Pre-Crowdsourcing: Predicting Wireless Propagation with Phone-Based Channel Quality Measurements

Rita Enami, Yan Shi, Dinesh Rajan, and Joseph Camp
Southern Methodist University
{renami,shiy,rajand,camp}@smu.edu

ABSTRACT

Conducting in-field performance analysis for wireless carrier coverage and capacity evaluation is extremely costly, in terms of equipment, manpower, and time. Hence, there is a growing number of opportunities that exist for crowdsourcing via smart applications, firmware, and cellular standards. These facilities offer carriers feedback about user-perceived wireless channel quality. Crowdsourcing provides the ability to rapidly collect feedback with dense levels of penetration using client smartphones. However, mobile phones often fail to capture the fidelity and high sampling rate of more advanced equipment (e.g., a channel scanner) used when drive testing for analysis of propagation characteristics. In this work, we study the impact of various effects induced by user equipment (UE), when sampling signal quality. These shortcomings include averaging over multiple samples, imprecise quantization, and non-uniform and/or less frequent channel sampling. We specifically, investigate the accuracy of characterizing large-scale fading using crowdsourced data in presence of the aforementioned phone measurement shortcomings. To do so, we conduct extensive in-field experiments across heterogeneous devices and environments to empirically quantify the perceived channel characteristics by phone measurements. Analyzing the quality of the smartphone measurements in LTE indicates that the inferred radio propagation models, is comparable with models obtained by advanced equipment.

KEYWORDS

crowdsourced mobile network measurement; LTE; path loss evaluation; radio propagation model

1 INTRODUCTION

Cellular network providers need to collect and analyze radio signal measurements continuously to improve network performance and optimize network configuration. Available methods to obtain the signal measurements consist of drive testing, network-side-only tools, dedicated testbeds, and crowdsourcing [12]. The former three methods are extremely resource intensive. For example, one common approach for capturing radio signal measurements is to outfit a backpack with six mobile phones running various applications and network protocols alongside an expensive mobile channel scanner (see Fig. 1) for network engineers to gather data on foot. Vehicles are often used for an even greater numbers of and potentially higher-powered and more costly devices and allowing higher levels of mobility in a targeted region. In congested areas with various technologies (*e.g.*, LTE, GSM, UMTS, and TETRA) the problem becomes worse: to get an acceptable quality of service, data collection should be repeated multiple times per roll out of each technology to appropriately configure the network [31]. Further complicating matters, physical changes to the environment such as construction of new buildings or highways can render the obtained data useless.

Figure 1: Typical Rohde & Schwarz backpack for walk/drive testing (left) and TSMW channel scanner (right) [8].

Crowdsourcing is an economical alternative to these resource-intensive methods that has the additional benefit of considering the in-situ performance at the end user device. Consequently, many carriers are rolling out smart applications, firmware, and standardization efforts to crowdsource perceived channel state by user equipment (UE). Furthermore, LTE release 10 in 3GPP TS 37.320 has developed a Minimization of Drive Test (MDT) specification to monitor the network Key Performance Indicators (KPIs) via crowdsourcing. While there is less control of the factors leading to a recorded channel quality, there are many advantages to crowdsourcing this information in terms of lessening the need for costly equipment, reduced in-field man hours, rapid scalability of data sets, and penetration into restricted physical locations. These advantages have sparked a number of works where crowdsourcing has been utilized to identify network topology [7], perform real-time network adaptation [28], characterize Internet traffic [27], detect network events [4], fingerprint and georeference physical locations [23], assess the quality of user experience [15], and study network neutrality [10]. To evaluate the wide area wireless network performance [11] and in-context performance [32], the bandwidth, latency, and throughput are previously-crowdsourced KPIs [25, 29].

However, mobile phones possess a number of shortcomings when compared to a channel scanner in reporting channel quality, such as: *(i)* averaging over multiple samples which can flatten channel fluctuations [29] with manufacturer-specific methodologies to estimate the received signal power [5], *(ii)* coarse quantization which can impose a unit step for minuscule changes, *(iii)* sampling

MSWiM '17, November 21–25, 2017, Miami, FL, USA

ACM ISBN 978-1-4503-5162-1/17/11…$15.00
https://doi.org/10.1145/3127540.3127563

at non-uniform intervals when crowdsourcing information as opposed to long, consecutive testing periods recorded when drive testing, and *(iv)* clipping that results from less sensitive receivers with fringe network connectivity.

The accuracy of the received signal reporting by mobile phones as compared to a channel scanner was evaluated in [17], but the effect of averaging was not considered. Hence, while a crowdsourcing framework for characterizing wireless environments would have tremendous impact on drive testing costs, we believe that a first step in doing so requires understanding the viability of mobile phones to replace more advanced measurement equipment.

In this work, we study the impact of various effects induced by user equipment, when sampling signal quality. These shortcomings include averaging over multiple samples, imprecise quantization, and non-uniform and/or less frequent channel sampling. We specifically, investigate the accuracy of characterizing large-scale fading using crowdsourced data in presence of the aforementioned phone measurement shortcomings. To do so, we perform extensive in-field experimentation to quantify the impact of each of these four effects when evaluating the viability of mobile phones to characterize large-scale fading effects. This metric is commonly used by carriers for deployment planning, frequency allocation, and network adaptation. Our results indicate that the inferred propagation parameters by smartphone measurements in GSM and LTE networks is comparable to those obtained by the advanced equipment that are frequently used by drive testers (*e.g.*, channel scanners). In particular, our work consists of the following three contributions.

First, we set forth a framework to evaluate the impact of strictly using mobile phones (as opposed to a channel scanner) in propagation prediction. As depicted in Fig. 2, we consider how the averaging, uniform and non-uniform downsampling over time and space, and quantization of mobile phone channel quality samples at both the firmware and API level affect the path loss characterization. At the API level, we have designed an Android application, called WiEye, which can be used by users globally as an economical spectrum analyzer. Additionally, WiEye functions as a crowdsourcing tool, which has captured over 250 million signal quality measurements from over 60 thousand users (protected by an IRB) and has a global footprint. At the firmware level, we are able to capture signal quality directly from the hardware via a Rohde & Schwarz tool called Qualipoc.

Second, we compare the perceived channel quality across the channel scanner, multiple mobile phone models, and various levels of the software stack. To do so, we perform extensive local experiments across downtown, single-family residential, and multi-family residential regions and directly compare the received channel quality as reported by the channel scanner to mobile phone firmware-level and API-level data, where each mobile phone measurement considered has a corresponding channel scanner measurement for comparison. We are also interested the range over which each user-side device and software is able to receive cellular base station transmissions (*i.e.*, their sensitivities) to understand where clipping of crowdsourced data might occur.

Third, we quantify the impact on inferring propagation characteristics from the various calculations and imperfections that mobile phones can perform on received channel quality before reporting it to the user. To do so, we consider numerous data sets

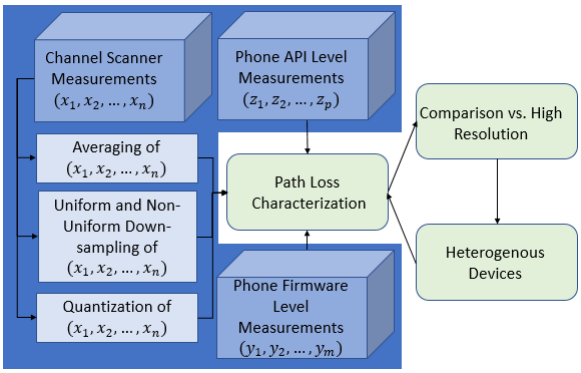


Figure 2: Pre-processing and post-processing of collected data by channel scanner and mobile phones.

from the channel scanner in the aforementioned environmental contexts and impose these imperfections to understand their role by evaluating against the root mean-squared error of path loss prediction from the original channel scanner data set in that region. Our results show that the fading parameters obtained by mobile phone samples are sufficiently comparable to the advanced drive testing equipment, paving the way for crowdsourcing as a viable solution for in-field performance analysis.

The remainder of the paper is organized as follows. In Section 2, we experimentally quantify the channel quality reporting differences of mobile phones versus a channel scanner. In Section 3, we analyze the role of mobile phone imperfections in terms of path loss prediction. We discuss related work in Section 4 and conclude in Section 5.

2 IN-FIELD CALIBRATION OF RECEIVED SIGNAL POWER FROM MOBILE PHONES

The purpose of this study is to compare the ability of mobile phone measurements, captured either at the API level or the firmware level, to an advanced measurement tool, the channel scanner, in characterizing wireless channels in terms of path loss. Before doing so, in this section, we compare and calibrate the raw measurements provided by diverse mobile phones at different levels of the software stack with data provided by a channel scanner.

API-Level Phone Data. At the API level, we modify our Android application WiEye, which we designed to crowdsource measurements, to log signal quality measurements at the highest sampling rate that the operating system will allow (1 Hz). Since WiEye can be installed on any Android-based phone, we can compare API-level measurements across a wide array of devices. In our study, we use four different mobile phones: (*i.*) Samsung S5, (*ii.*) Nexus 5, (*iii.*) Google Pixel, and (*iv.*) Samsung S8. While the former two phones are not the latest models, they provide a comparison across multiple generations, and the Samsung S5 is the phone that allows a firmware-based tool that we will now discuss.

Firmware-Level Phone Data. At the firmware level, we have purchased a software tool called Qualipoc from Rohde & Schwarz, which allows signal strength measurements to be reported directly from the chipset. Qualipoc can receive the channel quality information from many diverse technologies, such as LTE, GSM, and

WCDMA. The sampling rate of the Qualipoc is approximately 3 Hz. Unlike the channel scanner, the mobile phones continuously search for the best visible base station on which to camp by measuring the signal power received from base stations, affecting both the API-level and firmware-level measurements.

Channel Scanner Data. For a piece of equipment that is commonly used by drive testers to gather signal quality measurements in the field, we have acquired a Rohde & Schwarz TSMW Channel Scanner. The TSMW can passively and continuously monitor numerous technologies in 30 MHz - 6 GHz frequency range, with a sampling rate of 500 Hz. The scanner is controlled by Romes software (version 4.89), which is installed on a laptop connected via wire to the TSMW.

In-Field Measurement Setup and Calibration. Since the difference between the aforementioned devices might vary across regions, we conducted the measurement campaign in three diverse regions of Dallas, Texas with regards to their terrain type: single-family residential, multi-family residential, and downtown. All five device types are connected to the same network operator for direct comparison and perform measurements in parallel on a co-located roof of a car. In each region, we observed 11 base stations.

We first would like to understand the range of signal quality sensitivities of each device for measurements taken at the same time and location. To do so, we applied a post-processing procedure on the entire collected data set. Since the sampling rate of the channel scanner is higher than that of Qualipoc (firmware) or WiEye (API), we extract the samples from channel scanner data set, which are the closest in time to that of WiEye and Qualipoc. The matching process consists of two steps: (*i.*) grouping measurements based on the transmitting base station, and (*ii.*) downsampling channel scanner data to have the same number of samples as the Qualipoc and WiEye's data set, where each mobile phone sample has a corresponding channel scanner measurement in time. If the channel scanner did not report a measurement within one second of the mobile phone measurement, we do not consider that data point in our comparison.

Table 1 shows the minimum, maximum, and range of the received signal power for all of these measurements across all cell towers in each region. As it is seen from the results, the widest range (77) and greatest sensitivity (-134 dBm) is captured by the channel scanner with the least range (71) and sensitivity (-128 dBm) captured by WiEye. The reduced range experienced by the mobile phone will cause some clipping on the extreme ends of the connectivity ranges, especially with poor signal quality.

Table 1: Field-tested range of reported signal quality (dBm) from channel scanner (TSMW), Qualipoc, and WiEye.

Device	Min	Max	Range
Channel Scanner	-134	-56	77
Qualipoc Phone	-129	-55	74
WiEye Phone	-128	-57	71

Next, we again consider this downsampled data set which matches the time stamps across devices to consider the difference in reported signal quality per signal quality sample across devices. Table 2 shows the difference of WiEye compared to the matched channel scanner measurement and Qualipoc compared to the matched channel scanner measurement across the three region types. This measurement shows the bias that a mobile phone induces on a crowdsourced data set as compared to more advanced equipment. We also report the mean reported signal strength per region for completeness.

Table 2: Average signal quality bias reported from Qualipoc and WiEye with matched channel scanner measurement.

Device	Qualipoc	WiEye
Location	dBm Diff. (Mean)	dBm Diff. (Mean)
Downtown	-1.5 (-75.6)	-4.4 (-78.5)
Single-Family	-1.3 (-82.5)	-3.8 (-85.0)
Multi-Family	-1.9 (-78.4)	-4.1 (-80.3)

We observe that the difference in reported received signal level is on average 1.57 dBm higher on the channel scanner versus Qualipoc across the three regions with a range of 1.3 to 1.9. In contrast, the difference in reported received signal level is on average 4.43 dBm higher on the channel scanner versus WiEye across the three regions with a range of 4.1 to 4.8. These biases directly affect the path loss characterization as a higher reported channel quality will lower the path loss exponent versus a lower reported channel quality will raise the path loss exponent. In the following section, we will consider the role of this bias as well as multiple other mobile phone imperfections.

3 EVALUATING MOBILE PHONE IMPERFECTIONS ON PATH LOSS PREDICTION

One of the most common metrics which drive testers use to evaluate a given region is path loss. Since we ultimately want to use mobile phone measurements in a crowdsourcing manner to obtain the same metric, we need to understand the role of mobile phone imperfections on evaluating the path loss of a given environment. In particular, reported signal quality from mobile phones will have the following effects: averaging, uniform and non-uniform downsampling, and different resolutions caused by quantization. In this section, we first provide some background with path loss modeling and then will experimentally evaluate the role of these mobile phone imperfections on path loss modeling.

3.1 Modeling Large-Scale Fading: Path Loss

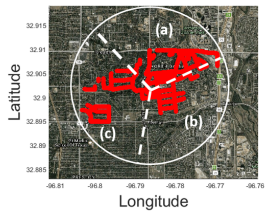
Large-scale fading refers to the average attenuation in a given environment to transmission through and around obstacles in an environment for a given distance [24]. Path loss prediction models are defined in three different categories: empirical, deterministic, and semi-deterministic. Empirical models such as [14] and [22] are based on measurements and use statistical properties. However, the accuracy of these models is not as satisfying as deterministic models to estimate the channel characteristics. These models are widely-used because of their low computational complexity and simplicity. Deterministic models or geometrical models using the Geometrical Theory of Diffraction to predict the path loss. To consider the losses due to diffraction, detailed knowledge of the terrain is needed to calculate the signal strength such as [16]

and [30]. These models are accurate, however their computational complexity is high and need detailed information about the region of interest. Semi-deterministic models applied in [6] and [9] are based on empirical models and deterministic aspects. In our study, we use the empirical method since it is the type of modeling that would be most appropriate to leverage crowdsourcing. The large-scale fading is a function of distance (d) between the transmitter and the receiver and the γ as the path loss exponent, where the path loss exponent varies due to the environmental type from 2 in free space to 6 in indoor environments. Some typical values are 2.7-3.5 in typical urban scenarios and 3-5 in heavily shadowed urban environments [24]. In this work, we focus on the inferred path loss exponent from mobile phone measurements, where a linear regression model is used to calculate the path loss exponent.

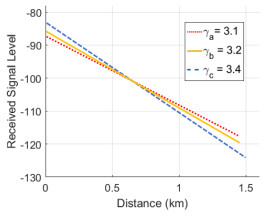
3.2 In-Field Per Sector Analysis of Inferred Path Loss Exponent Across Devices

As discussed in Section 2, our experimental analysis spans three region types (single-family residential, multi-family residential, and downtown) with multiple mobile phone types at the API-level (WiEye), with mobile phones at the firmware level (Qualipoc), and with a channel scanner (TSMW). All of these devices report which base station sector is transmitting the received signal. Since prior work has shown per sector performance can differ [11], we first consider the path loss exponent from each sector in a single-family residential region from the channel scanner to show an example of the diversity a single base station can have across sectors.

Fig. 3a depicts the spatial distribution of signal strength measurements from a channel scanner for a base station in the single-family residential region. The measurement locations across the three sectors are represented by red dots. We perform linear regression on each sector's signal strength measurements independently to find the path loss exponent for that sector. In Fig 3b, we see that the path loss exponent of sector (a) to sector (c) ranges from 3.1 to 3.4, even from the same base station.



(a) Collected measurement from a base station in the single family region.



(b) The regression line fit to the measurements of each sector.

Figure 3: Collected measurements from three sectors around a base station (left) related path loss exponents of each (right).

We now focus on a single mobile phone (Samsung Galaxy S5) to directly compare the path loss exponent inferred from the received signal quality reported at the API and firmware levels to that reported by the channel scanner in the same environment. We consider the most densely measured sector from each region type in our comparison and calculate three different path loss exponents. First, we consider the path loss exponent γ_X as calculated from all measurements in the chosen sector for device X, where X is T for TSMW, Q for Qualipoc, or W for WiEye. Second, we downsample the TSMW measurements according to the matching process mentioned in Section 2, where the TSMW measurement with the closest time stamp to the mobile phone measurement is chosen for Qualipoc and then for WiEye. This second calculated path loss exponent is represented by $\gamma_{Q'}$ and $\gamma_{W'}$, respectively and allows the path loss exponent to be considered for the same number of measurements as Qualipoc and WiEye but with the signal strength readings from the TSMW. Third, we consider a similar path loss exponent as $\gamma_{X'}$ but considers the average bias that each tool would induce on each of the TSMW's signal strength measurements from Table 1 and denoted as $\gamma_{Q''}$ and $\gamma_{W''}$, respectively. The third calculated path loss exponent would subtract this average bias to each of the channel scanner's measurements before calculating the path loss exponent.

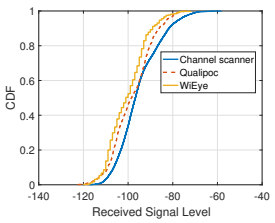
These three γ values are shown in Table 3. There are a few interesting effects that we observe in the comparison across these path loss exponents. First, by comparing γ_T with $\gamma_{Q'}$ and $\gamma_{W'}$, even when the same device is used (TSMW) to capture the signal strength measurements, downsampling the number to match the mobile phones raises the estimate of the path loss exponent in every environment. This effect could be explained by the inclusion of lower quality measurements (*i.e.*, considering the measurements that were clipped from the mobile phone measurements), which in turn lowers the path loss exponent. Second, by comparing each mobile phone measurement type (Q and W) across $\gamma_{X'}$ and $\gamma_{X''}$, we observe that the consideration of the bias brings the estimate even closer to that of the TSMW (γ_T). For the firmware level measurements, the average difference between γ_Q and γ_T in actual, matched, and compensated calculations are 0.1, 0.08, and 0.03, respectively. For the API level measurements, the average difference between the γ_W and γ_T in actual, matched and compensated calculations are 0.15, 0.05, and 0.03, respectively. Hence, the API level measurements have the largest error in path loss calculation from its raw measurements and require the most compensation.

To depict the difference in received signal power between the channel scanner, firmware, and API level, we depict the distribution of the Received Signal Received Power (RSRP) values obtained by each tool for a specific base station sector in Fig. 4a. To evaluate the distance between each curve, we can use the KolmogorovâĂŞSmirnov test (KS test). The ks distance between channel scanner curve and Qualipoc and WiEye curves are 0.11 and 0.13, respectively. Also, the difference between the CDF's median of the channel scanner (-77 dBm), Qualipoc (-78.8 dBm), and WiEye (-79.5 dBm) are about 1.8 dB and 2.5 dB, respectively. This is along the line of bias discussed in Table 1, especially for the firmware measurements but shows that the API level samples are subject to other effects such as averaging of samples, which will be explored in greater depth in the following section.

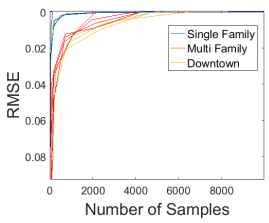
Table 3 also brings up an important issue with crowdsourcing regarding how many measurement samples are required to form an accurate estimate of the path loss exponent. Assume that X_i is the channel scanner signal measurements corresponded to a

Table 3: Path loss characteristics obtained by three devices in three modes: actual, matched, compensated mode.

Region	TSMW		Qualipoc					WiEye				
	Samples	γ_T	△Q&T (dB)	Samples	γ_Q	γ_Q'	γ_Q"	△W&T (dB)	Samples	γ_W	γ_W'	γ_W"
Single-Family	2063	3.1	1.4	620	3.23	3.17	3.15	3.8	293	3.31	3.19	3.15
Multi-Family	1961	3.41	0.9	970	3.48	3.42	3.44	3.1	350	3.51	3.38	3.42
Downtown	11634	3.85	1.2	512	3.97	3.87	3.85	3.5	225	4.00	3.89	3.88



(a) CDF of RSRP of channel scanner versus Qualipoc and WiEye.



(b) The impact of decreasing the number of samples on the ks-distance

Figure 4: Verifying the number of measurements in each region to estimate the γ accurately.

sector, where i shows the number of samples. Then, if we select different number of samples from our reference data set, then we can calculate the ks-distance between the new dataset, Y_m, and the reference distribution, where m represents the number of samples of the new data set.

The difference between the path loss exponent obtained by the reference data set and downsampled data set in terms of ks-distance is depicted in Fig. 4b. The starting point for m is 50 and increases by 50 to 3000 total number of samples for six densely-measured sectors across the three regions. Here, the error is depicted as the Root Mean-Squared Error (RMSE). We observe that the distance between the downsampled data set and the reference data set increases when the number of measurements drops below approximately 800 to 1000 samples. Also, the maximum RMSE between the path loss exponent is about 0.03 when the number of samples is less than 1000. Furthermore, the figure shows that decreasing the number of measurements in a single-family area has a lower impact on the ks-distance as compared to multi-family or downtown area. This effect can be credited to the relative homogeneity of the geographical features in the single-family area as opposed to the more heterogeneous multi-family or downtown regions. In the following section, we will explore the issue of downsampling uniformly and non-uniformly over time and space to understand another imperfection that is introduced with crowdsourcing signal strengths from mobile phones.

3.3 Impact of Mobile Phone and Crowdsourcing Limitations on Path Loss Estimation

In this section the impact of different shortcomings with mobile phone measurements (averaging, temporal downsampling, and quantization) and imperfections that arise with crowdsourcing wireless signal strengths (non-uniform downsampling in both time and space) as opposed to drive testing in a known physical pattern with a known periodic sampling frequency in a particular region under test. In this subsection (3.3) and Section 3.4, we use signal strength samples from the channel scanner exclusively in our analysis and emulate each mobile phone imperfection in isolation to evaluate the impact of that effect.

3.3.1 Averaging of the Received Signal Power. Network interfaces often use some form of hysteresis to suppress sudden fluctuations in channel state that might lead to overcompensation in adaptive protocols. Many times this hysteresis is performed by averaging multiple received signal qualities before reporting it to the higher layers (*e.g.*, within the firmware) and/or the user (*e.g.*, within the operating system in support of API calls). Each device uses its own policy (often proprietary) to take a specific number of samples over a certain period of time.

Even if two devices are in the same environment in close proximity and experience virtually the same channel quality fluctuations, differences in averaging window sizes could be interpreted as diverse fading behaviors. More importantly, when crowdsourcing signal strengths, we are forced to accept the averaging behavior of a broad range of devices. Hence, we seek to characterize the impact of differing averaging windows on the interpretation of large-scale fading. For the purposes of our work, we will be comparing multiple devices (the aforementioned heterogeneous mobile phones against a channel scanner).

A mobile phone in an LTE network is required to measure the Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ) level of a serving cell at least every Discontinuous Reception (DRX) cycle to see if the cell selection criteria is satisfied [3]. To do so, a filter is applied on the RSRP and RSRQ of the serving cell to continuously keep tracking of the quality of the received signal. Within the set of measurements used for the filtering, two measurements shall be spaced by no longer than DRX cycle/2 [2]. On the other hand, a mobile phone receives multiple resource elements and measures the average power of resource elements. However, the number of resource elements in the considered measurement frequency and period over which measurements are taken to determine RSRP by the mobile phone depends on the manufacturer.

Hence, we seek to empirically quantify the degree to which a range of averaging windows (*i.e.*, the number of samples used in the average reported) affects the calculation of the path loss exponent parameter. We depict the variation of the γ parameter in Fig. 5 when we vary the averaging window from 0.25 to 6s on the collected measurements by the channel scanner, which corresponds to a window size of 0 to 200 samples. We averaged the RMSE corresponded to each window size over multiple base stations in

each region. As we see, by increasing the filter size, the maximum error in three regions is about 0.1.

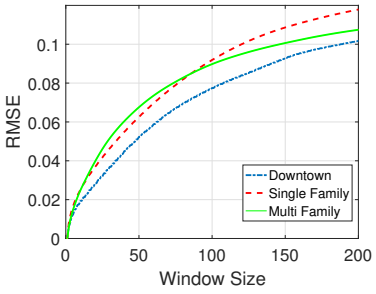


Figure 5: The impact of averaging on the path loss exponent (γ).

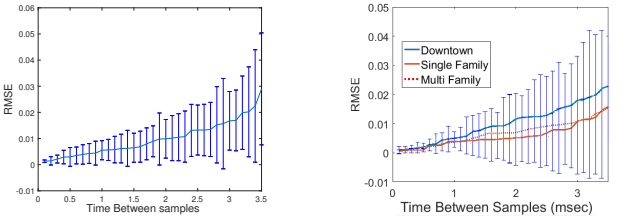
3.3.2 Non-Continuous Measurement Periods. When crowdsourcing information, users may be willing to participate in offering their data. However, there are data usage and battery consumption issues that preclude prolonged, continuous measurements of detailed signal strength values. One option may be to uniformly reduce the number of samples per unit time for a given user over an extended period. Another option could be to aggregate small numbers of samples at different time periods and space from one or more users to compose an aggregate channel effect. We now study both the former (uniform downsampling) and latter (non-uniform downsampling).

Uniform Downsampling Impact on The Channel Characteristic. The channel scanner samples the channel quality at approximately 500 times per second as opposed to about 3 and 1 Hz with the Qualipoc and WiEye, respectively. In this scenario, as the mobile phone preserves energy and/or data usage the question becomes: how would the γ parameters further diverge from the results shown in Table 3? In other words, the previous result showed the extreme of matching the same number of samples or having a very different number of samples, but not the trend in between.

To study this issue, we first examine the calculated γ parameter from a particular sector of a base station in each region, when using uniform and non-uniform down-sampling. We gradually reduce the number of samples obtained by channel scanner to eventually reach the same number of samples recorded by WiEye. At each step, we calculate the error in path loss exponent calculation with respect to our reference value, which is obtained by considering the highest resolution in channel scanner data set. To do so, we reduce the number of samples by i, where $i \in 1, .., n$ and $n = \frac{\#Channel\ scanner's\ records}{\#Phone\ records'}$. As we reduce the data set by i samples, we are able to leverage i data sets for a given i to increase the confidence in the result and show the variation of error in the figure.

Fig. 6a shows the error in terms of path loss calculation by reducing the signal samples received from a cell sector of a base station in the downtown area. By increasing the time interval between samples, the γ and resulting variation thereof are affected. We observe that the error caused by uniformly downsampling can reach up to 0.03 in this specific cell. We see that it is not getting very far from the reference γ. Although the RMSE over each 10 steps has some variation, it does not increase the error dramatically. Furthermore, by decreasing the number of samples, the variation of channel characteristic estimation is not as stable as when we have more data points.

Fig. 6b shows the impact of uniformly downsampling on the channel characteristics on each of the three different regions (single-family residential, multi-family residential, and downtown). The maximum variation over all three regions is depicted as the variation of the RMSE at each point. Of particular note in this result is that downtown shows more sensitivity to downsampling and the single-family residential region shows the least sensitivity.



(a) Uniformly Downsampling (Downtown).

(b) Uniform Downsampling from a Sector in Three Regions.

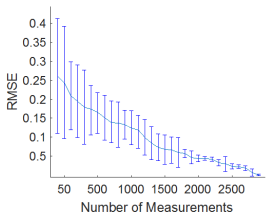
Figure 6: Uniformly downsampling the measurements of a sector in downtown (left) and across all three regions (right).

Non-Uniformly Downsampled Data Sets. In a second scenario, perhaps the crowdsourced measurements are not coming from a single user which has uniformly throttled the number of measurements recorded or reported but from multiple users in the same area. Controlling for device differences for now (we will study this issue in Section 3.5), the newly composed data set for mobile phone measurements Y has a non-uniform sampling period in time and space compared to drive testing the region with a channel scanner. As before, how far would the estimate of γ_Y be to that of the estimated γ when mobile phone signal strength readings are dispersed through time and space? Assume that a sufficiently large number of users in a similar area have crowdsourced measurements. Assuming the number of measurements from the non-uniformly sampled data set matches that of the uniformly sampled data set, what would be the effect of the difference downsampling types?

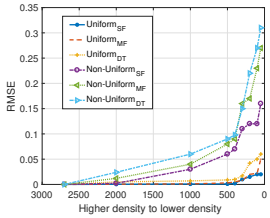
The non-uniform distributed measurements are studied with two types of distributions: *(a)* temporal and *(b)* spatial. For non-uniform temporal downsampling, we reduce the number of samples randomly based on the time stamp of the received signal measurements from the channel scanner dataset. Fig. 7a depicts the impact of the non-uniform downsampling data respect to time on the path loss exponent from a cell sector in downtown. It shows that by increasing the number of samples, the error respect to the reference value decreases. However, in general it has caused a higher value in terms of RMSE for the same number of measurements as compared to uniform downsampling.

For non-uniform spatial downsampling, we select the most populated sector in each region. Then, we chose the measurements based on three clusters which are randomly distributed over the region. Then, we increased the number of the selected measurements in each cluster. Finally, we compared the path loss exponent of the aggregated samples from non-uniformly distributed clusters with the γ computed from all measurements from the channel

scanner in the same region. A comparison between the uniform downsampling and non-uniform distributed measurements in space for three regions is depicted in Fig. 7b. The clustered scenario shows a higher error than the uniformly-distributed one. In addition, we observe that the corresponded error to the downtown is higher than two other regions. We have found that the location of the selected



(a) Non-Uniformly downsampling selection in time.

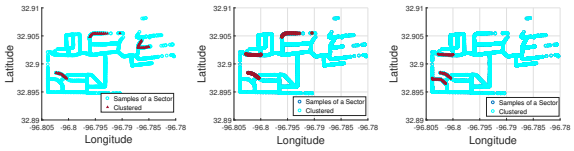


(b) Non-Uniform (space) vs. Uniform (space).

Figure 7: Non-uniform downsampling of a sector in downtown (left) and non-uniformly downsampling in space compared to uniformly (right).

clusters in the non-uniform scenario is significant and attempts to depict the results here. To do so, we again select the most populated sector in a region. Then, we determine the location of three clusters of measurements based on Fig. 8. We start by choosing 50 measurements in each cluster and we increase it by 50 until we have 3000 measurements. The left figure shows a model which is more dispersed through a sector. The middle scenario covers the left and top left area of the sector. In the right scenario, all measurements have a grouping on the left of the sector. We measured the average of the RMSE for each scenario. The results show the 0.083, 0.15, and 0.3 as the average of the RMSE for each aforementioned scenario. In other words, a spatially well-distributed group of user measurements would contribute to a better result to predict the path loss exponent. Also, the type of cluster distribution has impact on the number of measurements that are needed to estimate the channel condition. With this result and the current developments in the LTE standard (10) about the Minimization of Drive Test function [1], a carrier could more strategically poll users in a given area and/or at a certain time to reduce the resources necessary for their users to crowdsource and increase the likelihood of success of such an effort.

3.3.3 Quantization of the Received Signal Power. Android reports the quality of the common pilot channel received signal quality for LTE in terms of Arbitrary Strength Units (ASU) with 98 quantized levels. The received signal level has a range of -44 dBm to -140 dBm and is mapped to "0 to 97" with the resolution of 1 dBm. Since the obtained signal strength by a channel scanner has much greater granularity, the question becomes: what role does quantization have on the path loss exponent? We have considered the quantization impact on path loss estimation as defined as the difference between the estimated γ compared to the highest resolution setting as measured by the channel scanner and found it to be negligible (*e.g.*, less than 1 percent of the RMSE). We will show this effect in Fig. 9a of the following subsection, which considers the joint effect of all of these imperfections.

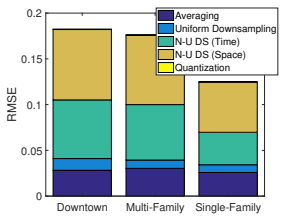


(a) RMSE = 0.083 (b) RMSE = 0.15 (c) RMSE = 0.39

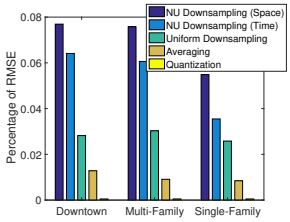
Figure 8: Impact of the non-uniformly downsampling in space.

3.4 Joint Analysis of Mobile Phone Factors on Path Loss

Up to this point, we applied each of the challenges with phone measurements individually. We now jointly consider the mobile phone imperfections impact (averaging, uniform and non-uniform downsampling in space and time, and quantization) on the γ estimation. To do so, we extract the collected data by the channel scanner obtained from a specific cell sector from three regions. Then, we apply the averaging on signal samples which are quantized already. Then, we downsampled (uniformly and non-uniformly in time and space) from the averaged and quantized values. At each step, we obtain the RSME from the path loss exponent calculated from the channel scanner's samples with the highest resolution. Fig. 9a depicts the relative error, caused by each shortcoming in compare with the other issues. Fig. 9b shows the percentage of RMSE caused by each individual issue respect to the reference γ. There are two inter-



(a) Relative impact of mobile phone factors on RMSE of γ.



(b) Impact of each factor on percentage of RMSE of γ.

Figure 9: Joint impact of mobile phone imperfections relatively (left) and per effect (right).

esting findings from these result: *(i)* either form of non-uniformly downsampling is clearly the most dominant effect in predicting the path loss exponent, and *(ii)* the two non-uniform downsampling techniques (time and space) have approximately equivalent performance (despite the noisiness of non-uniform downsampling noted earlier). The latter finding offers great hope for crowdsourced data sets to be influential in characterizing the path loss characteristics of an environment.

3.5 Impact of Heterogeneous Mobile Phones and Users on Path Loss Characterization

When crowdsourcing signal quality from mobile phone users, there is a diversity in hardware and software of the devices. Even two co-located mobile phones at the same time may report very different signal qualities due to different RF front ends. In this section, we study the impact of heterogeneous devices on the estimated path loss exponent. Up to this point, we have considered a single type of

mobile phone, Samsung Galaxy 5S, due to its ability to support both Qualipoc and WiEye. Here, we use WiEye across three other mobile phones (4 total) with a two-phase approach. First, we consider the signal strength samples from all the devices to calculate the path loss exponent and evaluate the accuracy compared to the path loss exponent from the channel scanner signal quality samples. Second, we account for the per-region-type bias introduced by each mobile phone in terms of dBm as compared to the raw measurements of the channel scanner. Lastly, we calculate the path loss exponent based on strictly crowdsourced data from WiEye users in different regions around the world and examine the geographical features of these areas.

3.5.1 Calibrating Diverse Phone Models and Setup. In this experiment, four Android phones described in Table 4 are used to collect signal strength data from the three aforementioned areas in Dallas (single-family residential, multi-family residential, and downtown). We installed our development version of WiEye, which logs signal strength samples at 1 Hz, on the following four phones: Samsung GS5, Nexus 5X, Samsung S8, and Google Pixel. Each phone was co-located alongside the channel scanner on the roof of a car. The duration of the experiment was 360 minutes.

Table 4: Measurement tools configuration and field-tested range of reported signal quality (dBm) from channel scanner (TSMW) and WiEye of four phones.

Tool	Model/OS	Chipset	Min	Max	Range
Channel Scanner	TSMW/-	-	-130	-52	78
W_1	Samsung GS5/A5	MSM8974AC	-118	-54	64
W_2	Nexus 5X/A5	MSM8974	-119	-58	61
W_3	Google Pixel/A7	MSM8996	-120	-57	63
W_4	Samsung GS8/A7	MSM8996	-121	-54	67

We first analyze the RSRP differences of the four phones in terms of the minimum, maximum, and resulting range of dBm reported across all measurements to understand the relative sensitivities. While a few hours of driving does not guarantee the full range of signal strengths, during this time, we observe that the greatest range of values is achieved by the Samsung S8 (67 dBm) as reported by WiEye and the least range of values belonged to the Nexus 5X (61 dBm). As a point of comparison, the TSMW Channel Scanner achieved a range of 78 dBm for the temporally-matched samples.

3.5.2 Inferring Path Loss Across Devices. We now will use each phone to predict γ for four observed base stations in aforementioned regions. The dBm offset bias between the average received signal level by each phone and the channel scanner is shown in Table.5 per region.

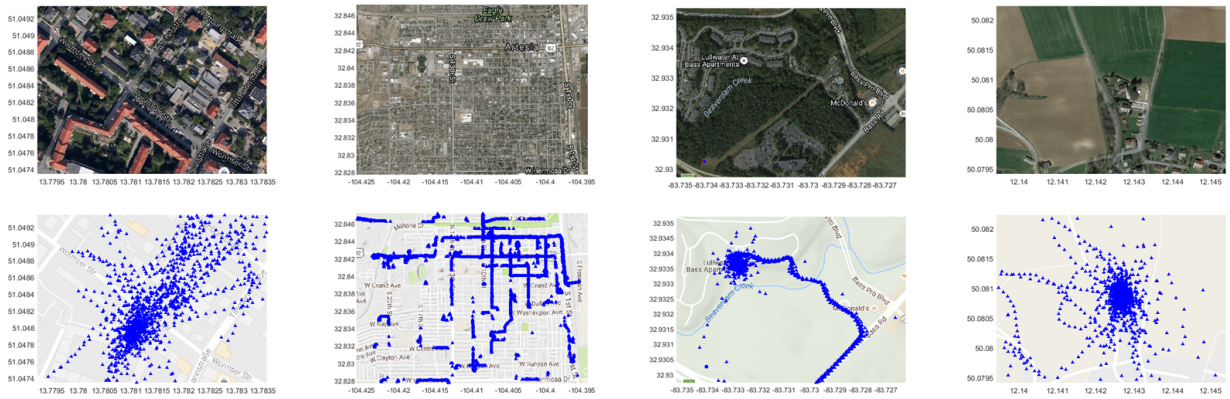
We observe that on average the difference in reported received signal level by the scanner is 3 dBm higher versus the phones across the three regions with a range of 1.46 to 4.1 dBm. As we depicted before, the biasses directly affect the path loss characterization. The lower reported channel quality corresponds to a higher value in obtained path loss exponent, while a higher reported channel quality corresponds to a lower path loss exponent. We now consider the calculated path loss exponent from the signal strength samples of each of the four phones, the calculated path loss exponent from the aggregated data set of the reported signal strength samples from all phones, and then the calculated path loss exponent from the compensated signal strength samples of all phones, considering the bias.

Table 6 shows the obtained path loss characteristics of one specific sector in three different regions, when we consider only a single phones' RSRP and all phones' RSRP. As a point of reference, we also include the γ from the channel scanner RSRP data. We observe that the obtained γ using the data set of each phone are relatively close to one another. We see that the Samsung S8 phone has the closest γ value between all four phones to the channel scanner. In other words, the device that receives the larger range is more accurate in terms of the γ estimation. The comparison shows that considering all RSRP data across device types actually increases the accuracy as compared to any given phone against the path loss exponent calculated from the channel scanner RSRP. Hence, we find that γ is predicted by using the RSRP from a *diverse* set of mobile phones. In addition, we compensated the signal strength of the aggregated dataset by using the 3 dBm obtained in the previous section. We find that the compensated results in terms of γ are extremely close (with 2.7 %, 0.3 %, and 0.6 % error for single-family residential, multi-family residential, and downtown, respectively) to the obtained results by the channel scanner.

3.5.3 Inferring the Path Loss from Crowdsourcing. We now use crowdsourced measurements taken from our widely-distributed WiEye application on the Google Play store. We estimate the path loss exponent of regions around the world without physically drive testing those areas. Based on some of our highest user density, we have selected four environments with diverse geographical features: (i) tall buildings and trees in Dresden, Germany, (ii) low buildings and no trees in Artesia, New Mexico, (iii) mostly trees with a few homes in Macon, Georgia, and (iv) mostly free space in Thiersheim, Germany. The aerial view of each of these environments can be seen in the top figures of Fig 10. In Fig. 10, the bottom figures show the number of crowdsourced signal strength samples and their spatial location as captured by our Android application overlayed on a more basic map of the same area displayed in the aerial view on the top. Using these signal quality measurements from each region, we have computed the path loss exponent γ, which can be seen in the caption of each subfigure. We have ordered the figures from left to right where we see the path loss exponent is decreasing from left to right. In particular, γ_a equals 3.3 with the most diverse and complex environment with tall buildings and trees, γ_b equals 2.7 with an environment that has similar, small building types but no trees, γ_c equals 2.5 with mostly trees and a few homes, and γ_d equals 2.1 with mostly free space. Therefore, the geographical features and complexity in the environment match the γ behavior we would expect, and the channel factors were derived strictly using crowdsourced measurements. Of particular note that in these measurements alone we saw a fairly dramatic change in the γ. In fact, we observed a range of 2.1 to 4.0 of the path loss exponent throughout this paper, which would constitute extremely different network designs across this range of propagation scenarios.

Table 5: Average signal quality bias reported from heterogeneous phones as reported by WiEye with matched channel scanner measurement.

Device	W_1 (GS5)	W_2 (N5X)	W_3 (Pixel)	W_4 (GS8)
Location	dBm Diff. (Mean)	dBm Diff. (Mean)	dBm Diff. (Mean)	dBm Diff. (Mean)
Downtown	4.4 (-78.5)	2.1(-76.2)	3.2 (-77.3)	1.6 (-75.7)
Single Family	3.8 (-85.0)	2.4 (-83.6)	2.5(-83.7)	1.7 (-82.9)
Multi Family	4.1 (-80.3)	2.7 (-79.1)	3.5 (-80)	1.1 (-77.6)



(a) Location: Dresden, Germany with γ_a=3.3

(b) Location: Artesia, New Mexico with γ_b=2.7

(c) Location: Macon, Georgia with γ_c=2.5

(d) Location: Thiersheim, Germany with γ_d=2.1

Figure 10: Path Loss Analysis for Crowdsourced Data Sets in Four Different Regions.

Table 6: Path loss characteristics obtained by four devices in three modes: matched, aggregated, compensated mode.

Device	Single Family	Multi-Family	Down-town
Channel Scanner	3.01	3.33	3.61
W_1 (GS5)	3.21	3.50	3.80
W_2 (N5X)	3.18	3.54	3.78
W_3 (Pixel)	3.38	3.58	3.90
W_4 (GS8)	3.19	3.47	3.75
Aggregated	3.27	3.53	3.83
Compensated	3.09	3.34	3.63

4 RELATED WORK

The Minimization of Drive Tests (MDT) initiative in the 3GPP standard has been created to exploit the ability of smartphones to collect radio measurements in a wide range of geographical areas to enhance the coverage, mobility, capacity optimization, and path loss prediction [1]. Also, a few measurements studies have been conducted to use API-level measurements to estimate different Key Performance Indicators (KPIs) of the cellular networks [13, 21, 25, 29]. They each measured KPIs in terms of throughput, received signal power, and delay and involved regular users to provide the measurements (*i.e.* crowdsourcing) across a large geographical region in some cases. In contrast, we focus on characterizing the wireless channel using diverse end user devices at different levels of the software stack. Predicting the cellular network coverage by using the crowdsourced data has been studied in a few studies. For example, network coverage maps using crowdsourced data is studied in [18]. However, the authors provided the observed received signal level without a discussion of the differences across end user devices. In addition, anther work used a similar idea of using crowdsourced data along with interpolation techniques to predict the coverage area [19]. Although, the impact of location inaccuracy and data distribution of the interpolation techniques was investigated, the impact of the imperfections of end user devices was not explored. Furthermore, others proposed the Bayesian Prediction method to improve the coverage estimation obtained by drive test and MDT measurements, but the results were strictly based on advanced devices as opposed to mobile phone measurements [26]. The provided X-map's accuracy, from simulated data in [20] has been evaluated in terms of the position inaccuracy, UE inaccuracy, and number of measurements. However, to analyze crowdsourced data, using in-field experimentation is important to distinguish between the performance of more advanced equipment versus a mobile phone in channels similar to those experienced by user devices.

To estimate the channel quality, we are using RSRP as our metric from the LTE standard. It was previously observed by [5] that the reported value by a mobile phone in terms of RSRP is influenced by averaging but did not consider the compounding effects. Similarly, [17] depicts that the received signal power by commercial phones is comparable to an advanced tool. While this is close in

nature, we also consider many of the spatial and temporal downsampling effects that would cause imprecise estimation of the path loss estimation for a given environment.

5 CONCLUSION

In this work, we take a first step towards crowdsourcing wireless channel characteristics in LTE cellular networks (and beyond) by considering the relationship between received signal strength measurements of diverse mobile phones at the firmware and API level versus advanced drive testing equipment. In particular, we performed extensive experimentation across four mobile phone types, two pieces of software, and a channel scanner in three representative geographical regions: single-family residential, multi-family residential, and downtown. With these devices and in-field measurements, we evaluated the effects of averaging over multiple samples, uniform and non-uniform downsampling (in time and space), quantization, and crowdsourcing on the path loss exponent estimation. We showed that both types of non-uniform downsampling have the most dramatic effects on path loss calculation. Conversely, we showed the quantization impact can largely be ignored since it showed a negligible influence on our estimation. One key result of note stems from the spatial non-uniformity of clusters of measurements observed within our crowdsourcing database, which required far more measurements than more uniformly spaced measurements. Using the MDT specification of LTE release 10 carriers could request specific measurement locations and times from users to be far more efficient in polling signal quality. Finally, we showed four regions around the globe and predicted the channel characteristics of these regions from our crowdsourced data. In summary, we lay a strong foundation for understanding a large majority of the issues involved with crowdsourcing channel characteristics.

ACKNOWLEDGMENTS

This work was in part supported by NSF grants: CNS- 1150215, CNS-1320442, and CNS-1526269. Also, we would like to thank Rhode & Schwarz for their extensive support in this measurement campaign.

REFERENCES

[1] 3GPP. 2011. *ETSI TS 137 320 "Radio measurement collection for Minimization of Drive Tests (MDT)".*

[2] 3GPP. January 2011. *ETSI TS 136 133 "Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management ".*

[3] 3GPP. November 2011. *ETSI TS 136 304 "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode".*

[4] Zachary S Bischof, John S Otto, Mario A Sánchez, John P Rula, David R Choffnes, and Fabián E Bustamante. 2011. Crowdsourcing ISP characterization to the network edge. In *Proc. of ACM SIGCOMM Measurements Up the Stack.*

[5] Joe Cainey, Brendan Gill, Samuel Johnston, James Robinson, and Sam Westwood. 2014. Modelling download throughput of LTE networks. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on.* IEEE, 623–628.

[6] Gerald K Chan. 1991. Propagation and coverage prediction for cellular radio systems. *IEEE transactions on vehicular technology* 40, 4 (1991), 665–670.

[7] Alessandro Checco, Carlo Lancia, and Douglas J Leith. 2014. Using Crowdsourcing for Local Topology Discovery in Wireless Networks. *arXiv preprint arXiv:1401.1551* (2014).

[8] Rohde & Schwarz GmbH & Co.KG. [n. d.]. *Radio Network Analyzer Operating Manual.*

[9] U Dersch and WR Braun. 1991. *A physical radio channel model.* Technical Report. IEEE CH2944-7/91/0000/0289.

[10] Marcel Dischinger, Massimiliano Marcon, Saikat Guha, P Krishna Gummadi, Ratul Mahajan, and Stefan Saroiu. 2010. Glasnost: Enabling End Users to Detect Traffic Differentiation.. In *Proc. of USENIX NSDI.*

[11] Aaron Gember, Aditya Akella, Jeffrey Pang, Alexander Varshavsky, and Ramon Caceres. 2012. Obtaining incontext measurements of cellular network performance. In *Proceedings of the 2012 ACM conference on Internet measurement conference.* ACM, 287–300.

[12] Utkarsh Goel, Mike P Wittie, Kimberly C Claffy, and Andrew Le. 2016. Survey of end-to-end mobile network measurement testbeds, tools, and services. *IEEE Communications Surveys & Tutorials* 18, 1 (2016), 105–123.

[13] Nicolas Haderer, Fawaz Paraiso, Christophe Ribeiro, Philippe Merle, Romain Rouvoy, and Lionel Seinturier. 2015. A Cloud-Based Infrastructure for Crowdsourcing Data from Mobile Devices. In *Crowdsourcing.* Springer, 243–265.

[14] Masaharu Hata. 1980. Empirical formula for propagation loss in land mobile radio services. *IEEE transactions on Vehicular Technology* 29, 3 (1980), 317–325.

[15] Tobias Hoßfeld, Michael Seufert, Matthias Hirth, Thomas Zinner, Phuoc Tran-Gia, and Raimund Schatz. 2011. Quantification of YouTube QoE via crowdsourcing. In *Proc. of IEEE Multimedia (ISM).*

[16] Fumio Ikegami and Susumu Yoshida. 1980. Analysis of multipath propagation structure in urban mobile radio environments. *IEEE transactions on Antennas and Propagation* 28, 4 (1980), 531–537.

[17] Mads Lauridsen, Ignacio Rodriguez, Lucas Chavarria Gimenez, Preben Mogensen, et al. 2016. Verification of 3G and 4G received power measurements in a crowdsourcing Android app. In *Wireless Communications and Networking Conference (WCNC), 2016 IEEE.* IEEE, 1–6.

[18] Jaymin D Mankowitz and Andrew J Paverd. 2011. Mobile device-based cellular network coverage analysis using crowd sourcing. In *EUROCON-International Conference on Computer as a Tool (EUROCON), 2011 IEEE.* IEEE, 1–6.

[19] Massimiliano Molinari, Mah-Rukh Fida, Mahesh K Marina, and Antonio Pescape. 2015. Spatial interpolation based cellular coverage prediction with crowdsourced measurements. In *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data.* ACM, 33–38.

[20] Michaela Neuland, Thomas Kurner, and Mehdi Amirijoo. 2011. Influence of different factors on X-map estimation in LTE. In *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd.* IEEE, 1–5.

[21] Ashkan Nikravesh, Hongyi Yao, Shichang Xu, David Choffnes, and Z Morley Mao. 2015. Mobilyzer: An open platform for controllable mobile network measurements. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services.* ACM, 389–404.

[22] Yoshihisa Okumura, Eiji Ohmori, Tomihiko Kawano, and Kaneharu Fukuda. 1968. Field strength and its variability in VHF and UHF land-mobile radio service. *Rev. Elec. Commun. Lab* 16, 9 (1968), 825–73.

[23] Anshul Rai, Krishna Kant Chintalapudi, Venkata N Padmanabhan, and Rijurekha Sen. 2012. Zee: zero-effort crowdsourcing for indoor localization. In *Proc. of ACM MobiCom.*

[24] Theodore S Rappaport et al. 1996. *Wireless communications: principles and practice.* Vol. 2. Prentice Hall.

[25] Sanae Rosen, Sung-ju Lee, Jeongkeun Lee, Paul Congdon, Z Morley Mao, and Ken Burden. 2014. MCNet: Crowdsourcing wireless performance measurements through the eyes of mobile devices. *IEEE Communications Magazine* 52, 10 (2014), 86–91.

[26] Berna Sayrac, Janne Riihijärvi, Petri Mähönen, Sana Ben Jemaa, Eric Moulines, and Sébastien Grimoud. 2012. Improving coverage estimation for cellular networks with spatial bayesian prediction based on measurements. In *Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design.* ACM, 43–48.

[27] Yuval Shavitt and Eran Shir. 2005. DIMES: Let the Internet measure itself. *ACM SIGCOMM CCR* 35, 5 (2005), 71–74.

[28] Jinghao Shi, Zhangyu Guan, Chunming Qiao, Tommaso Melodia, Dimitrios Koutsonikolas, and Geoffrey Challen. 2014. Crowdsourcing access network spectrum allocation using smartphones. In *Proc. of ACM Hot Topics in Networks.*

[29] Sebastian Sonntag, Jukka Manner, and Lennart Schulte. 2013. Netradar-Measuring the wireless world. In *Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), 2013 11th International Symposium on.* IEEE, 29–34.

[30] Joram Walfisch and Henry L Bertoni. 1988. A theoretical model of UHF propagation in urban environments. *IEEE Transactions on antennas and propagation* 36, 12 (1988), 1788–1796.

[31] SeungJune Yi, SungDuck Chun, YoungDae Lee, SungJun Park, and SungHoon Jung. 2012. *Radio Protocols for LTE and LTE-advanced.* John Wiley & Sons.

[32] Jongwon Yoon, Sayandeep Sen, Joshua Hare, and Suman Banerjee. 2015. WiScape: A Framework for Measuring the Performance of Wide-Area Wireless Networks. *IEEE Transactions on Mobile Computing* 14, 8 (2015), 1751–1764.