

# Coverage Estimation in Outdoor Heterogeneous Propagation Environments

NIHESH RATHOD<sup>1,2</sup>, RENU SUBRAMANIAN<sup>1</sup>, AND RAJESH SUNDARESAN<sup>1,2</sup>

<sup>1</sup>Department of Electrical Communication Engineering, IISc, Bengaluru 560012, India

<sup>2</sup>Robert Bosch Center for Cyber-Physical Systems, IISc, Bengaluru 560012, India

Corresponding author: Nihesh Rathod (niheshkumar@iisc.ac.in)

This work was supported in part by the Robert Bosch Centre for Cyber-Physical Systems, Indian Institute of Science, Bengaluru. The work of Nihesh Rathod was supported by the Fellowship Grant from the Centre for Networked Intelligence (a Cisco CSR initiative) of the Indian Institute of Science, Bengaluru.

**ABSTRACT** This paper is on a coverage estimation procedure for the deployment of outdoor Internet of Things (IoT). In the first part of the paper, a data-driven coverage estimation technique is proposed. The estimation technique combines multiple machine-learning-based regression ideas. The proposed technique achieves two purposes. The first purpose is to reduce the bias in the estimated received signal strength arising from estimations performed only on the successfully received packets. The second purpose is to exploit commonality of physical parameters, e.g. antenna-gain, in measurements that are made across multiple propagation environments. It also provides the correct link function for performing a nonlinear regression in our communication systems context. In the second part of the paper, a method to use readily available geographic information system (GIS) data (for classifying geographic areas into various propagation environments) followed by an algorithm for estimating received signal strength (which is motivated by the first part of the paper) is proposed. Together they enable quick and automated estimation of coverage in outdoor environments. It is anticipated that these will lead to faster and more efficient deployment of outdoor Internet of Things.

**INDEX TERMS** Coverage, geographic information system (GIS), heterogeneous propagation environment, Internet of Things (IoT) deployment.

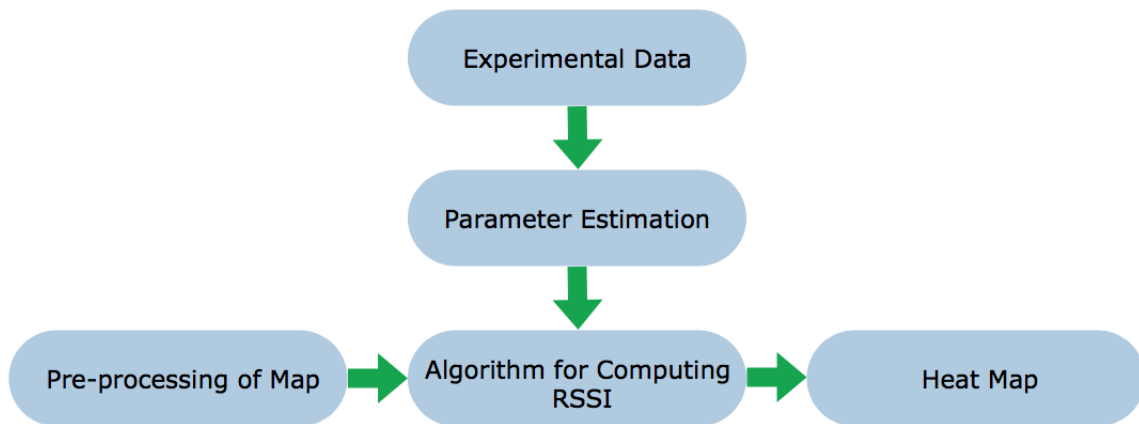
## I. INTRODUCTION

Given the anticipated expansion of the Internet of Things (IoT), traditional deployment strategies that involve “deploy first and fine-tune later” approaches are not scalable. One needs automated methods for large scale IoT network deployment. To enable this, one approach is to move away from the manpower-intensive measurement surveys of a particular deployment region, and instead utilize *prior knowledge of the terrain* and *prior measurements in typical environments* to arrive at estimations of signal coverage. The prior knowledge of the terrain can come from geographic information system (GIS) data. The prior measurements in typical environments can come from extensive prior experimentation in typical propagation environments. Such an approach, which is based on GIS data and prior measurements, can lead to quick and automated estimation of coverage at an early stage of network design. The resulting

estimations of received signal strength indication (RSSI) before actual deployment will save valuable human resources and can lead to rapid and more efficient network design and deployment. This paper demonstrates that quick estimation of coverage, based on GIS data and extensive prior measurements in typical propagation environments, is indeed possible.

Some networks designs are based on estimated channel parameters – path loss exponents, antenna-gain parameters for a specific transmitter-receiver antenna pair, frequency-dependent decay parameters, etc. – coming from extensive RSSI measurements. However, in typical operating system implementations, such as Contiki, these RSSI measurements are made available to the upper layers only on correctly received packets. It is then immediate that the estimation of RSSI on the link is *biased* because it depends only on correctly received packets. Some other network designs are based only on the packet error rate (PER) measurements without ever relying on the RSSI measurements. There is then an opportunity to improve coverage estimation by combining

The associate editor coordinating the review of this manuscript and approving it for publication was Ke Guan<sup>1</sup>.



**FIGURE 1.** The building blocks of the coverage estimation tool.

the two link quality indicators (RSSI on correctly received packets and PER).

It is also often the case that the same (or the same type of) transmitters and receivers are used across measurements. But these measurements may have been collected across multiple example propagation environments. There is then an added opportunity to exploit the knowledge that the antenna-gain parameters are the same (or similar) across measurements, even though the propagation environments across the measurements may have been different.

Our first goal in this paper is to propose a scheme that exploits the two opportunities highlighted above to come up with a better link quality estimation. We combine the RSSI measurements of correctly received packets and the PER due to lost packets to reduce the aforementioned bias in the estimated RSSI. Our scheme is a nonlinear regression scheme (akin to logistic regression) and works jointly with a regression-based estimation framework. The regression part minimizes the error between the measured RSSI on correctly received packets and the predicted RSSI. The logistic-like regression part takes into account the communication theoretic model of transmission over a Rayleigh fading channel or a Rician fading channel.<sup>1</sup> Together, they exploit the knowledge that the antenna-gain parameters are common across the measurements in the different propagation environments. The outcome is a more unbiased estimate of the received signal strength than the one that relies only on measured RSSI on the correctly received packets. Maltz *et al.* [2] considered the effect of lost packets in inferring network performance but in a different context of how the lost packets affect detection of changes in the network topology. Our work is towards improving the network design by coming up with a better link quality estimate via a more unbiased estimate of RSSI. We have consciously stayed away from neural network and SVM-based approach for channel modeling. The reason is that our physics-based models have fewer parameters that

could be well estimated and easily interpreted. The general purpose neural network and SVM-based approaches do not afford this interpretability.

Our second goal in this paper is to demonstrate that quick estimation of coverage, based on available GIS data and extensive prior measurements in example propagation environments, is possible. We describe a tool which has been developed in-house by the authors. The tool takes the open-source GIS data for the (heterogeneous) deployment region under consideration as input. The tool can classify the deployment area into various regions with different propagation characteristics. The methods of the first part of this paper are then applied to each of the smaller component regions to get the local propagation parameters. The tool then stitches these local estimates together to estimate the overall RSSI between any candidate transmitter and receiver pair in the deployment region. Note that we must take into consideration the heterogeneity in the propagation environment in arriving at the RSSI estimates. The tool then provides a *heat map* that enables easy visualization of the coverage and the coverage holes. Figure 1 shows the building blocks of our tool. We have used the Indian Institute of Science (IISc) campus as a vehicle to describe the key ideas and the algorithms, and also to highlight the outcomes. See Figures 5-7 at the end for a quick preview of the outcome.

There are many classical outdoor propagation models, for example the Longley-Rice model [3]–[5] and the Edwards-Durkin model [6], [7]. These involve sophisticated knife-edge diffraction techniques to estimate path loss and require very detailed topography information. Other models such as the Okumura model [8], the Hata model [9], the COST-231 model [10] are homogeneous models that work for large coverage areas (medium-sized city, metropolitan, suburban). The Walfisch-Bertoni model [11] handles rooftop-to-street diffraction and scatter and is suited for metropolitan areas with rows of buildings but requires detailed building profile data. The wideband-PCS-microcell model based on the work of Feuerstein *et al.* [12] categorizes a link as ‘line-of-sight’ or ‘obstructed’ and then fits a simple

<sup>1</sup>Extension to other fading channels, e.g., Nakagami-*m*, are straightforward; we do not pursue them here for the sake of brevity.

path loss model for the categorized type. Our method is a little more fine-grained than the Okumura, Hata, COST-231, or wideband-PCS-microcell models because it takes into account component path losses in smaller regions, but is coarser than the Longley-Rice, the Edwards-Durkin, or the Walfisch-Bertoni models in that only coarse-grain categorization of the deployment region into smaller component regions is done followed by a simple stitching strategy to arrive at a final prediction for link quality. See Rappaport [13, Sec. 3.10] for a detailed discussion of these models.

Some of the above mentioned models are used in typical network planning tools. See for example Götz [14], Teoco RAN Solutions [15] and Intermap [16]. These works do provide network planning tools with options for a user to pick a suitable channel model for the scenario of interest. But the user has to make a choice, and the choice is restricted to a single one. In particular, there is no automated tuning of the parameters to specific locations. In our work, we are able to do automated scenario tuning because of our automated partitioning of the region of interest into various subregions of differing types. Moysen *et al.* [17] provided a data-driven ML framework for locationing of base stations for a microcell. Our goal, also data driven, is however different in that we want to provide RSSI estimates in a heterogeneous environment. Chall *et al.* [18] proposed a large-scale radio propagation model. But, once again, it is a blanket model for the entire region, and does not handle heterogeneity. Also, they use only 20 packets per link, which is much lower than our 1200 packets per link described in our experimental methodology and data collection section. Hosseinzadeh *et al.* [19] proposed a neural network based correction to the COST-231 model. Similarly, Dobrilović *et al.* [20] proposed an optimisation of the Lee propagation model. However, both are homogeneous models (city-scale) and do not handle heterogeneity, which our work does.

There are many indoor models as well. Again, see Rappaport [13, Sec. 3.11]. For more recent work, see Agrawal *et al.* [21] who characterized links in an indoor factory environment and focused on a single model for the entire factory. See also Rath *et al.* [22] for a model that involves the number of intervening walls. These differ from the heterogeneous outdoor setting considered in our current paper. In the same spirit as our work, which is one of network design based on predicted measurements in the outdoor environment, Bhattacharya and Kumar [23] considered an indoor homogeneous setting and used a coarse-grained quantization of a link's quality to come up with relay placements. While the homogeneity assumption may work for short links in the indoor environment, our work significantly differs since it deals with heterogeneity issues coming from outdoor environments.

The renewed interest in this topic of outdoor channel modelling is for two current and relevant reasons: how to enable the IoT deployment expansion efficiently and how to use machine learning ideas in getting better predictions, in our

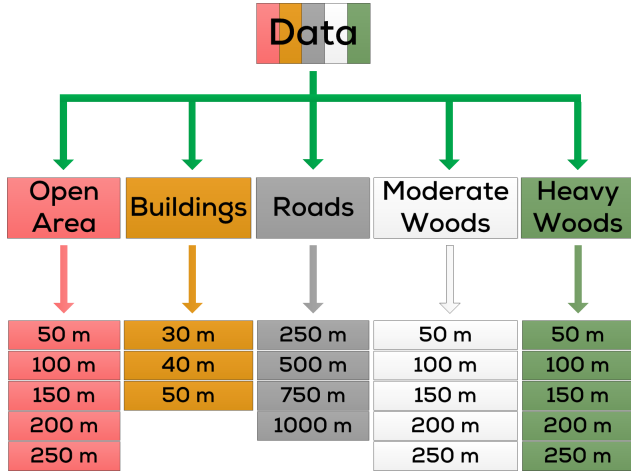
case to reduce bias arising from the lost RSSI information on the lost packets.

Our work opens up many new and interesting possibilities. As one example, Yang *et al.* [24] studied optimal downward titles in downlink cellular networks. The downward tilt could be added as an extra experimental in our data-driven approach and could be used to get a better estimate of the coverage. As another example, Ren *et al.* [25] maximized coverage estimation with only a subset of base stations kept active for energy savings. They assumed circular and homogeneous coverage for the active transmitters, and our work shows the direction on how this could be extended to heterogeneous propagation settings.

We now provide an outline of the rest of the paper. Section II explains our data-driven approach for joint parameter estimation and shows its superiority in terms of bias reduction over two simpler estimation schemes. One of these is based only on RSSI-from-correctly-received-packets. The other is based only on the packet error rate. Section III extends the approach of Section II to Rician fading channel and shows the effectiveness of the proposed scheme. Section IV describes the inner workings of our tool which provides quick estimations of coverage. This section also shows how to extract useful terrain information from a GIS database and how to tessellate the deployment area into various propagation environments. It then provides the RSSI computing algorithm with examples and demonstrates the tool's outcome in the form of a heat map for one example deployment. Section V provides some concluding remarks.

## II. THE DATA-DRIVEN APPROACH WITH RAYLEIGH FADING

The proposed data-driven methodology is based on combining multiple regression methods from the domain of machine learning (ML). Each data point is associated with several *factors* which can potentially affect a composite *outcome* (or multi-valued target) that indicates whether the packet was received and if received, the quality of the reception. The factors we consider are the following: the transmitter power, the transmitter height, the receiver height, the carrier frequency used for transmission and reception, and the propagation environment. In a previous work [26], we (along with other coauthors) classified our IISc campus into five distinct propagation environments or regions with different propagation characteristics. These were open areas (O), buildings (B), roads (R), moderately wooded areas (M), heavily wooded areas (H). (See also Figure 2 and Table 3.) We use that same classification in this paper to arrive at the propagation environment factor. Each data point is also associated with a composite outcome or the multi-valued target – a boolean value that tells whether the packet was received correctly and, if yes, the real value of the received signal strength indication (RSSI). (The latter is often quantized, but we shall treat it as a real-valued quantity.) We will study four



**FIGURE 2.** Measurement regions and the associated distances between transmitters and receivers. The colors correspond to the colors in Figure 3.

(example) regression-based approaches to estimate how these factors affect the composite outcome. We will also compare their respective estimation capabilities. As highlighted in the introduction, network design and deployment strategies often involve use of either only the packet error rate or only the RSSI of correctly received packets, but not both. As benchmarks, the first two regression approaches that we study use only the RSSI of the correctly received packets and only the packet error rate, respectively. The third approach and a fourth variant use both packet error rate and RSSI. The third and the fourth approaches result in significantly reduced biases.

### A. THE REGRESSION METHODOLOGIES

We consider the following well-established model for the received energy at the receiving antenna [27, p. 83]. Suppose that the transmitter and the receiver are located in a particular homogeneous propagation environment indexed by a parameter  $r$ . The quantity  $r$  will take one of five values and will stand for one of the regions specified in Figure 2. The received energy is modelled as:

$$P_{Rx} = C \cdot P_{Tx} \cdot h_{Tx}^2 \cdot h_{Rx}^\gamma \cdot d^{-\eta_r} \cdot f^{-\kappa_r}, \quad (1)$$

where  $P_{Rx}$  denotes the received power,  $C$  refers to a constant that depends on the transmitter and the receiver antenna gain factors,  $P_{Tx}$  denotes the transmitted power,  $h_{Tx}$  and  $h_{Rx}$  refer to the transmitter and receiver heights, respectively,  $\gamma$  denotes the exponent that specifies how the received power improves with receiver antenna height,  $d$  refers to the distance between the transmitter and the receiver,  $\eta_r$  which is typically between 2 and 6 denotes the region-dependent path-loss exponent for the region indexed by  $r$ , and  $f$  refers to the carrier frequency of operation. Finally  $\kappa_r$  is a region-dependent parameter (between 2 and 3) that tells how fast the received energy decays with increasing frequency in the region  $r$ .

Observe that some of these parameters, specifically  $C$  and  $\gamma$ , depend on the nature and the type of antennas used. If the same transmitter-receiver pair or devices of the same

type are used for making the measurements, these parameters are common across regions and therefore common across data points. Other parameters are of course region-specific, for e.g.,  $\eta_r$  and  $\kappa_r$ . Our regression approach, while accounting for the differences, exploits the commonality of the common parameters across the data points.

Assuming  $N_0$  is the thermal noise power, the signal-to-noise ratio (SNR) is, see [28, p. 173]:

$$\text{SNR} = \frac{P_{Rx}}{N_0} = \frac{C \cdot P_{Tx} \cdot h_{Tx}^2 \cdot h_{Rx}^\gamma \cdot d^{-\eta_r} \cdot f^{-\kappa_r}}{N_0}. \quad (2)$$

On top this, in this Section, we assume that the uncoded transmitted symbols undergo Rayleigh fading. Extension to Rician fading is done in Section III. Extensions to other fading models, e.g. Nakagami-m, are straightforward with associated changes to the parameters of the fading model. We restrict attention to Rayleigh and Rician fading in this paper mainly to highlight our approach in the simplest of settings. We may then view the SNR as the average signal-to-noise ratio, averaged across fading instances.

We now explain our regression methods, all of which have been tested on the same data set. Our approaches are designed to work even on data which may have been collected over different regions, over different periods, and perhaps without time stamps.

#### 1) RSSI FROM ONLY CORRECTLY RECEIVED PACKETS

In the first approach, included mainly for comparison purposes, RSSI measurements of only the correctly received packets are taken into account. This is often the case in common implementations of the Zigbee protocol, for e.g., implementations in the TelosB motes and in the RE-Mote [29]. Suppose that there are  $M_r$  correctly received packets in region  $r$ , where  $r = 1, \dots, 5$  is one of the five regions listed in Figure 2. Let  $\text{RSSI}(n)$  be the measured received power for the  $n$ th correctly received packet. Denote by  $P_{Rx}(n)$  the true received power when the transmit parameters are  $P_{Tx}(n)$ ,  $h_{Tx}(n)$ , when the receiver height, the receiver distance, and the frequency of operation are  $h_{Rx}(n)$ ,  $d(n)$ , and  $f(n)$ , respectively, and the region of operation is  $r(n)$ . Let us collectively denote all these factors by  $z(n)$ , i.e.

$$z(n) = (P_{Tx}(n), h_{Tx}(n), h_{Rx}(n), d(n), f(n), \gamma(n)). \quad (3)$$

Using these factors, we obtain  $P_{Rx}(n)$  from the formula (1). We then solve the regression problem:

$$\min \sum_{r=1}^5 \sum_{n=1}^{M_r} |\text{RSSI}(n) - P_{Rx}(n)|^\zeta, \quad (4)$$

where the minimization is over parameters  $C > 0$ ,  $\gamma \in [1, 2]$ ,  $\eta_r \in [2, 6]$ ,  $\kappa_r \in [2, 3]$ ,  $r = 1, \dots, 5$ . Let us reiterate that this involves a joint optimization across all collected data.  $\zeta = 1$  yields the absolute error loss between the predicted and the measured RSSI while  $\zeta = 2$  yields the squared error loss. (The approach extends to other loss functions such as  $|\log \text{RSSI}(n) - \log P_{Rx}(n)|$ .)



## 2) USE OF PER ALONE

In the second approach, we take inspiration from the machine learning technique of logistic regression [30, Ch. 4.4], although we emphasize that our technique is a more general nonlinear regression scheme, to exploit the information available on whether each individual transmitted packet was received correctly or not. Note that this approach uses finer information than just PER since each transmitted packet could have been transmitted at a different power, from a different transmitter height, etc.

Under the Rayleigh fading assumption, the probability of error of an uncoded BPSK transmission with an average signal-to-noise ratio of  $\text{SNR}$  is, see [31, eqn. (3.19)],

$$\Pr\{\text{Error}\} = \frac{1}{2} \left( 1 - \sqrt{\frac{\text{SNR}}{1 + \text{SNR}}} \right) \sim \frac{1}{4\text{SNR}}, \quad (5)$$

where the approximation holds when the  $\text{SNR}$  is high. Let us denote the signal-to-noise ratio by  $\text{SNR}(z)$  when the factor vector is  $z$ , and let us define

$$p(z) := \frac{1}{2} \left( 1 - \sqrt{\frac{\text{SNR}(z)}{1 + \text{SNR}(z)}} \right). \quad (6)$$

Then

$$\ln \left( \frac{p(z)}{1 - p(z)} \right) = \ln \left( 1 - \sqrt{\frac{\text{SNR}(z)}{1 + \text{SNR}(z)}} \right) - \ln \left( 1 + \sqrt{\frac{\text{SNR}(z)}{1 + \text{SNR}(z)}} \right)$$

which is not a linear function of the factors or other transformations. This is where our method differs from the standard logistic regression. However, notice that, on account of (5) and (2), as the  $\text{SNR}$  increases the proposed regression approaches the classical logistic regression. So we may view our proposed method as providing the appropriate generalization of a “link function” for our communication systems context in nonlinear regression (see [30, p.258]).

Suppose that the transmission factor is  $z(n)$  for the  $n$ th packet. Let  $y(n)$  take the value 1 when this  $n$ th packet is in error and let it take the value 0 otherwise. Assume independent receptions. This is a good assumption when there is sufficient time separation between receptions or when the data is randomly reordered and without time-stamps. Then  $\{y(n)\}_{n \geq 1}$  is an independent sequence of Bernoulli random variables with parameters  $\{p(z(n))\}_{n \geq 1}$ . Note that this sequence is not necessarily identically distributed since the Bernoulli parameter  $p(z(n))$  may vary with  $n$  on account of the variation in the factor  $z(n)$  with  $n$ . The likelihood of the observed sequence of packet errors corresponding to the sequence of factors  $\{z(n)\}_{n \geq 1}$  is:

$$\Pr\{y(1), \dots, y(N)\} = \prod_{n=1}^N p(z(n))^{y(n)} (1 - p(z(n)))^{1-y(n)},$$

where  $N$  is the total number of transmitted packets, taking both correctly and incorrectly received packets into account. The negative log-likelihood is then:

$$-\ln \Pr\{y(1), \dots, y(N)\} = \sum_{n=1}^N (-y(n) \ln(p(z(n))) - (1 - y(n)) \ln(1 - p(z(n)))) \quad (7)$$

In the second approach under discussion, the goal is to maximize the likelihood (or minimize the negative log likelihood) over parameters  $C > 0$ ,  $\gamma \in [1, 2]$ ,  $\eta_r \in [2, 6]$ ,  $\kappa_r \in [2, 3]$ ,  $r = 1, \dots, 5$ .

## 3) USING BOTH RSSI AND PER

In our third approach, we combine the objectives of maximizing the likelihood and of minimizing the RSSI estimation error, as follows:

$$\min \sum_{n=1}^N \left( -y(n) \ln(p(z(n))) - (1 - y(n)) \ln(1 - p(z(n))) + (1 - y(n)) \cdot |\text{RSSI}(n) - P_{\text{Rx}}(n)|^\zeta \right) \quad (8)$$

Yet again the minimization is over the parameters  $C > 0$ ,  $\gamma \in [1, 2]$ ,  $\eta_r \in [2, 6]$ ,  $\kappa_r \in [2, 3]$ ,  $r = 1, \dots, 5$ . We also consider a fourth approach, a variant, where we further optimize the relative weights assigned to these two objectives. The objective in equation (8) then gets modified to:

$$\min \sum_{n=1}^N \left( w_1 \cdot (-y(n) \ln(p(z(n))) - (1 - y(n)) \ln(1 - p(z(n)))) + w_2 \cdot (1 - y(n)) \cdot |\text{RSSI}(n) - P_{\text{Rx}}(n)|^\zeta \right) \quad (9)$$

(We refer the reader to the Rician Section III and Table 2 for the associated results when the weights  $w_1$  and  $w_2$  are optimized.)

Let us note, in passing, that if all the  $z(n)$  were the same, then the maximum likelihood estimation procedure chooses the parameters  $C$ ,  $\gamma$ ,  $\eta_r$ ,  $\kappa_r$  to bring  $p(z(n))$  as close to  $\frac{1}{N} \sum_{n=1}^N y(n)$  as possible in relative entropy distance measure (also known as the Kullback-Leibler divergence), i.e.  $\min D(\frac{1}{N} \sum_{n=1}^N y(n) \parallel p(z(n)))$ , where the quantity  $D(p \parallel q)$  is the binary relative entropy. Given the nature of the dependence of  $p(z(n))$  on the parameters, full flexibility is not available to make  $p(z(n))$  equal  $\frac{1}{N} \sum_{n=1}^N y(n)$ . The minimization tries to pick the parameters so that they are as close to each other as possible. In the general case,  $z(n)$  varies from the sample point to sample point and our approaches account for this variation appropriately by exploiting the commonality in the antenna gains and in the device-specific quantities.

## B. EXPERIMENTAL METHODOLOGY AND DATA COLLECTION

We now describe our experimental and data collection methodologies. We conducted our field experiments in five example propagation environments. See Figure 2 for a listing

of the regions. In each region, the transmitters and receivers were placed at different distances, as indicated in Figure 2. With a transmitter kept at a height of 1 m, three receivers placed at heights 1 m, 2 m, and 3 m, at a given location, listened simultaneously to the transmissions. A total of 1200 packets were transmitted from that transmitter height. Only one transmitter was allowed to transmit in any collection period to avoid packet collisions. The same procedure was then repeated for two other transmitter heights, namely 2 m and 3 m. There are thus nine combinations of transmitter and receiver heights for a given distance. The entire setup was then moved to a new transmitter-receiver pair of locations, and the experiment was repeated. Given that there are 22 distance-region pairs (see Figure 2), the number of data points is  $N = 22 \times 9 \times 1200 = 237,600$ .

Each transceiver is a RE-Mote which has the Texas Instruments CC1200 chip (sub-GHz radio operating on 865-868 MHz ISM band) [32]. The payload in each of the 1200 packets consisted of 16 Bytes with a header size of 9 Bytes. The physical layer parameters were as follows. These configurations are derived from an earlier work, Rathod et al. [26], and are used for data coherency. Rathod et al. [26, Table V] also provides information on the measurement accuracy through lab characterization of the device.

- Tx power: 14 dBm
- Symbol rate: 4.6 ksp/s
- Bit rate: 4.6 kbps
- Modulation Scheme: 2-FSK
- Deviation: 17.967224 kHz
- Centre frequency: 868 MHz
- Rx filter bandwidth: 128.205128 kHz.

### C. CROSS-VALIDATION

Cross-validation results are presented in Table 1. The gathered data was divided into ten random subsets for each environment, nine of which were used for estimating the parameters and the tenth was used for testing. This is called ten-fold cross-validation. The results of this procedure are listed in Table 1. The errors reported are the average of the measured values minus the predicted values. The values in the “Error I” column are observed when using the logistic-like regression without the RSSI term, i.e. optimization of (7). Columns “Error II” and “Error V” have error values corresponding to the regression on just the RSSI, i.e. optimization of (4) with absolute error loss ( $\zeta = 1$ ) and squared error loss ( $\zeta = 2$ ), respectively. Columns “Error III” and “Error VI” show the errors under the combined optimization in (8) with absolute error loss ( $\zeta = 1$ ) and squared error loss ( $\zeta = 2$ ), respectively. We also optimized over the  $w_1$  and  $w_2$  for the objective function in (9). These are termed “Error IV” and “Error VII”. The description of error columns, for ease of reference, is as follows:

- Error I: Logistic-like optimization of (7)
- Error II: Optimization of only RSSI term with  $\zeta = 1$  (4)
- Error III: Combined optimization with  $\zeta = 1$  (8)

**TABLE 1. Bias comparison. O = open area, B = buildings, R = roads, M = moderate woods, H = heavy woods. “Error IV” is not reported below and is for optimized  $w_1$  and  $w_2$  in (9). The highlighted columns “Error III” and “Error VI” indicate the methods with superior performance.**

Reg.	Dist. (m)	Meas. (dBm)	Error I (dB)	Error II (dB)	Error III (dB)	Error V (dB)	Error VI (dB)
O	50	-59.4	-47.4	4.7	<b>2.4</b>	5	<b>1.6</b>
O	100	-63.1	-53.3	4.5	<b>-1.1</b>	5.1	<b>-2.2</b>
O	150	-69.4	-52.5	3.7	<b>1.1</b>	3.6	<b>0.3</b>
O	200	-72.9	-53	3.1	<b>1.3</b>	3	<b>0.2</b>
O	250	-75.7	-53.4	4.7	<b>2.1</b>	4.8	<b>1</b>
B	30	-87.6	-7.6	4.3	<b>1.2</b>	4.6	<b>0.3</b>
B	40	-92.8	-4.7	5.3	<b>2.4</b>	5.1	<b>2.5</b>
B	50	-92.9	-8.8	4.7	<b>-2.3</b>	4.9	<b>-2.3</b>
R	250	-79.4	-15.6	-6.1	<b>-4.2</b>	-5.1	<b>-2</b>
R	500	-91.2	-7.6	3.3	<b>1.0</b>	3.9	<b>2.3</b>
R	750	-97.8	-5.1	2.2	<b>2.0</b>	3	<b>2.9</b>
R	1000	-99.0	-3.7	3.0	<b>2.8</b>	3.3	<b>2.8</b>
M	50	-65.7	-27.2	4.8	<b>-2.3</b>	4.8	<b>-2</b>
M	100	-69.2	-26.7	3.8	<b>-4.1</b>	4.5	<b>-4.7</b>
M	150	-82.9	-19.6	4.5	<b>1.5</b>	4.4	<b>0.1</b>
M	200	-88.2	-7.5	3.6	<b>2.5</b>	2.4	<b>0.6</b>
M	250	-98.3	-2.4	12.9	<b>12.3</b>	10.4	<b>10.3</b>
H	50	-86.1	-2.4	4.9	<b>2.5</b>	5.5	<b>3.1</b>
H	100	-90.2	-3.8	4.7	<b>-0.7</b>	4.7	<b>-0.2</b>
H	150	-93.4	-3.4	4.9	<b>-1.4</b>	5	<b>-0.6</b>
H	200	-95.1	-5.1	5.0	<b>-2.7</b>	5	<b>-0.2</b>
H	250	-99.0	0.7	2.3	<b>-0.4</b>	1.7	<b>-1.5</b>

- Error IV: Combined optimization with  $w_1$ ,  $w_2$  and  $\zeta = 1$  (9)
- Error V: Optimization of only RSSI term with  $\zeta = 2$  (4)
- Error VI: Combined optimization with  $\zeta = 2$  (8)
- Error VII: Combined optimization with  $w_1$ ,  $w_2$  and  $\zeta = 2$  (8)

Table 1 does not include columns “Error IV” and “Error VII” because these cases did not show significant improvement. These cases and their errors are referred to here for use later in Table 2 where the data for Rician fading is presented. The regression method that exploit boths sources of information (PER and RSSI) are far superior to those that rely on only one of these, except in 2 out of the 22 cases. The large “Error III” and “Error VI” of 12.3 dB and 10.3 dB for  $\zeta = 1$  and  $\zeta = 2$ , respectively, in the moderately wooded area at 250 m is due to a high packet error rate encountered there. This might have been due to some shadowing although we noticed no visible blockage at the physical location. This was consistently seen even in our second round of measurements made after the submission of our conference paper [1].

### D. REMARKS

Let us first highlight the need for a heterogeneous model. Suppose we restrict ourselves to a homogeneous model for the distance 50 m. The results in Table 1 indicate that the mean received RSSIs are:  $-59.4$  dBm (open area),  $-92.9$  dBm (buildings), less than  $-79.4$  dBm (roads),  $-65.7$  dBm (moderately wooded),  $-86.1$  dBm (heavily wooded). From this data, it is strikingly clear that a single distance-based prediction will have an error of at least  $\pm 16.7$  dB. Our errors (see Error VI column) range from  $-2.3$  to  $+3.1$  dB for the 50 m link distances. No single

**TABLE 2.** Bias comparison. O = open area, B = buildings, R = roads, M = moderate woods, H = heavy woods. The highlighted columns provide superior performance. The root-mean-squared error (RMSE) is also indicated for each method.

Reg.	Dist. (m)	Meas. (dBm)	Error I (dB)	RMSE (dB)	Error II (dB)	RMSE (dB)	Error III (dB)	RMSE (dB)	Error IV (dB)	RMSE (dB)	Error V (dB)	RMSE (dB)	Error VI (dB)	RMSE (dB)	Error VII (dB)	RMSE (dB)
O	50	-59.4	6.1	9.3	3.0	8.1	<b>3.0</b>	<b>8.1</b>	4.7	8.4	1.4	8.1	<b>1.3</b>	<b>8.1</b>	3.9	8.2
O	100	-63.1	1.3	5.4	-1.0	5.2	<b>-1.0</b>	<b>5.2</b>	-0.0	5.3	-2.3	5.3	<b>-2.3</b>	<b>5.3</b>	-1.0	5.3
O	150	-69.4	2.7	5.6	1.0	4.5	<b>1.0</b>	<b>4.5</b>	1.5	4.8	-0.2	4.5	<b>-0.2</b>	<b>4.5</b>	0.3	4.6
O	200	-72.9	2.6	3.4	1.3	4.0	<b>1.2</b>	<b>4.0</b>	1.4	3.8	0.2	4.0	<b>0.2</b>	<b>4.0</b>	0.2	3.9
O	250	-75.7	3.0	7.5	1.9	6.4	<b>1.8</b>	<b>6.4</b>	1.8	6.7	0.9	6.4	<b>0.9</b>	<b>6.4</b>	0.5	6.6
B	30	-87.6	46.7	6.0	0.8	5.2	<b>0.9</b>	<b>5.2</b>	9.2	5.4	0.3	5.2	<b>0.3</b>	<b>5.2</b>	-0.5	5.3
B	40	-92.8	50.9	6.1	3.1	5.4	<b>3.2</b>	<b>5.4</b>	10.8	5.6	2.5	5.4	<b>2.5</b>	<b>5.4</b>	1.2	5.5
B	50	-92.9	46.8	6.6	-1.4	5.2	<b>-1.3</b>	<b>5.2</b>	5.4	5.6	-2.1	5.2	<b>-2.1</b>	<b>5.2</b>	-3.8	5.4
R	250	-79.4	-0.2	4.3	-3.1	5.2	<b>-3.1</b>	<b>5.2</b>	1.2	5.0	-2.2	5.2	<b>-2.0</b>	<b>5.2</b>	-0.5	5.1
R	500	-91.2	5.2	4.4	1.0	3.7	<b>1.0</b>	<b>3.7</b>	6.6	3.8	2.4	3.7	<b>2.2</b>	<b>3.7</b>	3.9	3.8
R	750	-97.8	6.9	2.8	1.7	1.6	<b>1.7</b>	<b>1.6</b>	8.1	2.0	3.4	1.6	<b>3.0</b>	<b>1.6</b>	4.8	1.8
R	1000	-99.0	8.9	1.0	2.0	0.8	<b>1.9</b>	<b>0.8</b>	9.2	0.9	3.8	0.8	<b>3.3</b>	<b>0.8</b>	5.2	0.9
M	50	-65.7	1.9	6.3	-0.7	5.6	<b>-0.7</b>	<b>5.6</b>	0.1	5.8	-1.7	5.6	<b>-1.7</b>	<b>5.6</b>	-0.7	5.7
M	100	-69.2	-1.4	4.3	-3.5	3.6	<b>-3.5</b>	<b>3.6</b>	-2.8	3.8	-4.6	3.6	<b>-4.6</b>	<b>3.6</b>	3.7	3.7
M	150	-82.9	4.4	6.9	1.6	6.0	<b>1.6</b>	<b>5.9</b>	2.5	6.2	0.4	5.9	<b>0.4</b>	<b>5.9</b>	1.5	6.0
M	200	-88.2	5.3	2.9	2.0	3.6	<b>2.0</b>	<b>3.6</b>	3.1	3.4	0.8	3.6	<b>0.7</b>	<b>3.6</b>	1.9	3.5
M	250	-98.3	13.9	3.3	11.6	2.2	<b>11.6</b>	<b>2.2</b>	12.4	2.4	10.3	2.2	<b>10.3</b>	<b>2.2</b>	11.4	2.3
H	50	-86.1	18.9	6.1	2.7	5.8	<b>2.7</b>	<b>5.8</b>	8.8	5.9	3.2	5.8	<b>3.2</b>	<b>5.8</b>	5.5	5.8
H	100	-90.2	13.3	7.3	-0.5	6.0	<b>-0.5</b>	<b>6.0</b>	4.5	6.3	0.1	6.0	<b>0.1</b>	<b>6.0</b>	1.6	6.1
H	150	-93.4	12.1	6.7	-1.1	5.7	<b>-1.0</b>	<b>5.7</b>	3.5	6.0	-0.4	5.7	<b>-0.4</b>	<b>5.7</b>	0.8	5.9
H	200	-95.1	9.5	4.6	-3.9	4.2	<b>-3.9</b>	<b>4.2</b>	0.6	4.3	-3.3	4.2	<b>-3.3</b>	<b>4.2</b>	-2.3	4.2
H	250	-99.0	13.0	2.7	0.9	2.1	<b>0.9</b>	<b>2.0</b>	5.0	2.3	1.6	2.0	<b>1.6</b>	<b>2.1</b>	2.3	2.1

distance-based prediction can provide such a performance and cover the range from  $-59.4$  dBm to  $-92.9$  dBm.

The results in Table 1 also indicate that the bias is significantly reduced by employing all the information available at our disposal (RSSI and PER). Our data is made of Boolean-valued and real-valued observations. A combination of logistic-inspired regression and mean-absolute-error or mean-squared-error minimization enables a good use of both forms of information. The loss functions used are only examples, and one could equally well explore other loss functions. While we have demonstrated that equal weights for the RSSI estimation loss and the negative log-likelihood loss already helps in reducing the bias, we further optimized the weights assigned as in (9) to these objectives, “Error IV”, but did not see significant improvement over “Error III” or “Error VI” and have therefore not reported them in Table 1. See the larger Table 2 where “Error IV” is reported for the more general Rician fading. For mean-squared-errors, we refer the reader once again to Table 2 for Rician fading. Finally, as more data arrives, the above technique is easily amenable to incremental updates – one can make an incremental move from the current set of best parameters to a new set à la stochastic approximation [33].

### III. DATA-DRIVEN APPROACH WITH RICIAN FADING

In the previous section, we studied a regression problem where the transmission factors for the  $n^{\text{th}}$  data point are

$$z(n) = (P_{\text{Tx}}(n), h_{\text{Tx}}(n), h_{\text{Rx}}(n), d(n), f(n), \gamma(n)). \quad (10)$$

The factors in the equation (10) are described in the paragraph containing (3). Under Rayleigh fading, the probability of error is as in (6). We now extend this to Rician fading.

The probability of error for Rician fading is given by a generalization of (5) derived by Lindsey [34, eqn. (19)]:

$$\Pr\{\text{Error}\} = Q(u, w) - \frac{1}{2} \left( 1 + \sqrt{\frac{\text{SNR}}{1 + \text{SNR}}} \right) \cdot \exp\left(-\frac{u^2 + w^2}{2}\right) \cdot I_0(u \cdot w) \quad (11)$$

where

$$u = \rho \sqrt{\frac{1 + 2 \cdot \text{SNR} - 2\sqrt{\text{SNR}(1 + \text{SNR})}}{2 \cdot (1 + \text{SNR})}},$$

$$w = \rho \sqrt{\frac{1 + 2 \cdot \text{SNR} + 2\sqrt{\text{SNR}(1 + \text{SNR})}}{2 \cdot (1 + \text{SNR})}}, \quad (12)$$

the function  $Q(u, w)$  is Marcum’s Q-function given by

$$Q(u, w) = \int_{\beta}^{\infty} z \cdot \exp\left[-\frac{z^2 + \alpha^2}{2}\right] \cdot I_0(\alpha z) dz, \quad (13)$$

$I_0$  is the modified Bessel function of the zeroth order, and  $\rho$  in equation (12) is a parameter  $\geq 0$  that indicates the ratio of the energy in the specular component and the scattered component, i.e., the Rician factor.

Observe that when  $\rho = 0$ ,  $u = w = 0$ , hence  $Q(0, 0) = 1$ ,  $I_0(0) = 1$ , and thus equation (11) reduces to equation (5).

The above motivates the use of an enhanced set of transmission factors, extending (3), as follows:

$$z_R(n) = (P_{\text{Tx}}(n), h_{\text{Tx}}(n), h_{\text{Rx}}(n), d(n), f(n), \gamma(n), \rho(n)). \quad (14)$$

We can now write an analog of (6). The probability of error  $q(z_R)$  when the transmission factor is  $z_R$  is

given by:

$$q(z_R) = Q(u(z_R), w(z_R)) - \frac{1}{2} \left( 1 + \sqrt{\frac{\text{SNR}(z_R)}{1 + \text{SNR}(z_R)}} \right) \cdot \exp \left( \frac{u(z_R)^2 + w(z_R)^2}{2} \right) \cdot I_0(u(z_R) \cdot w(z_R)), \quad (15)$$

where  $u(z_R)$  and  $w(z_R)$  are defined via equation (12) with  $\rho$  taken to be the last component of  $z_R$  and  $\text{SNR}$  taken to be the average  $\text{SNR}$  under the set of transmission factor  $z_R$ .

The likelihood ratio, the negative log likelihood ratio, and the combined objective function which is a combination of negative likelihood and RSSI estimation error are extended in analogy to equations (4), (7) and (8), respectively.

#### A. REGRESSION AND CROSS-VALIDATION

We performed the above logistic-like regression with the same data under the constraint that  $\rho(n) \geq 0$  depends only on the region  $r(n)$ . This introduces 5 new parameters for the five regions. The ten-fold cross-validation outcome is presented in Table 2 below when  $w_1 = w_2$ . The approaches are as outlined in Section II-A except that  $p(z(n))$  in (3) is replaced with  $q(z_R(n))$  given by equation (15). Following the same convention used in Table 1, the errors are measured values minus the estimated values. For this parameter estimation, we also give the values of the root mean-squared errors along with the mean errors. The error values in the column “Error I” are observed when using the logistic-like regression (associated with Rician fading) without the RSSI term. Columns “Error II” and “Error V” show the error values observed when only the RSSI error term in equation (4) is minimized with absolute error loss ( $\zeta = 1$ ) and squared error loss ( $\zeta = 2$ ), respectively. Columns “Error III” and “Error VI” has the error values observed when solving the combined optimization given in equation (8) with absolute error loss ( $\zeta = 1$ ) and squared error loss ( $\zeta = 2$ ), respectively. Additionally, Table 2 also has columns “Error IV” and “Error VII” showing the errors when  $w_1$  and  $w_2$  are optimized in equation (9).

Table 2 also contains root-mean-squared errors for each of the methods. The error is defined as the measured RSSI minus the predicted RSSI on each of the test data points. Since this is done on a per-packet basis, and the RSSI is a nontrivial random variable (e.g., Rayleigh or Rician), the root-mean-square error reflects the variance (when RSSI is estimated in dBm scale).

The observations on cross-validation made in Section II-C under the assumption of Rayleigh fading are valid for Rician fading as well. Here too, the regression on the combined optimization “Error VI” yields the best results among the others. Since the regression was performed on the same data, the large packet error rate in the moderately wooded area at 250 m causes large estimation errors of around 11 dBm. The optimal estimated parameters for combined estimation (equation (9)) with  $\zeta = 2$  (“Error VI”) are reported in Table 3 below.

**TABLE 3.** Optimal estimated parameters for equally weighted negative log-likelihood and RSSI squared error loss functions (“Error VI”). O = open area, B = buildings, R = roads, M = moderate woods, H = heavy woods.

Region	C	$\gamma$	$\eta$	$\kappa$	$\rho$
O	995.81	1.43	2.85	2.03	0.0057
B			3.30	2.98	0.0022
R			2.00	2.98	0.1572
M			3.45	2.00	0.1787
H			2.84	2.84	0.2208

From Table 3, it is reassuring that the path loss exponents are reasonable. The best fit  $\rho$  are however somewhat counter-intuitive in some cases: the best  $\rho$  for the open ground is nearly Rayleigh whereas the best  $\rho$  for heavy woods is 0.2208 which indicates a significant specular component.

#### B. ADDITIONAL REMARKS

The logistic-like regression term’s contribution to (9) is much lower than the RSSI term. To equalize the contributions, while optimizing  $w_1$  and  $w_2$ , we scaled the errors using the approach given below in equation (16):

$$Error_{scaled} = \frac{|RSSI(n) - P_{Rx}(n)|^\zeta - Error_{min}}{Error_{max} - Error_{min}} \quad (16)$$

where

$$Error_{max} = \max_{i=1 \dots N} |RSSI(n) - P_{Rx}(n)|^\zeta,$$

$$Error_{min} = \min_{i=1 \dots N} |RSSI(n) - P_{Rx}(n)|^\zeta.$$

Though it improved the estimates, it did not give better estimates than the ones obtained with equal weights to both the terms.

#### IV. RSSI COMPUTATION IN A HETEROGENEOUS REGION

In the first part of this paper, we discussed how our data-driven approach estimates the parameters for transmission and reception within a homogeneous region. In this second part of the paper, we develop an RSSI estimation tool that uses the first part and extends it to heterogeneous propagation environments. We also indicate how these estimations have been automated, based on GIS data, in our tool. We then demonstrate its working on a test region which is the IISc campus.

**Overview:** See Figure 1 for an overview of the building blocks of our tool.

**Input:** Our tool takes the following as input from the user:

- the geographic information system (GIS) data which provides high-level information about the area of deployment;
- user clicks on the map that indicate one or more transmitter locations;
- the frequency of operation;
- the transmission power;
- the receiver noise power ( $N_0$ );



- a spatial resolution parameter that can be set based on the carrier frequency.

The tool also has as a separate input extensive experimental data from measurements in example propagation environments.

**Processing:** The GIS data is processed by a map pre-processor that partitions or segments the deployment region into appropriate smaller and, most importantly, approximately homogeneous propagation environments or regions. The data-driven approach takes the experimental data and identifies useful antenna-related, frequency-related, and path-loss-related propagation model parameters. We have already discussed this aspect in the previous section for a homogeneous propagation environment. The central RSSI computing engine then extends the RSSI computation scheme to a heterogeneous region. Finally, a heat map is generated and is superimposed on a suitable color-coded map for visualization. In the subsections that follow, we provide details on each of these aforementioned building blocks.

#### A. PRE-PROCESSING OF MAP

In one of our earlier works, the IISc campus (Figure 3A) was broadly classified into five different regions namely, Open area (O), Buildings (B), Roads (R), Moderate woods (sparsely thick trees, M) and Heavy woods (denser thin trees, H); see [26]. To predict signal strength in such a diverse area (at any given point for a given transmitter location and power), we will need to segment the entire region into these component regions. We will also need good measurements either in each of these areas or in other similar areas. In the first part of this paper, we discussed the extensive measurements carried out by us and how they were used to generate good models and propagation parameters. See Table 3.

Figure 3 shows the “Base Image” (A), the extracted “Color Image” (B) and the generated “Grayscale Image” (C). The base map of the area of interest can be obtained from different sources, for example, Google Maps, Bing Maps, Open Street Map (OSM), etc. We used OSM data because it was free and widely supported. We then used Quantum GIS – a free, open-source GIS application which supports the creation, editing and visualization of geospatial data – to extract the five regions from the base map. In Figure 3, steps 1-4 indicate the extraction of open areas, buildings, roads and heavy woods, respectively. We take the remaining areas to be moderately wooded. We then assign unique colors to each of these identified regions. Combining all these layers, we obtain the “Color Image” (B). The white parts in the image represent the moderately wooded regions. Each color in this image has three components (Red, Green, and Blue). For ease of computation, we convert this color image to a “Grayscale Image” (C), which is the last image in Figure 3. In this image, each of the five regions is represented by different shades of gray.

Even though the color image of the sectionalized map is displayed to the user, Figure 3B, it is the grayscale image that is used for performing the RSSI calculations. This map is then

resized to decrease the number of pixels in the image, based on the carrier frequency of operation, in order to speed up RSSI computation. Map representations may have different  $x$  and  $y$  scales. The actual distance between two points may differ from the distance between their respective pixel locations on the map. Our tool rescales the map distance to the actual distance via a suitable rescaling.

The above is a summary of the subblocks in the map pre-processing block. As we will soon see, the variation of RSSI across space is rather significant, and can be attributed to the heterogeneity of the region through which the signal propagates. The pre-processing on the base image and its segmentation into various component regions is a crucial step for estimating the overall RSSI, as we will see next.

#### B. ALGORITHM FOR RSSI COMPUTATION IN A HETEROGENEOUS REGION

Using the data-driven approach described in Section II, the model parameters  $C$ ,  $\gamma$ ,  $\eta_r$ , and  $\kappa_r$  are first estimated for each of the five regions, for example via extensive measurements carried out in each model region. Our measurements were taken in the IISc campus itself. We now come to the RSSI estimation in the heterogeneous region of deployment.

As indicated earlier, the user can input one or more transmitter locations by clicking on the map interface. The predicted received power  $P_{Rx}$  is then computed at each pixel and for each transmitter, as follows. Some computation savings in arriving at these predictions will be highlighted.

From (1), the received power  $P_{Rx}$  is

$$P_{Rx} = C \cdot P_{Tx} \cdot h_{Tx}^2 \cdot h_{Rx}^\gamma \cdot d^{-\eta_r} \cdot f^{-\kappa_r} \quad (17)$$

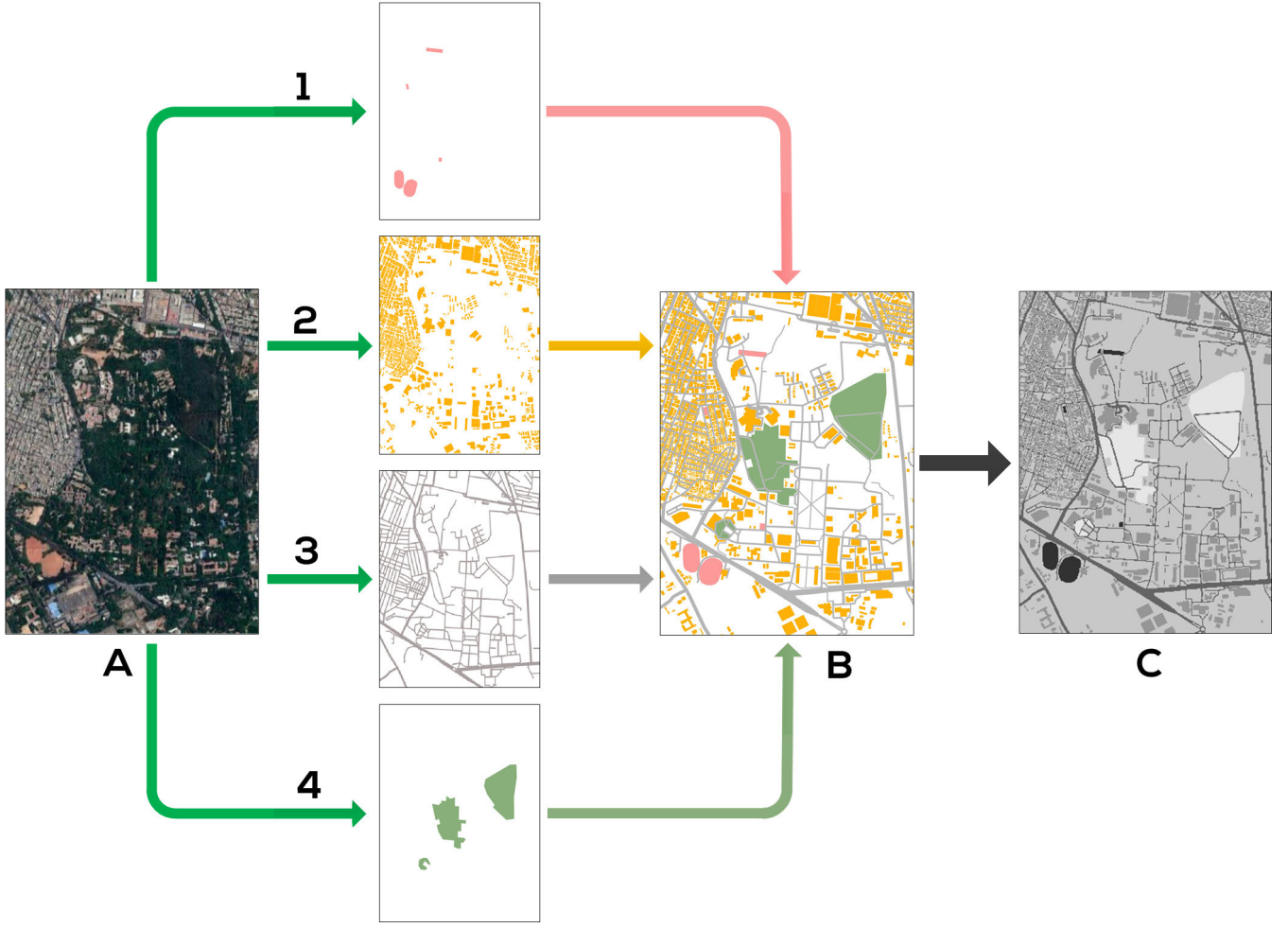
Let the transmitter  $T_x$  be located at a certain pixel, say pixel 1. We assume that  $T_x$  is located at the centre of this pixel. To identify the received power at a candidate pixel of interest, we assume that the receiver is at the centre of this pixel. Let  $d$  be the distance between the transmitter location and this candidate receiver location.

To predict the received power at this pixel, we apply (17) with  $C$  and  $\gamma$  as per the inferred values in Table 3, but with  $\eta_{eff}$  and  $\kappa_{eff}$  as given in (18) and (19), respectively, given below. The effective values handle the heterogeneity in the propagation conditions. The two examples in Figure 4 illustrate how to arrive at these effective values, which we now describe.

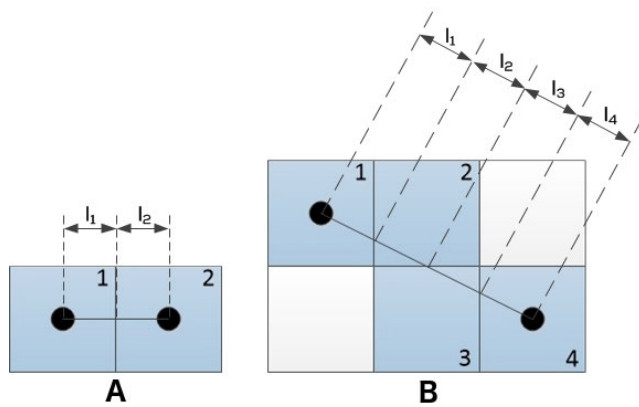
Draw a line between the transmitter and the receiver and identify all the pixels through which the line passes. Suppose there are  $i$  such pixels. Identify the lengths of the line segments in each pixel. Let these be  $l_k$ ,  $k = 1, \dots, i$ . Associate with each pixel a region  $r_k$ ,  $k = 1, \dots, i$ . In Figure 4(A), we have  $i = 2$ , and in Figure 4(B), we have  $i = 4$ .

We take  $\kappa_{eff}$  to be the weighted average of the individual  $\kappa_{r_k}$ 's, weighted by the lengths of the line segments:

$$\kappa_{eff} := \frac{\sum_{k=1}^i l_k \kappa_{r_k}}{\sum_{k=1}^i l_k} \quad (18)$$



**FIGURE 3.** Map pre-processing stages: A = Base Image, B = Color Image and C = Grayscale Image.



**FIGURE 4.** Ray tracing (A) transmitter and receiver are in adjacent pixels and (B) transmitter and receiver are in non-adjacent pixels.

To compute the effective pathloss parameter in Figure 4(A) with  $i = 2$  segments, we take

$$d^{-\eta_{\text{eff}}} = l_1^{-\eta_{r1}} \cdot \left( \frac{l_1 + l_2}{l_1} \right)^{\eta_{r2}},$$

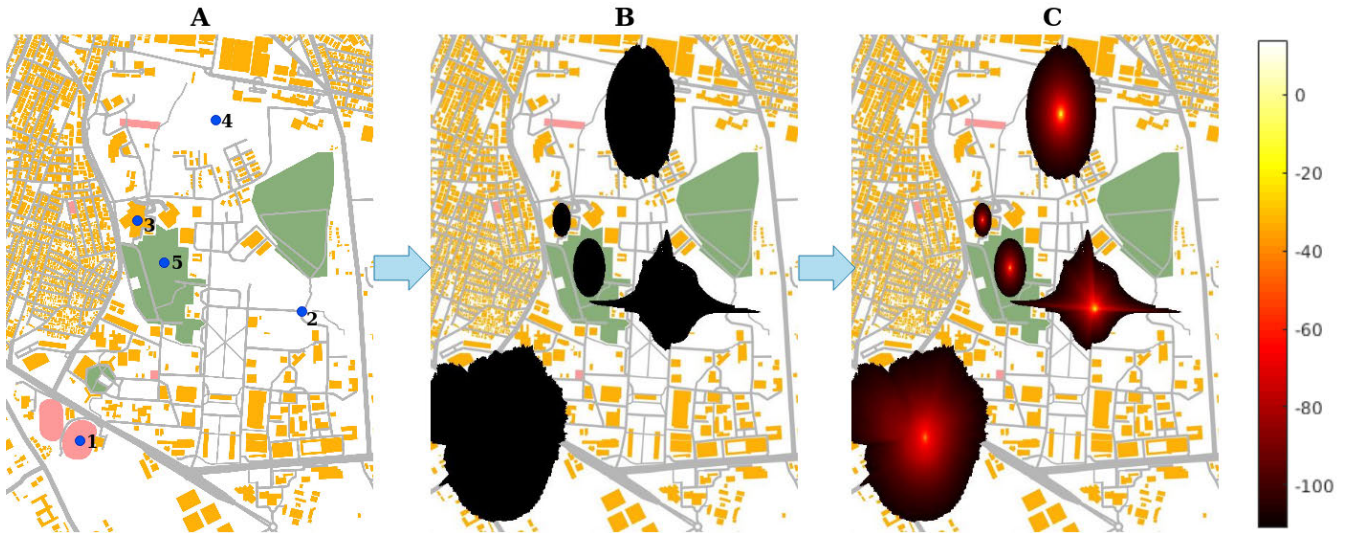
with  $d = l_1 + l_2$ , see for e.g. [27, p. 85]. The intuition for this comes from Huygen's principle that there is an imaginary transmitter at the pixel interface that radiates into the second region at exactly that power which it receives from the first region. This intuition can be extended. To compute the effective pathloss parameter in Figure 4(B) with  $i = 4$  segments, we deduce, this time with  $d = l_1 + l_2 + l_3 + l_4$ ,

$$d^{-\eta_{\text{eff}}} = l_1^{-\eta_{r1}} \cdot \left( \frac{l_1 + l_2}{l_1} \right)^{-\eta_{r2}} \cdot \left( \frac{l_1 + l_2 + l_3}{l_1 + l_2} \right)^{-\eta_{r3}} \cdot \left( \frac{l_1 + l_2 + l_3 + l_4}{l_1 + l_2 + l_3} \right)^{-\eta_{r4}}.$$

Generalizing, for a line segment that passes through  $i$  pixels, we take  $\eta_{\text{eff}}$  to be as calculated from

$$d^{-\eta_{\text{eff}}} = l_1^{-\eta_{r1}} \cdot \prod_{j=2}^i \left( \frac{\sum_{k=1}^j l_k}{\sum_{k=1}^{j-1} l_k} \right)^{-\eta_{rj}}. \quad (19)$$

For each transmitter, we compute the RSSI at each pixel as above. We then take the maximum value across the transmitters and associate the pixel to the corresponding



**FIGURE 5.** Coverage for five transmitters in a heterogeneous environment (IISc), under Rician fading, for transmitter and receiver heights of 1 m each. Image A indicates the Tx Locations. Image B has pixels with RSSI < -110 dBm shaded black. Image C is the “heat map” indicating regions of good coverage.

### Algorithm for RSSI Computation

**Data:** A set of  $N$  transmitters

$T = \{(y_1, x_1), (y_2, x_2), \dots, (y_N, x_N)\}$  where each Cartesian pair  $(y_i, x_i)$ ,  $1 \leq i \leq N$  represents a transmitter location in the Euclidean plane.

**Result:** **MatrixRSSI** $_{Y \times X \times N}$  where  $Y$  = number of pixels in the  $y$ -direction  $X$  = number of pixels in the  $x$ -direction

Initialization;

**for**  $i$  **arrow** 1 **to**  $N$  **do**

Set  $S = \{(y_i, x_i)\}$ ;

**for**  $j$  **arrow** 1 **to**  $\max\{X, Y\}$  **do**

Create a set  $R_1$  of all the pixels  $(y, x)$  such that  $x \in [x_{i-j}, x_{i+j}]$  and  $y \in [y_{i-j}, y_{i+j}]$ ;

$R_2 = R_1 \setminus S$ ;

**for** each pixel  $p = (p_y, p_x) \in R_2$  **do**

**MatrixRSSI** $(u, v, i)$  **arrow** **RSSI** $((y_i, x_i), (p_y, p_x))$ ;

**end**

$S = R_1$ ;

**if**  $\forall p = (p_y, p_x) \in R_2$ , **RSSI** $((y_i, x_i), (p_y, p_x)) \leq -110$  dBm **then**

break;

**end**

**end**

**end**

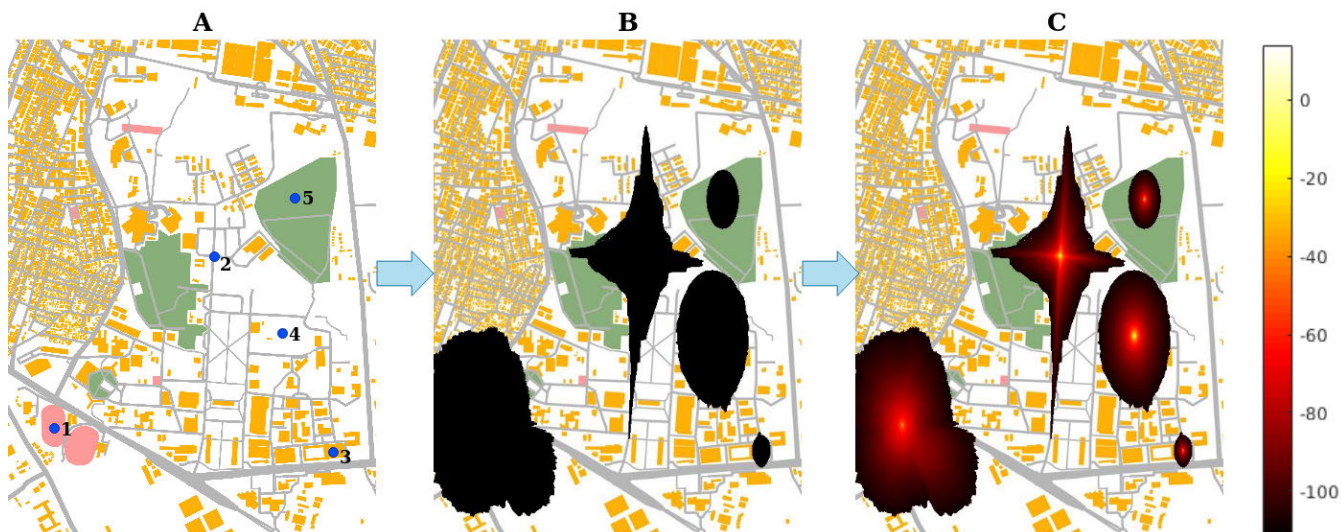
maximum-achieving transmitter. A receiver in this pixel will receive the strongest signal from that associated transmitter. The pixel is then colored according to the predicted

received power. The actual algorithm proceeds by expanding around the transmitter in concentric  $\ell_\infty$ -circles, and then stops when the RSSI is below a threshold, say -110 dBm, in an attempt to minimize the computational load. The following algorithm summarizes the steps. It assumes the existence of a subroutine **RSSI** $((y, x), (y', x'))$  which returns the RSSI (as obtained using the method above) when the transmitter is at the pixel  $(y, x)$  and the receiver is at the pixel  $(y', x')$ .

### C. HEAT MAP

Figures 5, 6, and 7 show the coverages for five transmitters marked 1-5 in Figures 5(A), 6(A), and 7(A), respectively. Figures 5 and 6 correspond to locations inside the IISc campus and Figure 7 corresponds to locations inside a different campus, the IIT Bombay campus, to test our learning's transferability to another setting. In Figure 5, transmitters 1-5 are kept, respectively, in an open area, on a road junction, inside a building, in a moderately wooded area, and in a heavily wooded area. In Figure 6, the transmitters are at different locations than those in Figure 5, but the environments are the same. In IIT Bombay too, the transmitters are at locations of similar environments as in Figures 5 and 6 with just one exception – transmitter 5 in Figure 7 is kept afloat over a water body and the propagation characteristics of the surrounding environment is assumed to be same as an open area environment. (Transmitter 5 in the other two figures are placed in heavily wooded areas). The receiver sensitivity was set to -110 dBm while generating all three heat maps. Locations with an RSSI of less than -110 dBm from all the five transmitters were taken to be out of coverage. The black pixels in Figures 5(B), 6(B), and 7(B) are the areas whose RSSIs from at least one of the transmitters is higher than -110 dBm. One can interpret the (B) images above as photographic negative images.





**FIGURE 6.** Coverage for five transmitters in a heterogeneous environment (IISc), under Rician fading, for transmitter and receiver heights of 1 m each. Image A indicates the Tx Locations. Image B has pixels with RSSI < -110 dBm shaded black. Image C is the “heat map” indicating regions of good coverage.



**FIGURE 7.** Coverage for five transmitters in a heterogeneous environment (IIT Bombay), under Rician fading, for transmitter and receiver heights of 1 m each. Image A indicates the Tx Locations. Image B has pixels with RSSI < -110 dBm shaded black. Image C is the “heat map” indicating regions of good coverage.

Heat maps are generated for only these locations, as shown in Figures 5(C), 6(C), and 7(C). The color-coding is displayed on the right-hand side of these figures.

As the five transmitters are placed in the five different propagation environments, their coverage areas, as well as their coverage patterns, are all significantly different. Being placed in an open area, transmitter 1’s coverage area is larger than those of the other transmitters (except in Figure 7(C) as one would expect). Transmitter 2 shows longer range along the roads, but the RSSI degrades faster in the other directions. As transmitter 3 is located inside a building, its coverage area is the least. Transmitter 4, in Figure 5(C), has a higher coverage area as compared to transmitter 5 because the area around transmitter 5 is more thickly wooded than the area around transmitter 4. The same is the case in Figure 6(C). In Figure 7(C), however, it is transmitter 5 that has the highest coverage area, as expected, because it is surrounded by a large water body which is taken to be similar to an open area environment.

All the above observations are *qualitatively appealing*. It is also evident from the coverage pattern that heterogeneity

significantly affects signal propagation and in turn the coverage area. Our GIS-enabled data-driven RSSI estimation tool has *captured this heterogeneity in a quantitative fashion*. Indeed, the heat map provides a visualization of the quantitative estimate of the RSSI, which is the output of our tool.

## V. CONCLUSION

We demonstrated that coverage estimation can improve significantly by properly utilizing all the available information. In particular, we used both the RSSI on the correctly received packets and additional information on the fraction of lost packets (PER). We also used several factors that are associated with each transmission. We proposed a nonlinear regression scheme, which was inspired by (yet is distinct from) the logistic regression that is popular in the machine learning literature, to make joint use of the packet error rate information as well as the RSSI measurements on the correctly received packets. The nonlinear regression scheme used a link function that is most appropriate for our communication systems context with Rayleigh fading. With Rician fading, a new link function with more parameters was used.



Interestingly, as the RSSI increases, our proposed scheme approaches the classical logistic regression scheme in the sense that the link function approaches the classical logit link function. The RSSI estimation on the correctly received packets involves a loss function. We studied the absolute error and squared error losses. But our method can be easily adapted to other loss functions. It can also accommodate unequal weights for the packet error rate and the RSSI-estimation-error loss functions. Our methodology is also amenable to incremental updating of the parameters. It will be reassuring to prove that, under the received power model (1), under the packet error rate model (6), and under our regression methodologies (8), under an additional assumption that there is positive variance in the independent variables, the estimates of the parameters converge to the true parameter values as the number of samples  $N \rightarrow \infty$ .

We then showed how the proposed estimation procedure can be effectively used, along with readily available open-source GIS data and automated classification of regions into various propagation environments, to estimate coverage in a heterogeneous propagation environment. The heat map that our tool generates enables easy visualization of coverage as well as coverage holes before actual deployment. This can lead to better, more efficient, and faster deployment of outdoor IoT networks.

## ACKNOWLEDGMENT

This article was presented in part at the 9th International Conference on Communication Systems & Networks, Bengaluru, India, January 2017, and in part at the 2018 IEEE Global Communications Conference: Selected Areas in Communications—Internet of Things, Abu Dhabi, December 2018. The authors would like to thank R. Ballamajalu of the Indian Institute of Science for help with the field experiments.

## REFERENCES

- [1] N. Rathod, R. Subramanian, and R. Sundaresan, "Data-driven and GIS-based coverage estimation in a heterogeneous propagation environment," in *Proc. IEEE Global Commun. Conf. Sel. Areas Commun. Internet Things*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.
- [2] D. A. Maltz, J. Broch, and D. B. Johnson, "Lessons from a full-scale multihop wireless ad hoc network testbed," *IEEE Pers. Commun.*, vol. 8, no. 1, pp. 8–15, Feb. 2001.
- [3] P. Rice, A. Longley, K. Norton, and A. Barsis, "Transmission loss predictions for tropospheric communication circuits," Nat. Bur. Standards, Washington, DC, USA, NBS Tech. Note No. 101, Jan. 1967, vol. 1.
- [4] A. Longley and P. L. Rice, "Prediction of tropospheric radio transmission loss over irregular terrain," Inst. Telecommun. Services, Boulder, CO, USA, ESSA Tech. Rep. ERL 79-ITS 67, Jul. 1968.
- [5] A. G. Longley, "Radio propagation in urban areas," in *Proc. 28th IEEE Veh. Technol. Conf.*, vol. 28, Mar. 1978, pp. 503–511.
- [6] R. Edwards and J. Durkin, "Computer prediction of service areas for VHF mobile radio networks," *Proc. Inst. Elect. Eng.*, vol. 116, no. 9, pp. 1493–1500, 1969.
- [7] C. E. Dadson, J. Durkin, and R. E. Martin, "Computer prediction of field strength in the planning of radio systems," *IEEE Trans. Veh. Technol.*, vol. VT-24, no. 1, pp. 1–8, Feb. 1975.
- [8] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Rev. Elect. Commun. Lab.*, vol. 16, nos. 9–10, pp. 825–873, 1968.
- [9] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. VT-29, no. 3, pp. 317–325, Aug. 1980.
- [10] *Urban Transmission Loss Models for Mobile Radio in the 900 and 1800 MHz Bands*, document COST 231, COST 231 TD(90) 119 Rev2, The Hague, The Netherlands, Sep. 1991.
- [11] J. Walfisch and H. L. Bertoni, "A theoretical model of UHF propagation in urban environments," *IEEE Trans. Antennas Propag.*, vol. AP-36, no. 12, pp. 1788–1796, Dec. 1988.
- [12] M. J. Feuerstein, K. L. Blackard, T. S. Rappaport, S. Y. Seidel, and H. H. Xia, "Path loss, delay spread, and outage models as functions of antenna height for microcellular system design," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 487–498, Aug. 1994.
- [13] T. S. Rappaport, *Wireless Communications: Principles and Practice*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
- [14] R. Götz, "Radio network planning tools basics, practical examples & demonstration on NGN network planning part I," LS telcom AG, Germany, Tech. Rep., 2005.
- [15] *Teoco RAN Solutions*. Accessed: 2019. [Online]. Available: <https://www.teoco.com/products/planning-optimization/>
- [16] *Intermap Telecommunications*. Accessed: 2019. [Online]. Available: <https://www.intermap.com/telecommunications>
- [17] J. Moysen, L. Giupponi, and J. Manges-Bafalluy, "A mobile network planning tool based on data analytics," *Mobile Inf. Syst.*, vol. 2017, pp. 1–16, Feb. 2017.
- [18] R. El Chall, S. Lahoud, and M. El Helou, "LoRaWAN network: Radio propagation models and performance evaluation in various environments in Lebanon," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2366–2378, Apr. 2019.
- [19] S. Hosseinzadeh, M. Almoathen, H. Larijani, and K. Curtis, "A neural network propagation model for LoRaWAN and critical analysis with real-world measurements," *Big Data Cogn. Comput.*, vol. 1, no. 1, p. 7, Dec. 2017.
- [20] D. Dobrilović, M. Malić, D. Malić, and S. Sladojević, "Analyses and optimization of lee propagation model for Lora 868 MHz network deployments in urban areas," *J. Eng. Manage. Competitiveness*, vol. 7, no. 1, pp. 55–62, 2017.
- [21] P. Agrawal, A. Ahlen, T. Olofsson, and M. Gidlund, "Long term channel characterization for energy efficient transmission in industrial environments," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 3004–3014, Aug. 2014.
- [22] H. K. Rath, S. Timmadasari, B. Panigrahi, and A. Simha, "Realistic indoor path loss modeling for regular wifi operations in india," in *Proc. 23rd Nat. Conf. Commun. (NCC)*, Mar. 2017, pp. 1–6.
- [23] A. Bhattacharya and A. Kumar, "A shortest path tree based algorithm for relay placement in a wireless sensor network and its performance analysis," *Comput. Netw.*, vol. 71, pp. 48–62, Oct. 2014.
- [24] J. Yang, M. Ding, G. Mao, Z. Lin, D.-G. Zhang, and T. H. Luan, "Optimal base station antenna downtilt in downlink cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1779–1791, Mar. 2019.
- [25] Y. Ren, W. Liu, J. Dong, H. Wang, Y. Liu, and H. Wei, "Genetic algorithm for base station on/off optimization with fast coverage estimation and probability scaling for green communications," in *Proc. Int. Conf. Signal Inf. Process., Netw. Comput.* Qingdao, China: Springer, 2018, pp. 78–88.
- [26] N. Rathod, P. Jain, R. Subramanian, S. Yawalkar, M. Sunkenapally, B. Amrutur, and R. Sundaresan, "Performance analysis of wireless devices for a campus-wide IoT network," in *Proc. 13th Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, May 2015, pp. 84–89.
- [27] M. D. Yacoub, *Foundations of Mobile Radio Engineering*. New York, NY, USA: Taylor & Francis, 1993.
- [28] A. Goldsmith, *Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [29] *Zolertia RE-Mote Platform*. Accessed: 2019. [Online]. Available: <https://github.com/Zolertia/Resources/wiki/RE-Mote>
- [30] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics). New York, NY, USA: Springer, 2013.
- [31] D. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [32] *CC1200 Datasheet*. Accessed: 2019. [Online]. Available: <http://www.ti.com/lit/ds/symlink/cc1200.pdf>
- [33] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint* (Texts and Readings in Mathematics). New Delhi, India: Hindustan Book Agency, 2009.
- [34] W. Lindsey, "Error probabilities for Rician fading multichannel reception of binary N-ary signals," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 4, pp. 339–350, Oct. 1964.



in design and implementation of the Internet of Things (IoT).



developing algorithms for smart networks. Her main interests are network design, architecture, and implementation of the Internet of Things.

**NIHESH RATHOD** received the B.E. degree in electronics and communication from Dharmasinh Desai University, Nadiad, in 2012, and the M.E. degree in telecommunication from the Indian Institute of Science, Bengaluru, in 2015, where he is currently pursuing the Ph.D. degree in communication and networks. Since 2015, he has been a Cisco Research Scholar with the Department of Electrical Communication Engineering, Indian Institute of Science. His interest includes the areas



a Professor with the Department of Electrical Communication Engineering and an Associate Faculty with the Robert Bosch Centre for Cyber-Physical Systems. His interests include the areas of communication, computation, and control over networks. He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, from 2012 to 2015.

**RAJESH SUNDARESAN** received the B.Tech. degree in electronics and communication from IIT Madras, India, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA, in 1996 and 1999, respectively. From 1999 to 2005, he worked with Qualcomm Inc., where he was involved in the design of communication algorithms for wireless modems. Since 2005, he has been with the Indian Institute of Science, Bengaluru, India, where he is currently

...