Radio Environment Map of an LTE Deployment Based on Machine Learning Estimation of Signal Levels

Yosvany Hervis Santana
Ghent University/IMEC
LACETEL
Havana, Cuba
yosvany.hervissantana@ugent.be

David Plets
Ghent University/IMEC
Ghent, Belgium
david.plets@ugent.be

Rodney Martinez Alonso
Ghent University/IMEC
Havana, Cuba
rodney.martinezalonso@ugent.be

Glauco Guillen Nieto
LACETEL
Havana, Cuba
glauco@lacetel.cu

Luc Martens
Ghent University/IMEC
Ghent, Belgium
luc1.martens@ugent.be

Wout Joseph
Ghent University/IMEC
Ghent, Belgium
wout.joseph@ugent.be

***Abstract*— Accurate estimation of Propagation Path Loss is important for reliable and optimized coverage of a service. In literature, a diversity of theoretically or experimentally based propagation models have been documented to estimate the received signal level. The goal of this work is to estimate the effective coverage area of service, predict the Path Loss, and build a Radio Environment Map (REM) using a sensor network. To this end, a sensor's correlation area is defined. By using Machine Learning (ML), the received signal level variation in this area can be estimated correctly 92.3% of the time, with a Mean Absolute Error (MAE) of 1.57 dB. Finally, a proper distribution of sensors based on the correlation area, and ML tools leads to building a REM for the effective coverage area. This approach is applied to a Long-Term Evolution network.**

Keywords*— *Coverage, Estimation, Machine Learning, Received Signal, REM

I. Introduction

Accurate estimation of Propagation Path Loss (PL) is important for reliable and optimized coverage of service to grant access to the radiofrequency spectrum. To improve the utilization of radio resources Radio Environment Maps (REMs) can contribute to representing and understanding the propagation environment. REMs can be used for a variety of analyses, such as available services, geographical features, grades of Quality of Service (QoS), and areas served under a specific service. However, the main challenge in building a REM is the method to collect the information for a large area.

In literature, a diversity of theoretically or experimentally based propagation models have been documented to estimate signal coverage path loss [1]. Several research works have investigated the proper way to deploy sensor nodes, the accurate measurement of the spectrum, and the prediction of signal levels [2][3][4][5][6].

In [2], the authors proposed an analytic framework based on spectrum data gathered by spatially distributed sensors to construct a REM. In places without sensors, they used signal strength interpolation by composition sampling. The estimation was made using two distributions of the sensors, i.e., randomly distributed, and square lattice located distribution. The results revealed that both sensor distribution methods are suitable for Received Signal Strength (RSS) estimation. It was proven that better performance was obtained with the sensors distributed in squared lattice mode in terms of the average inference performance, but the performance was not analyzed when the sensors are close to the transmitter. In our work, we analyze the performance when the sensors are close to each other and close to the source.

In [3], the authors presented how to place sensor nodes (SNs) to guarantee the performance of machine learning based on cooperative sensing schemes. They proposed a strategy on how to place a few SNs to cover the whole area of the PU. In the research, the problem caused by a hidden transmitter (is when the transmitted signal is not covered by a sensor) appeared because the authors do not prevent the deployment of sensors in an area where the transmitter can be placed. In [4], a heterogeneous network formed by a traditional cellular network and a Wireless Sensor Network (WSN) was considered. In this scenario, the role of the WSN was to estimate the REM of the cell using a Kriging geostatistical interpolation technique. The results show that for reducing the Minimum Square Error more nodes (sensors) had to be added. To reduce the number of sensors to deploy we introduce and define in our research the "sensor correlation area". It allows us to cover the entire area and used the measured information by the sensors to estimate the received signal level in places where no sensors are deployed.

In [5], a method for constructing a REM combining residual maximum likelihood-based radio propagation parameter estimation with Kriging-based transmission power prediction was proposed. With the Monte Carlo simulation, the result indicates that the authors' proposal provides a standard deviation of 4.8 dB, and the performance compared with different methods was improved by 2 dB. In [6], the authors built a REM using Mobile Crowd Sensing (MCS) to collect the information. To infer the missing information from the radio environment in the unsensed areas they used the Kriging algorithm. The results have shown that the interpolation error with the Kriging algorithm was around 5 dB outperforming by 2 dB and 5 dB the Nearest Neighbor and IDW algorithms respectively. In [5] and [6] was demonstrated that it is possible to use a Machine Learning (ML) algorithm to estimate the received signal level and build a REM. Furthermore, in our research, the performance is improved by at least 3 dB compared to the aforementioned research.

In the present research, the sensor's correlation area is obtained. This is an area where the received signal level at any point inside it, correlates equal to or higher than 0.5 with the received signal level measured by a sensor. Then, a proper distribution of sensors based on the correlation area, and ML tools will lead to a reduction in the number of sensors needed to reliably characterize the services' coverage. In this way, fewer sensors will be required to estimate the effective coverage area of service, predict the PL, and build a REM. This approach will be used to estimate the effective coverage area of a Long-Term Evolution (LTE) network in a suburban area.

The rest of the paper is organized as follows. Section II presents the methods used to collect the information with the sensor and obtains the sensor coverage area. It presents how the ML is trained/validated, and how the resulting model is combined with the sensor coverage area to estimate the received signal level for building a REM of an LTE network. Section III presents the main results of our research. Finally, Section IV presents the conclusions and future work.

II. METHODS

In this section, we present a methodology to build a REM of an LTE network using a sensor deployment and ML. We define the sensor correlation area, train/validate an ML model to estimate the received signal level in the areas without sensors, and analyze where to deploy sensors to properly estimate the effective coverage area of service. To define the sensor correlation area and to train/validate and test the ML model we conduct the field measurement of a Digital Television Broadcast transmitter working in a frequency close to the low LTE Band. Then, we used the results to build the REM in a simulated LTE network.

A. *Study Area*

We conduct the field measurements based on a Digital Television (DTV) transmitter. Field measurements were conducted in channel 45, central frequency 659 MHz, bandwidth BW 6 MHz in Jagüey Grande (suburban area), Matanzas province, Cuba. Using the sensor measurements and the measurements in the vicinity, we account for the variation of the received signal level as a function of the distance over 24 hours. Table I defines the link budget for the TV technology in the selected area. As can be noticed, the Receiver Height is above the average ground terrain (5m) for which at least 60% of the first Fresnel zone is not obstructed. For these conditions, direct visibility (Line of Sight) can be fairly assumed [7].

TABLE I: Link budget parameters

Parameter	***Value***
Tx Height [m]	346
BS Frequency [MHz]	509
Tx Power [dBm]	67
TV Standard	DTMB
Bandwidth [MHz]	6
Rx Height [m]	5
Receiver Antenna Gain [dBi]	0
Transmitter Antenna Gain [dBi]	6

For measurements, we used the sensor developed in [8]. It measures the received signal level in dBm, Signal to Noise Ratio (SNR) in dB, and calculates the signal level variation (fading) in dB. To calculate the fading, we follow the procedure described in [9], where the received signal level is measured 26 times in one second (it was proved by laboratory tests) to avoid the Doppler effect caused by vehicles. Then, assuming that the mean value of the fading is zero dB, the fading in one second is calculated as the difference between the 50-percentile minus the first sample. Finally, the variation in one minute is obtained as the average of the 60 fading calculated each second.

B. *Sensor Correlation Area: Definition*

First, we define the sensor correlation area. The idea is to determine the correlation between the signal measured at the sensor location and the signal measured in its vicinity. Then the received signal level at any point inside this area can be estimated using the sensor measurement. The three available sensors were re-used to take measurements at all 22 locations, thus taking place over several days. For the analysis, the measured values need to be organized according to the timestamp of the day they were taken, i.e., we guarantee that all measurements were taken at the same timestamp of the day in every location. Finally, to find the maximum distance at which two measurements of the received signal level are still correlated, we analyze the variance of the differences between the measurements on the sensor and the measurements in the vicinity. The variance gives an idea of the difference between the reference (the place where the sensor is located) and the sampling points (measured locations around the reference), then to have a good performance we look for a variance lower or equal to 3 dB. This only can be achieved if the correlation coefficient (relationship between the measured signal level at the sensor location and the measured signal level in the vicinity) is higher than or equal to 0.5 [10].

C. *Measurement and Data Collection*

The methodology proposed in this research has been developed using field measurements in an area of 3.5 km^2.

Fig. 1 shows the DTV measurement locations with 22 measurement locations around the sensor at distances from 50 m to 2500 m.

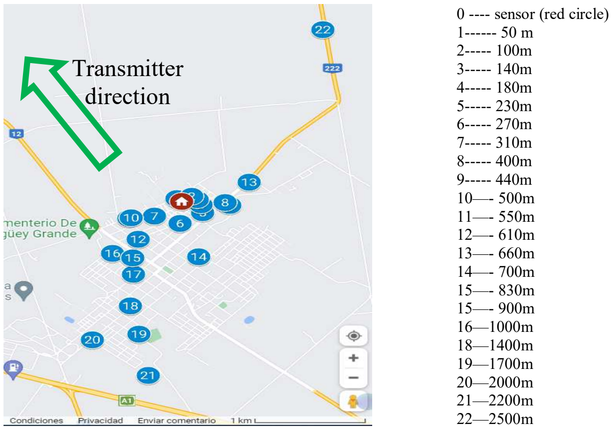


Fig. 1. DTV field measurement points distribution. The red circle represents the sensor, the blue circles are the measurement points, and the green arrow shows the transmitter direction, 42 km far away (Northwest direction).

For the measurement process, we used three sensors in every iteration, always keeping one sensor in the same position while the other two were moved among the 22 locations represented in Fig. 1. The sensors were adjusted and calibrated, so measurements were comparable with the professional instruments. We measured and registered in time the received signal level, the SNR, and the fading (calculation procedure explained in [9]). Then we organize the samples in our dataset so, we can process samples taken at the same timestamp of the day on different days at each of the 22 locations together.

D. Proposed model

To estimate the temporal and spatial signal variation around the sensor, we propose a method based on ML. In (1) we present the formula to model the received signal level.

$$RxV_{(d,t)} = RxS_{(t)} + \beta_{(d,Dt)} \quad (1)$$

Where, $RxV_{(d,t)}$is the received signal at a certain point in the vicinity at a distance d from the sensor and at a given time t, $RxS_{(t)}$ is the sensor's received signal at the same time t, and $\beta_{(d,Dt)}$ is the correction factor for the received signal level as a function of the distance and the different time of the day, obtained as part of the process to determine the correlation area.

1) Machine Learning Algorithm

To model $\beta_{(d,Dt)}$, we use different ML Classification and Regression algorithms [11]. Fig. 2 shows the diagram used for the $\beta_{(d,Dt)}$ calculation. The most important supervised learning algorithms are detailed in [12], and they are, among others, k Nearest Neighbor (kNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and Neural Networks (NN). As inputs, we used the measured signal level vectors.at the same timestamp of the day in all sensor locations and the distance between them. The whole dataset has 1440 samples in 24 hours per each of 22 sensors' locations (one sample per minute containing the measured signal level, SNR, and fading), then we divided it into samples for training (80%), samples for validation (15%), and samples for testing (5%). We compared the performance of 10 ML algorithms. The Classification algorithms are used to estimate if the correction should be negative or positive. On the other hand, we use a Regression algorithm to estimate the correction value ($\beta_{(d,Dt)}$).

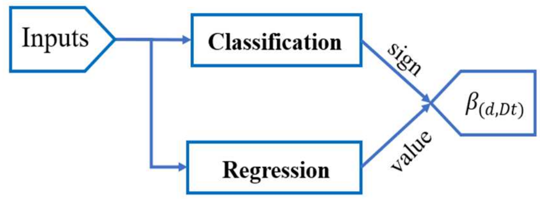


Fig. 2. Machine learning diagram for ($\beta_{(d,Dt)}$) estimation

E. LTE network

To evaluate this approach in an LTE network, the received signal level at the sensors' locations ($RxS_{(t)}$) is simulated using the deployment tool GRAND [13] and the $\beta_{(d,Dt)}$ correction factor is estimated using the ML model trained/validated with the field measurements in Section II-A. The GRAND tool is capacity-based, which means, that the traffic density and end-devices density are input parameters. The software also receives as input parameters the target area and all possible BS geo-locations including the BS antenna height. Fig. 3 shows the LTE network deployment.

The real BS settings are adopted in the tool: the radiated power of the BS was 36 dBm, Okumura-Hata as PL model for a suburban area [14], and a frequency of 800 MHz. The LTE network simulation is executed for an extraordinary situation (i.e., festivity day) where there are many users (i.e., 650@2Mbps) connected at the same time in a small area (4.8 km^2). Finally, 3 BSs are needed to cover 95% of the users.

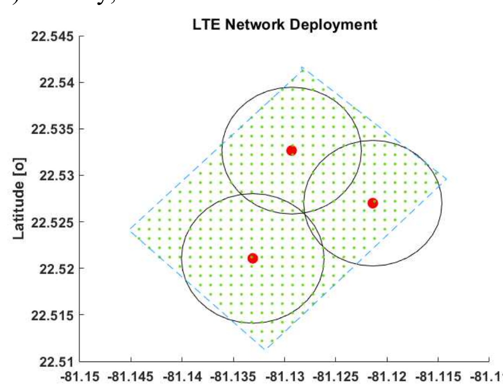


Fig. 3. LTE network deployment with 3 BS (red dots) for covering 4.8 km^2 and 650 users at 2Mbps. Green small dots represent the points where the received signal is estimated using equation 1.

The optimal sensors' deployment location to build a REM depends on the area to be covered, the sensor correlation area, and the distance between the BSs and the sensors. The signal level varies much faster for the distances closest to the BS than far away from it (log-distance dependence). Finally, we use the simulated received signal on the sensors and equation (1) to estimate the received signal level in 456 grid points (green small dots with 200 m spanning size) inside the 4.8 km^2 and build the REM.

III. RESULTS

A. Sensor Correlation Area, Result

In this section, we analyze the correlation coefficient to properly define the sensor correlation area. Fig. 4 shows the correlation coefficient between the DTV received signal level measurements made by the sensor and the ones made in its vicinity (at the same time instance). We consider here the trend value as a correlation model, represented with a continuous line, where the correlation coefficient behavior is a function of the distance between the sensor location and the measurements in the vicinity. The correlation coefficient is always higher than 0.5 up to 830 m.

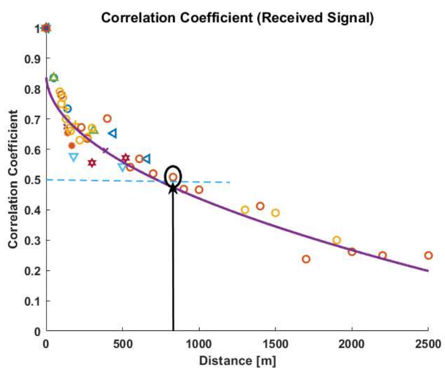


Fig. 4. Correlation between the received signal level measurements on the sensor place and the measurements in the vicinity. Plots (different shapes) represent the correlation coefficient between the measurements in the sensor and the ones in the vicinity. The continuous line represents the trending value of the correlation coefficient as a function of the distance. The black circle represents the distance at which the correlation coefficient is around 0.5 defining the sensor correlation radius.

B. Received Signal Estimation

For estimating the received signal level at any point inside the sensor correlation area (defined in Section III-A) we use equation 1 (Section II-D). In this section, we analyze the estimation of the $\beta_{(d,Dt)}$ correction factor as a function of the distance. A combination of 10 ML algorithms is compared in the Classification and Regression process. For testing the model, we used a dataset (72 samples, 5%) not included in the training/validating process. TABLE II shows the results of the four algorithms. The algorithm with the best performance is Naïve Bayes [15], which made a correct classification in 92.3% of the samples in our dataset.

TABLE II: ESTIMATED RESULTS IN THE CLASSIFICATION PROCESS.

Algorithm	*Correctly Classification Samples (%)*	*Incorrectly Classification Samples (%)*
NB	**66 (92.3%)**	**6 (8.0%)**
SVM	58 (81.0%)	14 (19.0%)
LR	61 (85.0%)	11 (15.0%)
MLP	64 (89.0%)	8 (11.0%)

For estimating the $\beta_{(d,Dt)}$ values, we compared six algorithms, Gaussian Process (GP), LR, MLP, Simple Linear Regression (SLR), SMOreg, and IBK (commonly known as the k-nearest neighbor algorithm). To evaluate the performance, we used the same dataset with 72 samples. TABLE III summarizes the results for every algorithm. The best performance up to 830 m is achieved with the SMOreg (Support Vector Machine algorithm for regression) [16] algorithm with an 85% of correlation between the target (calculated $\beta_{(d,Dt)}$) and responses (estimated $\beta_{(d,Dt)}$ with ML), and a Mean Absolute Error (MAE) of 1.57 dB.

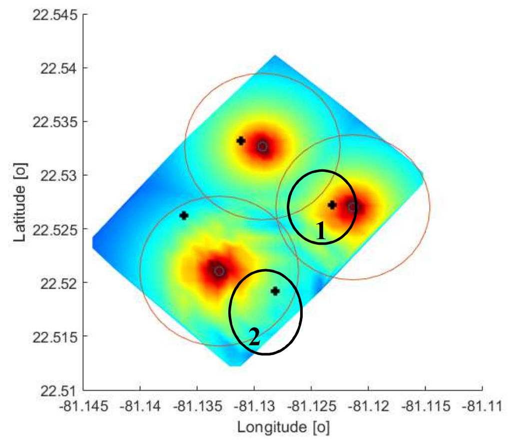
TABLE III: ESTIMATED RESULTS IN THE REGRESSION PROCESS FOR 830 M

Algorithm	*Correlation [%]*	*MAE [dB]*
GP	72	2.44
LR	81	1.70
MLP	52	2.10
SLR	62	2.05
SMOreg	**85**	**1.57**
IBK	79	1.65

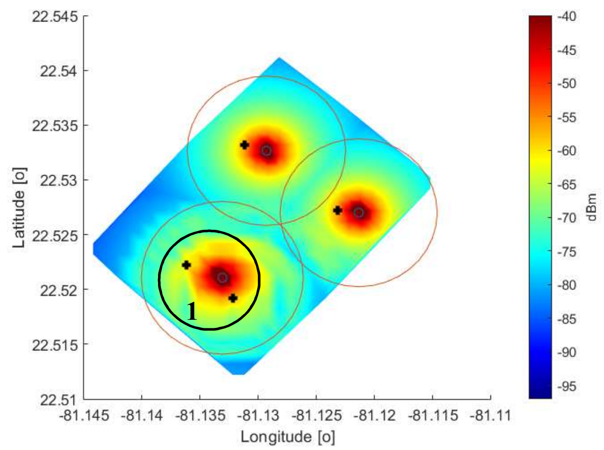
C. REM construction

From Section II-E 4 sensors are required to properly make predictions in the 456 grid points using the proposed approach in this paper. To decide the optimal sensors' deployment, we proposed three different setups and analyze the contribution of every deployment. Fig. 5 shows the REM for the estimated $\beta_{(d,Dt)}$ at 9:00 am with the information obtained by the sensors and the estimation with ML.

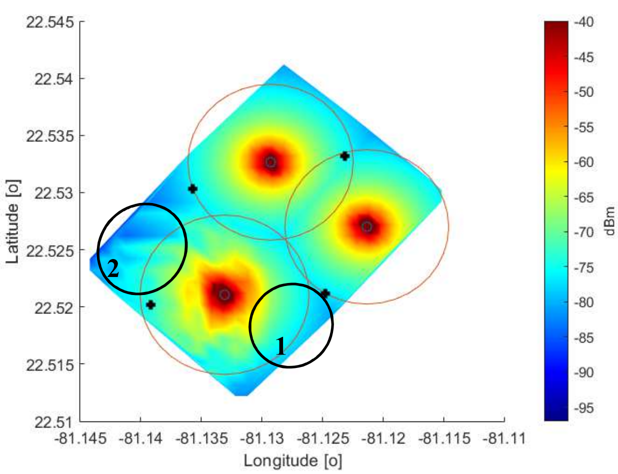
Fig. 5a shows the obtained REM with sensors uniformly distributed. In this scenario, each sensor collects information from several BSs. In this way, the sensors can be reused (i.e., zone 1, Fig. 5a). In the case when the sensor is placed close to the BS, it does not effectively contribute to detecting the edge of the coverage area. In opposite, positioning sensors closest to the expected edge of the BSs' coverage area (zone 2, fig. 5a) will increase the effectiveness in determining the actual covered area, reduce the error in the estimation process, and thus the precision in the estimation using ML from the inputs collected by the sensors. Fig. 5b shows the obtained REM when the sensors are all close to the BSs. As can be seen, in zone 1 (Fig. 5b) there are two sensors close to the BS. In this case, there are no sensors close to the edge of the BS's expected coverage area, which means there is no effective contribution to detecting the behavior at the edge of the actual coverage area. Finally, Fig. 5c shows the obtained REM with sensors deployed near the theoretical BS's coverage area. In this case, all the sensors were positioned close to the theoretical BS's coverage area, which contributes to the detection of served areas (zone 1 and 2, Fig. 5c). In this scenario, the proposed model (ML + sensors) shows that more than 90% of the area could be covered in terms of received signal level with three LTE BSs.



(a)



(b)



(c)

Fig. 5. REM of an LTE deployment based on ML to estimate the signal levels in the Sensors' correlation area. The dashed line (circles) represent the LTE BSs' and served areas, both obtained with the GRAND tool, and the black crosses are the sensors. (a) Sensors are uniformly distributed. (b) Sensors are close to the BSs (c) Sensors near the theoretical BS's coverage area.

IV. Conclusions

In this paper, a procedure to properly estimate the sensor's correlation area was defined. From measurements merging distances up to 2500 m from the sensor, we found a correlation around 0.5 up to 830 m. Using ML algorithms, the received signal level variation around a sensor can be correctly estimated 92.3% of the time, with an MAE of 1.57 dB. Finally, three different sensor deployments were analyzed showing that the best approach is to deploy the sensors near the edge of the BS expected coverage area.

Future works will consist of grouping measurements in time intervals (i.e., 4, 6, or 8 hours) the variation of $\beta_{(Dt)}$along the time of the day is assessed for every distance value.

References

[1] K. Pahlavan and A. H. Levesque, *Wireless Information Networks*, 2nd ed. 2005.

[2] Y.-Q. Xu, B. Zhang, B. Zhao, and D. Guo, “Radio Environment Map Construction with Spatially Distributed Sensors,” in *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2021, pp. 1–7. doi: 10.1109/WCNCW49093.2021.9419993.

[3] J. Kim and J. P. Choi, “Sensing Coverage-Based Cooperative Spectrum Detection in Cognitive Radio Networks,” *IEEE Sens. J.*, vol. 19, no. 13, pp. 5325–5332, 2019, doi: 10.1109/JSEN.2019.2903408.

[4] F. Frantzis, V. Chowdappa, C. Botella, J. J. Samper, and R. J. Martinez, “Radio Environment Map Estimation Based on Communication Cost Modeling for Heterogeneous Networks,” in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 2017, pp. 1–6. doi: 10.1109/VTCSpring.2017.8108227.

[5] D. Mao, W. Shao, Z. Qian, H. Xue, X. Lu, and H. Wu, “Constructing accurate Radio Environment Maps with Kriging Interpolation in Cognitive Radio Networks,” in *2018 Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, 2018, pp. 1–3. doi: 10.1109/CSQRWC.2018.8455448.

[6] Z. Han, J. Liao, Q. Qi, H. Sun, and J. Wang, “Radio Environment Map Construction by Kriging Algorithm Based on Mobile Crowd Sensing,” *Wirel. Commun. Mob. Comput.*, vol. 2019, p. 4064201, 2019, doi: 10.1155/2019/4064201.

[7] App, “APPLICATION NOTE Line of Sight Obstruction”.

[8] Y. H. Santana, D. M. Betancourt, A. I. M. Valdes, Y. H. Sanchez, G. A. G. Nieto, and R. M. Alonso, “DTMB monitoring tool based on a commercial set-top box,” in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2017, pp. 1–6. doi: 10.1109/BMSB.2017.7986216.

[9] P. Angueira, M. Velez, D. D. La Vega, A. Arrinda, and J. L. Ordiales, “Fading caused by moving vehicles near the receiver on DTV (COFDM) 8-MHz signals,” *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 250–252, 2002, doi: 10.1109/LCOMM.2002.1010871.

[10] S. Wang and L. Zhang, “A Supervised Correlation Coefficient Method: Detection of Different Correlation,” in *2020 12th International Conference on Advanced Computational Intelligence (ICACI)*, 2020, pp. 408–411. doi: 10.1109/ICACI49185.2020.9177709.

[11] O. Rozenblit, Y. Haddad, Y. Mirsky, and R. Azoulay, “Machine Learning Methods for SIR Prediction in Cellular Networks,” *Eur. Coop. Sci. Technol.*, 2019.

[12] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. O'Reilly Media, Inc., 2019.

[13] R. M. Alonso, D. Plets, M. Deruyck, L. Martens, G. G. Nieto, and W. Joseph, “TV White Space and LTE Network Optimization Toward Energy Efficiency in Suburban and Rural Scenarios,” *IEEE Trans. Broadcast.*, vol. 64, no. 1, pp. 164–171, 2018, doi: 10.1109/TBC.2017.2731043.

[14] T. Mahmood, H. K. AL-Qaysi, and A. S. Hameed, “The Effect of Antenna Height on the Performance of the Okumura/Hata Model Under Different Environments Propagation,” in *2021 International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1–4. doi: 10.1109/CONIT51480.2021.9498395.

[15] J. Van Der Herten, I. Couckuyt, D. Deschrijver, and T. Dhaene, “Adaptive classification under computational budget constraints using sequential data gathering,” *Adv. Eng. Softw.*, vol. 99, pp. 137–146, 2016, doi: 10.1016/j.advengsoft.2016.05.016.

[16] S. M. A. M. Gadal and R. A. Mokhtar, “Anomaly detection approach using hybrid algorithm of data mining technique,” in *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, 2017, pp. 1–6. doi: 10.1109/ICCCCEE.2017.7867661.