

LAIK: Location-Specific Analysis to Infer Key Performance Indicators

Rita Enami, Sabyasachi Gupta, Dinesh Rajan, *Senior Member, IEEE*, and Joseph Camp, *Member, IEEE*

Abstract—Key Performance Indicators (KPIs) are important measures of the quality of service in cellular networks. There are multiple efforts by cellular carriers and 5G standardization on the use of crowdsourcing to minimize drive tests (MDT) and self-organize the network while improving KPIs via a user feedback loop. Since propagation highly depends upon the environment, readily-available geographical data could be coupled with the crowdsourced user data to infer performance. In this paper, we build a framework to infer KPIs by establishing a relationship between geographical data and crowdsourced channel measurements via neural networks. In particular, for a specific user location, we leverage delay spread measurements in the region to design a cone-shaped filter for the geographical and user data extraction. Then, a location-specific received signal power prediction is obtained via the neural network trained using the extracted geographical and user data. We study the impact of the angle chosen for the cone and various features selected on location-specific KPI prediction. We then leverage the location-specific inference by repeating the prediction over a set of locations in a region to infer the path loss in a given environment. In both types of KPI inference, we compare against state-of-the-art solutions and show that significant improvement in KPI prediction accuracy is achieved using the proposed strategy. Furthermore, for network planners, we show that our framework can use only geographical information to predict KPIs with a negligible error in user locations that lack signal quality data. By employing the proposed framework to predict location-specific and regional KPIs, we achieve an accurate estimation of network coverage and a 7-fold reduction in throughput estimation error compared to a state-of-the-art solution.

Index Terms—Crowdsourced mobile network measurement, Long-Term Evolution (LTE), received signal power prediction, radio propagation model, neural network.

I. INTRODUCTION

Network planning in the deployment phase plays a critical role in improving coverage estimates and resulting user performance. Models for predicting large-scale fading or path loss have traditionally been classified into three different categories: empirical, deterministic, and semi-deterministic. Empirical or statistical models such as Hata [1], and COST-231 Walfisch-Ikegami [2] are obtained by using in-field measurements that correspond to a specific environment type. Although an empirical model is easy to implement and its computational complexity is low, the accuracy is often lacking. In deterministic models such as with ray tracing, the received

signal strength at a particular location is calculated using the Geometrical Theory of Diffraction. In these models, direct, reflected, and diffracted paths are considered to predict coverage. Hence, the computational complexity of these approaches is high. Also, to obtain the propagation model of an area of interest, detailed knowledge of the clutter is required, which can be very costly. Semi-deterministic models straddle the line between empirical and deterministic models, improving the accuracy of empirical models while reducing the computational complexity of deterministic models. An alternative approach to predict the path loss of an environment is emerging via the use of neural networks. Such a model can process a large amount of data in a reasonable time with the ability to learn the characteristics of a new environment. The results show that the model can provide a close estimation for the signal attenuation in an area. In [3], the proposed artificial neural network (ANN) model uses the free geodata (light detection and ranging or LiDAR) to infer the signal quality of a region using proper geographical features as the model's input. Detailed information such as the vegetation type and density have also been used to train the predictor [4] to infer the path loss, and Sotiroudis *et al.* [5] implemented an ANN model to predict the path loss using a minimized input to increase its accuracy. ANN model has been used to predict the indoor propagation as well [6].

A key question to consider in the case of each of these models is how signal measurements can be obtained. Historically, drive testing has been widely used by carriers and third-party entities to collect a sufficient density of Key Performance Indicator (KPI) data to accurately characterize the network performance. To track network performance, network operators use KPIs such as received signal power, received signal quality, throughput, and delay. Despite providing detailed information at certain locations, this approach is costly in terms of manpower, time, and equipment. Even with the high costs associated with drive testing, carriers do not have access to some regions and often can not anticipate the breadth of user devices, contexts, and functionalities with an in-field operation. Further complicating the problem, drive testing may have to be repeated with changes to the physical environment, such as the construction of new buildings or highways, seasonal variations, or modifications to the spatial distribution of users in the network [7].

Crowdsourcing has begun to be used as a less costly mechanism by which KPIs may be captured, as outlined in the Minimization of Drive Test (MDT) effort of Long-Term Evolution (LTE) release 10 in 3GPP TS 37.320 [8]. MDT allows carriers to monitor the in-situ network performance

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

R. Enami is with Qualcomm, San Diego, CA (e-mail: renami@smu.edu). S. Gupta, J. Camp, and D. Rajan are with the Department Electrical and Computer Engineering, Southern Methodist University, Dallas, TX 75275 USA (e-mail: sabyasachig@smu.edu; rajand@smu.edu; camp@smu.edu).

of end-users to detect variations of the provided quality of service (QoS). Depending on the nature of the problem, handover could be a solution for a single user, whereas self-organization could address an issue with one or more towers. Self-organization can alter antennae configuration in terms of transmit power, tilt, or height, but more persistent effects could necessitate smaller cell deployment in detected network holes. To make efficient use of the crowdsourced data (to preserve bandwidth and battery life of users), a natural extension of MDT is to interpolate the region's performance from discrete user locations using propagation models [1], [9] and coverage maps [10], [11]. Crowdsourced data have used to measure network metrics in [12]–[16]. However, none of these approaches directly consider the geographical features of an environment to predict the propagation characteristics and resulting KPIs.

In this paper, we build a framework to predict location specific and regional KPIs based on the available geographical data and crowdsourced channel measurements. Since location specific received signal power information can be used to infer many regional KPIs, *e.g.*, path loss, coverage area probability, we first aim to develop a framework that can predict the location specific received signal power with high accuracy by establishing a relationship between geographical data and crowdsourced channel measurements via neural networks. To do so, we use LiDAR data to represent the physical characteristics of an environment and an Android-based crowdsourcing infrastructure for in-field signal measurements. Then, a cone shaped filter with the apex at the transmit location is applied to extract the useful information from these two data sets to obtain the receiver location specific KPI estimation using Location-Specific Analysis to Infer Key Performance Indicators (LAIK). The relationships are formed by the use of ANNs consisting of a feed-forward, back-propagation model, which employs multi-layer perceptrons. Using LAIK, we can iterate the KPI prediction over a region to infer the path loss exponent in that type of area, spanning single-family and multi-family residential neighborhoods and downtown regions. We then evaluate our framework with various filter shapes and features selected as well as against state of the art Kriging Algorithm, a method of interpolation applied to cellular coverage prediction [10], [17], [18]. Our main contributions are:

- We introduce a neural-network based framework for mapping geographical information to crowdsourced signal quality measurements to infer location-specific KPIs.
- We design a novel cone-shaped filter with the apex at the transmitter location to extract the geographical data corresponding to the crowdsourced signal data over a region of interest. For this purpose, we investigate the role that delay spread has on adjusting the cone shape angle from the transmitter. We investigate KPI prediction performance using cone shapes that correspond to the minimum, maximum, and mean of the delay spread across single-family, multi-family, and downtown areas. Among the three delay spread values, we find that using the mean delay spread to form the cone shape minimizes the KPI prediction error.

- We consider the accuracy of predicting KPIs in areas in which the LAIK framework lacks any signal quality training, relying solely on the geographical features of the area. This reliance on geographical features allows our approach to have significant performance improvements over state-of-the-art schemes in areas for which measurement data is not available.
- In contrast to a square-tile approach that we previously used [3], the proposed LAIK framework aims to establish a relationship between geographical features over each signal path and corresponding crowdsourced signal measurement using cone shaped filter (tuned according to delay spread values) and neural network to obtain location specific KPI prediction. We use the LAIK ANN to predict location-based performance at various locations in a given area and then use linear regression to predict the path loss in the area. We show that the proposed LAIK model improves path loss prediction compared to the regional analysis to infer KPIs (RAIK) model [3].

The remainder of the paper is organized as follows. Section II reviews the related work. Section III presents the proposed LAIK framework. In Section IV, we analyze the impact of different angle sizes of the cone shape filter and the features selection strategies on location-specific KPI prediction. In Section V, we describe the procedure to obtain the path loss prediction using LAIK, and compare our proposed scheme against state-of-the-art methods in terms of accuracy of location-specific received signal strength prediction and path loss prediction. Finally, conclusions are drawn in Section VI.

II. RELATED WORK

Many different propagation models have been used to predict the coverage area of a network, such as Okumura-Hata [1], [9] and the Longley-Rice irregular terrain model [19]. In these models, one must collect radio signal measurements from a specific region to calibrate the model for that region and find the appropriate constants.

Recently, machine learning algorithms have emerged as an alternative approach to overcome the low accuracy of empirical models and the complexity of deterministic models while predicting the path loss propagation of a given region. Sohrabi *et al.* [20] used regression clustering and K-Nearest Neighbor (KNN) algorithms to construct the reference signal's received power (RSRP) maps from a sparse set of MDT measurements. This work assumed a fixed grid size. In [21], a reliable mobile coverage map generation framework was proposed that tolerated a reduced sample size. Based on the similarity and differences in signal strength in different sub-areas of a region, the region was divided into multiple clusters and non-uniform sampling was employed to build coverage maps. Chou *et al.* [22] proposed a crowdsourced data-driven indoor diagnostic framework for fault detection in cellular-based IoT networks. RSRP maps are constructed using machine learning algorithms such as random forest, neural network, linear regression, and decision tree for diagnosis purposes, and the performance of these algorithms are compared. None of these works [20]–[22] considered geographical features of the environment to

construct the RSRP map or coverage map. In [23], satellite two-dimensional (2-D) images were used to predict the path loss distribution of a region using deep learning methods. In this case, the height information of the objects was neglected. However, the height of the objects in between the transmitter and receiver can have an impact on the received signal strength due to reflections of the received signal.

Several studies have addressed the prediction of urban, suburban [4], [24], and indoor environments [6], [25] using Neural Network algorithms. Wu *et al.* [26] considered just the distance between the transmitter and the receiver as the input features to predict the path loss. Alotaibi *et al.* [27] proposed an adaptive neuro-fuzzy inference system to reduce the radial basis function neural network. The author simplified the input by considering the distance between the transmitter and receiver because the percentages of the area covered by buildings for the urban and suburban base stations were fixed. Support vector machine (SVM) and ANN models were compared in [28], considering the antenna-separation distance, terrain elevation, horizontal angle, vertical angle, latitude, longitude, horizontal, and vertical attenuation of the antenna as the input features. Geodata was not considered in this model.

A multilayer perceptron (MLP) based ANN technique was implemented to predict the path loss at 900 MHz [29]. A 2-D map was extracted from aerial photography to obtain geographical data. A General Regression Neural Network to predict the propagation path loss was used in [30]. Compared to the empirical models, a high prediction accuracy improvement was obtained due to considering the street width, rooftop height, and building block spacing as input features to the neural network. In [31], a neural network based pathloss prediction scheme was proposed that achieved low prediction error in multi-band scenarios.

All the above methods failed to consider the following essential items to predict received signal quality: (i.) Foliage (ii.) geographical features along the direct path between the base station and the user equipment (UE) and along multiple paths that are formed by reflectors or scatterers; (iii.) Geographical features in the effective area surrounding the UE.

In [3], we proposed the RAIK framework to predict KPIs. For this purpose, an MLP based neural network was implemented which was trained using crowdsourced measurements and geographical data. We showed that the accuracy of KPI prediction in a region depends upon the size of square-shaped tiles (effective area) for which to consider geographical data and RSRP to train the neural network. In contrast to the aforementioned works, to the best of the authors' knowledge, this is the first work that jointly considers the angle of the direct path and multipath components between the base station and the UE, and the effective area surrounding the UE to extract the geographical data (e.g., building and foliage) to train a neural network for location-specific and regional inference of KPIs.

The other method to predict propagation coverage over an area is utilizing geostatistical modeling techniques, where the measurements are collected strategically and different interpolation techniques are applied to predict the propagation model of the uncovered locations. For example, geo-

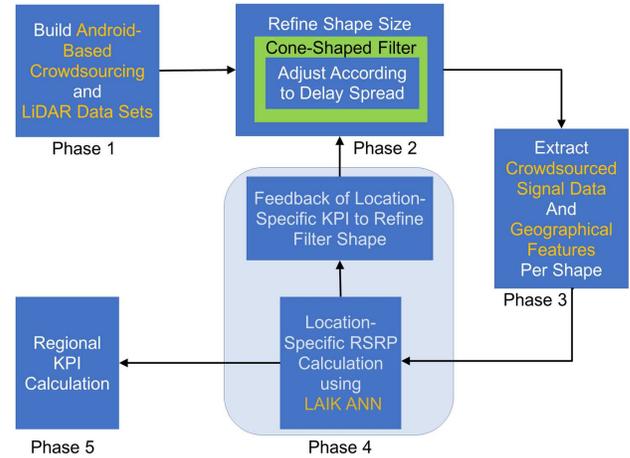


Fig. 1: Location-Specific Analysis to Infer KPIs LAIK Framework.

statistical modeling and interpolation technique is utilized to construct a radio environment map of 2.5 GHz worldwide interoperability for microwave access (WiMax) in [10]. The geostatistical modeling and interpolation technique is termed ‘Kriging’ in the geostatistical literature. The Kriging method has high computational complexity. Therefore, several works investigated low complexity implementation of the Kriging algorithm [11], [18]. Liu *et al.* [32] proposed a radio map construction method in which a triangulation scheme is applied to divide the region of interest into triangular sub-regions with measurements available at the vertex locations of the triangular sub-regions and then linear interpolation is applied to construct the radio map for each subregion. The triangulation method is adaptive to the base station location to reduce the measurement cost. In contrast to the radio map construction methods based on geostatistical modeling and interpolation techniques, we specifically target a relationship between the signal quality of a network at a given location and the geographical features in that area to predict the KPIs of that region and regions that lack accessibility or crowdsourced measurements.

III. FRAMEWORK TO INFER KPIs

In this section, we describe the LAIK framework, various data sources that are used in our approach, and the MLP components used.

A. Inferring KPIs at a Specific Location and over a Region

To predict KPIs for a specific location and ultimately for a specific region, we build a framework depicted in Fig. 1.

1) Phase 1: Building Android-based Crowdsourcing and LiDAR Data Sets

- a. We first build an android-based crowdsourcing infrastructure, which allows the widespread collection of in-field signal quality data coupled with the location of that user at the time of the measurement.
- b. Since the received signal attenuation is affected by foliage and buildings surrounding the UE, we

consider 3-dimensional geographical data from the region of interest. For this purpose, we use LiDAR data, which includes detailed information of buildings and foliage such as height and surface area (see Section III-B for more details). Also, the procedure to extract geographical features is presented in Fig. 2. In particular, a LiDAR point cloud from United States Geological Survey (USGS) is processed to consider terrain, buildings, and vegetation in three dimensions.

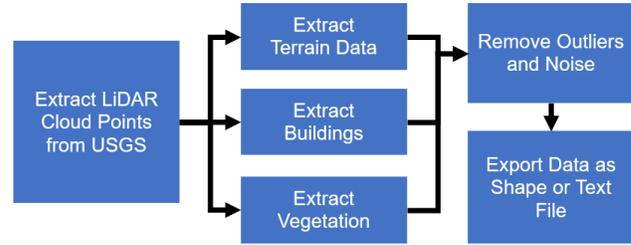


Fig. 2: Extracting Geographical Features from LiDAR Data.

- 2) *Phase 2:* We build a cone-shaped filter to extract geographical data corresponding to crowdsourced signal data over the relevant region. Given a particular environment, we decide the angle of the cone based on minimum, maximum, or average delay spread in the region. For this purpose, the performance of the LAIK is obtained using cone angles corresponding to three delay spread metrics in phase four, and this information is fed back to tune the cone angle such that the geographical data is extracted efficiently.
- 3) *Phase 3:* We extract information about all the objects that lie within the cone-shaped filter, to capture their impact on the signal propagation from the transmitter to the UE. Also, we extract all the geographical objects that lie in the circle created by the cone-shaped filter around the UE.
- 4) *Phase 4: Location-specific KPI Prediction and Tuning the Cone-Shaped Filter According to a Specific Region*
 - a. Location-specific RSRP in the given region is predicted by feeding the extracted geographical features and corresponding signal strength information to the LAIK ANN model.
 - b. Then, the predicted KPI is refined by setting the cone-shaped filter angle (and therefore the extracted geographical features) according to the maximum prediction accuracy of LAIK among the angles corresponding to three different delay spread metrics.
- 5) *Phase 5: Path-Loss Prediction in a Region.* Once the cone angle is decided, the LAIK model can be applied to predict the received signal strength at different locations in the region in which it is trained or in other regions in the same environment type. After predicting the received signal strength in different locations of the region, the path loss estimate is obtained.

B. Android-Based Crowdsourcing and LiDAR Data Sets

In this section, we describe the procedure for building the data sets in Phase 1. The two data sets on which our LAIK model is based are: (a.) Received signal quality data collected by Android phones; (b.) LiDAR, which describes the geographical features in the area. We consider these two data sets because the geographical features directly impact the received signal quality in a given region.

Android-Based Crowdsourced Data. We have a crowd-sourced data set, which is built from voluntary participants

who installed our publicly-available Android application (WiEye) to collect global radio measurements [33]. To limit the power and bandwidth consumed by our app, signal levels from all visible cellular and WiFi base stations are recorded 10 times per day. We have a development version of our app that captures measurements at a frequency of 1 Hz, which we have used to emulate a more concentrated user base in relevant geographical regions in this paper. We specifically record received signal strength across all available technologies, Global Positioning System (GPS) coordinates, Mobile Country and Network Codes, base station identification (CellId, Location Area Code), device identification, and velocity of the receiver (when locally collecting data).

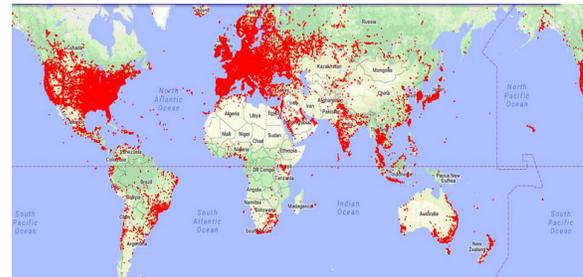


Fig. 3: LTE measurements collected from WiEye.

We have acquired hundreds of millions of crowdsourced signal strength data points using an Android-based application called WiEye.¹ Locally, we collected an additional 10 million measurements with greater densities in three representative geographical regions in Dallas: downtown, single-family, and multi-family residential areas. Due to propagation differences that arise across carrier frequencies, we extract the signal measurements for 2.1 GHz.

Generally, the density of the foliage in the single-family area is higher than the other two regions, the downtown area is mainly covered by tall buildings, and the multi-family area has a mixture of vegetation and moderately-sized buildings (e.g., 2-3 stories). Fig. 4 depicts the collected RSRP in three representative geographical regions: downtown, single-family residential, and multi-family residential. In each region, RSRP values are based on signals received from a single base station. It can be observed that the variation of received signal

¹WiEye collects signal measurements from different technologies such as WiFi, GSM, UMTS, and LTE. It offers users a free WiFi scanner and each user can contribute to our measurement campaign according to their choice (protected by an Institutional Review Board (IRB)). Users' data will be submitted to a central database server at SMU. Fig. 3 shows the spatial distribution of WiEye measurements.

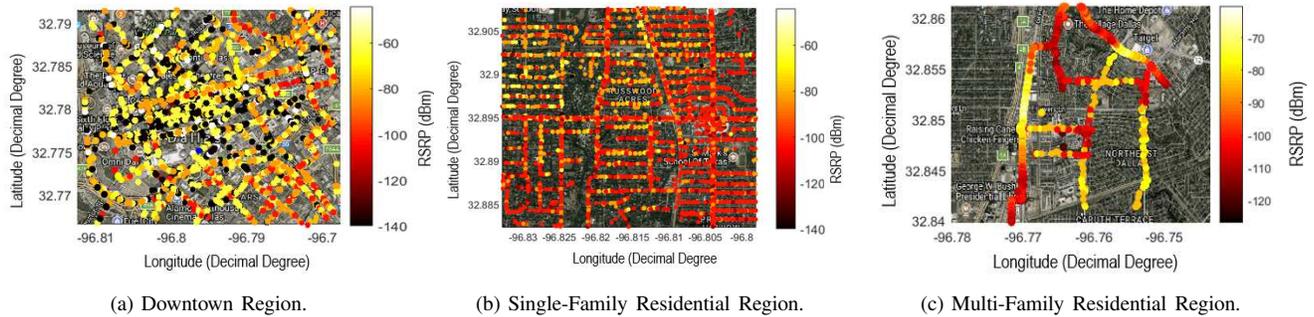


Fig. 4: RSRP from Downtown (left), single-family residential (middle), and multi-family residential (right) regions.

strength is different in each region due to the diversity of the geographical features across region types. The variation of the received signal strength in the downtown area is substantial from street to street. In particular, we observe very strong signals adjacent to dead zones with respect to the RSRP. In Section IV, we have investigated the optimal angle size of the cone shape filter that can characterize the multipath propagation environment of a wireless signal for each region.

In building a crowdsourcing data set, we specifically target the RSRP since: (i.) Network providers seek to provide coverage over an area to deliver sufficient quality of service to customers; (ii.) A well-known relationship exists between the received signal power and the throughput [34]; (iii.) UEs regularly measure the received signal power to keep track of visible base stations in case of handovers, even if the phone is idle. Thus, the battery consumption to measure RSRP is low and conducive to MDT efforts.

LiDAR-Based Geographical Features. To consider the vertical and horizontal footprints of trees and buildings, we use LiDAR data, which creates a 3-dimensional (3-D) point cloud of the Earth’s surface. LiDAR employs a remote sensing method from airplanes or helicopters that transmits pulses of light to detect the distance from the earth. The laser sends these pulses and measures the time delay between the transmitted and the received pulse to calculate the elevation. LiDAR systems are equipped with a laser scanner that measures the angle of each transmitted pulse and the returned pulse from the surface, high precision clocks which record the time that the laser pulse leaves and returns to the scanner, an Inertial Navigation Measurement Unit to measure the angular orientation of the scanner relative to the ground (pitch, roll, yaw), a data storage and management system, and a GPS detector.

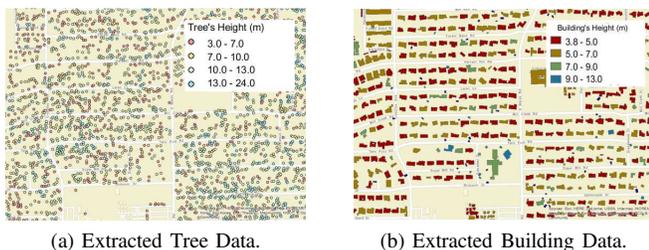


Fig. 5: 3-D map from same region using LiDAR.

The sampling rate of the LiDAR data that we use is 400×10^3 pulses per second, with an accuracy of about 15 cm

vertically and 40 cm horizontally. Hence, LiDAR systems provide a high-resolution 3-D geometric model for the earth, clutter, and foliage, with applicability across a broad range of fields such as archeology, geology, and seismology [35]. Relevant to our work, we use LiDAR to represent a 3-D map of building and tree data in the three Dallas regions under test. The procedure to extract geographical features is presented in Fig. 2. In particular, a LiDAR point cloud from USGS is processed to consider terrain, buildings, and vegetation in three dimensions. Outliers and noise are removed from each of these geographical features, and the result is output to a shape or text file. For wireless propagation characteristics, the crowdsourced received signal data is extracted for the filtered region.

Each record that corresponds to a tree in our 3-D map includes coordinates of the object, height, and area. We have the same information for buildings. Fig. 5 shows the detected trees and buildings in a suburban region in Dallas. The background of each figure is from OpenStreetMaps to verify the accuracy of the LiDAR information from the same area.

C. Multi-Layer Perceptron Components Used

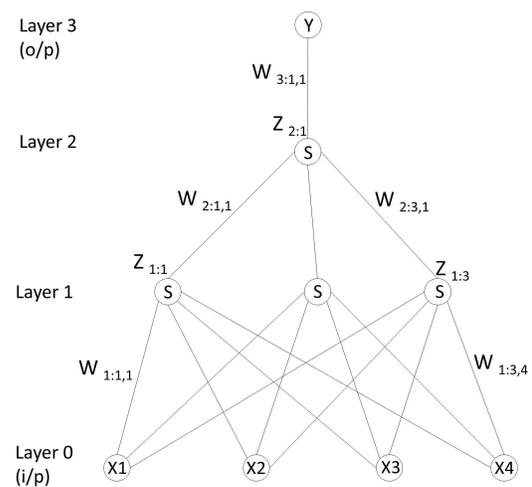


Fig. 6: Feed-forward neural network.

Neural network algorithms have been widely applied to predict the channel propagation in wireless networks [30], [36]–[38]. In our study, we use an MLP based ANN introduced

in [39] and [40]. MLP performs the Levenberg-Marquardt (LM) algorithm as a supervised-learning technique for training the network [41]. MLP consists of input, output, and hidden layers. The output of the j th node of the i th layer in a neural network can be expressed as:

$$Z_{i,j} = S \left(\sum_{k=1}^{N_{i-1}} w_{i,j,k} Z_{i-1:k} + \beta_{i,j} \right). \quad (1)$$

Here, N_{i-1} denotes the number of nodes in layer $i-1$, $w_{i,j,k}$ is the weight of the connection between k th node in layer $i-1$ and j th node in layer i , $\beta_{i,j}$ is the bias term for the j th node in layer i , and $S(\cdot)$ is the activation function. A typical feed-forward neural network is shown in Fig. 6.

In the present study, the sigmoid function [42] is mainly used as the activation function, which is easily differentiable with respect to the network parameters. The sigmoid function is expressed as:

$$S(x) = \frac{1}{1 + \exp(-x)}. \quad (2)$$

We have also investigated performance of the proposed LAIK framework with another commonly used activation function in Section IV-C.

In neural network algorithms, the goal is to find the best weights and biases ($\{w_{i,j,k}\} \cup \{\beta_{i,j}\}$) such that the difference between the predicted and target values is minimized.

The LM algorithm is adopted to update the weight vector of the neural network since it can be implemented efficiently and is considered to be one of the faster training methods with relatively good convergence performance [43]. The algorithm uses the sum of squares error as the objective function and uses the Jacobian matrix as an approximation of the Hessian matrix with a Newton-based updating rule [44]. Let $\mathbf{w} = [w_1, w_2, \dots, w_n]$ be the vector composed of all weights and biases of the neural network, *i.e.*, each element $w_l \in (\{w_{i,j,k}\} \cup \{\beta_{i,j}\})$, $l \in \{1, \dots, n\}$, and n is the total number of weights and biases. Then, the updating rule in iteration m is:

$$\mathbf{w}(m+1) = \mathbf{w}(m) - (\mathbf{J}^T(m) \cdot \mathbf{J}(m) + \mu \mathbf{I})^{-1} \cdot \mathbf{J}^T(m) e(\mathbf{w}(m)). \quad (3)$$

Here, \mathbf{I} is the identity matrix and

- $e(\mathbf{w}(m)) = \frac{1}{2} \sum_{j=1}^P e_i^2$ with P being the total number of training records, and e_i being the difference between the output of the MLP and the target for the i th record in the training set, in the m th iteration.
- $\mathbf{J}(m)$ is the Jacobian matrix of $e(\mathbf{w}(m))$. The Jacobian of $e(\mathbf{w}(m))$ can be computed using a back-propagation approach [44]. The dimension of the matrix is $P \times n$, and each element in row i and column j can be expressed as $\mathbf{J}[i, j] = \frac{\partial e_i}{\partial w_j}$.
- $\mu > 0$ is a damping factor that ensures that the matrix inversion will always produce a result. If $\mu = 0$, the updating rule is equivalent to Newton's method. If μ is high, the algorithm behaves as gradient descent with a small step size. At each iteration, μ is adjusted in the following manner: after each successful step, μ is decreased, and if

a step results in increase in the performance function, μ is increased.

IV. LOCATION-SPECIFIC KPI INFERENCE EVALUATION

In this section, we describe how we include all relevant geographical features in the analysis by leveraging the delay spread information to construct a cone-shaped filter from the tower to the surrounding region of a specific location. To do so, we evaluate the angle size of the cone shape and the features selected in the training process to understand their respective role in reducing KPI prediction error at a given location.

A. Calibration of the Cone-Shape Filter

We first aim to extract the data in a cone shaped geographical region with the apex at the transmit location that affects the signal propagation from the transmitter to the receiver. In this analysis, we consider two important factors: (*i.*) The obstacles along the direct path from the base station to the UE that have a significant effect on the propagation characteristics; (*ii.*) A broader region of relevant geographical features in addition to the direct path due to the existence of multiple paths from the base station to UE. Along with the line of sight signal reception through the direct path, multiple copies of the signal are received at the receiver due to reflection and diffraction of the transmitted signal in an environment with buildings and foliage. The multipath propagation environment of a wireless signal can be characterized with help of a cone-shaped filter in which the transmitter is situated at the apex of the cone and only the geographical features that lie within the cone-shaped filter create multiple copies of the signal at the receiver. The angle of the cone around the direct path might be larger or smaller depending on the degree to which multipath delay spread exists in the environment. We expect that if the delay spread is large, the relevant angle along the direct path would be larger, representing more distant structures that produce significant reflections. In contrast, we expect that if the delay spread is small, the relevant angle would likely be smaller.

We study the shape of the cone in terms of the angle around the direct path as a function of the: (*i.*) Delay spread in that physical location; (*ii.*) Environment type; (*iii.*) Distance from the base station to the receiver. By providing delay spread information to our prediction model, we can adapt the cone shape to understand the accuracy of location-specific KPI prediction. Considering the above-mentioned cases, we aim to design a cone-shaped filter that contains the objects along the direct path from the base station to the UE with an angle specific to the environment type.

Acquiring Delay Spread Information. As we mentioned before, the delay spread of the received signals implicitly shows the area which contains the objects that affect a signal while it travels from the transmitter to the receiver. The impulse response for the fading multipath channel can be expressed as:

$$h(t) = \sum_{i=0}^{L-1} a_i \delta(t - \tau_i). \quad (4)$$

Here, L is the number of paths, and a_i and τ_i are the attenuation and delay of the i^{th} path, respectively. The delay

between the paths depends on the height and relative spatial locations of the geographical features in the environment. The difference between the arrival time of the last path and the first path is called delay spread, which can be obtained by:

$$\tau_d = \frac{\Delta x}{C}. \quad (5)$$

Here, Δx is the difference in length between the last path and the first path (in meters), and C is the speed of light (3×10^8 m/sec.).

We use a TSMW channel scanner from Rohde & Schwarz, to measure the delay spread in each of the regions where we collect data, i.e., downtown, single-family, and multi-family residential areas. The TSMW provides the average, minimum, and maximum values of the delay spread of the received signals at a particular measurement location which is depicted in Table I. Also, each delay spread value in time is converted into an equivalent distance using (5). As expected, the largest and the smallest values correspond to the downtown and single-family regions, respectively.

B. Impact of the Cone-Shaped Filter Angle Size

In this section, we determine an appropriate angle for the cone-shaped filter that captures the impact of obstacles which scatter, reflect, and diffract the transmitted signal before it reaches the receiver. To do so, we perform the following steps:

Step 1. The first step is to find the angle of the cone-shape filter as a function of the delay spread value. We consider an isosceles triangle² to obtain the angle between the reflected beam and direct signal as shown in Fig. 7. The direct path between the transmitter and the receiver has a distance of b m. The two equal sides of the isosceles triangle create the longest path, of length $2a$ m, corresponding to the delay spread. Therefore, we have $\Delta x = 2a - b$ and $\tau_d = \frac{2a-b}{C}$.

Let θ be the angle of the cone-shaped filter, and therefore, $\theta/2$ is the equal angle of the isosceles triangle. Therefore, the angle, θ , of the cone-shaped filter can be computed as:

$$\theta = 2\cos^{-1}\left(\frac{b}{\tau_d C + b}\right). \quad (6)$$

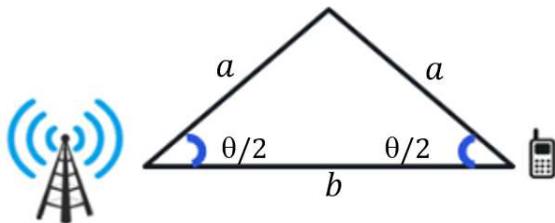


Fig. 7: Filter's Angle Sizes for Different Regions.

Step 2. The estimated angle would be different at each location due to the variation of the observed delay spread. Since users would not have the ability to measure delay spread

²We consider an isosceles triangle to build the cone-shaped filter since a direct relationship between the cone angle and the delay spread can easily be obtained, and LAIK can achieve high prediction accuracy based on geographical feature extraction using such a cone-shaped filter. However, our framework is flexible enough to use other triangle types.

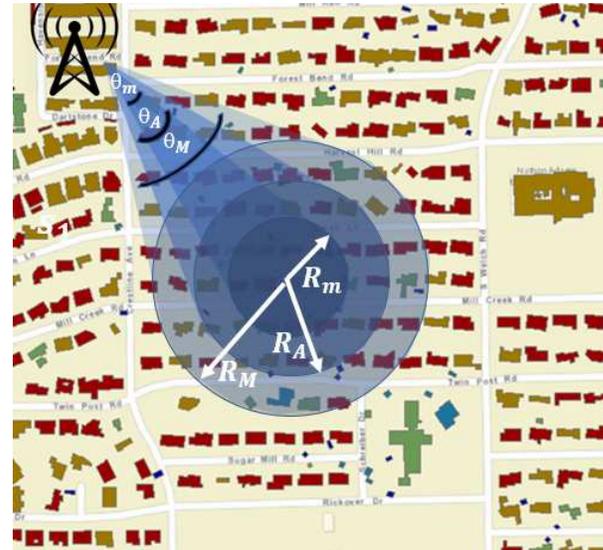


Fig. 8: The reference area vs smaller and larger areas.

from the UE, we find the most appropriate angle for each region type so that a carrier may then use crowdsourced signal strength measurements for location-based KPI prediction. For this purpose, we consider three different angles which correspond to the minimum, average, and maximum delay spread observed in that region.

The minimum, maximum, and average angles for the cone-shaped filter in a region are calculated using the following procedure. We find the maximum delay spread corresponding to the farthest measurements and based on this value, we calculate the maximum angle θ_M using (6). Similarly, we find the minimum delay spread corresponding to the nearest measurements and we calculate the minimum angle θ_m using (6). Then, we find the average angle θ_A for the cone-shaped filter as the mean of θ_m and θ_M . In Table II, we provide the values of the angles θ_m , θ_M , and θ_A obtained based on this approach in three different regions. It can be observed that the filter angle sizes for the downtown area are larger than the single-family and multi-family area due to the higher density of the buildings in the downtown area as compared to the other two areas.

The 3-D filter with an angle tuned according to θ_m , θ_A , or θ_M creates a circular footprint around each target location. The objects surrounding the receiver lie within this circular footprint, as depicted in Fig. 8.

Step 3. Next, for each signal measurement in the given area, we extract all the objects located within the range of the θ_m , θ_A , and θ_M . Each surrounding object may or may not intersect with the cone-shaped filter in the vertical plane, depending on the object's height. For example, in Fig. 9, the objects O_2 , O_3 , and O_4 intersect with the cone-shaped filter with angle size θ_A in the vertical plane. Therefore, these objects are extracted to train the model when the angle of the cone-shaped filter is set according to the average delay spread of the region. Note that we consider an error margin of 15 cm for each object's height to compensate for the vertical error of the LiDAR data set. We repeat the process for all the signal measurements

TABLE I: Delay Spread Ranges for Three Different Environment Types.

Area Type	Min Delay Spread		Max Delay Spread		Mean Delay Spread	
	in sec.	in meter	in sec.	in meter	in sec.	in meter
Single Family	0.01	3	1.9	570	0.33	99
Multi Family	0.01	3	2.1	630	0.34	102
Downtown	0.2	60	3.1	930	0.8	240

TABLE II: Delay Spread Corresponds to Three Different Environment Types.

Region	Single Family	Multi Family	Downtown
θ_m	22	22	28
θ_M	74	84	110
θ_A	48	53	69

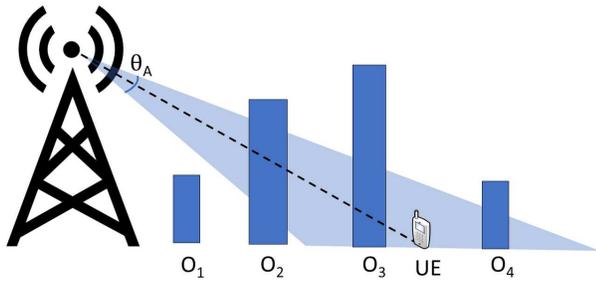


Fig. 9: Extracting the objects using cone-shape filter.



Fig. 10: Extracting geodata using cone-shaped filter.

by moving the cone-shaped filter across the collected signal measurements in the area of interest as depicted in Fig. 10. Here, S_1 , S_2 , and S_3 are the signal measurement locations.

Step 4. Lastly, we train the LAIK framework with three different data sets obtained from three different area types to predict the received signal level as the output of the model. In this experiment, we only consider the objects that intersect the cone shaped filter to train the network and therefore, evaluate the impact of the objects that intersect the direct path from the transmitter to the receiver on received signal strength prediction. The selected input features are: (i.) distance between the transmitter and the receiver, (ii.) number of buildings and trees in the direct path, (iii.) average height of the buildings and trees in the direct path, and (iv.) standard deviation of the heights of the buildings and trees in the direct path.

Table III depicts the performance of the ANN model with different cone-shaped filter angles. Here, R is the correlation coefficient between the predicted signal strength and the corresponding actual value. The correlation factor varies between +1 and -1. It can be observed that the mean absolute error (MAE) in dB and standard deviation of error reduces and the correlation factor increases if the cone angle is increased to θ_M or θ_A from θ_m for a given region. With the smaller angle size (θ_m) of the filter, some of the objects that affect the transmitted signal are ignored. The standard deviation of error and MAE is minimized, and the correlation factor is maximized when the angle of the cone-shaped filter is calibrated according to the average delay spread of the channel. In this case, the filter with cone angle θ_A more accurately extracts the objects that affect the transmitted signal compared to a filter with a wider angle θ_M or smaller angle θ_m . Furthermore, the improvement in the correlation factor when the filter angle size increases from θ_m to θ_A is higher in the downtown region as compared to the single-family or multi-family residential areas. This effect is due to the presence of more tall buildings in the downtown area that affect the transmitted signal from greater distances.

C. Impact of the Feature Selection Strategies

In this section, we show the importance of the geographical feature selection strategies on the performance of the prediction model to estimate the signal strength in a particular area. The angle size of the cone-shaped filter in each region is calibrated according to the average delay spread since we showed in the previous section that the performance of LAIK is improved when the angle size of the filter is set in this manner. To investigate the importance of various inputs on the prediction performance, we consider multiple ANN, each with a different subset of input features as follows: a) ANN_a only considers the distance between the transmitter and receiver. b) ANN_b uses the distance (between transmitter and receiver) and the extracted objects along the direct path using the cone-shaped filter. c) ANN_c uses the distance and the obstacles surrounding the UE. d) Finally, ANN_d uses the distance, the extracted objects along the direct path using the cone-shaped filter, and the objects surrounding the UE.

Given an input features selection strategy ANN_b , ANN_c or ANN_d , the input features to the LAIK model are: (i.) distance between the transmitter and the receiver, (ii.) percentage of the area covered by buildings (i.e., footprint) [29], trees (i.e., canopy or crown), and free space (i.e., unoccupied by trees or buildings), (iii.) number of buildings and trees, (iv.) average height of the buildings and trees, and (v.) standard deviation of the heights of the buildings and trees. The input parameters for a feature selection strategy have been extracted from a 3-D LIDAR database using the cone-shaped filter. The model's

TABLE III: The impact of the one-shape filter angle on the LAIK performance.

Area Type	Single Family			Multi Family			Downtown		
Filter Angle Size	θ_m	θ_A	θ_M	θ_m	θ_A	θ_M	θ_m	θ_A	θ_M
Std	6.2	4.7	5.7	6.4	4.8	5.5	6.5	4.9	5.8
MAE	5.2	4.5	4.7	5.0	3.9	4.1	5.4	4.5	4.9
R	0.86	0.89	0.88	0.84	0.87	0.85	0.77	0.88	0.83

TABLE IV: The impact of the input features on the ANN performance.

Area	Single Family				Multi Family				Downtown			
Model	ANN_a	ANN_b	ANN_c	ANN_d	ANN_a	ANN_b	ANN_c	ANN_d	ANN_a	ANN_b	ANN_c	ANN_d
Std	7.2	4.3	3.8	2.1	9	4.4	4.6	2.5	9.8	4.4	5.5	3.1
MAE	5.5	3.6	3.3	2.9	6.7	3.3	4.3	3.0	6.9	4.1	4.8	3.2
R	0.76	0.91	0.92	0.94	0.73	0.89	0.83	0.93	0.72	0.87	0.8	0.9

output is the location-specific predicted received signal level (*i.e.*, RSRP). To increase the efficiency of the model, all features are normalized to fall in a range of [0, 1].

Table IV shows a comparison between the performance metrics of LAIK for four different feature selection strategies across three different environments. The results show that the training model that only considers the distance as the input feature has inferior performance compared to the other approaches. Therefore, geographical features play a critical role in determining the received signal strength. In a single-family environment, the surrounding objects affect the received signal strength prediction accuracy more compared to the obstacles along the direct path due to the density of trees around the UE, which scatter the received signal. Therefore, ANN_c performs better than ANN_b in this environment. The results show that the standard deviation and the MAE of the prediction model are minimized when we consider the distance, the objects along the direct path, and the objects surrounding the UE as the input to the model. Similarly, the correlation factor is maximized when we consider all the above-mentioned features to train the ANN model. In this case, a significant improvement in the performance metrics is achieved if, along with the distance between the transmitter and receiver, the information on the extracted objects along the direct path is also considered to train the neural network. For example, when ANN_b is employed instead of ANN_a , the reduction in MAE and the standard deviation are 34 – 51% and 40 – 55%, respectively, across different environments. An additional improvement is achieved by considering the objects surrounding the UE along with the distance, and the objects along the direct path as the input to the model. If ANN_d is employed instead of ANN_b for location-specific signal strength prediction in three different environments, the reduction in MAE and standard deviation are 10 – 20% and 30 – 51%, respectively, and the increase in correlation factor is 3 – 4%.

To show the impact of vegetation and building geographical data on the ANN model’s performance, we train the model without vegetation and building features. For this purpose, we consider two extreme environments of vegetation and building coverage when training the ANN model: downtown, and single-family residential areas. In both these regions, the

ANN model is trained with three geographical data types: (*i.*) only the presence of building data, *i.e.*, no vegetation, (*ii.*) only the presence of vegetation data, *i.e.*, no buildings, and (*iii.*) all data are included, which we refer as ANN_d . The accuracy of signal strength prediction using LAIK with each of these geographical data types is evaluated in terms of the standard deviation of error, MAE, and correlation factor, and the results are shown in Table V. The results show that the MAE performance of the single-family residential area by eliminating the vegetation and building data drops about 1.8 dB and 0.9 dB, respectively. Also, by ignoring the vegetation and building data, the performance of the downtown area in terms of the MAE decreases by 1.7 dB and 3.1 dB, respectively. Additionally, it can be observed that the building data has a greater impact in terms of accuracy of prediction compared to vegetation data in the downtown area while the effect of these two data types is reversed in the case of the single-family residential area.

Next, we compare the performance of the proposed LAIK ANN_d with the Rectified Linear Unit (ReLU) function as the activation instead of the sigmoid function. Both sigmoid and ReLU functions are commonly used as activation functions. The ReLU activation function can be expressed as $S_{ReLU}(x) = \max(0, x)$. In Table VI, we show the performance of LAIK ANN_d when the ReLU function is selected as the activation function of the neural network. We compare the results with the performance of ANN_d depicted in Table IV. When ReLU is used as the activation function instead of the sigmoid function, the correlation factor improves by 0.02 and MAE reduces by 0.4 dB in the case of the downtown area. Also, the correlation factor reduces by 0.01, and the MAE increases by 0.2 dB for the single-family and multi-family residential areas. Both the sigmoid and ReLU activation functions can be used in the learning framework to demonstrate the benefits of the proposed approach. Quantifying the cases where the activation functions are optimal can be studied in future work. We choose the sigmoid function as the activation function of the neural network for the rest of the analysis in this paper.

TABLE V: The impact of the buildings and vegetation on the ANN performance in different environments.

Area	Single Family			Downtown		
	ANN_d	No Vegetation	No Building	ANN_d	No Vegetation	No Building
Std	2.1	4.2	3.1	3.1	3.3	6.5
MAE	2.9	4.7	3.8	3.2	4.9	6.3
R	0.94	0.81	0.85	0.9	0.86	0.75

TABLE VI: Performance of LAIK ANN_d with activation function as ReLU function.

Area	Single Family	Multi Family	Downtown
MAE	3.1	3.2	2.8
R	0.93	0.92	0.92

V. PERFORMANCE ANALYSIS OF LAIK VERSUS EXISTING APPROACHES

In this section, we compare the LAIK framework with two state-of-the-art approaches, Kriging and RAIK [3]. First, we compare to Kriging in terms of location-specific received signal strength prediction. Then, we describe the process of iteratively using location-specific prediction for regional prediction and compare the performance against RAIK for path loss prediction in a given region.

A. Location-Specific Comparison: LAIK vs. Kriging

In this section, we evaluate performance of the LAIK when it is used to predict the RSRP of an arbitrary location. To evaluate LAIK in context of the most relevant related works in coverage prediction, we compare it with the Kriging algorithm, which is a common approach to address the spatial propagation prediction [10], [18], [45]. To predict the lost data in a region, Kriging employs regression of the surrounding values of that region by assigning weights to these values to capture the spatial correlation of field of interest. Many studies have used Kriging to estimate the path loss [10], [46], [47]. For example, an empirical Okumura-Hata model with Inverse Distance Weighting (IDW) and Kriging has been evaluated in prior work [47]. They have shown that the approach with Kriging achieves an improved performance versus the Okumura-Hata and IDW models.

To compare LAIK with the Kriging algorithm, we first apply the two algorithms to predict the received signal level of a certain coordinates, and then we compare the MAE of the RSRP prediction. We select the largest group of measurements corresponding to a base station in the downtown area and train two models using all the signal measurements. Even though the output of both methods is the predicted received signal strength, the models' inputs are different. The input of the Kriging algorithm is the signal measurement coordinate and the output is the received signal power. The input of LAIK is decided based on the feature selection strategy ANN_d . Since the Kriging algorithm is vulnerable to large distances, we design two special cases as follows to compare the LAIK and Kriging performance:

a) The distribution of the collected crowdsourced signal measurements usually does not follow a uniform pattern.

In real life, adjacent to a region for which a large set of measurement data is available, a region with a lack of information may exist. To model this scenario, we consider a square area with 200-m sides and then intentionally remove the measurements in the middle of our area of interest, as shown in Fig. 11. Fig. 11a shows the area of interest with all signal measurements. We provide a measurement hole, emulating the lack of crowdsourced measurements, as shown in Fig. 11b. The dots show the available measurements, and the lack of dots denotes the lack of signal measurements. Fig. 12a shows the estimated received signal level by the Kriging algorithm when all signal measurements are used. Fig. 12b shows the results of Kriging prediction in the absence of the selected signal quality measurements, from which we find the MAE to be 4.1. We conduct the same analysis on this region with LAIK and find the MAE to be 2.1. Therefore, the use of geographical data to predict the received signal level can reduce the error two-fold. However, note that this is a relatively small region (a subspace of a 200-m square region). In reality, the distances that a carrier might wish to infer location-specific propagation could be multiple km away from existing measurements. Correspondingly, in the next experiment, we show the LAIK and Kriging performance when the available adjacent measurements are far from the area that we want to predict the corresponding RSRP.

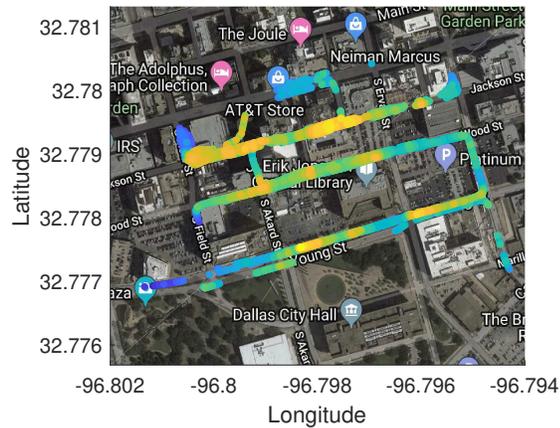
b) We now select a measurement route in the downtown region, which covers a distance of about 1000 m from a specific transmitter. We temporarily remove the measurements of the selected route from the data set for testing purposes. The signal measurements that are captured in a distance less than 400 m from the transmitter are considered as the training data set. Then we train the Kriging and LAIK models using the same training data set, and we evaluate the performance of both models using the same test data set.

In Table VII, we take a look at various distances away from an existing deployment to an area that we assume there is no measurements in an adjacent region, and we are trying to predict the performance of that area. The MAE of the predicted signal strength by the Kriging algorithm increases when the distance between the training points and the testing points increases. We observe that the largest MAE (the largest RSRP prediction error) occurs at distances of 800 to 1000 m.

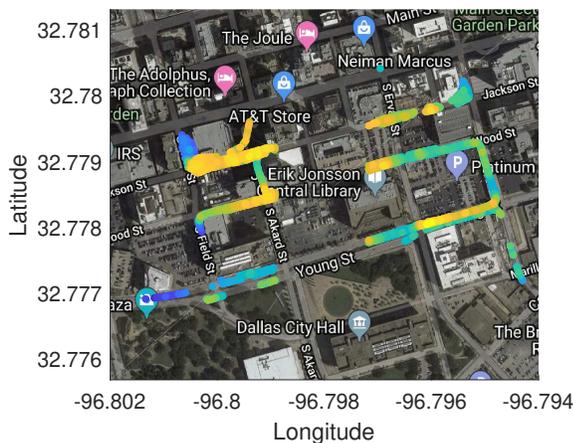
To interpret RSRP prediction error in terms of operational network performance, we depict the role of this prediction error on the channel throughput (bits/Hz) estimation. To do so, we use throughput in terms of the Signal-to-Noise Ratio (SNR) for various modulation and coding as measured in a previous work [48]. Since SNR is the ratio of transmit power to the noise power and a linear relationship exists between SNR and RSRP, we can use RSRP in the throughput prediction.

TABLE VII: Comparing LAIK and Kriging Performance in terms of Received Signal Strength Prediction.

Scheme	0-100m	100-200m	200-300m	300-400m	400-500m	500-600m	600-700m	700-800m	800-900m	900-1000m
MAE_{LAIK}	2.7	2.9	4.2	3.6	4	3.20	3.4	3.1	3.2	3.9
$MAE_{Kriging}$	2.2	2.5	4.5	4.7	4.8	5	10	10	15	15



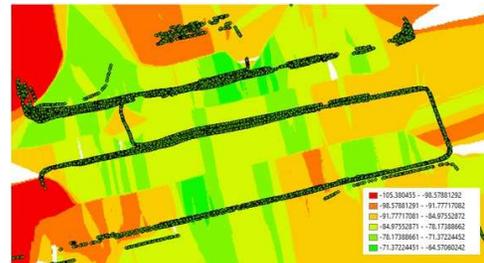
(a) Collected signal in downtown region.



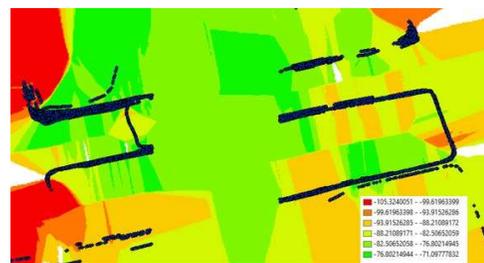
(b) Removing a group of data.

Fig. 11: Non-uniform data selection.

In Table VII, we observe that the error is approximately constant over various distances in case of LAIK, while in case of the Kriging scheme, it grows very rapidly. In fact, at the largest distance, the difference between the LAIK and Kriging algorithms is about 12 dB (3.9 vs. 15). Considering [48, Fig. 19], if we revisit the impact of the user performance based upon the error in RSRP prediction, we observe that the estimated throughput that would relate to the ground truth of real measurements is 4 bits/Hz whereas LAIK would estimate 3.5 bits/Hz for that user. However, Kriging would estimate a throughput performance of 0.5 bits/Hz. As we can observe, by applying the LAIK model, the error ratio is decreased 7-fold. Note that this could lead to incorrect estimation of link quality at the carrier in two ways. Firstly, an error could exist in the form of overestimating the throughput performance of a user in a certain area where the carrier thinks



(a) Kriging interpolation on all signal measurements.



(b) Kriging interpolation on selected signal measurements.

Fig. 12: The impact of the clustered data on Kriging prediction performance.

that the throughput performance at the user is satisfactory, while actually there is a problem and the carrier doesn't know about it. Secondly, an error could exist in the form of underestimating the performance where the carrier thinks there is a problem that needs fixing, but the throughput performance at the user is satisfactory. The overall conclusion that can be obtained from Table VII is that the dependence on distance from existing measurements is vastly reduced in LAIK by relying on the underlying geographical features to form this prediction. To this point, the trained model using Kriging can not be generalized for any other region. The LAIK framework uses every measurement to train and improve the prediction model and a generalized LAIK model can be applied to any similar area to which the model has been trained.

B. Regional Comparison: LAIK vs. RAIK

In this section, we evaluate the performance of LAIK in terms of regional KPI inference. To do so, we compare the LAIK performance with our previous model [3]. Using the RAIK framework [3], we map between geographical features in a region of interest and the corresponding path loss exponent using square-shaped tiles around the receiver. However, since the output of LAIK is the location-specific received signal power (RSRP), we need to add one more step to calculate the path loss exponent of certain area using the predicted RSRPs by LAIK. To do so, we iteratively use various locations and

the resulting predicted RSRP in a particular region to calculate the path loss exponent of that region using a linear regression algorithm.

In this experiment, we train both models with the extracted information from the tiles with more than 700 signal measurements and a tile size of 200 m square. More detailed information regarding the selected tile size and the number of measurements can be found in [3]. To increase the size of the training data, the tile is shifted with a step of 50 m. The results show that LAIK improves the path loss exponent prediction on average about 0.2 compared to the RAIK model because it considers more detailed geographical data (of the full path with consideration to multipath components) that affect the received signal strength to train the model.

To interpret the reduction in MAE with LAIK compared to RAIK in terms of operational network performance, we analyze the performance of these strategies in terms of coverage area probability [49]. The impact of the error in path loss exponent prediction in case of overestimation and underestimation on the probability of cell area coverage estimation is shown in [49, Fig. 15(a)]. We consider the case where the actual average path loss exponent in the multi-family area is 3.3. Considering the case in which the predicted path loss exponent using RAIK is greater than actual, the coverage area probability drops from 47% to 29%, in case the predicted path loss exponent is 3.6. In other words, the estimated error in predicting γ is +0.3. Considering LAIK, the path loss exponent prediction error is +0.2, which results in a drop in the coverage area probability from 47% to 41%. Therefore with a fixed transmission power, the coverage area prediction accuracy is improved by 12% when the LAIK framework is employed for path-loss prediction versus the RAIK framework.

VI. CONCLUSION

In this paper, we have proposed a framework to infer location specific KPIs and regional KPI by establishing a relationship between geographical data and crowdsource channel measurement using neural networks. For this purpose, we built an Android based crowdsourcing infrastructure and performed in-field measurements to create a high density of signal strength measurements in three representative region types: downtown, single-family, and multi-family residential. Then, delay spread measurements corresponding to a region were used to build a cone-shaped filter, and the geographical and user data corresponding to each signal measurement location were extracted using the filter to build a neural network based KPI prediction model. We have shown that the cone-shaped filter angle should be set according to the mean delay spread of the region. We have also shown that for the best performance, the input features to the ANN are the distance between the transmitter and UE, the objects along the direct path and multipath components, and the objects surrounding the UE. Using LAIK, we iterate the KPI prediction over a region to predict the path loss exponent in that type of area, spanning the aforementioned three region types. It has been shown that compared to the state-of-the-art Kriging algorithm and RAIK, significant improvement in KPI prediction accuracy can be

achieved using the proposed LAIK framework. Furthermore, we have shown that LAIK can predict the KPIs with high accuracy in areas in which it lacks any signal quality training, relying solely on the geographical features of the area. Lastly, we have developed an intuition for the importance of crowdsourcing-based propagation prediction by evaluating the effect on coverage estimation when deploying operational networks. For example, we have shown that the RSRP prediction error would affect the channel throughput, and a 7-fold improvement in throughput estimation can be achieved using LAIK. Also, we have shown that LAIK can improve the coverage area prediction accuracy by 12% compared to the RAIK. We believe that this work will serve as a fundamental step in extending the reach of MDT measurements taken by carriers and thereby reduce the load on users and their devices. The overall analysis was carried out at a particular carrier frequency. However, the methodology described in this paper is general and can be applied for other lower or higher frequencies, and the effect of various attributes in those cases can be quantified.

ACKNOWLEDGMENT

This work was in part supported by NSF grants CNS-1150215 and CNS-1526269 and Air Force Office of Scientific Research grant FA9550-19-1-0375. Also, we would like to thank Textron Systems for their support in this measurement campaign.

REFERENCES

- [1] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, 1980.
- [2] *Digital mobile radio towards future generation systems*. COST Action 231 Final Report, EUR 18957, pp. 5, ch.4 pp. 21–26, 1999.
- [3] R. Enami, D. Rajan, and J. Camp, "RAIK: Regional analysis with geodata and crowdsourcing to infer key performance indicators," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017.
- [4] E. Östlin, H.-J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2735–2747, 2010.
- [5] S. P. Sotiroudis and K. Siakavara, "Mobile radio propagation path loss prediction using artificial neural networks with optimal input information for urban environments," *AEU-International Journal of Electronics and Communications*, vol. 69, no. 10, pp. 1453–1463, 2015.
- [6] A. Neskovic, N. Neskovic, and D. Paunovic, "Indoor electric field level prediction model based on the artificial neural networks," *IEEE Commun. Lett.*, vol. 4, no. 6, pp. 190–192, 2000.
- [7] S. Yi, S. Chun, Y. Lee, S. Park, and S. Jung, *Radio Protocols for LTE and LTE-advanced*. John Wiley & Sons, 2012.
- [8] *ETSI TS 137 320 "Radio measurement collection for Minimization of Drive Tests (MDT)"*, 2011.
- [9] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Rev. Elec. Commun. Lab.*, vol. 16, no. 9, pp. 825–73, 1968.
- [10] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Oct. 2012.
- [11] H. Braham, S. B. Jemaa, B. Sayrac, G. Fort, and E. Moulines, "Low complexity spatial interpolation for cellular coverage analysis," in *Proc. WiOpt*, May 2014.
- [12] S. Sonntag, J. Manner, and L. Schulte, "Netradar-measuring the wireless world," in *Proc. WiOpt*, May 2013, pp. 29–34.
- [13] A. Nikraves, H. Yao, S. Xu, D. Choffnes, and Z. M. Mao, "Mobilizer: An open platform for controllable mobile network measurements," in *Proc. 13th Annual Int. Conf. Mobile Systems, Applications, and Services*, 2015, pp. 389–404.

- [14] J. Yoon, S. Sen, J. Hare, and S. Banerjee, "WiScape: A framework for measuring the performance of wide-area wireless networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 8, pp. 1751–1764, 2015.
- [15] J. Huang, C. Chen, and Y. Pei, "MobiPerf: Mobile network measurement system," 2011.
- [16] S. Rosen, S. Lee, J. Lee, P. Congdon, Z. M. Mao, and K. Burden, "MCNet: Crowdsourcing wireless performance measurements through the eyes of mobile devices," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 86–91, 2014.
- [17] Y. Hu and R. Zhang, "A spatiotemporal approach for secure crowd-sourced radio environment map construction," *IEEE/ACM Trans. Netw.*, vol. 28, no. 4, pp. 1790–1803, 2020.
- [18] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, and B. Sayrac, "Fixed rank kriging for cellular coverage analysis," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4212–4222, 2017.
- [19] G. A. Hufford, A. G. Longley, and W. A. Kissick, *A guide to the use of the ITS irregular terrain model in the area prediction mode*. Nat. Telecommun. Inf. Admin., Washington, DC, USA, 1982.
- [20] F. Sohrabi and E. Kuehn, "Construction of the RSRP map using sparse MDT measurements by regression clustering," in *IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
- [21] M. Fida, A. Lutu, M. K. Marina, and . Alay, "ZipWeave: Towards efficient and reliable measurement based mobile coverage maps," in *IEEE INFOCOM*, May 2017, pp. 1–9.
- [22] S. Chou, H. Yen, and A. Pang, "A REM-Enabled diagnostic framework in cellular-based IoT networks," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5273–5284, 2019.
- [23] O. Ahmadien, H. F. Ates, T. Baykas, and B. K. Gunturk, "Predicting path loss distribution of an area from satellite images using deep learning," *IEEE Access*, vol. 8, pp. 64 982–64 991, 2020.
- [24] K. Stocker, B. Gschwendtner, and F. Landstorfer, "Neural network approach to prediction of terrestrial wave propagation for mobile radio," in *IEE Proc. H-Microw. Antennas Propag.*, vol. 140, no. 4, 1993, pp. 315–320.
- [25] I. Popescu, P. Constantinou, M. Nafornita, and I. Nafornita, "Generalized regression neural network prediction model for indoor environment," in *Proc. Int. Symposium Computers and Commun.*, Jun. 2004, pp. 657–661.
- [26] D. Wu, G. Zhu, and B. Ai, "Application of artificial neural networks for path loss prediction in railway environments," in *Proc. CHINACOM*, Aug. 2010, pp. 1–5.
- [27] F. D. Alotaibi, A. Abdenour, and A. A. Ali, "A robust prediction model using ANFIS based on recent TETRA outdoor RF measurements conducted in Riyadh city–Saudi Arabia," *AEU-International Journal of Electronics and Communications*, vol. 62, no. 9, pp. 674–682, 2008.
- [28] R. D. Timoteo, D. C. Cunha, and G. D. Cavalcanti, "A proposal for path loss prediction in urban environments using support vector regression," in *Proc. Advanced Int. Conf. Telecommun.*, Oct. 2014, pp. 1–5.
- [29] A. Neskovic and N. Neskovic, "Microcell electric field strength prediction model based upon artificial neural networks," *AEU-International Journal of Electronics and Communications*, vol. 64, no. 8, pp. 733–738, 2010.
- [30] I. Popescu, A. Kanstas, E. Angelou, L. Nafornita, and P. Constantinou, "Applications of generalized RBF-NN for path loss prediction," in *Proc. IEEE Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sept. 2002, pp. 484–488.
- [31] M. Ayadi, A. Ben Zineb, and S. Tabbane, "A UHF path loss model using learning machine for heterogeneous networks," *IEEE Trans. Antennas Propag.*, vol. 65, no. 7, pp. 3675–3683, 2017.
- [32] Y. Liu, W. Huangfu, H. Zhang, and K. Long, "Multi-criteria coverage map construction based on adaptive triangulation-induced interpolation for cellular networks," *IEEE Access*, vol. 7, pp. 80 767–80 777, 2019.
- [33] R. Meikle and J. Camp, "A global measurement study of context-based propagation and user mobility," in *Proc. 4th ACM Int. workshop on Hot topics in planet-scale measurement (ACM HotPlanet)*, Jun. 2012.
- [34] A. Nikravesh, D. R. Choffnes, E. Katz-Bassett, Z. M. Mao, and M. Welsh, "Mobile network performance from user devices: A longitudinal, multidimensional analysis." in *PAM*, vol. 14. Springer, 2014, pp. 12–22.
- [35] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing*. Guilford Press, 2011.
- [36] T. Balandier, A. Caminada, V. Lemoine, and F. Alexandre, "170 MHz field strength prediction in urban environment using neural nets," in *Proc. IEEE Personal, Indoor and Mobile Radio Commun. (PIMRC)*, Sept. 1995, pp. 120–124.
- [37] R. Fraile and N. Cardona, "Macrocellular coverage prediction for all ranges of antenna height using neural networks," in *Proc. IEEE Int. Conf. Universal Personal Communications (ICUPC)*, Oct. 1998, pp. 21–25.
- [38] P.-R. Chang and W.-H. Yang, "Environment-adaptation mobile radio propagation prediction using radial basis function neural networks," *IEEE Trans. Veh. Technol.*, vol. 46, no. 1, pp. 155–160, 1997.
- [39] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.
- [40] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ:Prentice-Hall, 1999.
- [41] M. H. Hassoun, *Fundamentals of artificial neural networks*. MIT press, 1995.
- [42] R. Rojas, *Neural networks: A systematic introduction*. Springer Science & Business Media, 2013.
- [43] H. Yu and B. M. Wilamowski, *Industrial Electronics Handbook*. CRC Press, 2011, ch. Levenberg–Marquardt training, pp. 1–15.
- [44] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, 1994.
- [45] M. A. Oliver and R. Webster, "A tutorial guide to geostatistics: Computing and modelling variograms and kriging," *CATENA*, vol. 113, pp. 56–69, 2014.
- [46] A. Konak, "Estimating path loss in wireless local area networks using ordinary kriging," in *Proc. Winter Simulation Conference*, Dec. 2010, pp. 2888–2896.
- [47] S. Kolyaie and M. Yaghooti, "Evaluation of geostatistical analysis capability in wireless signal propagation modeling," in *Proc. 11th International Conference on GeoComputation*, Jul. 2011.
- [48] N. Zarka, A. Khalil, and A. Assimi, "Adaptive modulation and coding simulations for mobile communication networks," in *Proc. Int. Conf. for Young Researchers in Informatics, Mathematics and Engg.*, Jun. 2016.
- [49] R. Enami, Y. Shi, D. Rajan, and J. Camp, "Pre-crowdsourcing: Predicting wireless propagation with phone-based channel quality measurements," *Computer Communications*, vol. 132, pp. 96–110, 2018.



Rita Enami received the Ph.D. degree from Southern Methodist University, Dallas, USA in 2019. She is currently working as a Senior System Engineer at Qualcomm, San Diego, USA.



Sabyasachi Gupta received the M.Tech degree from the National Institute of Technology (NIT), Durgapur, India and the Ph.D. degree from the Indian Institute of Technology Delhi, New Delhi, India. He has held research positions with the University of Pompeu Fabra, Barcelona, Spain and with the University of Alabama, Tuscaloosa, USA. He is currently a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, Southern Methodist University, Dallas, USA. He was the recipient of the Institute Gold Medal from NIT Durgapur in 2010. His research interests are in resource allocation for wireless networks, mobile-edge computing networks, investigating application of optimization technique, machine learning and graph theory for wireless communication.



Dinesh Rajan (Senior Member, IEEE) is currently Department Chair and Cecil and Ida Green Professor in the Electrical and Computer Engineering Department at Southern Methodist University. He received the B.Tech. degree in Electrical Engineering from Indian Institute of Technology (IIT), Madras. He received his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Rice University, Houston, Texas. He joined the Electrical Engineering Department at Southern Methodist University, Dallas, Texas in August 2002 as an Assistant Professor. His

current research interests include communications theory, wireless networks, information theory and computational imaging. He received a NSF CAREER award for his work on applying information theory to the design of mobile wireless networks. He is also a recipient of the Golden Mustang Outstanding faculty award and the Senior Ford Research Fellowship from SMU.



Joseph Camp (Member, IEEE) received the B.S. (Hons.) degree in electrical and computer engineering from the UT-Austin, Austin, TX, USA, and the M.S. and Ph.D. degrees in electrical and computer engineering from Rice University, Houston, TX, USA. He is currently an Associate Professor of Electrical and Computer Engineering with Southern Methodist University, Dallas, TX, USA. He joined the SMU Faculty in 2009. His research team has performed more than 200 million in-field wireless

measurements around the world via Android deployment and local characterization via drones, campus buses, vehicles, and buildings. His research interests are wireless communications and networking, crowdsourcing, and drones, specifically focused on the deployment, measurement, and analysis of large-scale systems and development of embedded protocols. He received the National Science Foundation CAREER Award in 2012 and the Golden Mustang Teaching Award in 2014.