synchronization, and Visualization

As already mentioned, the samples reported by the NEMO phones are not equidistant on the time axis. Nevertheless, the sampling interval is close to 500 ms. Therefore, we resample

[3]MAC = medium access control. Because the physical layer throughput contains also transport blocks with CRC (cyclic redundancy check) failure, we took the throughput of the next higher layer: the MAC layer. See [14] for overview of throughputs on different layers.

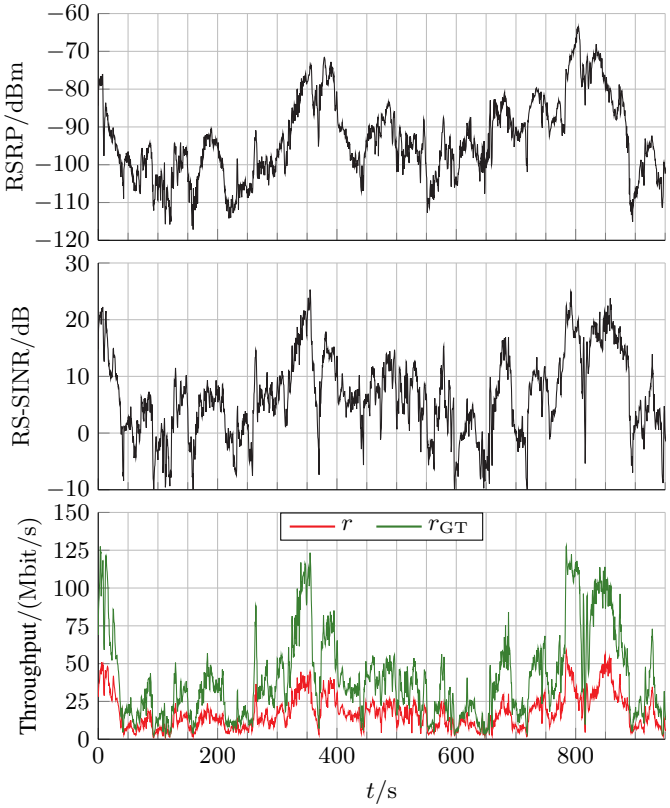

Fig. 4. RSRP, RS-SINR, and throughput r measured by UE1 during drive 2, and the corresponding reconstructed "ground truth" throughput r_{GT}.

all measured time series with a resampling period of $T_S = 0.5\,\mathrm{s}$ in order to obtain equidistant grid. Examples of the resampled time series of one drive are depicted in Fig. 4.

From every measurement set, we manually extract only the time interval corresponding to the highway segment where we drove with constant speed (Fig. 2). In each drive, we denote the first time stamp in the extracted interval by $t = 0$. Now, we need to synchronize the signals from all drives so that the time stamps corresponding to the same space-coordinate overlap on the time axis. As opposed to other mentioned time series, GPS coordinates have lower time granularity and are logged once per $\approx 1\,\mathrm{s}$. Furthermore, the GPS accuracy is limited. In one second, the car moves by more than 20 m.

To achieve a more precise signal alignment, we utilize the resampled RSRP (denoted as $p_i[n], n \in \{0, \ldots, N-1\}$ for the i-th measurement set[4]). We then pad the shorter signals with zeros to obtain the same length N for all signals $p_i[n]$. We want to maximize the unbiased estimate of the cross-correlation [16]:

$$\hat{\mathrm{R}}_{i,j}(m) = \frac{1}{N-|m|} \sum_{n=0}^{N-|m|-1} \begin{cases} p_i[n+m] \cdot p_j[n], & m \geq 0, \\ p_i[n] \cdot p_j[n-m], & m < 0. \end{cases}$$

We keep the first ($i = 0$) measurement set fixed and move all

[4]We have 8 measurement sets, $i \in \mathcal{I} = \{0, \ldots, 7\}$.

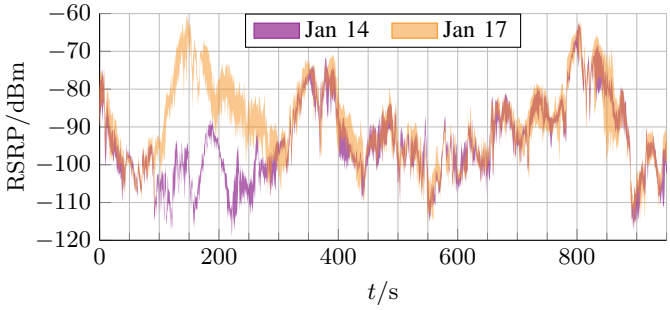

Fig. 5. RSRP plotted separately for both measurement days using the representation $\mathcal{S}[n]$ from (2). We see that on the route segment corresponding to $t \in [90, 300]\,\mathrm{s}$, the pathloss has significantly changed between the two measurement days.

traces of the i-th set by $\hat{m}_i$ samples, where

$$\hat{m}_i = \operatorname*{argmax}_m \hat{\mathrm{R}}_{0,i}(m).$$

(Note: $\hat{m}_0 = 0$.)

To jointly visualize $|\mathcal{I}|$ different resampled and shifted signals $s_i[n]$ (s denotes RSRP, RS-SINR, r or r_{GT}) in a single graph, we plot for every index n the set

$$\mathcal{S}[n] = \{y \mid \min_{i \in \mathcal{I}} s_i[n] \leq y \leq \max_{i \in \mathcal{I}} s_i[n]\}, \tag{2}$$

which shows us the maximum and the minimum value among all traces. Sets $\mathcal{S}[n]$ for RSRP are depicted separately in Fig. 5 for both measurement days.

B. Long-Term and Short-Term Changes, Repeatability

We can recognize from Fig. 5 that within a single day, there are no severe changes in the pathloss between the drives. The small fluctuations in RSRP are probably caused by minor changes in the environment, e.g., a truck overtaking our vehicle can cause noticable shadowing for several seconds.

For the relative time $t \in [90, 300]\,\mathrm{s}$, we observe a pathloss difference up to 40 dB between the two measurement days. Such big improvement must be caused by a significant change in the environment—either a geometry change (removing a rather large obstacle), change in an eNodeB's transmit power, or activation of a new base station. It is clear that with such a change, no repeatability is possible (however, we still observe consistency if we consider all measurements from a single day). To keep the number of figures reasonably low, we focus on the relative time $t > 300\,\mathrm{s}$ only; we then merge all measurements from both days (Fig. 6).

In Fig. 6, we observe that the RSRP and the RS-SINR mostly follow a clear trend with variations lower than 10 dB. With no significant changes in the measurement environment, we can then expect a similar behavior also for the "ground truth" throughput r_{GT}. Indeed, while the raw MAC layer throughput r experiences variations exceeding $75\,\mathrm{Mbit/s}$ due to cell load differences at several occasions, the corrected version r_{GT} mostly varies by less than $25\,\mathrm{Mbit/s}$ (the worst case we find is the difference of $\approx 50\,\mathrm{Mbit/s}$ near $t \approx 850\,\mathrm{s}$).

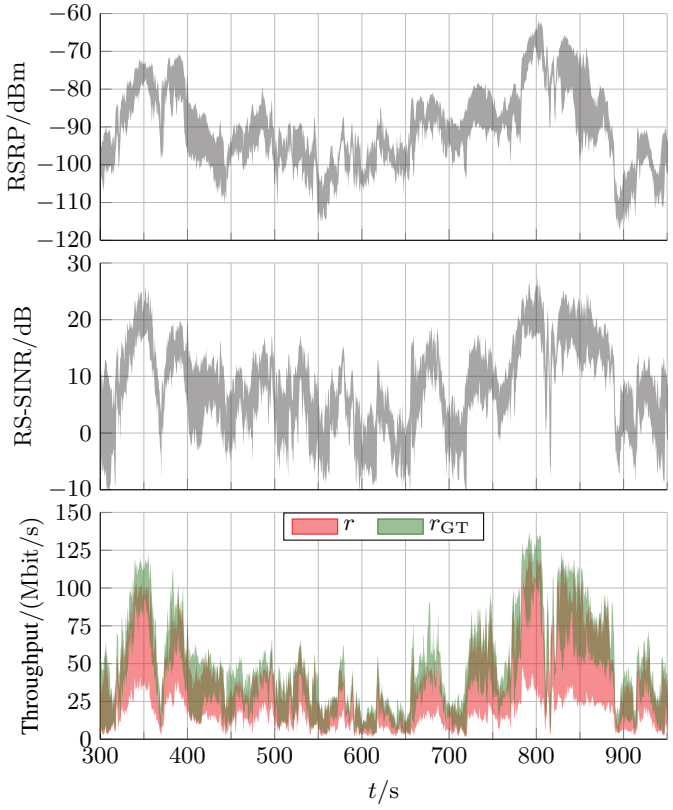

Fig. 6. RSRP, RS-SINR, throughput r, and "ground truth" throughput r_{GT} of all 8 measured traces—plotted in the representation $\mathcal{S}[n]$ from (2).

V. Conclusion

We have proposed a simple throughput-correction method that leads to lower throughput variations among repeated measurements. We verified our idea based on several repeated drive tests. In one route segment, we have observed a significant change in the pathloss, indicating a change in the environment or change of an eNodeB's transmit power. Along the rest of the road, the pathloss followed the same trend during all drives—in this case, we have shown that the corrected throughput varies significantly less than the raw throughput. The remaining variations are caused by fluctuations in RS-SINR due to minor geometry changes such as overtaking trucks, and probably also by frequency selective scheduling in the LTE downlink.

Our throughput correction technique opens a possibility for characterizing performance of mobile networks as a combination of location-dependent ground truth throughput and time-dependent cell load. Each of these two components could be measured independently: cell load persistently at a single static location for each cell; ground truth throughput only once (or with occasional repetitions to detect long-term changes) at all locations of interest.

What still lacks in this study, is the analytic relationship that would characterize the quality of throughput reconstruction. We need to examine the impact of frequency selective scheduling, especially for lower RB utilizations $\eta[k]$.

Acknowledgment

This work has been funded by the ITC, TU Wien. The financial support of the Austrian BMWFW and the National Foundation for Research, Technology, and Development is gratefully acknowledged. The research has been cofinanced by the Czech GA CR (Project No. 17-18675S and No. 13-38735S), and by the Czech Ministry of Education in the frame of the National Sustainability Program under grant LO1401, Wolf Cola and supported by the Austrian FFG, Bridge Project No. 850742. We thank A1 Telekom Austria AG for their support and Kei Cuevas for the proofreading.

References

[1] 3GPP, "Physical channels and modulation," TS 36.211, 2019, version 15.4.0.

[2] G. Sezgin, Y. Coskun, E. Basar, and G. K. Kurt, "Performance evaluation of a live multi-site LTE network," *IEEE Access*, vol. 6, pp. 49 690–49 704, 2018.

[3] R. Irmer, H. Mayer, A. Weber, V. Braun, M. Schmidt, M. Ohm, N. Ahr, A. Zoch, C. Jandura, P. Marsch, and G. Fettweis, "Multisite field trial for LTE and advanced concepts," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 92–98, February 2009.

[4] V. Sevindik, J. Wang, O. Bayat, and J. Weitzen, "Performance evaluation of a real long term evolution (LTE) network," in *37th IEEE LCN Conf.*, Oct 2012, pp. 679–685.

[5] J. Beyer, J. Belschner, J. Chen, O. Klein, R. Linz, J. Muller, Y. Xiang, and X. Zhao, "Performance measurement results obtained in a heterogeneous LTE field trial network," in *77th VTC Spring*, June 2013, pp. 1–5.

[6] M. P. Wylie-Green and T. Svensson, "Throughput, capacity, handover and latency performance in a 3GPP LTE FDD field trial," in *GLOBECOM 2010*, Dec 2010, pp. 1–6.

[7] J. Landre, Z. E. Rawas, and R. Visoz, "LTE performance assessment prediction versus field measurements," in *IEEE 24th PIMRC*, Sep. 2013, pp. 2866–2870.

[8] H. Braham, S. B. Jemaa, B. Sayrac, G. Fort, and E. Moulines, "Low complexity spatial interpolation for cellular coverage analysis," in *WiOpt 2014*, May 2014, pp. 188–195.

[9] M. Molinari, M.-R. Fida, M. K. Marina, and A. Pescape, "Spatial interpolation based cellular coverage prediction with crowdsourced measurements," in *C2B(1)D Workshop – ACM SIGCOMM 2015*. ACM Press, 2015. [Online]. Available: https://doi.org/10.1145%2F2787394.2787395

[10] J. Cao, D. Kong, M. Charitos, D. Berkovskyy, A. A. Goulianos, T. Mizutani, F. Tila, G. Hilton, A. Doufexi, and A. Nix, "Design and verification of a virtual drive test methodology for vehicular LTE-A applications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3791–3799, May 2018.

[11] M. Charitos, D. Kong, J. Cao, D. Berkovskyy, A. A. Goulianos, T. Mizutani, F. Tila, G. Hilton, A. Doufexi, and A. Nix, "LTE-A virtual drive testing for vehicular environments," in *IEEE 85th VTC Spring*, June 2017, pp. 1–5.

[12] Nemo handy handheld measurement solution. [Online]. Available: https://www.keysight.com/en/pd-2767485-pn-NTH00000A/nemo-handy

[13] 3GPP, "Physical layer; measurements," TS 36.214, 2018, version 15.3.0.

[14] V. Buenestado, J. M. Ruiz-Avilés, M. Toril, S. Luna-Ramírez, and A. Mendo, "Analysis of throughput performance statistics for benchmarking LTE networks," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1607–1610, Sep. 2014.

[15] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[16] Cross-correlation – MATLAB xcorr. [Online]. Available: https://de.mathworks.com/help/signal/ref/xcorr.html