Compensation Method of Received Signal Power observed by Smartphone for Crowdsensed Spectrum Database

Taiki Matsushima* and Takeo Fujii*
* Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications
1–5–1 Chofugaoka, Chofu, Tokyo 182–8585, Japan
E-mail: {matsushima, fujii}@awcc.uec.ac.jp

***Abstract*—Crowdsensing is a technique for observing a wide range of radio environments at high density. Collecting radio environment information using sensors built into smartphones, a huge data set can be acquired while reducing the installation cost of new measurement equipment. By accumulating the data set in Measurement-based Spectrum Database (MSD), a wide range and high-precision Radio Environment Map (REM) can be generated. However, crowdsensing is impossible to measure the received power accurately without considering the power fluctuation due to the difference of the working scenario, e.g. specific conditions of human such as putting a wireless terminal in a pocket, and the error tendency of each terminal. In this research, we have a measurement campaign focusing on measurement errors caused by working scenarios and smartphone types in the cellular environment. Also, we propose the compensation method of observed values using simple linear regression. In the proposed method, the observed value of the terminal and one of the measurement equipment for area evaluation for LTE system are compared. Those data are classified in the MSD, and the mesh IDs are obtained. Then, those data matching mesh IDs are plotted to obtain a simple regression line. The predicted value is compensated by substituting the observed value of the terminal into the regression line. Comparing the difference between the ground-truth and the predicted value, the proposed method can improve the accuracy of REM under all working scenarios and smartphone types.**

Index Terms*—*Spectrum Database, Crowdsensing, Linear Regression, Working Scenario, Smartphone

I. Introduction

With the rapid increase in demand for wireless communication, shortage of spectrum resource becomes a serious problem [1]. To save the spectrum resource shortage, spectrum sharing [2], the technology that Secondary Users (SU) use the temporal and spatial unoccupied band which is not used by Primary Users (PU) is proposed. It is expected the improvement of utilization efficiency of frequency by making effective use of temporal or spatial unoccupied bands. Also, cognitive radio technology [3] has attracted attention as an example of using spectrum sharing. Cognitive radio is a technology that sensors recognize the surrounding wireless environment and set communication parameters to suit environment. It is expected that the improvement of utilization efficiency of frequency will be improved by adapting communication parameters by utilizing cognitive radio, but it is necessary to accurately estimate the radio environment to use decide appropriate parameters.

Then, radio environment estimation methods using the MSD have been proposed [4][5]. The spectrum database stores wireless environment information such as received power and is used to estimate the spectrum usage. SU can know the spectrum usage by using the spectrum database. SU receives instructions to change communication parameters such as power and frequency. Then, SU starts transmission after changing parameters. As a result, both PU and SU can share spectrum resources in the same spectrum band. However, it is necessary to collect a large amount of data with sensors to construct the database. If the high-precision sensor is used, the cost is high, and it is not suitable for a wide range measurements. As an alternative proposal, crowdsensing attracts much attention [6][7]. Crowdsensing is a technology that collects various types of information in wireless environments using personal devices such as smartphones. As a result, low cost and large-scale radio environment information can be collected from personal users. However, in crowdsensing, since the quality of the sensor is low, the observation error varies according to the observation situation and devices. To reduce the influence of observation error, an estimation method of the received signal strength when humans behave in various patterns was proposed [8]. Ref. [8] reports that the parameters of the measurement model are different if the working scenario is different when the same smartphones are used and that different types of smartphones have different parameters when we observed data in the same working scenario. Then, compensation methods for these problems have been proposed. Specifically, by performing the iterative estimation of model parameters based on the Davidon-Fletcher-Powell (DFP) algorithm and construction of Model-driven REM construction, the accuracy of the average Received Signal Strength (RSS) error between the predicted value and the ground-truth is improved by 57.2 % compared with the conventional research. Also, [9] proposed a method for calibrating observation data for each statistical area. Performing estimation of observation error for each mesh and transmission power of PU by using the Expectation-Maximization algorithm and using least squares method and the average value in mesh, [9] can be removed the effect by fading component of radio propagation and construct high-precision databases.

However, in [8], observation points are the building rooftop without obstacle and propagation characteristics in the cellular environment are unknown. Also, [9] exclude instantaneous fluctuation of power for the difference of working scenarios and measurement errors of the smartphone. Moreover, [9] does not use measurement data of the actual environment for error estimation. In crowdsensing, there are various ways to hold and place wireless terminals. Then, it is assumed that the observation error varies depending on the smartphone type and working scenario. To realize accurate radio propagation characteristics prediction, an observation value compensation formula according to the working scenario of general users is required.

Therefore, to confirm the results in the general environment, we conduct measurement campaign to collect data in an actual cellular environment and propose a new estimation of the method received signal strength. As a result, the error between the estimated value and ground-truth (accurate value measured by high precision equipment) is reduced. The main contributions of our work are proposing a compensation method of the observed value collected by smartphone in crowdsensing. Finally, we evaluate the error between ground-truth and observed value of smartphone or predicted value. Then we find out that suppression of error in each working scenario.

The rest of the paper is organized as follows. The radio environment recognition technology and sensing technology for realizing spectrum sharing are described in Sec. Ⅱ. In Sec. Ⅲ, we give an overview of the proposed method in this study and explain the simple linear regression used as a method of compensation of observed values. The outline and results of the observation experiments in the cellular environment using the measurement equipment for area evaluation and smartphones, focusing on working scenarios and errors caused by the type of smartphones are described in Sec. Ⅳ. Finally, V concludes the paper and describes future works.

II. Radio Environment Recognition and Sensing Technologies

In this section, the radio environment recognition technology and the sensing technology using the database that plays an important role in collecting wireless environment information are described as wireless environment estimation methods.

A. Measurement-based Spectrum Database

As an example of using the spectrum database, MSD has been proposed [4][5]. Fig. 1 shows the schematic diagram of the MSD. In MSD, the wireless terminal owned by the SU plays the role of a sensor and observes the wireless environment such as received power and location information at the arbitrary receiving position. After wireless terminal reports observed information to MSD, MSD divides observation data into two-dimensional meshes (a division of the ground surface into many squares according to certain rules) based

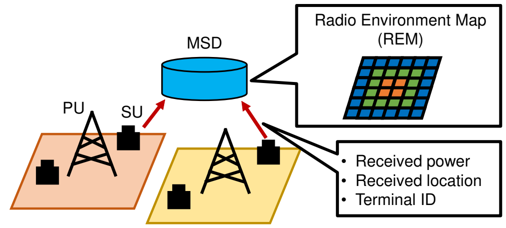

Fig. 1. Schematic diagram of MSD.

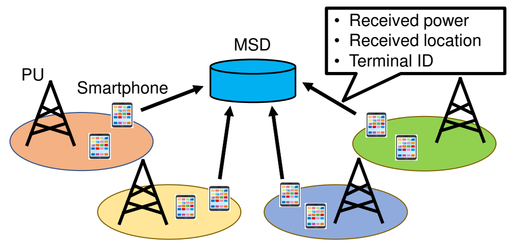

Fig. 2. Schematic diagram of crowdsensing.

on accumulated information, and performs statistical processing such as averaging. Since the MSD is constructed with information from actual observations considering the effects such as shadowing, the wireless terminal can recognize radio environments with high accuracy. Also, since the accumulated data is based on actual observations, more accurate parameters can be transmitted to SU, and SU can adaptively determine the transmission power that does not affect the PU. Therefore, PU does not excessively limit the spectrum sharing of PU, and the improvement of utilization efficiency of frequency can be expected.

To build a high-precision MSD, wireless terminals as sensors require to have high reliability. However, it is not practical to observe the wide range of wireless environment at high density because it is expensive to use a spectrum measurement device such as a spectrum analyzer with a small error. Therefore, an idea has been proposed that a large amount of information can be obtained in various environments by using low cost personal terminals such as smartphones instead of expensive spectrum measurement equipment.

B. Crowdsensing

As a method to collect low cost and large-scale radio environment information, crowdsensing, using the reception function of the smartphone, is attracting attention [6] [7]. The smartphone is equipped with various sensors, e.g., acceleration sensor used to change the orientation of the display, Global Positioning System (GPS) that displays its position information, and a magnetic sensor that detect the direction and received power [8]. Fig. 2 shows the schematic diagram of crowdsensing. Crowdsensing is a mechanism for observing and collecting wireless environment information using inexpensive sensors such as smartphones. Information collected by crowdsensing is accumulated in the database, and processing such as extraction of information (latitude/longitude, received signal strength) required by the smartphone is performed on the database. When the processing is completed, distribution of statistics is sent to the smartphone. Then, the smartphone grasps the distribution and usage of the surrounding radio wave environment. Finally, the smartphone can perform spectrum sharing in unused bands and spaces by adapting based on information. The advantage of crowdsensing is that observation information can be obtained from personal users who have a smartphone, and low cost and large-scale radio environment information can be collected from these users.

III. Generating Predicted Values from Residuals in Linear Simple Regression

This section describes the proposed method for improving the accuracy of observed values. III-A describes the flow of the proposed compensation method. Here, observed values of a measurement instrument for area evaluation is treated as ground-truth when generating the predicted values. III-B describes the simple linear regression and the residual used in the method to compensate errors between ground-truth and the observed value of the android terminal.

A. The Flow of The Proposal Compensation Method

First, to accumulate a large amount of wireless environment information in the database, the observation data for each working scenario is uploaded in the database at each android terminal and the measurement instrument for area evaluation, which is the equipment that can obtain highly reliable radio wave propagation characteristics even in an environment with a lot of interference. Then, the uploaded data is divided into 2D meshes on the database. At this time, data processing such as the acquisition of mesh ID given for each mesh and averaging of RSRP is performed. After that, we prepare a filter designed based on the request of the user to limit the center frequency, latitude and longitude, and Physical Cell ID (PCI), and extract the data that pass the filter. The processed RSRP is output to REM. We treat this RSRP as the observed value of the android terminal before prediction.

Second, we derive a simple regression line using the processed data. In the filtered data, observed values of the android terminal and the measurement instrument for area evaluation with the same working scenario are compared, and only the data whose mesh ID matches are extracted. Next, these data are plotted for each smartphone and working scenario. The horizontal axis of the plotted data is the RSRP of the smartphone, and the vertical axis is the RSRP of the measuring instrument for area evaluation. Then, based on the relationship between the RSRP of the smartphone and the measuring instrument for area evaluation, we derive the regression line by using simple linear regression. As an example, the plot of the data and the regression line are shown in Fig. 3. After that, the observed value of a smartphone is substituted into the regression line created by simple linear regression to generate temporary RSRP.

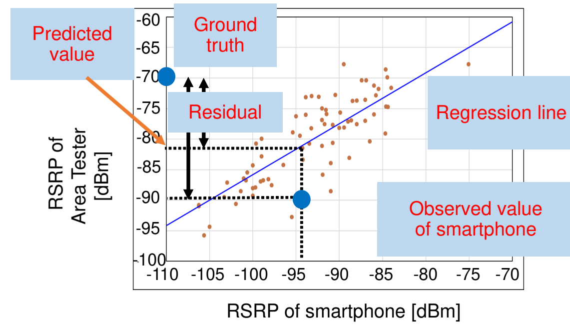

Fig. 3. The plot of the data and the regression line.

Third, the data obtained by replacing the data observed at each terminal with the predicted value is uploaded to the database, and we output REM. Then, we compare the accuracy of REM including the error of predicted value and ground-truth.

B. Simple Linear Regression

In this paper, we use simple linear regression as a method for compensating observed values. Simple linear regression is that the data that is scattered on the scatter diagram is represented by a linear model represented by the formula $y = ax + b$. Then, let us consider how to determine the parameter a and b of the linear model $y = ax + b$. In this research, we represent (x_i, y_i) as the i th measurement data set. If the relationship between the measured data is completely determined by the model $y = ax + b$, a relational expression $y = ax_i + b$ should be established when $x = x_i$. However, in fact, the relationship $y = y_i$ is true, so it cannot be said that $y = ax_i+b$ is false when $x = x_i$. Therefore, the parameter that minimizes the sum of the squares of the difference between y_i and $y(= ax_i + b)$ for all (x_i, y_i) is the most appropriate. When these are expressed using mathematical formulas, a and b are decided by minimizing L,

$$L = \sum_{i=1}^{n}(y_i - (ax_i + b))^2, \tag{1}$$

where n is the number of data sets. To derive a and b, (1) is a partial derivative concerning a and b, and solve an equation with the partial derivative value set to 0. As a result, a and b is expressed as:

$$\begin{aligned} a &= \frac{s_{xy}}{{s_x}^2} \\ &= \frac{\sum_{n=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{n=1}^{n}(x_i - \overline{x})^2}, \end{aligned} \tag{2}$$

$$b = \overline{y} - a\overline{x}, \tag{3}$$

where s_{xy} is the covariance of x and y, ${s_x}^2$ is the variance of x, and $\bar{x}$ and $\bar{y}$ is the mean value of x and y, respectively. The method of deriving such parameters is called the least-squares method, and the linear equation $y = ax + b$ obtained by least-squares method is called the regression line of y to x. By substituting (3) to $y = ax + b$, the detailed formula for the regression line is obtained as,

$$y - \overline{y} = a(x - \overline{x}). \tag{4}$$

Therefore, the regression line represents a straight line with gradient a passing through point (x, y).

Here, for each x_i, the corresponding y on the regression line is represented as $\hat{y}_i$, and this $\hat{y}_i$ is called the predicted value. At this time, the difference between the actual data y_i and the predicted value $\hat{y}_i$ is called the residual d_i, which is expressed as,

$$d_i = y_i - \hat{y}_i = y_i - (ax_i + b). \tag{5}$$

The closer the residual d_i for each i is to 0, the better the fit to the regression line. This research evaluates the accuracy based on the distribution of the residuals. In this paper, d_i is represented as,

$$d_i = y_g - y_p, \tag{6}$$

where y_g is the RSRP (Reference Signal Received Power) of ground-truth and y_p is the predicted value derived from the regression line.

The flow of simple linear regression was explained above, but we will explain the reasons for using the linear simple regression are described below. First, since only the data whose latitude and longitude match between the area tester and the terminal are extracted, it is easy to grasp the correlation of the data. Second, simple linear regression is a method that minimizes the contradiction between the predicted value and the actual observed value, so it is easy to compensate for all the observed data. Finally, simple linear regression is a long-established statistical method that guarantees the reliability of the data after statistical processing.

IV. Measurement Campaign

This section describes the observation experiment using smartphones and the measurement instrument for area evaluation, which was performed to evaluate the compensation method of radio propagation characteristics in a cellular environment focusing on working scenario and smartphone-specific observation errors. Next, we describe the results of implementing the proposed method and compare data before and after compensation.

A. *Observation Experiment in a Cellular Environment*

On December 26, 2019, the measurement campaign using a measurement instrument for area evaluation and smartphones was performed at Fujimi-Cho, Chofu-shi, Tokyo. From the viewpoint of crowdsensing, to consider the outdoor sensing of personal users, we observed the received power in a cellular environment. Receivers include a measurement instrument for area evaluation called Area Tester that is manufactured by Anritsu. We used Essential Phone PH-1 (manufactured by Essential), Zenfone Live (manufactured by ASUS), and Pixel3 (manufactured by Google) for smartphone devices. To accurately record the location information of the wireless terminal, DG-PRO1RW (manufactured by Drogger) was prepared as GPS and paired using Drogger GPS Android application. Then, we used the “Self Implemented Measurement Application (SIMA)” as the android application to collect observation value. SIMA has the function to accumulate information (latitude and longitude, measurement time, RSRP, etc.) collected using the sensor of smartphones and upload the accumulated information to the database after the observation is completed. About the observation experiment, we walked the observation area in three raps at normal walking speed. The actual walking route is shown in Fig. 4. Then, for the base station, the center frequency was 2137.6 [MHz] and the frequency bandwidth was from 2110 to 2170 [MHz]. When we use the Area Tester, one person operates cart carrying the Area Tester, and the antenna direction was changed as three directions in consideration of directivity. Fig. 5 shows the example of condition of antenna connected with the Area Tester. Also, we observed received power by using android terminals in three working scenarios: 1. In hand, 2. In the pocket of clothes, 3. In bag.

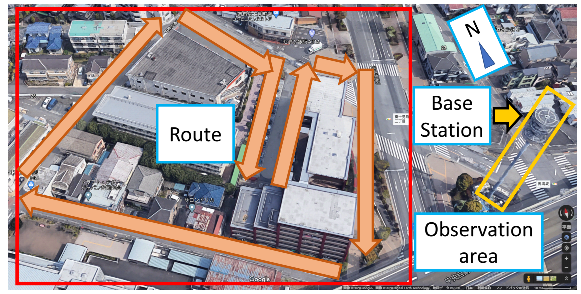

Fig. 4. Walking route.

Fig. 5. Antenna connected with the Area Tester.

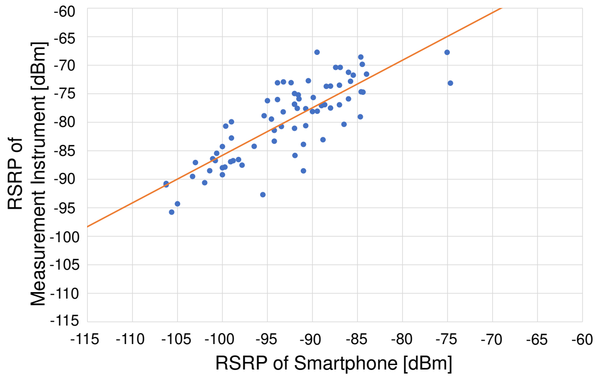

Fig. 6. Relationship between the RSRP of android terminals and Area Tester when we use Pixel3 in pocket.

B. Results

First, as an example, Fig. 6 shows the relationship between the RSRP of android terminals and the Area Tester when we use Pixel3 in the pocket and the regression line is inserted in the same graph. Note that the relationship between RSRP when Essential Phone PH-1 put in the bag is not manage in this research because of nothing of the mesh IDs on the smartphone and on the Area Tester match.

Next, the REM of the ground-truth is shown in Fig. 7. Besides, REMs of RSRP when Pixel 3 places in the pocket is shown in Figs. 8 to 11. These REMs show distribution of the observed value of Pixel3, the predicted value, the error between the ground-truth and the observed value of Pixel3, and the error between ground-truth and the predicted value.

Compared to Figs. 8 and 9, Fig. 9 is more similar to Fig. 7 than Fig. 8. So, we confirm error reduction by simple linear regression. Besides, compared to Fig. 10, average RSRP per mesh of Fig. 11 approaches to 0 [dB]. So, this indicates the usefulness of compensation by the proposed method. The right and left end in Fig. 10 are close to each other at about -20 [dB], but Fig. 11 shows that there are differences in the degree of compensation. This is because in Fig. 7, differences between the right and left end values are about 5 [dB], so it is considered that the difference appeared during the subtraction to find the error.

To evaluate the accuracy of the proposed method, we output histograms of the error between ground-truth and the observed value of Pixle3 and the error between ground-truth and the predicted value as Fig. 12 and Fig. 13. Comparing Fig. 12 and Fig. 13, we can find that errors of RSRP decrease and the average value approaches to 0 [dB]. Here, there is no RSRP in the range from -5 to -10 [dB] in the histogram before compensation, but in the histogram after compensation, errors were in the range from -5 to -10 [dB]. This means that the observed value is higher than the ground-truth. From these results, the proposed simple linear regression method can reduce the error, and we can understand more than 75% of errors between ground-truth and the predicted value is within ±5 [dB] for all working scenarios and smartphone types.

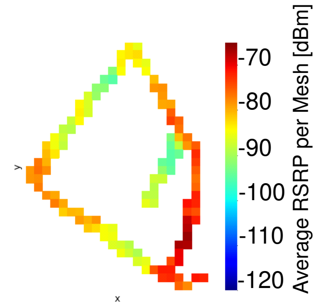

Fig. 7. RSRP of ground-truth.

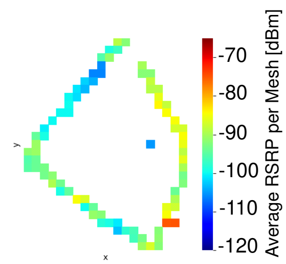

Fig. 8. RSRP of smartphome.

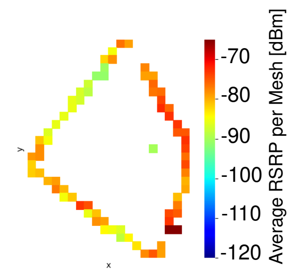

Fig. 9. RSRP of predicted value.

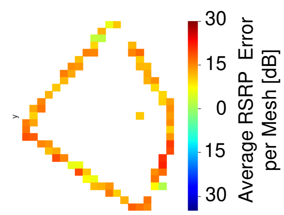

Fig. 10. The error between Fig. 7 and Fig. 8.

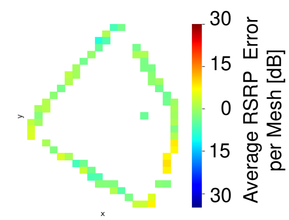

Fig. 11. The error between Fig. 7 and 9.

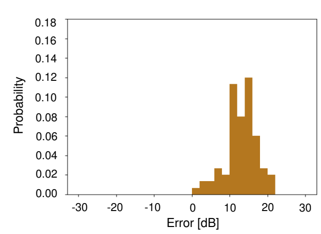

Fig. 12. The error between ground-truth and the observed value.

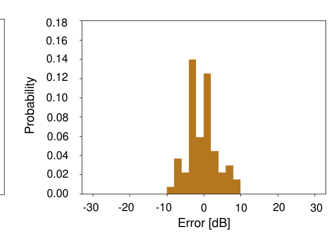

Fig. 13. The error between ground-truth and the predicted value.

Here, Table I shows statistical data on the error between ground-truth and the observed value of the smartphone. Also, from (6), if the error is a positive value, ground-truth is a higher value than the observed value of the smartphone. This indicates that the observation is performed properly because

TABLE I
STATISTICAL DATA ON THE ERROR BETWEEN GROUND-TRUTH AND THE OBSERVED VALUE OF SMARTPHONE.

Android terminal	Working scenario	Average [dB]	Average of absolute value [dB]	Median [dB]
Zenfone	In hand	5.0	5.1	4.8
	In pocket	7.9	7.9	7.9
	In bag	6.6	8.6	8.6
Pixel	In hand	14.5	6.7	14.5
	In pocket	13.0	13.0	13.1
	In bag	4.7	5.1	4.3
Essential	In hand	14.5	15.0	14.5
	In pocket	15.5	15.5	14.5

TABLE II
STATISTICAL DATA ON THE ERROR BETWEEN GROUND-TRUTH AND PREDICTED VALUE.

Android terminal	Working scenario	Average [dB]	Average of absolute value [dB]	Median [dB]
Zenfone	In hand	0.2	2.9	-0.5
	In pocket	-0.2	2.0	0.06
	In bag	1.2	3.0	-0.09
Pixel	In hand	0.4	3.1	-0.08
	In pocket	-0.4	3.1	-0.5
	In bag	1.3	3.4	0.4
Essential	In hand	0.9	3.3	0.01
	In pocket	0.01	3.1	0.7

Area testers have better quality sensor than the smartphones. On the other hand, if the error is a negative value, the observed value of the smartphone is a higher value than ground-truth. This indicates that radio waves may have been blocked by the effect of fading when we performed observation experiments with the Area Tester.

Table II shows statistical data on the error between ground-truth and the predicted value. From (6), if the error is a positive value, ground-truth is a higher value than the predicted value. This indicates that there is a high possibility that the compensation has been performed appropriately. On the other hand, if the error takes a negative value, the predicted value is a higher value than ground-truth. This indicates that the observed value was overcompensated, and if this result is reflected in REM, the smartphone may select the wrong parameter.

When considering the change from Tables I and II, the difference of the average value indicates the degree of compensation, the absolute value of the average value indicates the degree of compensation of the entire mesh, and the median indicates the degree of compensation of one mesh. From these results, the final average value of the error was within 1.5 [dB], and we can confirm that the error can be reduced and compensated for both the entire mesh and one mesh.

V. CONCLUSION

In this paper, we investigated radio observation and accuracy improvement of distribution of RSRP in the cellular environment in the actual environment using smartphones by crowdsensing. In the proposed method, the regression line and residuals were derived by simple linear regression, and we evaluated the accuracy of the residual of observed value using the histogram. From the evaluation results, we confirmed that 75% or more the error between the predicted value and ground-truth was distributed within only ± 5 [dB] for all android terminal and working scenarios. Then we can contribute the suppression of error of the observed value in crowdsensing. This contribution allows us to realize that the proposed method may have a high degree of application for estimation of the radio environment.

We plan to confirm the performance of regression analysis with the order of 2 or 3 considering new variables for future work.

ACKNOWLEDGMENT

This work was supported by the European Commission in the framework of the H2020-EUJ-02-2018 project 5GEnhance (Grant agreement no.815056) and the Ministry of Internal Affairs and Communications (MIC) of Japan (Grant no. JPJ000595).

REFERENCES

[1] Cisco, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Up-date, 2017-2022 White Paper,” [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.pdf, accessed June 17, 2020.
[2] S. Yrjola, P. Ahokangas, and M. Matinmikko, “Evaluation of recent spectrum sharing concepts from business model scalability point of view, ” in *Proc. IEEE Int. Symp. Dyn. Spectrum Access Netw,* pp. 241-250, Sept. 2015.
[3] J. Mitola, J. Maguire, G. Q. Maguire, “Cognitive radio: making software radios more personal, ” *IEEE Personal Communications,* vol. 6, no. 4, pp. 13-18, Aug. 1999.
[4] S. Sodagari, “A secure radio environment map database to share spectrum,” *IEEE J. Sel. Topics Signal Process,* vol. 9, no. 7, pp. 1298-1305, Oct. 2015.
[5] R. Hasegawa, T. Fujii, “Measurement-based spectrum database with mesh clustering,” *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN),* pp. 956-960, July 2017
[6] R. K. Ganti, F. Ye, and H. Lei, “Mobile crowdsensing: Current state and future challenges,” *IEEE Commun. Mag,* vol. 49, no. 11, pp. 32-39, Nov. 2011.
[7] I. Carreras, D. Miorandi, A. Tamilin, E. Ssebaggala, and N. Conci, “Crowd-sensing: Why context matters, ” in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.* Workshops, pp.368-371, Mar. 2013.
[8] C. Xiang, P. Yang, C. Tian, L. Zhang, H. Lin, F. Xiao, M. Zhang and Y. Liu, “CARM: crowd-sensing accurate outdoor RSS maps with error-prone smartphone measurements,” *IEEE Trans. Mobile Comput,* vol. 15, no.11, pp.2669-2681, Nov. 2016.
[9] M. Ito, K. Onose, K. Sato, K. Inage, T. Fujii, “A calibration method for crowdsensing based radio environment database,” *IEICE technical report,* SR2017-103, pp. 57-62, Jan. 2018.