Update and Compensation of Radio Map Based on Statistical Inference

Keita Katagiri

Department of Communication Engineering and Informatics
The University of Electro-Communications

This dissertation is submitted for the degree of
Doctor of Philosophy

March 2022

Supervisory Committee

Chairperson: Professor Takeo Fujii

1. Member: Professor Kenta Umebayashi

2. Member: Professor Koji Ishibashi

3. Member: Associate Professor Koichi Adachi

4. Member: Associate Professor Celimuge Wu

Day of the Predefense: November 9, 2021.

Day of the Final Defense: MM DD 2022.

Acknowledgements

This dissertation comprehensively describes the research contents of my doctoral study at The University of Electro-Communications, Tokyo, Japan. I am grateful to a large number of people who have helped me to accomplish my work.

First of all, I would like to express my deepest gratitude to my supervisor, Professor Takeo Fujii, for guiding my throughout my work. His motivations and encouragements immensely helped me throughout my life in The University of Electro-Communications. I also would like to give my special thanks to emeritus Professor Yasushi Yamao, Professor Koji Ishibashi, and Associate Professor Koichi Adachi for their valuable guidance, instructions and continuous support.

I also grateful to all the members of Advanced Wireless and Communication Research Center (AWCC), especially Assistant Professors Koya Sato and Kei Inage, Mrs. Koji Ichikawa, and Takanori Hara for their discussions and kindness.

Finally, I would like to thank for their words, attitudes, and every supports.

Abstract

The number of terminals has increased rapidly with the development of wireless technology. In future wireless communications, massive numbers of terminals will communicate with each other using finite frequency bands. However, the quality of communication in such an environment degrades because of various factors, such as packet collision and uncertainty of radio propagation characteristics. To guarantee the reliability of each wireless system, it is necessary to determine the communication parameters according to the transmission timing of a packet and surrounding radio propagation characteristics appropriately.

Several researchers have studied radio-propagation estimation to improve communication reliability. It has been reported that communication quality can be significantly improved by estimating radio propagation with high accuracy. Radio maps are used in various wireless systems as a radio propagation estimation method. Based on actual measurements, the radio map accumulates comprehensive statistical data of the radio environment, such as the average received signal power and the frequency band occupancy. The average received signal power is often used as typical information in radio map construction. The constructed radio map is accumulated in a database, spectrum database. This map enabled us to estimate the path loss and shadowing at each position skillfully. However, conventional radio maps store the average received signal power in each location. In such a simple accumulation, the amount of statistical data may be enormous if the communication area is wide. Consequently, the registered data size of the radio map increases. We considered reducing the data size as the first primary task in our dissertation. Conventional studies assume that massive received signal power samples are obtained to estimate the average power in each location accurately. It is difficult to collect such a large number of samples owing to the enormous measurement times. The minimum number of required samples should be utilized to construct the radio map efficiently. Thus, we considered the determination of the minimum required sample size for calculating the average received signal power value as the second primary task.

Conventional radio maps have primarily been constructed for fixed transmitter locations. In this environment, the average received signal power exhibits time-invariant characteristics. Assuming this characteristic, many researchers have created radio maps using mainly spatial interpolation. However, if the location of a transmitter varies with time, the received signal

power exhibits time-variant characteristics. In a time-variant environment, the average received signal power values before and after the movement of the transmitter may be significantly different. Thus, it is necessary to update the radio map based on something logical. We considered updating the radio map as the first application task in our dissertation.

In addition, if multiple transmitters operating at the same frequency exist near the communication area, a target signal may not be obtained depending on the signal-to-interference-plus-noise ratio owing to inter-transmitter interference. Although spatial interpolation can be used to estimate the missing data, it estimates the missing data inside the known data. If not all target signals can cannot be observed, the spatial interpolation may not be applicable; thus, a novel spatial extrapolation method is required, which is the second application task in our dissertation.

Finally, conventional radio maps cannot be used in a wireless system in which both the transmitter and receiver move dynamically. The system is known as an ad-hoc network. To promote the utilization of radio maps in various systems, it is necessary to apply the radio map to ad-hoc networks, in addition to the fixed transmitter environment. We considered the construction of radio maps in a dynamic environment as the third application task in our dissertation.

This dissertation discusses radio map construction for solving the aforementioned tasks: reducing the registered data size, determining the required sample size, updating the radio map, extrapolating the missing data, and constructing radio maps in ad-hoc networks. Although massive received signal power samples are required to solve each task perfectly, it is extremely difficult to obtain such massive samples owing to various issues, such as enormous measurement times. To solve the tasks efficiently using only limited samples, we propose several radio-map construction methods based on *statistical inference*. Statistical inference represents the population as various values (e.g., the mean) that are calculated using several samples. Because of statistical inference, radio maps can be skillfully created with low registered data sizes by considering several factors, such as the updating and extrapolation of the average received signal power.

In Chapter 2, we explain the main conventional study and the tasks for the practical realization of the radio maps. Subsequently, statistical inference is introduced to solve the the tasks. Finally, the organization of the dissertation is presented.

Chapter 3 proposes a shadowing classifier to reduce the registered data size of the radio map. The shadowing classifier was constructed by quantizing the measured shadowing values by a certain size. Thereafter, similar shadowing values were unified into a single value using the constructed classifier. The emulation results indicate that the proposed

shadowing classifier can reduce the registered data size of the radio map while maintaining high estimation accuracy.

Chapter 4 formulates the minimum required sample size to estimate the average received signal power. Sample size was defined as the number of instantaneous received signal power samples used to derive the average power. Here, three statistical methods: the confidence interval, central limit theorem, and *t*-test, are used to determine the sample size. The simulation results confirmed that the average received signal power could be accurately calculated using statistical methods.

Subsequently, Chapter 5 proposes a method for updating a radio map based on statistical hypothesis testing. This chapter assumes that the location of a transmitter varies according to the elapsed time. To update the radio map in such an environment, it is necessary to infer the significant difference between the two average received signal power values before and after transmitter movement. Therefore, we used the Welch's *t*-test, which is an improved version of the *t*-test. The proposed method updates the radio map if a significant difference is detected using the Welch's *t*-test. The first simulation results clarify the relationship between statistical power and sample size used in testing. The second results indicate that the proposed method can accurately update the radio map compared to the conventional updating methods.

In Chapter 6, an empirical extrapolation method is proposed to compensate for the missing received signal power values in a multiple-transmitter environment. The proposed method first estimates the empirical cumulative distribution function (CDF) using measured datasets. Thereafter, the histogram of the empirical CDF is compensated by considering the number of missing data. Finally, the median received signal power is extrapolated using the compensated empirical CDF. We used the 3.5GHz band datasets to verify the effectiveness of the proposed method. The emulation results indicate that the missing received signal power values can be accurately extrapolated using the proposed method compared with the conventional interpolation and extrapolation methods.

Chapter 7 applies a radio map to ad-hoc networks, where the proposed method first collects the received signal power samples in each transmitter and receiver pair. Subsequently, the radio map was constructed for each transmission position by averaging the datasets. To confirm the usefulness of the proposed method, we conducted a measurement campaign for vehicle-to-vehicle communications in Japan and the United States. The emulation results indicate that the average received signal power can be skillfully estimated using the radio maps compared with conventional path loss models. Additionally, the packet delivery rate was derived in real environments through performance evaluation.

Finally, Chapter 8 concludes the dissertation. We summarize the research content of each chapter and explain the future works of our dissertation.

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

a_1	accumulated data size of the proposed classifier and conventional clustering
B	path loss offset
C	path loss index
d_m	link distance between a transmitter and the m-th mesh
d_{w}	degree of freedom in the non-central t distribution
e_{sme}	signed mean error
F	fading value in compuer simulation
f_0	center frequency of each transmitter in interference-limited observations
H	number of mobile terminals in computer simulation
H_0	null hypothesis of Welch's t-test
H_1	alternative hypothesis of Welch's t-test
h_{Rx}	receiver height including antenna height
h_{Tx}	transmitter height including antenna height
$I_{\mathrm{sum},m}$	aggregate interference power in the m-th mesh
I_{d}	desired confidence interval
K	number of quantized shadowing values
L_{fspl}	free space path loss

M	number of meshes
N_0	noise floor
n_{ini}	number of samples in each mesh of the initial radio map
N_m	number of instantaneous samples in the m-th mesh
n_{new}	number of new samples
N_{Tx}	number of radio maps in MANETs
O	number of meshes where have known average value
$\hat{P}$	median path loss
$\bar{P}_m$	average received signal power in the m-th mesh
$P_{m,i}$	i-th instantaneous received signal power in the m-th mesh
P_{Tx}	transmission power
Q_k	k-th instantaneous target power
R	rice factor
R_{b}	break point
R_i	number of added data points in the i-th bin
$\boldsymbol{R}_m$	time-series data in the m-th mesh
$S^2_{0,\mathrm{w}}$	unbiased sample variance of $\boldsymbol{W}_0$
S_{I}	number of interfering transmitters
$\hat{s}_k$	quantized shadowing in the k-th model
s_m	non-quantized shadowing in the m-th mesh
T	time length
t	t-value
V	number of datasets in MI
v	degree of freedom

W	shadowing value in computer simulation
w	class width
$\boldsymbol{W}_0, \boldsymbol{W}_1$	two datasets for Welch's t-test
w_{c}	quantization size of shadowing
(x_m, y_m)	two-dimensional coordinates
$\boldsymbol{x}_{\mathrm{Rx}}$	location of mobile terminal
$\boldsymbol{x}'_{\mathrm{Tx}}$	location of transmitter in phase 1
$\boldsymbol{x}_{\mathrm{Tx}}$	location of transmitter in phase 0
$\bar{X}_{\mathrm{w}}$	sample mean of $\boldsymbol{W}_0$
$\bar{Y}_{\mathrm{w}}$	sample mean of $\boldsymbol{W}_1$
Z_α	100α percentile of the standard normal distribution
z_k	latent variable of Gaussian mixture model

Greek Symbols

α	significance level
α_1, α_2	path loss indexes in MANETs
β_1, β_2	path loss offsets in MANETs
β	type II error
δ	difference between $\mu_{0,\mathrm{w}}$ and $\mu_{1,\mathrm{w}}$
η	forgetting factor in deletion of samples
$\Gamma(\cdot)$	gamma function
$\gamma_{g,m}$	instantaneous SINR for the g-th instantaneous target power in the m-th mesh
γ_{th}	SINR threshold
λ	wavelength
λ_{w}	non-centrality parameter of the non-central t distribution

μ_{Lag}	Lagrange multiplier
μ_1 and μ_2	population means for explanation of t-test
$\boldsymbol{\mu}_k$	centroid (mean) vector of the k-th cluster
ω	integral variable
ω_o	weight of Kriging in the 0-th mesh
π_k	mixture coefficient of the k-th normal distribution
$\Psi(\cdot)$	CDF of the standard normal distribution
ψ_{Lag}	objective function of Lagrange multiplier method
$\rho_{i,j}$	correlation coefficient of shadowing between the i-th and j-th terminals
$\sigma_{i,\mathrm{s}}$	standard deviation of the shadowing from the transmitter to the i-th terminal
$\chi(Q_k)$	an indicator function
τ	time index
θ_{n}^2	nugget
θ_{r}	range
θ_{sl}^2	sill
ξ	semivariogram
$\zeta_k^{(t_{\mathrm{em}})}$	posterior distribution of z_k in the t_{em}-th iteration

Acronyms / Abbreviations

AP	Access Point
BHT	Bayesian Hypothesis Testing
CDF	Cumulative Distribution Function
i.i.d.	Independent and Identically Distributed
CLT	Central Limit Theorem
CSV	Comma-Separated Values

D2D	Device-to-Device
EM	Expectation-Maximization
IQR	InterQuartile Range
FCC	Federal Communications Commission
GMM	Gaussian Mixture Model
GPS	Global Positioning System
IDW	Inverse Distance Weighted
KPCA	Kernel Principal Component Analysis
LOS	Line Of Sight
MAC	Media Access Control
MAE	Mean Absolute Error
MANETs	Mobile Ad-hoc NETworks
MIC	Ministry of Internal Affairs and Communications
MI	Multiple Imputation
ML	Machine Learning
NLOS	Non-Line-Of-Sight
OFDM	Orthogonal Frequency Division Multiplexing
PDF	Probability Density Function
PDR	Packet Delivery Rate
PHP	Hypertext Preprocessor
PU	Primary User
RB	Resource Block
RMSE	Root Mean Squared Error
SHT	Statistical Hypothesis Testing

SINR	Signal-to-Interference-plus-Noise Ratio
SIR	Signal-to-Interference power Ratio
SI	Single Imputation
SME	Signed Mean Error
SNR	Signal-to-Noise Ratio
SU	Secondary User
V2V	Vehicle-to-Vehicle
WLAN	Wireless Local Area Network

Chapter 1

Introduction

Here, we introduce our dissertation. The following sections first explain the allocation of frequency bands and its problems. Thereafter, the radio propagation characteristics and their estimation methods are described. Moreover, a radio map that accumulates the statistical information of radio propagation characteristics, such as the average received signal power values, is introduced. Finally, a spectrum database that stores the constructed radio maps is presented.

1.1 Allocation of Frequency Bands and Crucial Problems in Wireless Communications

With the rapid development of wireless technology, the number of wireless devices such as smartphones and tablets is increasing. Fig. 1.1 shows the number of devices in the entire world, as estimated by Cisco [1]. These results demonstrate that the number of wireless devices increases linearly by a few billion units every year. If the number of devices continues to increase, a massive number of devices will communicate with each other in various environments. Currently, wireless communication is essential in daily life.

Unfortunately, the available frequency bands are finite in wireless communication. The ministry of internal affairs and communications (MIC) published the spectrum charts that represent the current frequency allocation status in Japan [3]. Moreover, national telecommunications and information administration presents the frequency allocation chart of the United States [4]. Both charts show that most frequency bands have already been allocated to existing systems. In particular, because the frequency band between 30 [MHz] and 3 [GHz] is frequently used in wireless communication systems, very few bands are available in this range. In general, frequency bands can be classified into licensed and unlicensed. Licensed

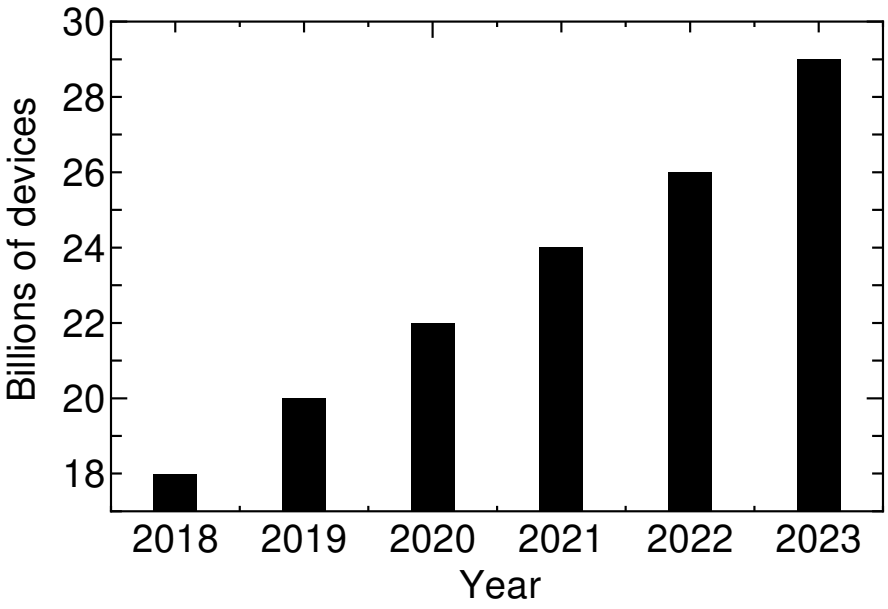


Fig. 1.1 The number of devices in the entire world estimated by Cisco [1].

bands are exclusively allocated to a specific wireless system, that is, a primary user (PU). However, unlicensed bands are often used in multiple wireless systems. For instance, the 920 MHz band is assigned to a low-power wide area [5], such as LoRa and Sigfox. In addition, 2.4GHz and 5GHz bands are allocated to wireless local area networks (WLANs) [6]. Because the unlicensed bands are much fewer than the licensed bands, current wireless systems need to exclusively use the licensed bands. Owing to this exclusive usage, the shortage of available frequency bands is an extreme problem in wireless communication. In this dissertation, we considered the use of the licensed bands.

Licensed bands are not always used by PUs. For instance, [7] showed that television networks contain temporally and spatially unused bands. Fig. 1.2 shows an overview of the two-dimensional spatial distribution of usage of licensed bands. Each PU uses a licensed band within a gray-colored two-dimensional space. Moreover, licensed bands are not used in other areas, which are referred to as white spaces. If a user other than PUs, a secondary user (SU), can utilize the licensed band within the white space, the utilization efficiency of the spatial frequency can be improved. The secondary usage of white space is known as spectrum-sharing [8]. In spectrum-sharing, the SU must avoid interference with the PU. Here, interference is defined as the signal strength of the SU received by the PU from the SU. Although the SU can utilize a larger white space by increasing its transmission power, it introduces more harmful interference to the PU. The interference strength depends on various radio propagation characteristics such as path loss. The SU should accurately estimate the

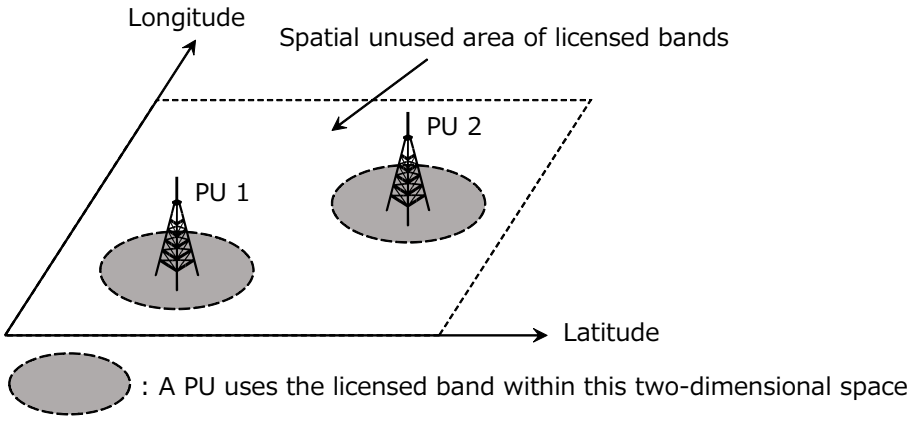


Fig. 1.2 Spatial unused space of licensed bands.

radio propagation characteristics in advance to enhance the frequency utilization efficiency while satisfying the interference avoidance constraints.

1.2 Relation Between Frequency Utilization Efficiency and Radio Propagation Estimation

Here, we first explain radio propagation characteristics. Thereafter, three typical radio propagation factors are presented.

- Path loss: This is the distance-dependent radio propagation characteristic. The path loss is calculated deterministically according to several parameters, such as the Euclidean distance between the transmission and reception positions. Several path loss models, including theoretical-based and empirical-based methods, have been constructed [9]. The theoretical method is based on geometric analysis, such as calculation of the effective area of a reception antenna. Moreover, massive observation data were statistically processed to create a path loss model using the empirical method. Each method can roughly estimate the global radio propagation characteristics. Thus, path loss is a fundamental factor in the radio propagation field.

- Shadowing: This phenomenon occurs when the propagation path is obstructed by structures around a transmitter and a receiver. Its occurrence depends on geographical conditions, such as the number of structures. As a typical value, [10] described that the shadowing value may vary every few meters up to several hundreds of meters. Because it is difficult to investigate the shadowing values for all types of structures, a measurement campaign was conducted in several representative environments. The

results indicate that the shadowing is empirically modeled as a log-normal distribution with a logarithmic mean of 0 [dB]. The standard deviation is determined based on the communication environment. This is generally approximately 4-13 [dB].

- Multipath fading: This is the instantaneous fluctuation of the received signal power owing to the scatters around a receiver. If there are several scatters in the communication area, multiple signals reach the receiver with a time lag. Under this condition, the received signal power fluctuates significantly by approximately 20–30 [dB] at the receiver. Because the fluctuation occurs for approximately milliseconds, it is difficult to estimate the multipath fading realizations perfectly.

In the radio propagation field, the received signal power is typically estimated by considering the above three factors. Path loss and shadowing can be accurately determined according to the relative positions of the transmission and reception positions.

In addition, the relationship between the spatial frequency utilization efficiency and radio propagation estimation is described in [11]. In [11], spectrum sharing between a single PU and a single SU was considered. The SU determines its transmission power such that the desired signal-to-interference power ratio (SIR) at the coverage edge of the PU is satisfied. Two methods were considered for the transmission power design. The first method estimates the path loss and shadowing at the coverage edge of the PU. The second method calculates only the path loss. Here, it is assumed that the path loss and shadowing of the SU are known. Each method determines the transmission power of the SU such that the desired SIR is probabilistically satisfied. The simulation results indicated that the first method could increase the transmission power of the SU by 12 [dB] while accurately guaranteeing the desired SIR, as compared with the second method. If the transmission power of the SU increases, the SU can use a larger white space; thus, the spatial frequency utilization efficiency can be improved. Because the first method can skillfully estimate the white space by considering path loss and shadowing, the transmission power of the SU can be increased. In addition to the transmission power evaluation, the root mean squared error (RMSE) between the true received signal power of the PU and the estimated power was calculated. The evaluation results indicated that the RMSE is decreased small by estimating the shadowing in addition to the path loss. In summary, the spatial frequency utilization efficiency can be enhanced by accurately estimating radio propagation characteristics. Therefore, this dissertation aims to estimate the radio propagation skillfully.

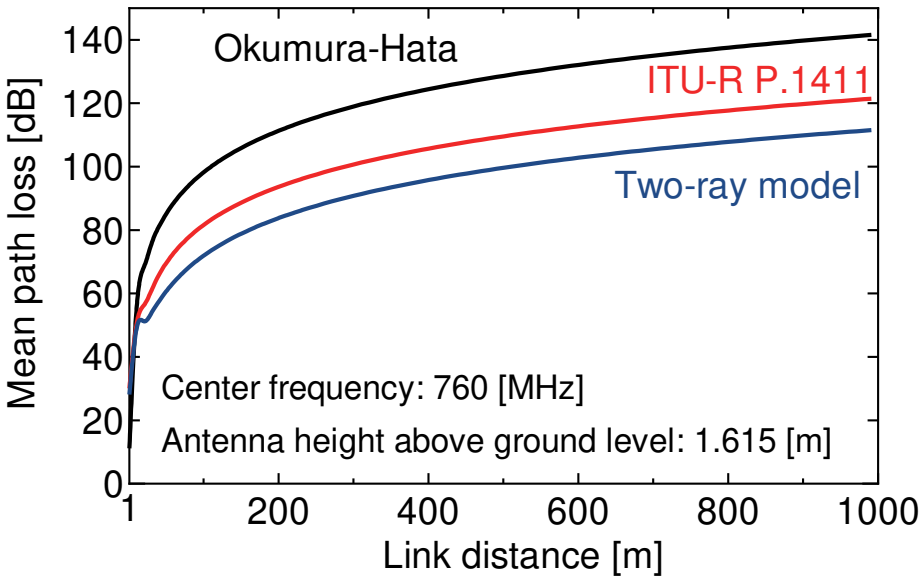


Fig. 1.3 Examples of the mean path loss values.

1.3 Radio Propagation Estimation Methods and the Focus of this Dissertation

This section explains the radio-propagation estimation methods. First, path-loss-based methods and their problems are shown using existing path-loss models. Subsequently, the modeling of the shadowing and the multipath fading is described. Finally, a measurement-based method utilizing actual observational data is explained.

1.3.1 Path Loss-Based Methods

Conventional studies have estimated the received signal power using path loss models, such as the Okumura–Hata model [12], ITU-R P.1411 model [13], and two-ray model [14]. As the first problem of path-loss-based methods, the accuracy of the calculated path loss value significantly depends on the path loss model used. Fig. 1.3 shows the examples of the mean path loss values in several models. In the figure, the horizontal axis denotes the Euclidean distance between the transmission and the reception positions, and the vertical axis represents the mean path loss value in each model. Here, the center frequency was 760 [MHz], and the antenna heights above the ground level of the transmitter and receiver were 1.615 [m]. As examples of path-loss-based methods, the Okumura–Hata, ITU-R P.1411, and two-ray models are used. These results indicate that the mean path loss differs significantly depending

on the path loss model. For instance, the difference between the Okumura–Hata and ITU-R P.1411 models was approximately 20 [dB]. If the SU designs its transmission power based on the overestimated path-loss value, the interference power is underestimated compared to the actual power. Consequently, harmful interference may occur at the PU. Therefore, an appropriate path loss model should be used to estimate the mean path loss with high accuracy. Thus, each system must understand the principles of each path loss model in advance.

As the second problem of path-loss models, the estimation accuracy of the received signal power may be low owing to the shadowing and multipath fading in a specific environment. [9] reported that the estimation accuracy of the received signal power may be at best 8 [dB] in any path loss model.

1.3.2 Modeling of Shadowing and Multipath Fading Based on Existing Probability Distribution

Shadowing and multipath fading fluctuate locally depending on geographical conditions, such as the number of buildings. It is difficult to obtain all realizations of these two factors in various environments. Thus, shadowing and multipath fading were modeled using the existing probability distributions. For instance, shadowing is empirically modeled as a log-normal distribution [15]. Furthermore, multipath fading is theoretically modeled as a Rayleigh distribution and a Nakagami–Rice distribution. Even if these models are used to estimate the shadowing and multipath fading, instantaneous realization at an arbitrary position may not be accurately estimated because the probability distribution only represents the histogram of realizations.

1.3.3 Measurement-Based Method

The path loss-based methods and existing probability distributions can not estimate a shadowing and fading realizations in each location; thus, the SU must suppress its transmission power excessively to avoid interference with the PU. This suppression may deteriorate the efficiency of spatial frequency utilization. The SU should accurately estimate not only the path loss but also the shadowing and the fading in each location.

Measurement-based methods have attracted attention for improving the estimation accuracy of the received signal power [16, 17]. This method collects radio environment information such as the received signal power from actual observations in a realistic environment. Thereafter, the observational data are processed using statistical methods, and representative values, such as the mean and median, are calculated. The representative values

were used to estimate the received signal power at an arbitrary position. Actual observation data can include the shadowing effects owing to various structures. This shadowing estimation is the most advantage of the measurement-based methods compared to the path loss-based methods. A well-known tool that accumulates representative values is the radio (environment) map [18–22]. This dissertation treats the radio map as a radio-propagation estimation method. The next section provides an overview of the radio map.

1.4 Radio Map

The definition of the radio map varies depending on the researchers and estimation target. As described in Sect. 1.2, this dissertation aims to estimate the radio propagation characteristics accurately. Therefore, we assume that the radio map accumulates the average received signal power values as the representative values. The following sections first explain the construction and utilization of the radio map. Subsequently, the assumptions in the radio map construction are presented.

1.4.1 Crowdsourcing

During the construction of the radio map, instantaneous received signal power samples are required to calculate the average values. The spectrum analyzer enables precise observation of the instantaneous received signal power because of its high observation capability. However, the production cost of the spectrum analyzer is extremely high (e.g., a few million); that is, the spectrum analyzer has not yet become popular among ordinary people. Therefore, a new measurement device is required to obtain the received signal power samples at a low production cost.

Many researchers have employed smartphones as the measurement device [23]. Although smartphones are not measurement devices for observing the received signal power, they can be useful because of the following reasons:

- The prices of smartphones are lower compared to those of the spectrum analyzers.
- Many ordinary people have used smartphones in their daily lives for various purposes. Thus, massive received signal power samples can be obtained if smartphones are used as the measurement devices.

Owing to the above reasons, smartphones may be utilized for radio map construction. This method is called *crowdsourcing* [24–28]. In crowdsourcing, a sensing program is

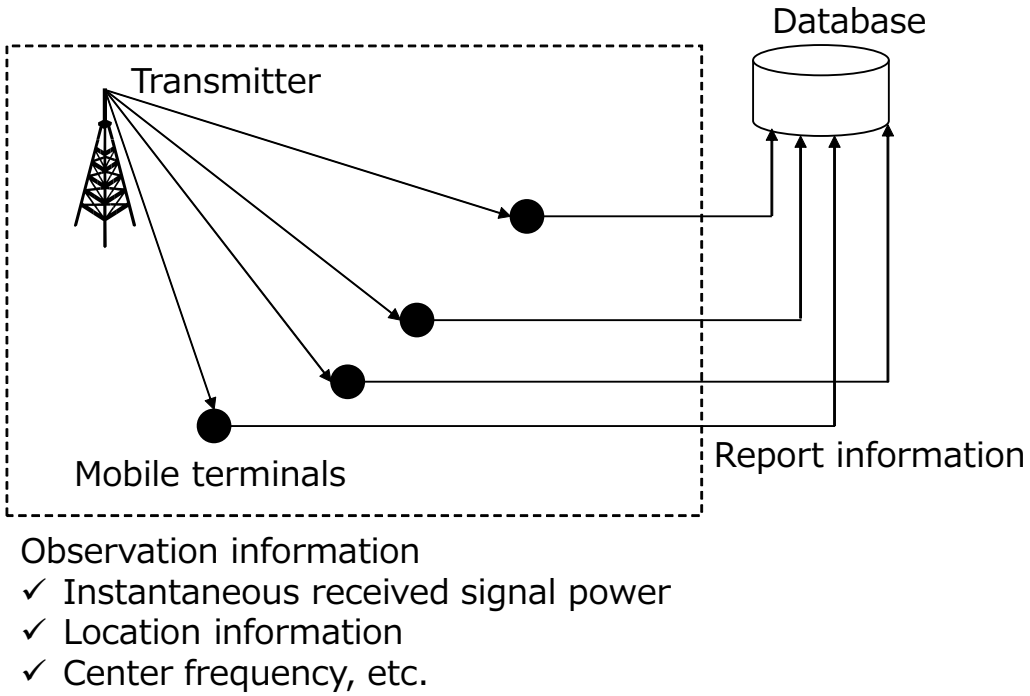


Fig. 1.4 Construction of radio map.

implemented on a smartphone to observe the received signal power. Although crowdsourcing-assisted radio map construction has not been realized owing to several issues, we believe that this method will be used in future wireless communications. Several issues are discussed in Chapter 8.

1.4.2 Measurements

Fig. 1.4 shows the crowdsourcing-assisted radio map construction. We consider the downlink communications in a fixed transmitter, such as the cellular systems. A transmitter exists in the communication area and it sends a signal, including own identification to each location. Mobile terminals, such as smartphones, receive the transmitted signals in each position and record instantaneous received signal power values in association with the location information. This dissertation considers that the location information can be obtained from Global Positioning System (GPS). Mobile terminals that implemented the sensing program enable us to record the location information easily; however, the measurement accuracy is low owing to simple specifications. Thus, the measured latitude and longitude values should be corrected using some method, such as the Quasi-Zenith satellite system [29].

In addition to the instantaneous received signal power and location information, center frequency, the reception time, and transmitter identification were recorded at each terminal. At the end of the day, each terminal reports the measured dataset to the database. The database accumulates the reported datasets on a database server, such as MySQL. The registration contents of the database server are explained in Sect. 3.2. After accumulating the reported datasets, the database statistically processed these datasets to create a radio map.

1.4.3 Mesh Definitions

As described at the beginning of this chapter, the task of the radio map is to estimate the average received signal power. Because the instantaneous received signal power contains a multipath fading factor, the average received signal power may not be accurately known. The database first calculates a mesh based on the latitude and longitude to remove the fading effects. Subsequently, the average received signal power is derived using the instantaneous received signal power samples of each mesh. A mesh is an equally divided geographical space. As an alternative expression, the grid is well known by several researchers. This dissertation uses the expression 'mesh'.

A mesh code is often used in radio map construction to identify each mesh. [30] defined the mesh code as an extension of the geographic information system, which is based on grid square code defined by the Japanese Industrial Standard X 0410 standard [2]. Table 1.1 lists the definitions of mesh codes. The primary mesh code is a 4-digit number calculated based on the integer parts of the latitude and longitude as,

$$1\text{st\_code} = \lfloor L_1 \times 1.5 \rfloor \times 100 + (L_2 - 100), \tag{1.1}$$

where L_1 [deg] and L_2 [deg] are the latitude and longitude, respectively. The secondary mesh code is a 2-digit number obtained by equally dividing the primary mesh into the latitude and longitude directions, that is, dividing the primary mesh into 64 secondary meshes. The third mesh code is a 2-digit number obtained by equally dividing the secondary mesh in the latitude and longitude directions, that is, the secondary mesh is divided into 100 tertiary meshes. The quaternary and quinary mesh codes are two-digit numbers obtained by dividing the previous mesh into 100 equal parts. Each mesh code is linked to a hyphen and managed in the database. For instance, the quinary mesh can be converted to "5538-23-54-15-76" as shown in Fig. 1.5.

Table 1.1 Definitions of the mesh codes [2].

Mesh type	Descriptions	Approximate size of one side
Primary	It consist of each longitude degrees, even latitude degrees, and three equally divided even latitude degrees, based on longitude 100 degrees east and latitude 0 degrees north.	80 [km]
Secondary	An area formed by eight equally divided primary mesh sections in the latitude and longitude directions.	10 [km]
Tertiary	An area formed by ten equally divided secondary mesh sections in the latitude and longitude directions.	1 [km]
1/10 subdivided mesh	An area formed by ten equally divided tertiary mesh sections in the latitude and longitude directions.	100 [m]
Ten m mesh	An area formed by 10 equally divided mesh sections in the latitude and longitude directions.	10 [m]

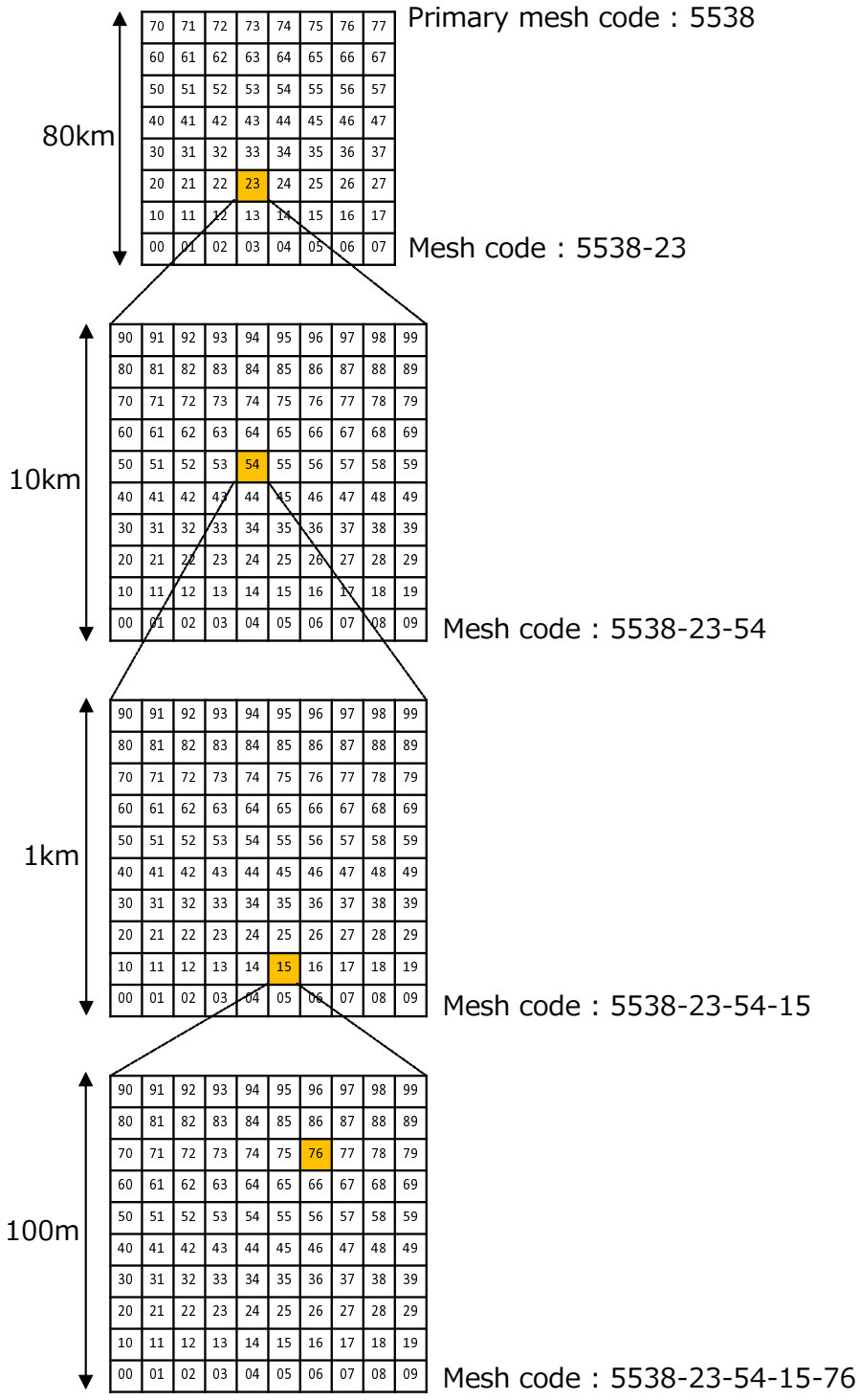


Fig. 1.5 Overview of meshes.

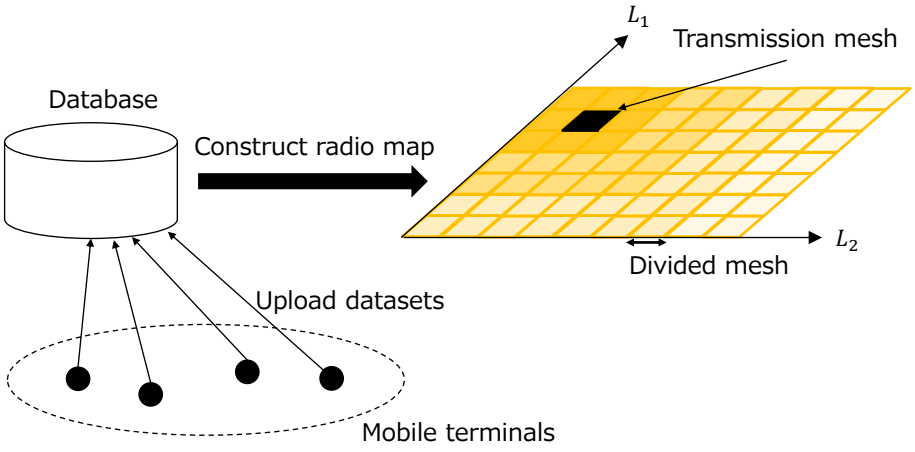


Fig. 1.6 Image of the radio map.

1.4.4 Statistical Processing on Database

The database first assigns a mesh code to the instantaneous received signal power. During this process, the mesh size is determined by the manager of the database. Subsequently, the average received signal power is derived for each divided mesh according to the following equation:

$$\bar{P}_m = \frac{1}{N_m} \sum_{i=0}^{N_m-1} P_{m,i} \ [\mathrm{mW}], \tag{1.2}$$

where $\bar{P}_m$ denotes the average received signal power in the m-th mesh, $P_{m,i}$ [mW] denotes the i-th instantaneous received signal power in the m-th mesh, and N_m denotes the number of instantaneous samples in the m-th mesh.

In general, the average value is almost equivalent to the population mean value by increasing N_m owing to the law of large numbers. However, if there are outliers in a small number of samples, the estimation accuracy of the average value may degrade. This is because the effect of outliers is highly sensitive in the milliwatt unit. There are several methods for eliminating the impact of outliers using statistical processing. First, we can use the median value because it is unaffected by outliers. Second, the effect of outliers was suppressed by calculating the average value in the logarithmic domain. This is because the number of digits can be further reduced in the logarithmic domain than in the milliwatt domain.

Fig. 1.6 shows the image of the radio map construction. The black mesh denotes the transmission mesh where the transmitter is located. The transmission mesh can be derived

using the location information of the transmitter. The surrounding meshes express the average received signal power calculated according to Eq. (1.2). The darker meshes around the transmitter denote the larger average value. Moreover, the average value decreases in the area edge.

By utilizing the radio map, we can estimate the two radio propagation characteristics: path loss and shadowing. As described in Sect. 1.2, the path loss is uniquely determined depending on the Euclidean distance between the transmitter and a receiver. Thus, the average received signal power contains the path-loss effect in each mesh. Conversely, the shadowing estimation accuracy varies according to the relationship between the mesh size and the shadowing correlation distance. The correlation distance is defined such that the autocorrelation of shadowing is 0.5. [31] reported that the distance is approximately 20 [m] in an urban area. If the mesh size is smaller than this distance, the shadowing realizations may be similar in a mesh. Under this condition, the shadowing effect can be estimated precisely by averaging the instantaneous received signal power samples. Because the number of meshes increases as the communication area becomes wider or the mesh size becomes small, the registered data size may be large. Moreover, the estimation accuracy of the average value may be degraded owing to fluctuations in the path loss and shadowing in a large mesh. There is a trade-off between the registered data size and estimation accuracy.

1.4.5 Utilization Procedures

Before transmitting a signal, the transmitter accesses the database via a communication line such as Wi-Fi to obtain the radio map information. The database searches for radio map information and provides it to the transmitter. The transmitter can understand path loss and shadowing effects with high accuracy using the radio map. The communication parameters (e.g., modulation format and transmission power) can be appropriately determined according to the average received signal power. For instance, if the average value is smaller compared to the desired received signal power value, the transmitter uses lower modulation formats to guarantee communication quality. Additionally, the transmitter can suppress the transmission power when the average received signal power is large to improve power efficiency.

If mobile terminals need to estimate the average received signal power in downlink communications, the database provides the radio map information to each terminal. Each terminal can determine the signal strength using the radio map information.

In spectrum sharing, if the radio map is created for a PU, the database provides the radio map information to the SU. The SU obtains the average received signal power value in each mesh from the radio map and calculates the SIR. The SU can appropriately determine its transmission power based on the permissible SIR. If the calculated SIR is smaller than the

permissible SIR, the SU must suppress its transmission power to satisfy the permissible SIR. Conversely, if the calculated SIR is larger than the permissible value, the SU can increase its own transmission power while guaranteeing the permissible SIR.

1.4.6 Application Examples

Radio maps are typically used to estimate television white space [32, 33]. [17] clarified that the estimation accuracy of the received signal power over the TV network can be achieved at 3.1–5.0 [dB] by the radio map. These values are superior to 8 [dB], which is the accuracy of the path loss-based methods [9]. Based on the skillful accuracy, several researchers have used radio maps for various applications, such as spectrum sharing [11], spectrum sensing [34], localization [35], determining the location of unmanned aerial vehicle [36], and frequency allocation [37]. These studies have revealed that the radio map enables us to precisely estimate the received signal power, as compared to the path loss-based methods.

1.4.7 Conventional Studies of Radio Maps and Main Mathematical Tool of this Dissertation

In radio map construction, the following assumptions are considered.

- Massive instantaneous received signal power samples are stored in each mesh to estimate the average value based on the law of large numbers accurately. Many researchers consider this assumption reasonable in crowdsourcing-assisted radio map construction because ordinary people can collect many samples.
- The location of the transmitter and geographical conditions, such as the number of buildings, do not vary before and after the radio map construction. That is, the actual average value in each mesh is time invariant.
- There are no instantaneous received signal power samples that are less than the noise floor of mobile terminals. As described in Sect. 1.4.2, only one transmitter sends a signal. Thus, no instantaneous sample is missing owing to the interference signals transmitted by other transmitters with the same frequency band.
- The downlink communications are performed between a fixed transmitter and mobile terminals. Hence, a radio map was not created between terminals.

Several researchers have continuously studied the construction methods of the accurate radio map by considering that the above assumptions do not hold. There are various methods, such

as the machine learning (ML) and the Bayesian inference. For instance, [38] has proposed the neural network-based method to construct radio maps between each terminal. Although this method can accurately estimate the average received signal power, the estimation accuracy depends on various hyperparameters, such as the number of epochs.

[39] has proposed the radio map construction using Bayesian inference. Bayesian-based method estimates population parameters by modeling the prior distribution based on some prior information. If the prior information, such as the variation of geographical conditions, can be known in advance, the average received signal power may be accurately estimated by using Bayesian inference. In the Bayesian inference, the estimation accuracy is significantly different according to the modeled prior distribution. The inaccurate radio map may be created owing to the inappropriate prior distribution.

We use *statistical inference* as the main mathematical tool to simply construct accurate radio maps. The statistical inference does not require complicated calculation and tuning the hyperparameters. Additionally, the prior information and modeling of the prior distribution is not necessary in statistical inference. Motivated by these facts, we will introduce statistical inference in Sect. 2.3.1, and explain the statistical inference-based radio map construction in Chapters 3–7.

1.5 Hierarchical Architecture of Spectrum Database

The constructed radio maps were stored as statistical tools in the spectrum database. The construction range of the radio map varies according to the size of the area, such as a city or town. We consider a hierarchical spectrum database [40] to manage the radio maps efficiently as shown in Fig. 1.7. The spectrum database consists of several hierarchies based on the area types. In the following section, we describe each database.

1.5.1 Regulation Database

As the first layer, the regulation database is installed using a spectrum regulator, such as federal communications commission (FCC) and MIC. This database is considered the highest hierarchy in this dissertation and it summarizes the radio map information reported from each global database. The FCC and MIC access the regulation database and research radio map information if the SU requires the secondary use of licensed bands. Using the regulation database, the average received signal power can be estimated for each local area or global area. Each institution then provides the radio map information for each area to the SU. The SU determines its own transmission power based on the procedures described in Sect. 1.4.5.

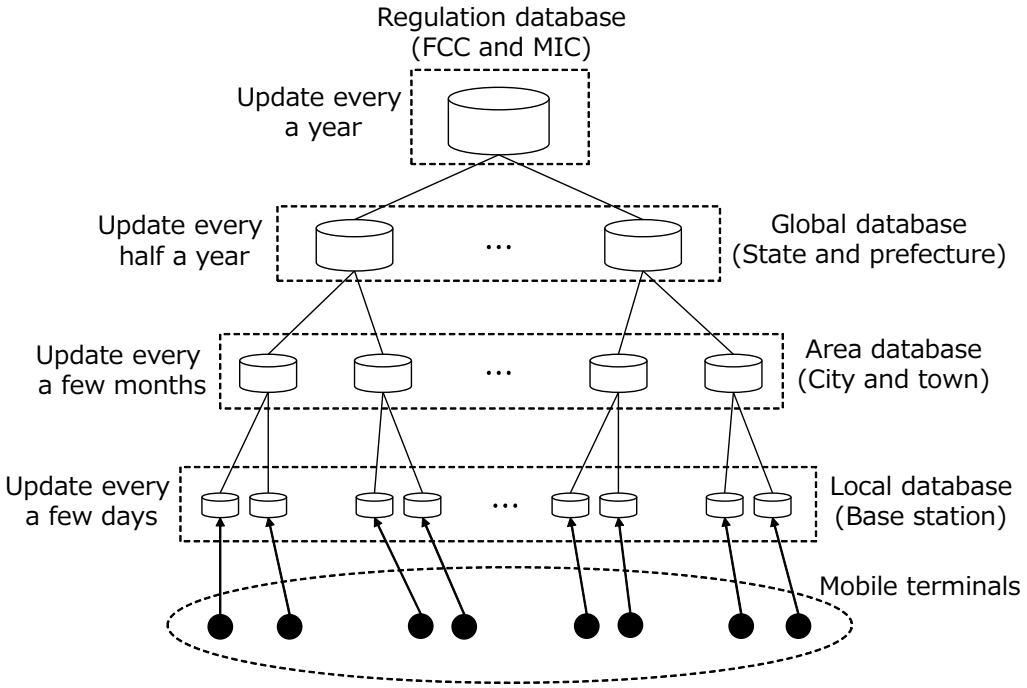


Fig. 1.7 Hierarchical architecture of spectrum database.

When the institution needs to know the average received signal power value in a few days, the manager of the database in each hierarchy sends the mean value to the regulation database. Each institution provides the required report cycle to the a manager of each hierarchy database.

1.5.2 Global Database

In the second layer, the constructed radio maps are accumulated in a global database. This database manages radio map information in each global area, such as the state in the US and the prefecture in Japan. The map information was uploaded from an area database. The manager of the global database regularly reports radio-map information to the regulation database to reflect fluctuations in the average received signal power.

1.5.3 Area and Local Databases

We considered an area database managed in each city or town and a local database to accumulate more detailed radio map information. In the former database, the radio map information of each local database is summarized. The latter database stores the measured datasets observed by the mobile terminals in each location. This database corresponds to a base station and a roadside unit. Using such a hierarchical architecture, a large number of

datasets can be efficiently summarized according to the area range. Notably, this dissertation considered a radio map in the local database.

1.6 Chapter Summary

This chapter introduces this dissertation. A shortage of frequency bands is observed. Subsequently, the relation between the frequency utilization efficiency and the radio propagation estimation is described. Finally, we introduce the radio map and spectrum database.

Chapter 2

Main Conventional Study and Tasks for Practical Realization of Radio Maps

Here, we describe the main conventional study and the remaining tasks of radio maps. Subsequently, this chapter introduces statistical inference to solve the remaining tasks. Finally, the organization of the dissertation is presented.

2.1 Main Conventional Study of Radio Maps

2.1.1 Spatial Interpolation-Based Methods

Conventional radio-map works have mainly studied spatial interpolation [41–46]. Instantaneous received signal power samples may not be obtained at several meshes because of the long measurement cycle of the mobile terminals. Although this issue can be solved by increasing the number of terminals, enormous production costs may be incurred depending on the number of terminals.

Therefore, many researchers have continuously discussed the spatial interpolation of radio maps for a fixed transmitter location (e.g., the cellular system). Spatial interpolation estimates unknown data using several known data. There are several interpolation methods, such as the inverse distance weighted (IDW) method [47] and spline interpolation [48]. In spatial interpolation, an unknown average received signal power value is interpolated by weighted averaging as follows:

$$\hat{P}_0 = \sum_{m=0}^{N_{\text{inter}}-1} w_m 10 \log_{10}(\bar{P}_m) \ \text{[dBm]}, \tag{2.1}$$

where $\hat{P}_0$ is the interpolated average value and w_m is the weighting factor of the known average received signal power in the m-th mesh. Notably, each known average power is a logarithmic value because the interpolation accuracy may be degraded owing to the outliers in the milliwatt domain. Spatial interpolation assumes that an unknown value exists within the known values. That is, the unknown value is larger than the minimum known value and smaller than the maximum known value. Additionally, spatial interpolation assumes that the spatial correlation is high among neighborhood meshes. In radio propagation, spatial correlation means that several received signal power values may be similar in the nearby meshes. Based on this property, each weight is determined according to the Euclidean distance between a mesh with an unknown value and that with a known value. The calculation procedures differed among the interpolation methods. For instance, IDW is the simplest method that uniquely determines a weight based on the reciprocal of the Euclidean distance. By considering the Euclidean distance, the path-loss effect can be considered in the interpolation. However, this method does not consider the shadowing effect in the weight calculation; thus, interpolation accuracy may be degraded in a realistic environment.

2.1.2 Spatial Correlation of Shadowing and Kriging Interpolation

Before explaining interpolation, we describe the spatial correlation of shadowing. Each weight should be determined by considering the spatial correlation of the shadowing values at different meshes to improve the interpolation accuracy. Shadowing correlation has been known for a long time. The exponential decay model is the most famous for shadowing correlations [31]. The model was constructed experimentally in a real environment. The autocorrelation of shadowing decays almost exponentially with the moving distance of the receiver. That is, several shadowing values may be similar in neighborhood meshes. In such a spatially correlated random field, the unknown average received signal power values can be precisely estimated using the Kriging interpolation [17]. The Kriging interpolation determines w_m by considering the spatial correlation of shadowing; therefore, the error between a true and an interpolated value is minimized. Although several methods have been constructed for the Kriging interpolation, ordinary Kriging is often used because of its simple modeling of the spatial covariance structure. This method estimates the spatial covariance structure based only on the Euclidean distance between data points. Using such an estimation, the path loss effect can be considered for determining each weight. Many researchers have confirmed that accurate radio maps can be constructed using the Kriging interpolation [17, 49, 11]. The details of the ordinary Kriging are shown in Sect. 6.5.1.

2.2 Tasks for Practical Realization of Radio Maps

2.2.1 Primary Tasks

First, we describe the two primary tasks of the radio map. As described in Sect. 1.4.4, the radio map accumulates the average received signal power in each mesh. In such a simple structure, the registered data size may be large, according to the mesh size and coverage of the transmitter. In the crowdsourcing-assisted radio map construction, massive meshes are stored over a wide area, such as prefectures or states. Conventional studies have not sufficiently studied this problem; thus, we consider the following content as the first primary task of the radio map:

- The novel method for reducing the registered data size of the radio map should be considered without compromising the high estimation accuracy.

As described in Sect. 1.4.7, enormous instantaneous samples are used to calculate the average received signal power in each mesh skillfully. If the variance in the received signal power values is small in a mesh, such massive samples may not be necessary. Because several costs, such as the observation period, may increase to obtain an enormous number of samples, the number of samples should be kept to the minimum. Hence, the following content is considered the second primary task of the radio map:

- The minimum required sample size should be determined for accurately estimating the average received signal power in each mesh.

2.2.2 Application Task 1

As explained in Sect. 1.4.7, conventional works assume that the location of the transmitter and the geographical conditions do not vary before and after the radio map construction. Under these conditions, an accurate radio map can be constructed using only an appropriate spatial interpolation method. Moreover, if the location of the transmitter varies according to the elapsed time, the average received signal power exhibits time-variant characteristics in each mesh. As a result, an initial radio map constructed before the transmitter moves may provide outdated information. Even if conventional interpolation-based methods are used, an accurate radio map cannot be constructed owing to fluctuations in radio propagation. That is, the use of spatial interpolation may be limited. Thus, we consider the following content as the first application task of the radio map:

- A radio map must be updated in the time-variant radio propagation environment. The necessity of updating should be judged correctly using a logical method.

2.2.3 Application Task 2

Conventional works assume that only one transmitter exists in the radio map construction as described in Sect. 1.4.7. Under this assumption, there are no missing received signal power samples owing to the interference from the other transmitters. Conversely, if multiple transmitters operating in the same frequency band exist near the communication area, all instantaneous samples that are less than a certain SIR cannot be obtained. Thus, unknown samples exist outside the known samples. This missing phenomenon differs from the assumption of the spatial interpolation. Hence, this dissertation considers the following content as the second application task of the radio map:

- A novel extrapolation method is necessary to create an accurate radio map. Specifically, the radio map should be empirically extrapolated without using the parametric approaches by considering site-specific fluctuations in the radio propagation.

2.2.4 Application Task 3

Finally, the third application task is described here. Conventional studies consider downlink communications between a fixed transmitter and mobile terminals. With the development of wireless technology, mobile terminals communicate with each other while dynamically moving their communication area. Hence, a large number of communication links are created between multiple transmitters and receivers. Such communication systems are known as mobile ad-hoc networks (MANETs). Vehicle-to-vehicle (V2V) communication is an example of MANETs. In MANETs, the instantaneous received signal power may fluctuate more intensely than in a fixed transmitter environment because of the dynamic movement of terminals. Because conventional methods do not construct a radio map in MANETs, this dissertation determines the third application task as follows:

- How to create the radio map in MANETs by considering multiple communication links?

2.3 Statistical Inference-based Radio Maps Construction

To solve the primary and application tasks, this dissertation utilizes *statistical inference*. Statistical inference is a statistic, and its basic principles have been established by *Sir Ronald Aylmer Fisher*. The following sections first provide an overview of the statistical inference. Subsequently, the motivation for using statistical inference is described. Finally, we explain the relationship between the statistical inference and each task.

2.3.1 Overview of Statistical Inference

Statistics aims to estimate population characteristics based on samples obtained using various methods. Descriptive statistics has long been known as one of these methods. The basic idea of descriptive statistics was proposed by Karl Pearson between the late 1800s and the 1900s. Descriptive statistics attempt to visualize the features of the obtained samples by calculating representative values, such as the mean, median, and mode. However, it is difficult to calculate representative values for all samples for a large population size.

To solve the problem of descriptive statistics, the statistical inference has been considered by various statisticians. Fig. 2.1 shows the overview of the statistical inference. The left circle denotes the population with massive samples. As shown in the right circle, several samples were extracted from the population, and the statistical characteristics of the population were estimated using the extracted samples. Statistical inference can be divided into statistical estimation and statistical hypothesis testing (SHT). In statistical estimation, the population parameters were estimated by calculating representative values using the extracted samples. Furthermore, the statistical estimation consisted of point and interval estimations. Point estimation infers a population parameter by calculating only one representative value. In the interval estimation, the population parameter was estimated within a certain range. SHT assesses two mutually exclusive hypotheses regarding the properties of a population using the extracted samples. For instance, hypothesis testing can infer a significant difference between the two population means. In summary, population characteristics are generally represented by certain values, such as the mean, in the statistical inference. The population parameter can be estimated efficiently, with low observation costs, using the statistical inference.

2.3.2 Relation between Statistical Inference and Each Task

Finally, we explain the motivation for using statistical inference to solve each primary and application task. For the first primary task, we considered reducing the registered data size for each mesh. As described in Sect. 2.1.2, neighboring meshes may exhibit similar shadowing values owing to the spatial correlation. Hence, the registered data size can be reduced by representing similar shadowing values as a single value using the statistical inference.

For the second primary task, we determine the minimum sample size required to estimate the average power in each mesh accurately. First, the true average received signal power was calculated using the massive received signal power samples in each mesh. Thereafter, we calculated the error between the true and estimated power values by changing the sample size to derive the estimated power. If the permissible error is satisfied for a specific sample size, we set the sample size to the minimum required size. The observation and emulation times

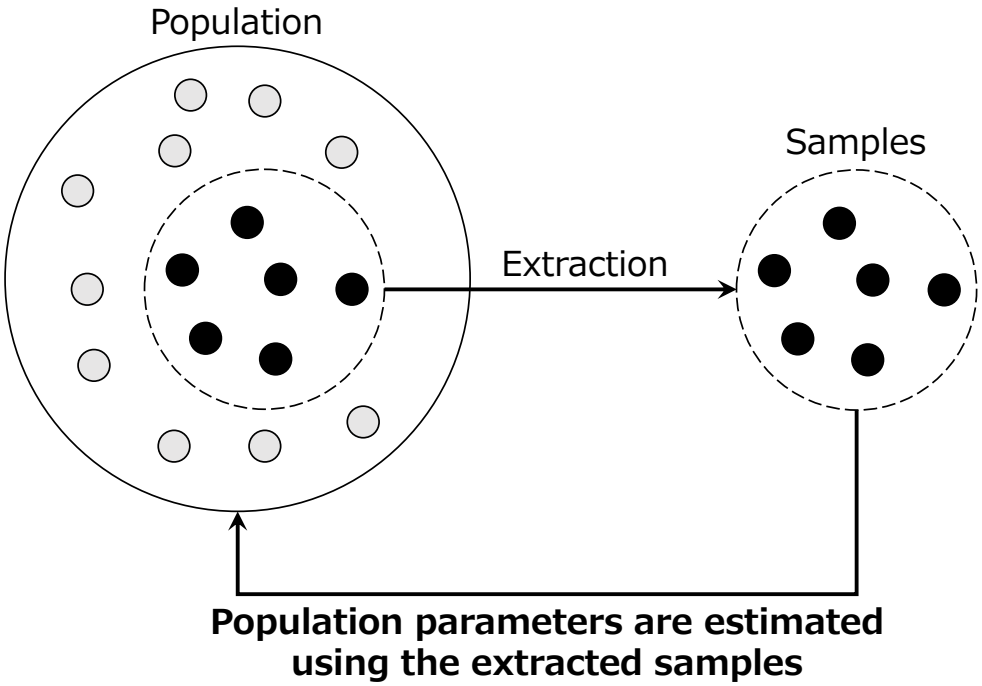


Fig. 2.1 Overview of statistical inference.

in this evaluation depend on the number of meshes. Thus, the true average power should first be represented as the estimated average power calculated using only a few samples. The required sample size is then determined to satisfy the permissible estimation error. When only a few samples are used for statistical inference, a large number of samples are not necessary for each mesh.

In the first application task, if the transmitter moves in the communication area, the path loss and shadowing may be significantly changed in several meshes. Hence, the movement of the transmitter can be detected by referring to the variation in the average received signal power using massive samples in each mesh. If the number of meshes is large, it is difficult to obtain massive samples in each mesh owing to the increase in the observation time. SHT can solve this problem because the variation in the true average power can be inferred using only a few samples.

The second application task aims to extrapolate the missing received signal power samples in multiple transmitter environments to construct an accurate radio map. In multiple-transmitter environments, there are no received signal power samples below a certain SIR owing to interference from the other transmitters; thus, not all missing samples can be perfectly extrapolated. By knowing the number of missing samples, we can construct a

histogram of the received signal power values. Once the histogram is created, we use the median received signal power obtained by statistical inference as the representative value of the missing samples from the collected histogram.

Additionally, we describe the relationship between the statistical inference and the third application task. In MANETs, because the transmitter and the receiver move dynamically, massive received signal power samples are required in each transmitter and receiver pair to estimate the radio propagation characteristics accurately. This observation is very difficult because of various costs such as enormous measurement times. As described in Sect. 1.2, the path loss and the shadowing may be uniquely determined according to the positional relation between a transmitter and a receiver. Based on this property, we calculate the average received signal power using several samples for each transmission and reception mesh pair. Thereafter, the average power is utilized as the representative value for efficiently estimating radio propagation.

2.4 Organization

Fig. 2.2 shows the overview of each chapter in this dissertation. Chapters. 3 and 4 focus on the reduction of the registered data size in the database. Chapter 5 considers the situation that the location of the transmitter varies. In this case, we deal with updating the radio map. Chapter 6 discusses the extrapolation of the radio map in multiple transmitter environments. Finally, Chapter 7 applies the radio map to ad-hoc networks. Here, we consider the downlink communications in Chapters 3, 4, 5, and 6. Notably, as described in Sect. 2.1, most conventional studies of the radio map only consider the spatial interpolation in the fixed transmitter location. Thus, the novelties of our researches are shown in the following by the *italic fonts*.

Chapter 3: Shadowing Classifier for Radio Map

To reduce the registered data size of the radio map, this chapter proposes a shadowing classifier. The classifier was constructed by quantizing shadowing realizations. Subsequently, a quantized value is assigned based on the proposed objective function. The proposed classifier can unify similar shadowing realizations into a single value. The emulation results using the measured datasets show that the proposed method can achieve a high estimation accuracy with a small registered data size. *The novelty of this chapter is to propose the shadowing classifier for creating the low storage radio map that has not been considered in most conventional works. In particular, the proposed classifier can classify shadowing values based on a new objective function that differs from conventional clustering methods, such as k-means++.*

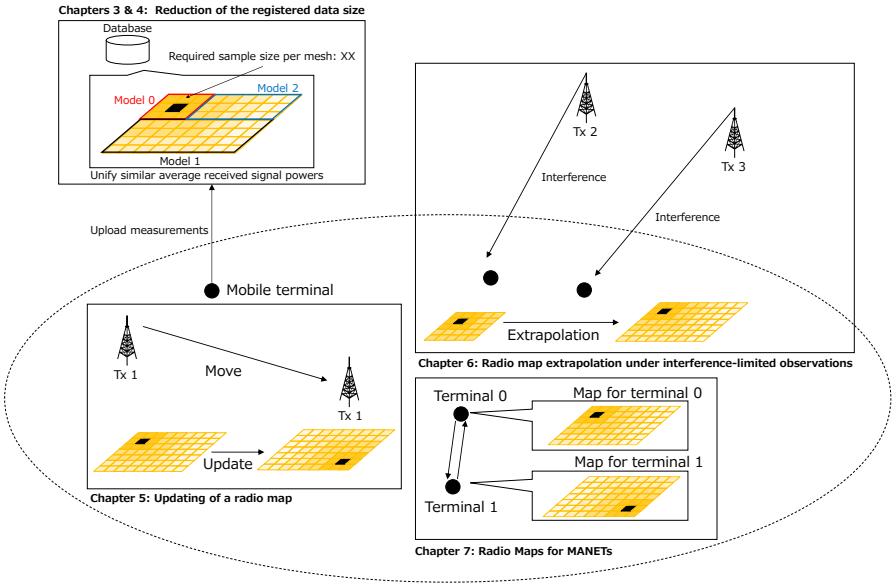


Fig. 2.2 Overview of each chapter in this dissertation.

Chapter 4: Sample Size Formulation using Statistical Inference for Radio Map

Here, we formulate the minimum required sample size to calculate the average received signal power in the radio map accurately. We utilize three statistical methods: the confidence interval, central limit theorem (CLT), and *t*-test. The simulation results indicate that the proposed formulation can significantly reduce the sample size for calculating the average power while estimating radio propagation with high accuracy. *The main contribution of this research is to formulate the required sample size for accurately estimating the average received signal power. Most conventional researchers have constructed radio maps based on spatial interpolation by assuming the law of large numbers for each average power value. That is, it was assumed that a large number of samples could be obtained. We do not consider this assumption in this chapter.*

Chapter 5: Radio Map Updating using the Welch's *t*-test

This chapter proposes a hypothesis-testing-based method as one of the solutions for radio map updating. Based on our investigations, we used the Welch's *t*-test to infer a significant difference between the two mean values. The proposed method first models the measurements as the time-series data for hypothesis testing. Thereafter, a theoretical analysis of the Welch's *t*-test was conducted to formulate statistical power. The relationship between the sample size and statistical power can be determined via computer simulations. Additionally, we performed radio map updating using the Welch's *t*-test in the simulation. The results indicate

that the radio map can be skillfully updated, even if the location of the transmitter varies. *The conventional works of the radio map have considered spatial interpolation in the static environment where the transmitter is fixed. Thus, the main novelty of this chapter is the skillful creation of a radio map based on the update of the average power, even if the transmitter moves. This update cannot be realized using conventional interpolation-based methods.*

Chapter 6: Radio Map Extrapolation under Interference-Limited Observations

Here, we propose the extrapolation of the radio map under interference-limited observations. Multiple transmitters operating at the same frequency were assumed here. In this situation, the target received signal power may be missing at several points where interference from the surrounding transmitters is dominant. To extrapolate the missing data, we first compensated for the empirical CDF of the target received signal power by considering the number of missing data. Thereafter, the median value is recalculated using the compensated CDF. The emulation results based on the measured datasets over 3.5GHz indicate that the extrapolation accuracy of the proposed method is superior to that of the conventional interpolation and extrapolation methods. *The main contribution of this chapter is to propose the spatial extrapolation method because most conventional works have only considered spatial interpolation. Although the conventional spatial interpolation cannot precisely compensate for missing data in interference-limited observations, the proposed extrapolation method can accurately estimate the missing data.*

Chapter 7: Crowdsourcing-Assisted Radio Maps for MANETs

We apply the radio map for MANETs in which both the transmitter and receiver move dynamically. The proposed method accumulates instantaneous received signal power samples in each transmitter and receiver pair. In the database, these datasets were statistically processed and a radio map was created for each transmission position. Two types of measured datasets were used to confirm the effectiveness of the proposed method. The evaluation results indicate that the average received signal power can be accurately estimated, as compared to the path loss-based methods, even in MANETs. In addition, we clarified that the packet loss rate is significantly different according to the communication environment. *The usage of radio maps in MANETs is the main novelty of this chapter because conventional works have only applied the radio map in the fixed transmitter environment, such as the cellular systems. Additionally, the effectiveness of radio maps can be clarified through measurement campaigns rather than simulation results. This clarification is the main contribution of our study.*

2.5 Chapter Summary

This chapter first explains the main conventional study of radio maps. Conventional methods continuously use spatial interpolation to create radio maps. After the five remaining tasks of the radio maps were described, statistical inference was introduced to solve them. Statistical inference may enable us to solve each task efficiently without observing massive received signal power samples.

Chapter 3

Shadowing Classifier for Radio Map

The radio map accumulates statistical information of the radio propagation in each mesh. However, the number of statistical information may be large according to the mesh size and area range. This chapter considers the reduction of the accumulated data size of the radio map.

In this chapter, we first explain the background of the proposed method. Subsequently, the shadowing classifier is proposed to unify similar average received signal power values. By the proposed classifier, the accumulated data size can be reduced while accurately estimating the radio propagation. We show the usefulness of the proposed method using the 3.5GHz band datasets observed in the real environment.

3.1 Background

The conventional radio map stores the average received signal power in each mesh, here, the average value is represented as a floating-point type. Thus, even if average values are similar in several meshes, the unique average value is registered in each mesh. As a result, the accumulated data size becomes large. Although several databases, such as SQL server, can reduce the registered data size by compressing the accumulated data, the compression-based method requires to execute the query when terminals access to the database. If the massive number of terminals use the radio map, the compression time may increase. In several use case, such as spectrum sharing, terminal may need to obtain radio map information rapidly for estimating the radio propagation in real time. In such the situation, the compression-based method may be not used owing to large compression time.

As a simple reduction method of the data size, we can utilize path loss models [50, 51]. Path loss models are divided into an empirical type (e.g., Okumura-Hata model) and a theoretical type (e.g., the free-space path loss model). Both models use simple parameters,

such as the communication distance, the center frequency, and the antenna height. By appropriately selecting these parameters, the median path loss can be estimated with high accuracy. Especially, the measurement-based path loss model based on the observation datasets is superior to an empirical model and a theoretical one [52]. When using these path loss models, the database calculates the communication distance between a transmitter and a receiver and estimates the median path loss. Hence, the accumulated data size can be drastically reduced because the local database does not need to register the average received signal power in each mesh.

However, it has been reported that the estimation accuracy of the radio propagation is limited around 8 [dB] in the path loss model because of the shadowing deviation [9]. To estimate the shadowing in addition to the path loss, we need to construct the radio map using a small mesh size that is less than the correlation distance of the shadowing. Obviously, there is a trade-off between the estimation accuracy and the mesh size. Under the trade-off, we should reduce the accumulated data size of the radio map while accurately forecasting the location-dependent radio propagation.

In radio propagation, the spatial correlation of the shadowing has been empirically found. [31] has clarified that the correlation value of the shadowing exponentially decreases depending on the moving distance of the receiver. This phenomenon implies that shadowing values are similar in several positions because the shadowing realization significantly changes whether the receiver is blocked by buildings. Based on these characteristics, we aim to unify the similar shadowing values to a single one by something algorithm for reducing the accumulated data size.

For the unification of similar shadowing values, we can accumulate each shadowing value in integer format. This method first truncates the shadowing value after the decimal point and registers the truncated value in each mesh. Then, we can reduce the accumulated data size by unifying the same truncated values to a single one. However, this method may not accurately estimate the detailed shadowing value owing to the accumulation in integer format.

Motivated by these facts, this chapter proposes the shadowing classifier. First, the proposed method collects the received signal power and reception location via a measurement campaign. After that, the local database constructs the radio map and the measurement-based path loss model for a moment using the measured datasets. To construct the shadowing classifier, we calculate the shadowing value in each mesh from the residual component between the average received signal power and the median path loss. Next, we generate the shadowing classifier by quantizing shadowing values with a certain size. The local database assigns the quantized shadowing value to each mesh. The proposed classifier enables us to unify similar shadowing values while considering the local variation of the radio

Table 3.1 Accumulation contents of the conventional radio map for fixed transmitter environment.

Item	Format	Size [Byte]
Measurement data and time	datetime	8
Latitude L_1	double	8
Longitude L_2	double	8
Altitude	double	8
Center frequency	double	8
Average received signal power	double	8
Mesh code (10m)	text	16

Table 3.2 Accumulation contents of the proposed classifier-based radio map for fixed transmitter environment.

(a) Mesh Table

Item	Format	Size [Byte]
Measurement data and time	datetime	8
Latitude L_1	double	8
Longitude L_2	double	8
Altitude	double	8
Center frequency	double	8
Shadowing label	int	4
Mesh code (10m)	text	16

(b) Shadowing Table

Item	Format	Size [Byte]
Shadowing label	int	4
Quantized shadowing	double	8

propagation. As a result, the accumulated data size can be significantly reduced with keeping the high estimation accuracy. To verify the effectiveness of the proposed classifier, we utilize measured datasets in cellular communications and vehicle-to-infrastructure communication. The emulation characteristics reveal that the proposed method is superior to the conventional radio map in terms of the accumulated data size.

3.2 Accumulation Contents of Radio Map

This section explains the accumulated contents of the radio map. Table 3.1 shows the accumulation contents of the conventional radio map for fixed transmitter environment. The

conventional radio map stores several items, such as the measurement date and time, the average received signal power, and mesh code. Considering the typical mesh size, 10m mesh code is utilized in my dissertation. It can be seen that the accumulated data size may be large because the local database registers the average received signal power as the double format.

To overcome this issue, the proposed method stores the shadowing label in integer format. Table 3.2 represents the accumulation contents of the proposed classifier-based radio map. The proposed method utilizes two tables; the mesh table and the shadowing one. The former table stores the shadowing label instead of the average received signal power as shown in Table 3.2(a). The shadowing label is used to identify the quantized shadowing value and these items are stored in the shadowing table represented in Table 3.2(b). Thus, the accumulated data size depends on the number of quantized shadowing values. It is necessary to quantize the shadowing more finely for accurately estimating the radio propagation; however, the accumulated data size is large. This trade-off is very important in the proposed classifier.

In an actual communication, the transmitter sends latitude and longitude of a location that exists the receiver to the local database. The local database finds the reception mesh corresponds to the reported reception latitude and longitude, and grasps the shadowing label accumulated in the mesh table. Then, the quantized shadowing related to the shadowing label is provided to the transmitter. Finally, the local database calculates the median path loss estimated via the linear regression and sends the value to the transmitter. The transmitter estimates the average received signal power from the sum of the median path loss and the quantized shadowing. Note that the estimation procedures of the median path loss will be shown in Sect. 3.3.2.

3.3 Shadowing Classifier

This section shows the overview of the shadowing classifier. We first explain the radio propagation model assumed in the proposed classifier-based radio map. Next, the generation procedures of the shadowing classifier and classification method of the quantized shadowing are represented.

3.3.1 Radio Propagation Model of Shadowing Classifier

We aim to reduce the accumulated data size while estimating the location-dependent radio propagation characteristics. Thus, the local database requires to quantize the shadowing with high accuracy and assign the quantized value to each mesh. The proposed classifier first generates K quantized shadowing values using the average received signal power and median

path loss. We define the k-th propagation model ($k = 0, 1, \cdots, K-1$) as follows:

$$\hat{P}_{\mathrm{c},k}(d_m) = B - 10C\log_{10}(d_m) + \hat{s}_k \quad [\mathrm{dBm}], \tag{3.1}$$

where $\hat{P}_{\mathrm{c},k}$ is the k-th propagation model in the proposed classifier, d_m [m] is the link distance between a transmitter and the m-th mesh, B [dBm] is the path loss offset that depends on the transmission power and antenna effects, C is the path loss index, and $\hat{s}_k$ [dB] is the k-th quantized shadowing value.

This propagation model consists of two factors: the path loss and the shadowing. In the proposed method, the local database constructs K models and stores it as the shadowing classifier. In the classifier-based radio map construction, the propagation model is assigned to each mesh based on the objective function. Each model has a different shadowing value that is why the proposed method is called a *shadowing classifier*. To generate the shadowing classifier, we need to calculate parameters B, C, and $\hat{s}_k$.

3.3.2 Estimation of B and C

We first explain the path loss estimation. In this dissertation, linear regression is utilized to estimate the path loss parameters. As the first step of the path loss estimation, the local database derives the average received signal power in each mesh using observation datasets. After calculating the link distance d_m, the local database generates the scatter plot as the horizontal axis being $\log_{10}(d_m)$ and the vertical axis being the average received signal power. Finally, we can obtain the estimated parameters B and C by fitting the average received signal power values to the first and second terms of Eq. (3.1).

The estimation accuracy of the path loss parameters depends on the number of meshes having observation datasets. The linear regression may not accurately derive the path loss parameters if there are few meshes. Thus, it is important to observe radio environment information in many positions.

If B and C are different in each mesh, the accumulated data size is large. Thus, we assume that B and C are constant values in all meshes. When it is necessary to consider the anisotropy of the path loss, we can compensate for the fluctuation of path loss by finer quantization of the shadowing value.

3.3.3 Estimation of $\hat{s}_k$

This subsection represents the modeling method of $\hat{s}_k$. We assume that M meshes ($m = 0, 1, \cdots, M-1$) are created by the local database using the mesh definitions. The local

database first estimates a non-quantized shadowing s_m in the m-th mesh as follows:

$$s_m = 10\log_{10}(\bar{P}_m) - \hat{P}(d_m) \quad [\mathrm{dB}], \tag{3.2}$$

where $\hat{P}(d_m) = B - 10C\log_{10}(d_m)$ [dBm] is the median path loss in the m-th mesh. The local database calculates the above equation for each mesh and generates the shadowing vector $\boldsymbol{s} = (s_0, s_1, \cdots, s_{M-1})$.

The local database then finds the maximum shadowing $s_{\max} = \max(\boldsymbol{s})$ [dB] and the minimum shadowing $s_{\min} = \min(\boldsymbol{s})$ [dB], where $\max(\boldsymbol{s})$ and $\min(\boldsymbol{s})$ find the maximum and minimum values from the vector $\boldsymbol{s}$. After that, the quantization is performed between $[s_{\min}, s_{\max}]$ using a quantization size w_{c} [dB]. We can express the k-th quantized shadowing value $\hat{s}_k$ as follows,

$$\hat{s}_k = s_{\min} + w_{\mathrm{c}}k \quad [\mathrm{dB}]. \tag{3.3}$$

3.3.4 Classification

The proposed classifier assigns the k-th propagation model to a mesh. It is assumed that N_m instantaneous received signal power samples $\boldsymbol{P}_m = (P_{m,0}, P_{m,1}, \cdots, P_{m,N_m-1})$ are accumulated in the mesh, here, $\boldsymbol{P}_m$ is the instantaneous received signal power vector in the m-th mesh. We are familiar that there is generally a high spatial correlation of the shadowing within the small scale mesh; however, the shadowing value may be significantly different in several meshes if the geographical conditions, such as the number of buildings, notably vary. Considering this property, we use the RMSE as the objective function for accurately classifying the propagation model. The local database stores the shadowing label k_m in the m-th mesh based on the following function:

$$k_m = \underset{k=0,1,\cdots,K-1}{\arg\min} \sqrt{\frac{1}{N_m}\sum_{j=0}^{N_m-1}\left[P_{m,j,\mathrm{dB}} - \hat{P}_{\mathrm{c},k}(d_m)\right]^2}, \tag{3.4}$$

where $P_{m,j,\mathrm{dB}} = 10\log_{10}(P_{m,j})$ [dBm]. This function enables us to accurately classify the quantized shadowing value in each mesh considering the instantaneous fluctuation of the received signal power. The local database accumulates the shadowing label k and quantized shadowing value on the shadowing table.

Fig. 3.1 shows the concept of the shadowing classifier. The local database has the shadowing classifier that consists of K propagation models. The black circle is an instantaneous received signal power in a mesh. Each colored line represents the average received signal power estimated by the shadowing classifier. Based on the objective function, we find the

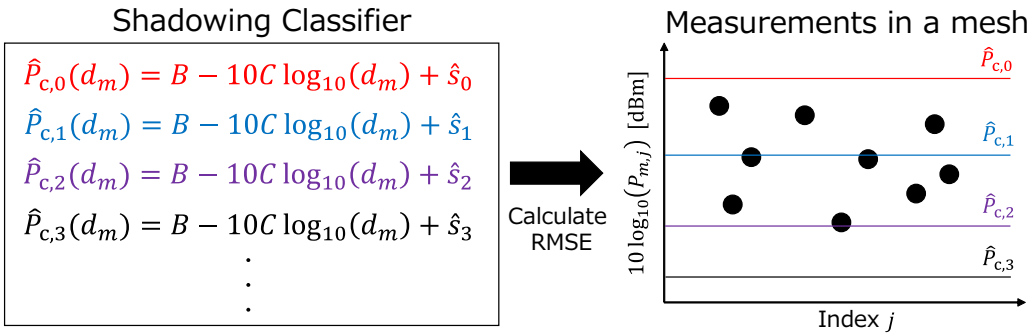


Fig. 3.1 Concept of the shadowing classifier.

propagation model having minimum RMSE. In this example, the RMSE can be minimized using the model $\hat{P}_{c,1}$; thus, the local database assigns this model to the mesh.

The originality of the proposed classifier is to classify propagation models by considering the instantaneous fluctuation of the received signal power. Conventional clustering methods that are described in the next section classify observation data based on the mean values; that is, the instantaneous fluctuation is not considered.

3.4 Comparison Methods of Shadowing Classification

If the shadowing exhibits high spatial correlation, the neighboring meshes may show the similar shadowing values. Based on this observation, we can utilize k-means++ [53] to unify similar shadowing values since it is assumed that several measurements are concentrically dense. This assumption matches the shadowing distribution on the two-dimensional meshes because of the spatial correlation.

Next, it is well known that the shadowing distribution is modeled as the log-normal distribution based on the empirical rule of thumb. Therefore, we utilize Gaussian Mixture Model (GMM) to cluster the shadowing values. This model assumes that an observation value follows the normal distribution; that is, the assumption is suitable in the shadowing classification.

To utilize the clustering algorithms, we first define the input data vector of the m-th mesh as follows:

$$\boldsymbol{x}_m = (x_m, y_m, s_m), \tag{3.5}$$

where (x_m, y_m) is the two-dimensional coordinates of the m-th mesh. These values are derived using the 10m mesh code.

(x_m, y_m) and s_m are the two-dimensional coordinates and logarithmic values; that is, these values are different scales. If the clustering is performed using these different scales, the accuracy of the clustering is degraded. Hence, we standardize each input data and multiply the weight before the classification. As the first step of the standardization, the local database first estimates the mean and standard deviation of each input data. After that, each mean value is subtracted from each value and divided by the standard deviation. These procedures are well known as the general standardization method in statistics. Note that B and C are constant in the k-means++ and GMM.

3.4.1 K-means++

The k-means may not accurately classify observation values if an Euclidean distance is short between initial centroids. K-means++ can solve this issue by determining initial centroids so that a Euclidean distance between each centroid is long. This method clusters each input data by calculating the following function E:

$$E = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} u_{mk} d_{\mathrm{e}}(\boldsymbol{x}_m, \boldsymbol{\mu}_k), \tag{3.6}$$

where k is the cluster label that corresponds to the shadowing label. K is the number of clusters and this is the same value as the number of quantized shadowing values in the proposed classifier. u_{mk} is the binary variable; that is, the value is 1 if $\boldsymbol{x}_m$ is included in k-th cluster; otherwise, 0. $\boldsymbol{\mu}_k$ is defined as the centroid vector of the k-th cluster. $d_{\mathrm{e}}(\boldsymbol{x}_m, \boldsymbol{\mu}_k)$ is an Euclidean distance between $\boldsymbol{x}_m$ and $\boldsymbol{\mu}_k$. k-means++ allocates the cluster label so that an input data vector belongs to the nearest cluster.

In the conventional k-means, the accuracy of the clustering degrades if the euclidean distance between initial centroids is too small. To overcome this issue, k-means++ selects each initial centroid as follows:

i). $\boldsymbol{x}_m$ is randomly selected as the first centroid $\boldsymbol{\mu}_1$ from the M input data vectors.

ii). The local database selects $\boldsymbol{x}'_m$ as another centroid $\boldsymbol{\mu}_k$ from the M input data vectors based on the following probability:

$$\frac{\min\limits_{0 \le k \le K-1} d_{\mathrm{e}}(\boldsymbol{x}'_m, \boldsymbol{\mu}_k)}{\sum_{m=0}^{M-1} \min\limits_{0 \le j \le K-1} d_{\mathrm{e}}(\boldsymbol{x}_m, \boldsymbol{\mu}_j)}. \tag{3.7}$$

iii). ii). is performed until K centroids are selected.

Using Eq. (3.7), the Euclidean distance may be large between centroids because a new centroid is selected relatively far from the existing centroids.

The shadowing values are classified into K clusters by k-means++. After linking the mesh index m and the cluster index k, the local database calculates the average shadowing value in each cluster. We accumulate the average value as $\hat{s}_k$ with the mesh index m. However, the performance of k-means++ may be low if several shadowing values are notably different from the mean values; that is, the outliers deteriorate the clustering accuracy. This is because k-means++ updates each centroid by calculating the mean value of input data.

3.4.2 GMM

GMM consists of the linear combination of several normal distributions. We can utilize this distribution as the clustering by estimating which normal distribution each data belongs to. We define the probability density function (PDF) of D-dimension normal distribution as follows:

$$p_{\mathrm{normal}}(\boldsymbol{X}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right), \tag{3.8}$$

where $\boldsymbol{X} = (X_1, \cdots, X_D)$ denotes the random variable vector, here, X_i the i-th random variable that follows the single normal distribution. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix of D-dimension normal distribution, respectively. The PDF of GMM is given by

$$p_{\mathrm{gmm}}(\boldsymbol{X}) = \sum_{k=0}^{K-1} \pi_k p_{\mathrm{normal}}(\boldsymbol{X}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{3.9}$$

where π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$ are the mixture coefficient, the mean vector, and the covariance matrix of the k-th normal distribution, respectively. Next, we define the log-likelihood function of GMM as follows:

$$\log p_{\mathrm{gmm}}(\boldsymbol{\chi}) = \log \prod_{i=0}^{N_{\mathrm{gmm}}-1} p_{\mathrm{gmm}}(\boldsymbol{X}_i) = \sum_{i=0}^{N_{\mathrm{gmm}}-1} \log \sum_{k=0}^{K-1} \pi_k p_{\mathrm{normal}}(\boldsymbol{X}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{3.10}$$

where $\boldsymbol{X}_i$ ($i = 0, \cdots, N_{\mathrm{gmm}} - 1$) is the i-th data vector that follows GMM, $\boldsymbol{\chi} = \{\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_{N_{\mathrm{gmm}}-1}\}$ and N_{gmm} is the number of data.

To maximize the above log-likelihood function, we use expectation-maximization (EM) algorithm and the latent variable z_k. z_k is 1 if $\boldsymbol{X}_i$ is subject to the k-th normal model; otherwise, 0. The procedures of EM algorithm are defined as follows:

i). We calculate the posterior distribution of z_k as follows (E-step):

$$\zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i) = \frac{\pi_k^{(t_{\mathrm{em}})} p_{\mathrm{normal}}\left[\boldsymbol{X}_i|\boldsymbol{\theta}_k^{(t_{\mathrm{em}}-1)}\right]}{\sum_{j=0}^{K-1} \pi_j^{(t_{\mathrm{em}})} p_{\mathrm{normal}}\left[\boldsymbol{X}_i|\boldsymbol{\theta}_j^{(t_{\mathrm{em}}-1)}\right]}, \tag{3.11}$$

where t_{em} is the iteration index, $\zeta_k^{(t_{\mathrm{em}})}$ is the posterior distribution of z_k in the t_{em}-th iteration, and $\boldsymbol{\theta}_k^{(t_{\mathrm{em}})} = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, respectively.

ii). Based on $\zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i)$, the following calculations are performed (M-step):

$$\pi_k^{(t_{\mathrm{em}})} = \frac{1}{N_{\mathrm{gmm}}} \sum_{i=0}^{N_{\mathrm{gmm}}-1} \zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i), \tag{3.12}$$

$$\boldsymbol{\mu}_k^{(t_{\mathrm{em}})} = \frac{\sum_{i=0}^{N_{\mathrm{gmm}}-1} \zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i)\boldsymbol{X}_i}{\sum_{i=0}^{N_{\mathrm{gmm}}-1} \zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i)}, \tag{3.13}$$

$$\boldsymbol{\Sigma}_k^{(t_{\mathrm{em}})} = \frac{\sum_{i=0}^{N_{\mathrm{gmm}}-1} \zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i)\left[\boldsymbol{X}_i - \boldsymbol{\mu}_k^{(t_{\mathrm{em}})}\right]\left[\boldsymbol{X}_i - \boldsymbol{\mu}_k^{(t_{\mathrm{em}})}\right]^{\mathrm{T}}}{\sum_{i=0}^{N_{\mathrm{gmm}}-1} \zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i)}. \tag{3.14}$$

The above steps are repeated while the following evaluation function Q converges:

$$Q = \sum_{i=0}^{N_{\mathrm{gmm}}-1} \sum_{j=0}^{K-1} \zeta_k^{(t_{\mathrm{em}})}(\boldsymbol{X}_i) \log \pi_j^{(t_{\mathrm{em}})} p_{\mathrm{normal}}\left[\boldsymbol{X}_i|\boldsymbol{\theta}_j^{(t_{\mathrm{em}})}\right]. \tag{3.15}$$

3.5 Emulation Setups of Shadowing Classification

This section describes the emulation setups for evaluating the effectiveness of the proposed classifier. We first show the measurement campaign of the 3.5GHz band datasets. Then, the 700MHz band datasets are explained.

3.5.1 Overview of Measurement for 3.5GHz Band

To obtain radio environment information of a cellular communication at an urban area, the measurement campaign was conducted in Kudanshita Chiyoda-ku in Tokyo, Japan. In the campaign, experimenters placed the transmission antenna in the area shown in Fig. 3.2. The transmitter sent the continuous wave to each location. The height of the transmitter above ground level was 17.5 [m]. The receiver fixed the antenna on the cart and traveled the area

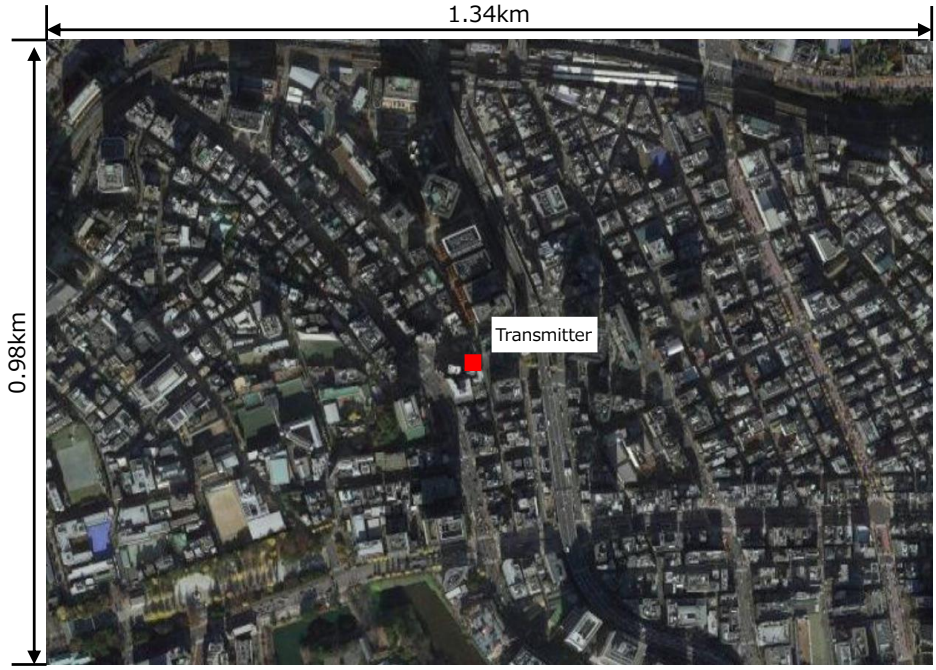


Fig. 3.2 Observation area (3.5GHz band).

Table 3.3 Observation parameters (3.5GHz band) © 2020 IEEE

Center frequency [GHz]	3.5
Transmission power [dBm]	29
Transmission antenna	Omnidirectional
Reception antenna	Omnidirectional

at 4 [km/h]. The height of the receiver above ground level is 1.5 [m]. The received time, reception location, and received signal power were measured and stored at every 0.15 [ms]. We present the observation parameters in Table 3.3. The 3.5GHz band has been assigned in the 5G systems. Considering the usage of the radio map in the 5G, we use a 3.5GHz as the center frequency in this emulation. The number of measured samples was 100423.

3.5.2 Overview of Measurement for 700MHz Band

Next, the observation data of a road-to-vehicle communication system [54] were utilized. We conducted the measurement campaign in Odaiba Minato-ku in Tokyo, Japan. This area is a typical suburban area and the road side unit is installed on the red square shown in Fig. 3.3 by Toyota Motor Corporation. We prepared three vehicles to measure the beacon signal transmitted from the road side unit. The measurement device was the spectrum analyzer

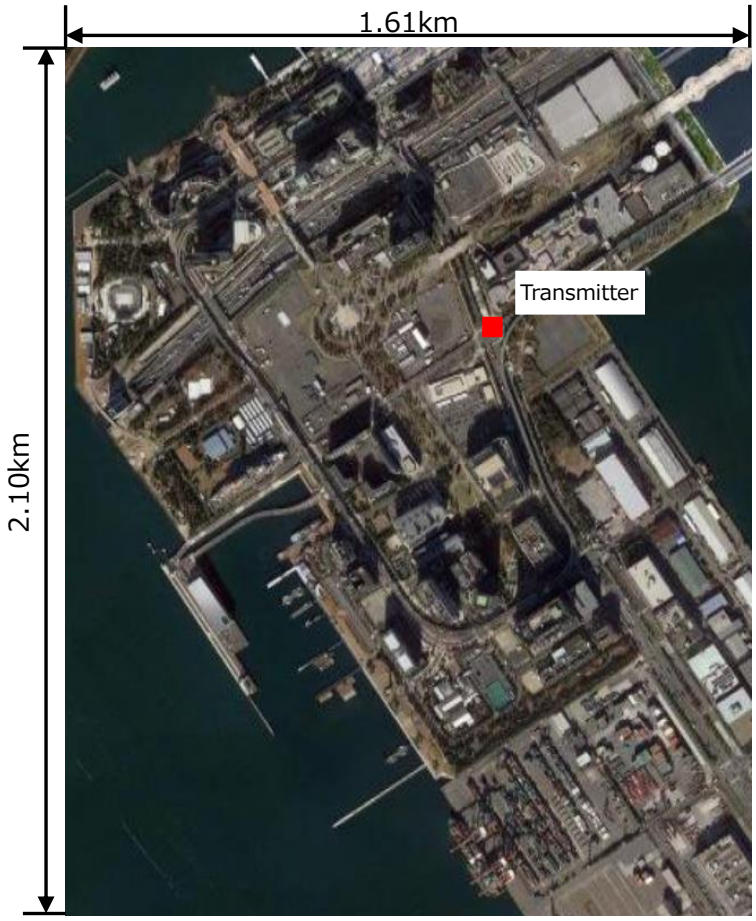


Fig. 3.3 Observation area (700MHz band).

and the observation program was implemented by the Python script. The spectrum analyzer enabled us to observe the received position and the received signal power in each position. Table 3.4 are the observation parameters in 700MHz band datasets. The number of datasets was 3775456.

3.6 Emulation Results of Shadowing Classification

We present the emulation results for the two kinds of datasets. First, the examples of the shadowing classification are explained. After depicting the outliers of the average received signal power, we evaluate the accumulated data size of the radio map. Each weight for (x_m, y_m) and s_m is 0.1 and 0.9.

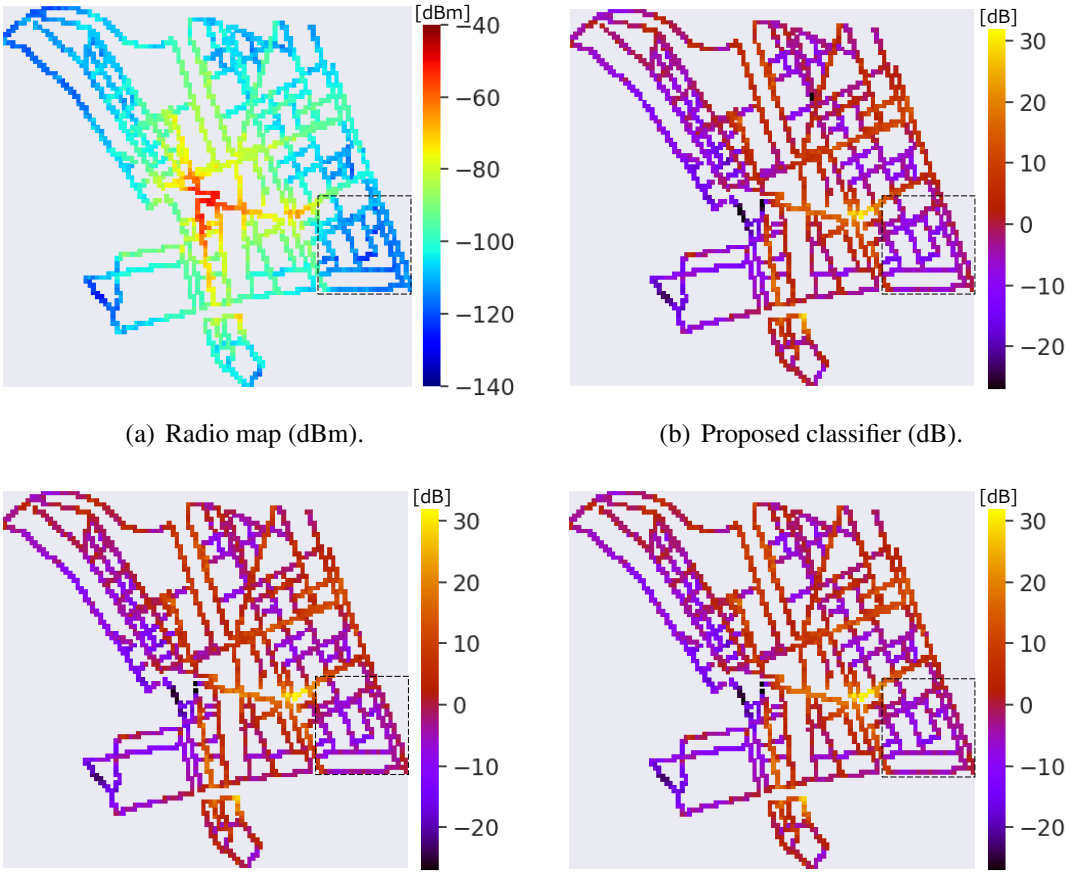
Table 3.4 Observation parameters (700MHz band) © 2020 IEEE

Communication standard	STD-T109 [55]
Center frequency [MHz]	760
Transmission power [dBm]	19.54
Modulation format	QPSK/OFDM
The number of subcarriers	52
Bandwidth [MHz]	9
Spectrum analyzer	RSA306@Tektronix

3.6.1 Examples of Shadowing Classification

Fig. 3.4 is the example of radio map and shadowing classification of the 3.5GHz band datasets. The mesh size is 10 [m] and the number of meshes is 2060. Additionally, $w_c = 0.2$ [dB] and $K = 217$. First, the average received signal power in each mesh is illustrated in Fig. 3.4(a). The others show the shadowing classification in each method as presented in Figs. 3.4(b), 3.4(c), and 3.4(d). We can see that each method accurately classifies similar shadowing values, especially it is well performed in the black dotted frame. This is because shadowing values are similar owing to dense structures in the area.

Next, we illustrate shadowing classification of the 700MHz band in Fig. 3.5. Each map consists of 2280 meshes. The classification is performed in $w_c = 0.5$ [dB] and $K = 49$. As represented in Fig. 3.5(a), there is a large difference between average received signal power values in the black dotted area; thus, the outliers exist in the measurement datasets. Fig. 3.5(c) shows that k-means++ incorrectly classifies shadowing value to 0 [dB] in the black dotted area despite the average received signal power is small in several meshes. A few structures only exist in this area; thus, the shadowing values are around 0 [dB] in most meshes. Since k-means++ updates each centroid using mean value, many shadowing values are clustered to 0 [dB]. Meanwhile, the proposed classifier can accurately classify shadowing values in the black dotted area as shown in Fig. 3.5(b). We assigned the shadowing label based on Eq. (3.4); that is, the classification is not affected by outliers in surrounding meshes. This is the advantage of our proposed method.



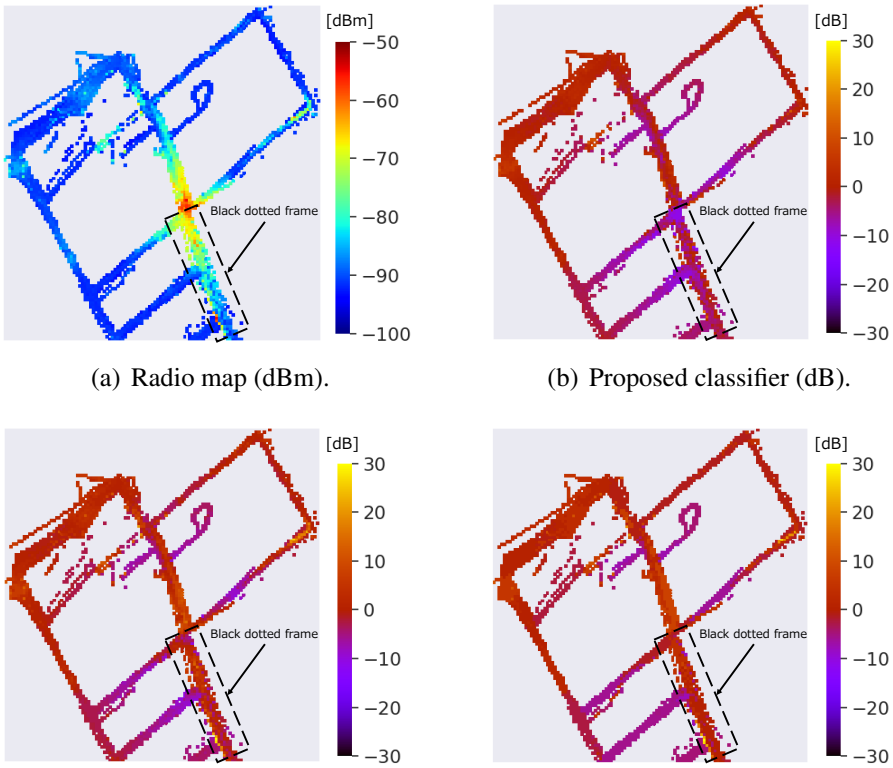
(a) Radio map (dBm).

(b) Proposed classifier (dB).

(c) *k*-means++ (dB).

(d) GMM (dB).

Fig. 3.4 Examples of shadowing classification (3.5GHz band).



(a) Radio map (dBm).

(b) Proposed classifier (dB).

(c) *k*-means++ (dB).

(d) GMM (dB).

Fig. 3.5 Examples of shadowing classification (700MHz band).

3.6.2 Statistical Analysis Using Box Plot

As described in the previous section, the shadowing classification is very sensitive to the outliers of the average received signal power. This section evaluates the outliers using a box plot. In this dissertation, we utilize an interquartile range (IQR) to find the outliers [56]. The IQR infers the $\bar{P}_m$ as the outlier if either of the following formulas is satisfied:

$$\bar{P}_m > q_3 + [1.5 \times (q_3 - q_1)], \tag{3.16}$$

$$\bar{P}_m < q_1 - [1.5 \times (q_3 - q_1)], \tag{3.17}$$

where q_1 [dBm] and q_3 [dBm] are the first quartile and the third quartile of the average received signal power, respectively. In this method, these occurrence probabilities of outliers are almost equivalent to upper and lower probabilities of the standard normal distribution using $3\sigma_{\text{normal}}$, here, σ_{normal} is a standard deviation.

Figs. 3.6 and 3.7 illustrate the outliers of the average received signal power using box plots. The horizontal axis is the communication distance between the transmitter and a mesh by 100 [m] interval. Additionally, we define the vertical axis as the average received signal power of the meshes where the communication distance is the range shown by the horizontal axis. In the box plot, the length of a box is the difference between q_3 and q_1; that is, the box contains half of the total data. Furthermore, the yellow line is the median. We show an outlier as a black circle.

Our analyses clarify that there are more outliers in the 700MHz band dataset compared with the 3.5GHz band. In the southern area shown in Fig. 3.3, another road side unit is installed; hence, the received signal power notably fluctuates because of the interference signals from another road side unit. We find 20 and 121 outliers in Figs. 3.6 and 3.7, respectively.

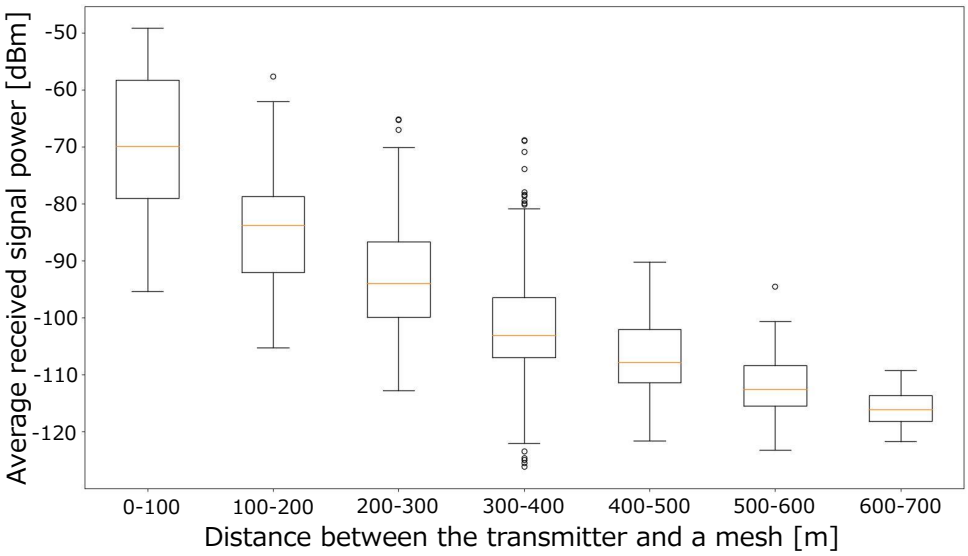


Fig. 3.6 Box plot of the average received signal power (3.5GHz band) © 2020 IEEE.

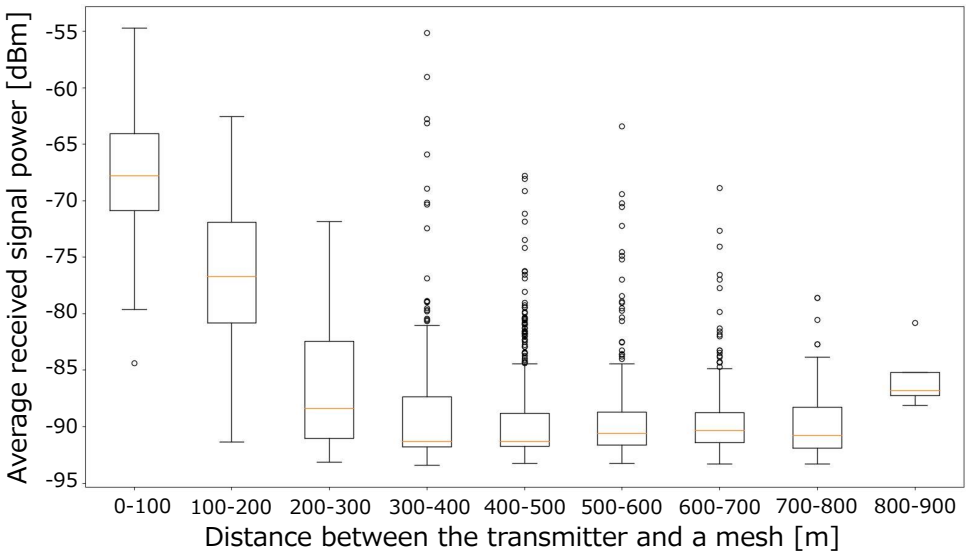


Fig. 3.7 Box plot of the average received signal power (700MHz band) © 2020 IEEE.

3.6.3 Estimation Accuracy of Shadowing Classifier

This section evaluates the RMSE using the following function e_1:

$$e_1 = \sqrt{\frac{1}{N_{\mathrm{eva}}} \sum_{m=0}^{M-1} \sum_{j=0}^{N_m-1} (P_{m,j,\mathrm{dB}} - \bar{P}_m)^2} \quad [\mathrm{dB}], \tag{3.18}$$

where e_1 is the RMSE between the instantaneous received signal power and the average received signal power estimated by each method in each mesh. N_{eva} is the number of instantaneous received signal power samples in all meshes.

We present the RMSE of each dataset in Figs. 3.8 and 3.9. Considering the more accurate performance, the conventional radio map stores the median received signal power in each mesh instead of the mean value. Fig. 3.8 reveals that the estimation accuracy are different between the proposed classifier and k-means++ according to K. The proposed method estimates $\hat{s}_k$ based on the quantization size w_{c} and the number of propagation models K is small in an increase of w_{c}. Hence, the RMSE is slightly degraded in the proposed classifier using small K owing to the quantization error. Meanwhile, k-measn++ can preciously estimate the radio propagation even in the small K. GMM may inaccurately calculate the average received signal power compared to the others. The geographical shape of the observation area shown in Fig. 3.2 is close to the square. Since GMM assumes the cluster shape as the elliptical, several shadowing values may be incorrectly clustered. Finally, we evaluate the RMSE of the measurement-based path loss model denoted by the black dotted line. The result shows that the RMSE is very poor because of the shadowing.

Fig. 3.9 is the RMSE of the 700MHz band dataset. These values show that the proposed classifier can realize the best performance compared to the others. Note that the error performance of the conventional radio map is a little poor. In the 700MHz band campaign, there is a large difference between the number of samples in several meshes. Due to the difference, the biased training and test data may be created in the cross-validation; therefore, the RMSE of the radio map is slightly degraded. Next, we explain the performance of k-means++ and GMM. Both methods cannot accurately estimate the radio propagation because there are many outliers in this area; that is, each cluster is inaccurately updated. Finally, the measurement-based path loss model can unexpectedly estimate the average received signal power with high accuracy. This is because the standard deviation of the shadowing is 4.69 [dB] that is smaller than the typical value (e.g., 8 [dB] in an urban area).

[57] has described that the RMSE may be around 4 [dB] if the path loss and shadowing can be perfectly estimated using the radio map. From Figs. 3.8 and 3.9, we can confirm that the proposed classifier reaches the performance limitation of the radio map. Although the

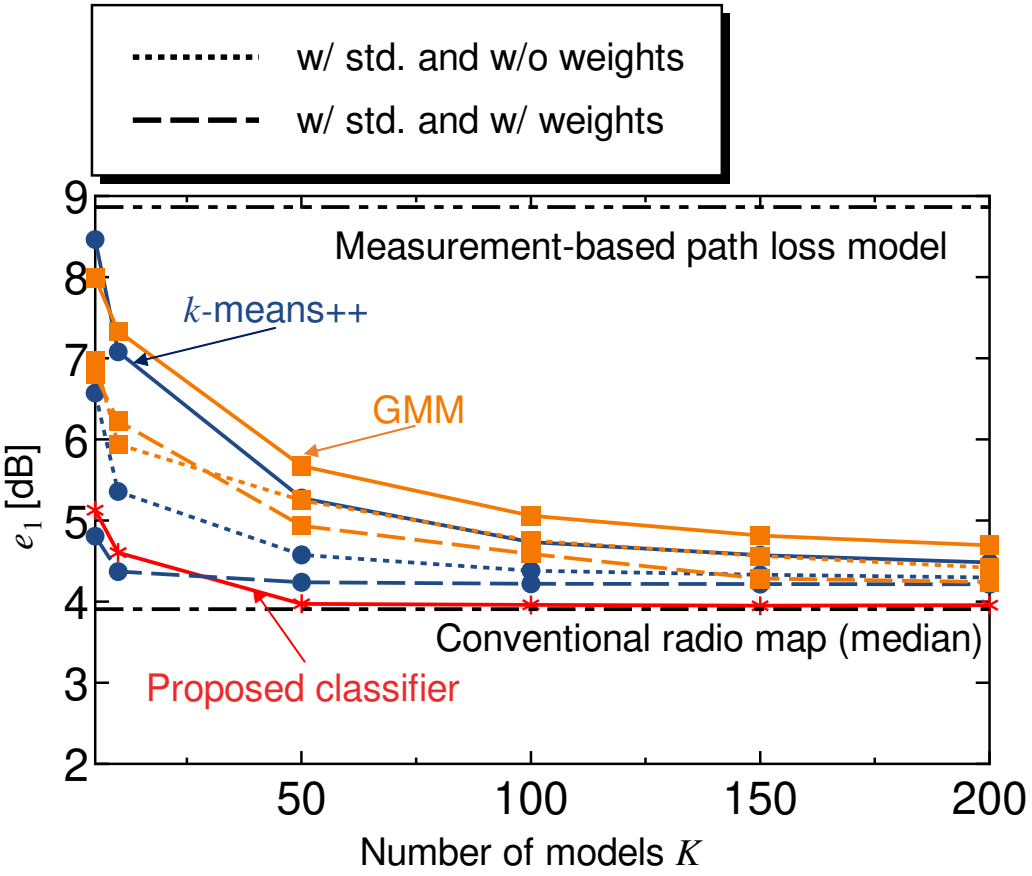


Fig. 3.8 RMSE of 3.5GHz datasets.

remained error occurs owing to the multipath fading, the radio map and classifier cannot estimate such an instantaneous fluctuation.

From Figs 3.8 and 3.9, we can confirm that the estimation accuracy of the k-means++ can be improved in an increase of K. In both figures, the average RMSE can be converged when K is around 25. Since the 3.5GHz band datasets were obtained in urban areas, the standard deviation of the shadowing was around 8 [dB]. This value is generally large in radio propagation. In other words, the estimation error may be small if K is around 25 in most environments, such as suburban.

3.6.4 Accumulated Data Size of Shadowing Classifier

Finally, this section calculates the accumulated data size of the radio map. We utilize the accumulation contents shown in Tables 3.1 and 3.2 for the conventional radio map and shadowing classifier, respectively. Note that the measurement data and time, latitude,

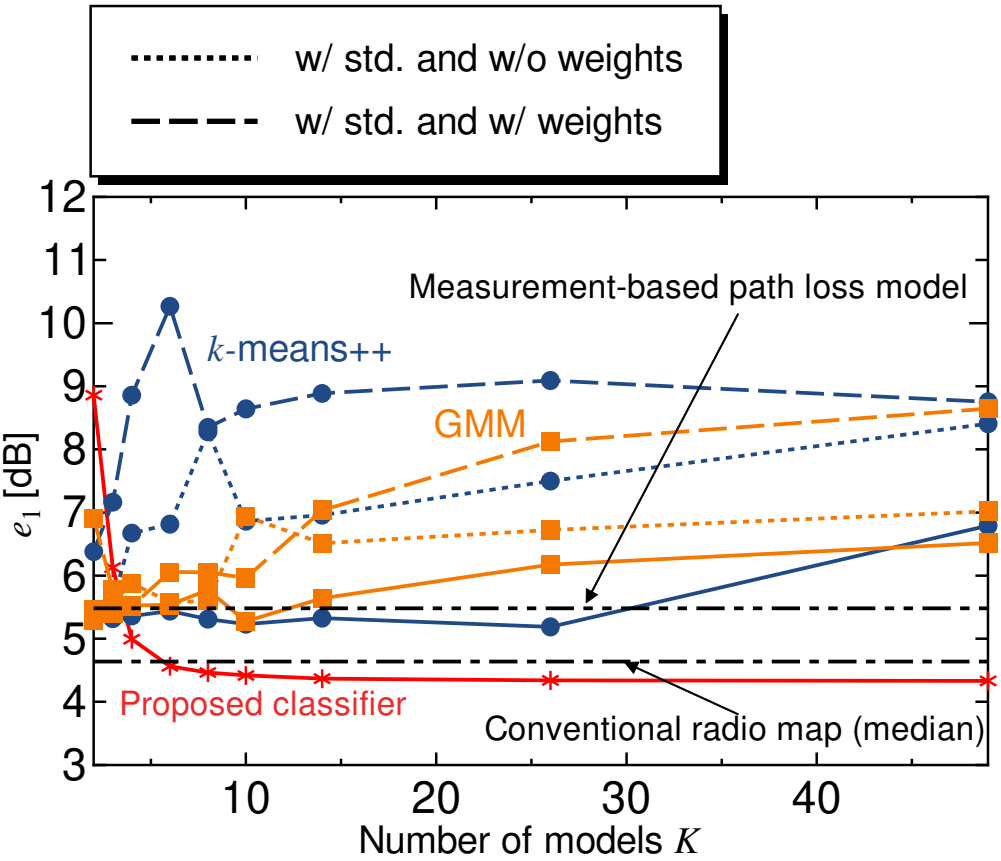


Fig. 3.9 RMSE of 700MHz datasets.

longitude, altitude, and center frequency are not considered in the accumulated data size since these items are registered in both conventional and proposed methods. Additionally, the first code of the reception mesh is the same value in the two communication areas; thus, this item is accumulated only 5 [Byte] in each method. The proposed classifier needs to calculate the path loss based on the parameters B, C, and communication distance. Considering this calculation, the proposed method uses the single table as shown in Table 3.5. L_1 and L_2 of the transmitter are utilized to derive the communication distance. From these tables, we can define the accumulated data size of the proposed classifier and the comparison methods as follows:

$$a_1 = (M \times 15) + (K \times 12) + 37 \quad \text{[Byte]}, \tag{3.19}$$

where a_1 is the accumulated data size in the proposed classifier and conventional clustering methods. Meanwhile, the accumulated data size of the conventional radio map a_2 is given by

$$a_2 = M \times 19 + 5 \quad \text{[Byte]}. \tag{3.20}$$

Table 3.5 Single table.

Item	Format	Size [Byte]
Path loss parameters B and C	double	16
L_1 and L_2 of transmitter	double	16
First code of reception mesh	text	5

In the 3.5GHz band dataset, we found $M = 2060$ and $K = 217$; thus, $a_1 = 33541$ [Byte] and $a_2 = 39145$ [Byte]. The rate between these values is $33541/39145 = 0.8568$; that is, 14.32 [%] reduction can be realized via the proposed method.

In the 700MHz band dataset, using $M = 2280$ and $K = 49$, $a_1 = 34825$ [Byte] and $a_2 = 43325$ [Byte] can be derived and the rate is $34825/43325 = 0.8038$. We can confirm that the number of accumulated data size is reduced about 19.62 [%].

3.6.5 Power Control

To clarify the improvement of communication efficiency, this section performs a power control. The transmission power is determined so that the desired outage probability is guaranteed. The outage event means that the instantaneous received power becomes smaller than the desired power value. The procedures of the power control are shown as follows:

i). The local database creates the radio map and shadowing classifier by processing the measured datasets in each 10m mesh.

ii). The multipath fading is calculated from the difference between an instantaneous power value and the average power value in each mesh.

iii). The local database estimates the empirical CDF of the multipath fading. This section assumes that the shape of the CDF is the same in all meshes.

iv). The percentage point that the lower-tail probability is the desired outage value is calculated from the empirical CDF.

v). The local database decreases the transmission power by the difference between the percentage point and the desired power value.

vi). The power control is performed in each mesh and the average outage probability is derived.

The performance evaluation was conducted using 3.5GHz band datasets. The desired power value is -130 [dBm], the number of shadowing models is 217, and the maximum transmit

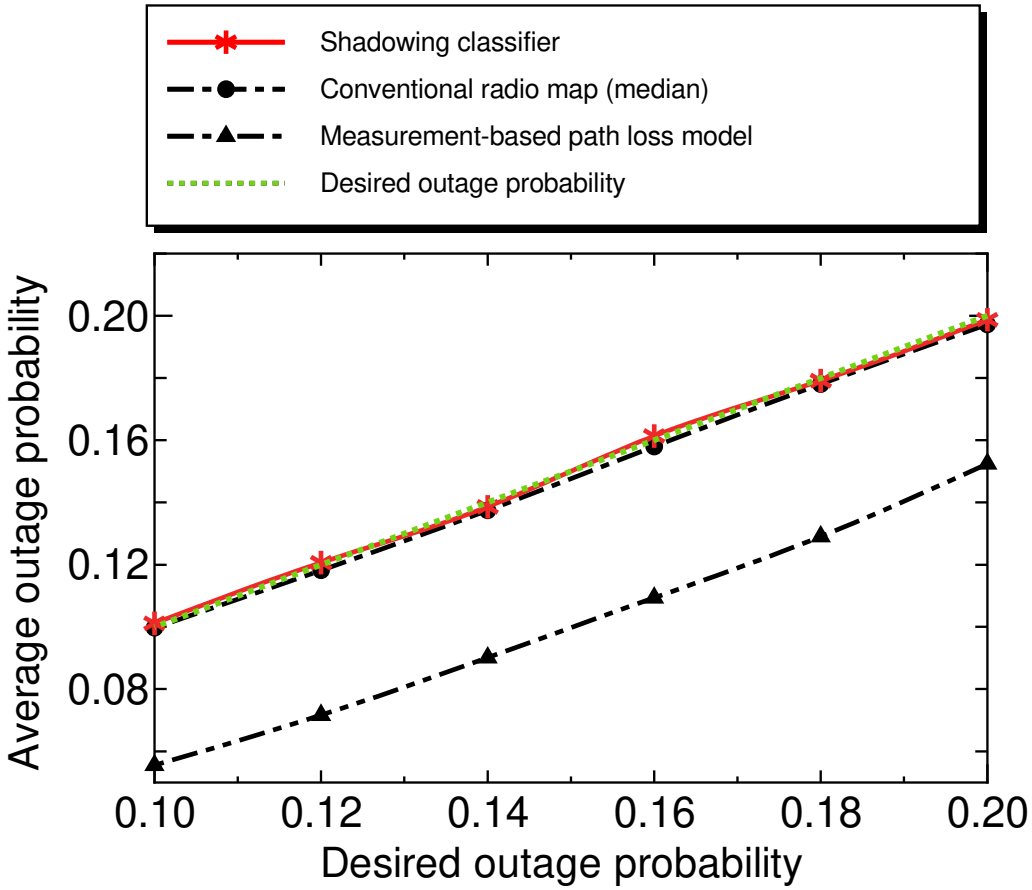


Fig. 3.10 The average outage probability.

power is 28.8 [dBm]. The average outage probability and average transmission power are shown Figs. 3.10 and 3.11, respectively. Each horizontal axis denotes the desired outage probability. The vertical axes denote the average outage probability and the average transmission power in Figs. 3.10 and 3.11, respectively. The green dotted line is the desired outage probability; that is, this line express the true value. Therefore, if the power control can be precisely performed, an obtained outage probability is approximately equivalent to the true value. As the comparison method, we used the measurement-based path loss model. This method estimates the average power in each mesh by considering only the path loss effect. Suzuki distribution is used as the mixture distribution of the shadowing and fading with a standard deviation of 8 [dB].

Fig. 3.10 reveals that the proposed classifier and conventional radio map can skillfully guarantee the desired outage probability compared to the path loss-based method. Because the proposed classifier and the radio map can accurately estimate the shadowing in addition

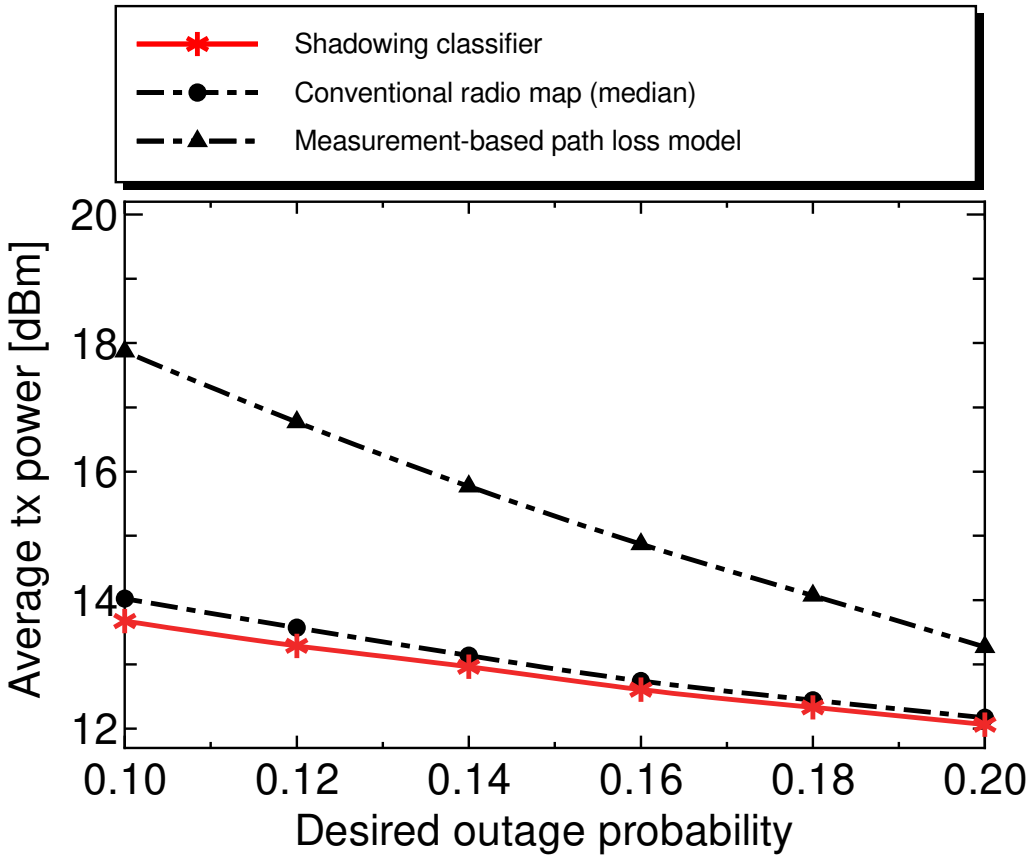


Fig. 3.11 The average transmission power.

to the path loss, the percentage point may be correctly derived. Consequently, the desired outage probability can be guaranteed. The path loss-based method may inaccurately the percentage point owing to the shadowing deviation.

Finally, Fig. 3.11 clarifies that the average transmission power can be lower than the maximum transmit power and the path loss-based method. From these results, we consider that the proposed classifier and the radio map can satisfy the desired communication quality while the power efficiency can be improved.

3.7 Chapter Summary

This chapter discussed the shadowing classification of the radio map to reduce the accumulated data size. We have proposed the shadowing classifier that unifies similar shadowing values to a representative value based on the objective function. The proposed classifier is generated by quantizing the measured shadowing value with the size w_c. As comparisons,

this chapter has explained two conventional clustering methods: k-means++ and GMM. The emulation results have clarified that it is necessary to adaptively utilize the proposed classifier and k-means++ according to whether the average received signal power values have outliers. Furthermore, the classification-based methods have enabled us to reduce the accumulated data size while accurately estimating the radio environment.

Chapter 4

Sample Size Formulation using Statistical Inference for Radio Map

Chapter 3 discussed the reduction of the accumulated data size in terms of the average received signal power in each mesh. The shadowing classifier and conventional clustering methods enable us to construct the low storage radio map while keeping the high estimation accuracy.

However, the proposed method focused on the classification of mean values only. Meanwhile, to accurately calculate the mean value, the local database should accumulate enough received signal power samples via crowdsourcing. In statistics, the required sample size depends on the variance of a population. Of course, the sample mean may become equivalent to the true mean using unlimited samples. Needless to say, it is the law of large numbers. However, the observation time and measurement cost may be enormous to realize the law of large numbers in all meshes when the area range is too large. Here, a crucial question arises:

- *How many* samples does it need to estimate the average received signal power with the desired accuracy?

To resolve this question, we formulate the required sample size to estimate the average received signal power for the radio map. This chapter presents three statistical methods: interval estimation, CLT, and *t*-test. These are venerable methods in statistics because of their simplicity.

This chapter first explains the detailed background of this study. Then, we formulate the required sample size based on the three statistical methods. To verify the usefulness of the statistical methods, computer simulations are utilized. The simulation results show that the average received signal power can be accurately estimated using minimum required samples.

4.1 Background

Crowdsourcing enables us to efficiently construct a radio map since a huge number of mobile terminals may observe radio environment information in various environments. Motivated by this fact, many researchers have continuously created radio maps assuming crowdsourcing.

However, conventional works do not consider the required sample size to accurately estimate the average received signal power in the radio map construction. We think the biggest reason is the abuse of the law of large numbers. Many researchers have believed that a huge number of samples can be obtained using crowdsourcing. Under this condition, the mean value may be calculated with high accuracy because of the law of large numbers. Actually, the law of large numbers is a very powerful law since we can improve the estimation accuracy of the first moment by only increasing the number of samples. However, this law should not be applied to crowdsourcing because ordinary people may need to observe the radio environment for a long time to collect a huge number of samples. To realize realistic crowdsourcing, it is necessary to collect the minimum required samples to estimate the average received signal power with high accuracy.

Focusing on the spatial correlation of the shadowing, an unknown received signal power can be estimated via spatial interpolation (e.g. kriging [11]). If the spatial interpolation is applied to the radio map construction, the radio environment observation is not necessary in several meshes having high correlation. As a result, the required sample size can be reduced. However, since the spatial correlation characteristics may be greatly different depending on the geographical conditions, such as the number of buildings, it is hard to appropriately determine the parameters of kriging (e.g., variogram model) for accurately considering these correlation characteristics [58].

To determine the required sample size, there are several methods in ML. For instance, the Vapnik–Chervonenkis dimension can be utilized. This method calculates the complexity of the learning method [59] and includes the sample size used in learning. Thus, the required sample size can be derived by giving the permissible learning error and complexity of the learning method. However, since the required sample size strongly depends on the complexity of the learning method, it is necessary to recalculate the sample size if the learning method used for estimation changes. As an alternative method, the amount of training data for constructing an accurate model is required ten times more than the number of parameters in the model [60]. However, this rule is based on the rule of thumb; thus, it is not an absolute and accurate index. The general method for determining the required sample size is necessary.

In this chapter, we consider three statistical methods for determining the required sample size to estimate the average received signal power. The first method is interval estimation that is used to estimate the confidence interval of the sample mean. This dissertation formulates

the sample size based on this confidence interval. Second, we use CLT. This method determines the required samples size so that the variance of the average received signal power is less than the desired value. Finally, the *t*-test in hypothesis testing is utilized. The simulation results reveal that the proposed methods can estimate the average received signal power with the desired accuracy while notably reducing the number of required sample sizes.

4.2 System Model

We assume that the radio map is created by crowdsourcing-based measurements as described in Sect. 1.4.2. To utilize three statistical methods described later, it is assumed that the initial radio map is generated by using enough instantaneous samples (e.g., one thousand) in each mesh. After the initial radio map is constructed, each mobile terminal observes the received signal power again and reports instantaneous samples to the local database. The local database randomly picks up samples in each mesh based on the required sample size determined in advance. Then, the average received signal power is recalculated using the picked samples. After that, when the observation data are reported from mobile terminals, the local database stores only required samples. The determination methods of this minimum required sample size will be described in Section 4.3.

In the actual measurement, the recorded received position information may include an observation error of 2-3 [m] due to the limitation of the GPS [61]. To suppress this error in the radio map construction, the local database sets the mesh size larger than the GPS error.

Furthermore, it is assumed that the measurement error of the instantaneous received signal power owing to the individual differences between mobile terminals [62] can be compensated by using the calibration process [62]. From [62], it can be found that the measurement error can be accurately calibrated by constructing the error model based on the linear regression. Although we need to consider the novel calibration method to accurately compensate for the measurement errors, the task is left as future work.

Intuitively, the instantaneous samples should be more accumulated in the meshes having low signal-to-noise ratio (SNR) because these samples may notably be missing. However, in such meshes, the average received signal power should be estimated using the extrapolation [63–65] to construct the accurate radio map. In general, radio propagation extrapolation is very difficult in such a noise-limited system because all measurement datasets are missing below the noise floor. Although Chapter 6 proposes the extrapolation method to estimate missing data, the extrapolation method can be performed well only in an interference-limited observations. Thus, we need to propose a novel extrapolation method for the noise-limited system before the formulation of the required sample size. Hence, the noise floor of the

terminal is assumed to be lower compared to the average received signal power in each location.

Moreover, this chapter assumes that geographical features, such as the transmitter position and the number of structures, are not changed. If geographical features vary, we need to update the average received signal power using hypothesis testing before the required sample size is calculated. This will be discussed in Chapter 5.

4.3 Formulation of Required Sample Size

We mathematically derive the required sample size for estimating the average received signal power. In the following, we mainly design the proposed methods based on statistical inference. Statistical inference is generally divided into the estimation and the test. The proposed methods consist of both approaches.

4.3.1 Interval Estimation-based Method

As the estimation method for a population mean, the interval estimation [66] is well known in statistics. The interval estimation is classified into parametric approach (e.g., the Student's t-distribution-based method [66]) and non-parametric approach (e.g., the bootstrap-based method [67]). In the bootstrap method, the population mean can be estimated based on the repetition of random sampling from the obtained samples. However, the required sample size cannot be theoretically derived from the confidence interval. This implies that the required sample size strongly depends on the shape of the probability distribution of the population.

Meanwhile, the required sample size is theoretically derived in the Student's t-distribution. Here, it is assumed that the population follows the normal distribution with mean μ_{normal} and variance $\sigma^2_{\mathrm{normal}}$. In this method, the confidence interval for $100(1-\alpha)\,[\%]$ of μ_{normal} is represented as follows:

$$\bar{x}_{\mathrm{normal}} - T_{\frac{\alpha}{2}} \frac{s_{\mathrm{normal}}}{\sqrt{n_{\mathrm{normal}}}} < \mu_{\mathrm{normal}} < \bar{x}_{\mathrm{normal}} + T_{\frac{\alpha}{2}} \frac{s_{\mathrm{normal}}}{\sqrt{n_{\mathrm{normal}}}}, \tag{4.1}$$

where $\bar{x}_{\mathrm{normal}}$ is the sample mean, s^2_{normal} is the unbiased variance, $\pm T_{\frac{\alpha}{2}}$ is the upper and lower critical values that are $100\frac{\alpha}{2}\,[\%]$ points of t-value, $100(1-\alpha)\,[\%]$ is the confidence coefficient, α is the significance level, and n_{normal} is the sample size. From this equation, if the sample mean $\bar{x}_{\mathrm{normal}}$ is included in the desired confidence interval I_{d}, the following inequality must be satisfied,

$$I_{\mathrm{d}} \geq 2T_{\frac{\alpha}{2}} \frac{s_{\mathrm{normal}}}{\sqrt{n_{\mathrm{normal}}}}. \tag{4.2}$$

The right term in Eq. (4.2) is derived by subtracting the first term from the third term in Eq. (4.1) that is the estimated confidence interval of $\bar{x}_{\text{normal}}$. If $\bar{x}_{\text{normal}}$ is included in the right term in Eq. (4.2), the desired confidence interval I_{d} should be equivalent or larger than the right term in Eq. (4.2). From this equation, the minimum required sample size $n_{\text{normal,min}}$ can be theoretically derived by solving the Eq. (4.2) for n_{normal} and its definition is

$$n_{\text{normal,min}} \triangleq n_{\text{normal}} \geq \left\lceil \left(\frac{2T_{\frac{\alpha}{2}} s_{\text{normal}}}{I_{\text{d}}} \right)^2 \right\rceil. \tag{4.3}$$

Because $n_{\text{normal,min}}$ must be calculated as an integer, the ceiling function is used. This equation means that if I_{d} is narrow or s^2_{normal} is large, $n_{\text{normal,min}}$ becomes large. To calculate $n_{\text{normal,min}}$, I_{d} and α are required to give in advance. Here, $T_{\frac{\alpha}{2}}$ can be calculated based on t-distribution tables. Additionally, s^2_{normal} should be calculated using enough sample size for accurately estimating the population variance.

The local database calculates Eq. (4.3) in each mesh and randomly extracts the instantaneous samples from the reported samples for $n_{\text{normal,min}}$. Note that if the sample size of the reported data is smaller than $n_{\text{normal,min}}$ in the m-th mesh, the aforementioned operation is not performed. In that case, the operation is performed when the number of newly accumulated samples becomes larger than $n_{\text{normal,min}}$. Here, s^2_{normal} is calculated by referring to an initial radio map constructed using enough samples. Here, s_{normal} and I_{d} are derived as the decibel domain rather than the milliwatt domain. If the milliwatt is used, the mean value may be inaccurately estimated due to the outliers.

Although $n_{\text{normal,min}}$ can be calculated using the CLT described later, the interval estimation is utilized in this dissertation. If the interval estimation is used, it can be derived not only $n_{\text{normal,min}}$ but also the confidence interval. The deviation range of the sample mean can be found by using the confidence interval. Meanwhile, this interval may not be accurately derived in the CLT if the sample size is small.

4.3.2 CLT-based Method

In the interval estimation using the Student's t distribution, it is assumed that the population follows the normal distribution. Thus, the Student's t distribution may not be appropriate if the instantaneous received signal power does not follow the log-normal distribution. Here, it is well known that the probability distribution of the sample mean follows the normal distribution with mean μ_{true} and variance $\frac{\sigma^2_{\text{true}}}{n_{\text{clt}}}$ in any probability distribution by increasing n_{clt}. Here, μ_{true}, σ^2_{true}, and n_{clt} are the population mean, population variance, and the sample size, respectively. By considering the variance as a desired standard deviation of the sample

mean, it can be defined as follows:

$$\sigma_{\mathrm{d}} \triangleq \frac{\sigma_{\mathrm{true}}}{\sqrt{n_{\mathrm{clt}}}}, \tag{4.4}$$

where σ_{d} is the desired standard deviation of the sample mean. From this equation, the minimum required sample size for satisfying σ_{d} can be formulated as follows:

$$n_{\mathrm{clt}} = \left\lceil \frac{s_{\mathrm{clt}}^2}{\sigma_{\mathrm{d}}^2} \right\rceil. \tag{4.5}$$

In the realistic environment, σ_{true}^2 may not be known in most cases. Thus, the unbiased variance s_{clt}^2 is used instead of σ_{true}^2. This value should be estimated using enough samples.

The local database calculates Eq. (4.5) in each mesh and randomly extracts the instantaneous samples from the reported samples for n_{clt}. If the sample size of the uploaded samples is smaller than n_{clt} in m-th mesh, the aforementioned operation is not performed. Here, s_{clt}^2 is calculated using an initial radio map constructed using enough samples. Note that σ_{d} and s_{clt} are given as the logarithmic values.

4.3.3 Normality Test Using Shapiro–Wilk Test

To use the two statistical methods, it is necessary to infer the normality of the obtained samples. Here, normality test methods have been proposed in statistics, and the accuracy of statistical power for these methods has been conducted [68]. From [68], it can be found that the Shapiro–Wilk test is the most powerful test if the number of samples used in testing is large. Thus, we utilize this test. The statistic of the m-th mesh in Shapiro–Wilk test is defined as follows:

$$S_m = \frac{\left[\sum_{i=0}^{N_m-1} c_i P_{(m,i,\mathrm{dB})}\right]^2}{\left(\sum_{i=0}^{N_m-1} P_{m,i,\mathrm{dB}} - \bar{P}_m\right)^2}, \tag{4.6}$$

where S_m is the statistic of the m-th mesh in Shapiro–Wilk test, c_i is the i-th tabulated coefficient [69], and $P_{(m,0,\mathrm{dB})} \leq P_{(m,1,\mathrm{dB})} \leq \cdots \leq P_{(m,N_m-1,\mathrm{dB})}$ are the ordered values of the samples $(P_{m,0,\mathrm{dB}}, P_{m,1,\mathrm{dB}}, \cdots, P_{m,N_m-1,\mathrm{dB}})$. S_m is always $0 < S_m \leq 1$ and the small S_m indicates abnormality. In this paper, the abnormality means that the obtained samples do not follow the log-normal distribution.

In this dissertation, Shapiro–Wilk test is performed using the initial radio map constructed using enough samples. Subsequently, $n_{\mathrm{normal,min}}$ is calculated in each mesh inferred as the normality. Meanwhile, n_{clt} is calculated in each mesh that not be inferred for normality.

4.3.4 T-Test-based Method

In the interval estimation, $n_{\text{normal,min}}$ is calculated more because it is determined in proportion to four times the variance s^2_{normal}. Here, the required sample size of t-test [70] can be utilized for solving this problem and its definition is given by

$$n_{\text{t}} = \left\lceil \frac{2s_{\text{t}}^2(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta})^2}{\Delta^2} \right\rceil, \tag{4.7}$$

where n_{t} and s_{t}^2 are the minimum required sample size and the unbiased variance using t-test-based hypothesis testing, respectively. $\Delta = \mu_1 - \mu_2$ is the difference between mean μ_1 and mean μ_2. Additionally, $Z_{1-\frac{\alpha}{2}}$ and $Z_{1-\beta}$ are the $100(1-\frac{\alpha}{2})$ and $100(1-\beta)$ percentiles of the standard normal distribution, respectively. These values can be obtained using z-tables. β is the type II error. If n_{t} is too small, the significant difference between μ_1 and μ_2 may not be inferred despite μ_1 and μ_2 are not equivalent. When the t-test is performed multiple times, the miss testing may occur with a certain probability. This probability corresponds to the β.

From Eq. (4.7), it can be found that if Δ is small or s_{t}^2 is large, many samples are required. Additionally, the required sample size can notably be reduced compared to Eq. (4.3) since n_{t} is determined in proportion to two times the variance s_{t}^2. In hypothesis testing, n_{t} is often calculated as $1-\beta = 0.8$ or 0.9 to infer the significant difference between μ_1 and μ_2. Meanwhile, in this paper, this significant difference should not be inferred to create an accurate radio map using the minimum required sample size. Thus, by setting smaller Δ and larger β, the minimum required sample size can be derived. Here, Δ is considered as the permissible difference between two means, and its value is required to be given in advance.

Eq. (4.7) is calculated for each mesh in the local database. Then, the cloud server randomly extracts the instantaneous samples from the reported samples for n_{t}. Note that if the sample size of the uploaded data is smaller than n_{t} in the m-th mesh, the aforementioned operation is not performed. Here, s_{t} and Δ are calculated in the decibel domain rather than the milliwatt domain. Additionally, s_{t} is estimated based on the initial radio map constructed using enough samples. Note that n_{t} is derived in all meshes regardless of the normality and abnormality. This is because t-test has the robustness [71] that the population does not need to follow the normal distribution.

Although the t-test is usually used in SHT, we attempt to utilize the t-test to calculate the required sample size in the radio map construction. Most researchers have not applied the t-test to the sample size determination; that is, the utilization of the t-test is the originality.

4.4 Computer simulation

Computer simulations are used to evaluate the effectiveness of our formulations. In the following, the simulation setups are shown.

4.4.1 Radio Propagation Model

In this simulation, the radio propagation model in the m-th mesh is defined as follow:

$$P(d_m) = P_{\mathrm{Tx}} - L_{\mathrm{fspl}}(d_0) - 10C\log_{10}\left(\frac{d_m}{d_0}\right) + W + F, \tag{4.8}$$

where $P(d_m)$ [dBm] is the instantaneous received signal power in the m-th mesh, P_{Tx} [dBm] is the transmit power of the transmitter, d_0 [m] is the reference distance, W [dB] is the shadowing value, and F [dB] is the small-scale fading. Additionally, we define L_{fspl} [dB] as the free-space path loss as follows:

$$L_{\mathrm{fspl}}(d_0) = 10\log_{10}\left(\frac{4\pi d_0}{\lambda}\right)^2, \tag{4.9}$$

where λ [m] is a wavelength.

In this simulation, we consider spatially-correlated shadowing. It is well known that the spatial correlation model of the shadowing is represented as the exponential decay model [31] and its definition is given by:

$$\rho_{i,j} = \exp\left(-\frac{\delta_{i,j}}{d_{\mathrm{cor}}}\ln 2\right), \tag{4.10}$$

where $\rho_{i,j}$ and $\delta_{i,j}$ [m] are the shadowing correlation and distance between two terminals i and j, respectively. d_{cor} [m] is the correlation distance, defined as a point on $\rho_{i,j} = 0.5$. The correlation distance of approximately 20 [m] is experimentally determined in an urban area [31]. Under spatial correlation shadowing, the probability distribution is modeled as the multivariate normal distribution [72] as follows:

$$f_{\mathrm{s}} = \frac{1}{\sqrt{(2\pi)^D \boldsymbol{\Sigma}_{\mathrm{s}}}} \exp\left(-\frac{1}{2}(\vec{r}-\vec{p})^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{s}}^{-1}(\vec{r}-\vec{p})\right), \tag{4.11}$$

where f_{s} is the PDF of the multivariate normal distribution, $\vec{r} = (r_0, \cdots, r_{D-1})^{\mathrm{T}}$ is the vector of the average received signal power values, $\vec{p} = (p_0, \cdots, p_{D-1})^{\mathrm{T}}$ is the median path loss vector, and D corresponds to the number of meshes. $\boldsymbol{\Sigma}_{\mathrm{s}}$ denotes the $D \times D$ shadowing covariance

Table 4.1 The common simulation parameters

Communication area [m^2]	200×200
Mesh size [m^2]	10×10
The number of meshes D	400
The number of samples n_{get}	10
The number of terminals H	40
Frequency [MHz]	3500
Transmit power P_{Tx} [dBm]	29
Reference distance d_0 [m]	10
Path loss index C	3.5
Standard deviation of W	8
Correlation distance d_{cor} [m]	20
Confidence coefficient $(1-\alpha)$	0.9, 0.99
Desired confidence interval I_{d} [dB]	2
Desired standard deviation σ_{d} [dB]	0.5, 1.0, 1.5
Significance level of Shapiro–Wilk test	0.05
Permissible difference Δ [dB]	1
Type II error β	0.9
Significance level of t-test α	0.1, 0.01

matrix, as follows:

$$\mathbf{\Sigma}_{\mathrm{S}} = \begin{pmatrix} \sigma_{1,\mathrm{s}}^2 & \cdots & \rho_{1,D}\sigma_{1,\mathrm{s}}\sigma_{D,\mathrm{s}} \\ \vdots & \ddots & \vdots \\ \rho_{D,1}\sigma_{D,\mathrm{s}}\sigma_{1,\mathrm{s}} & \cdots & \sigma_{D,\mathrm{s}}^2 \end{pmatrix}, \tag{4.12}$$

where $\sigma_{i,\mathrm{s}}$ [dB] is the standard deviation of the shadowing from the transmitter to the i-th terminal. In this simulation, W is obtained by calculating the random values that follow the multivariate normal distribution. Here, by using the smaller mesh size than d_{cor}, the deviation of W can be almost negligible in each mesh [73]. Thus, the mesh size is set to 10 [m] in this simulation.

Finally, the small-scale fading models are described. In the proposed methods, the minimum required sample size strongly depends on the variance of the received signal power. Thus, it is necessary to create various fading realizations in each mesh. In this simulation, the Rayleigh and Nakagami–Rice fading are utilized. The rice factor R is randomly determined in each mesh, and random samples are obtained based on the assigned rice factor.

4.4.2 Simulation Procedures

The simulation procedures are described as follows:

i). The log-normal shadowing or Rayleigh or Nakagami–Rice fading are randomly assigned in each mesh. Here, in the realistic environment, the shadowing deviation occurs even in the line-of-sight (LOS) environment with the small standard deviation [74]. To consider the shadowing effect, the log-normal shadowing is utilized in this simulation. The standard deviation of log-normal shadowing is determined from $[1,3]$ based on the measurement results [74]. Additionally, rice factor R is determined from $[0,10]$ [74]. Note that the units of these values are decibels.

ii). The initial radio map is generated by averaging 1,000 instantaneous samples in each mesh. Subsequently, the Shapiro–Wilk test is performed in each mesh. According to the testing results, $n_{\mathrm{normal,min}}$ or n_{clt} are calculated in each mesh. Note that n_{t} is calculated in all meshes regardless of the normality and abnormality.

iii). H terminals are randomly placed in communication area and each terminal observes n_{get} instantaneous samples represented in Eq. (4.8).

iv). The local database randomly extracts instantaneous samples for the number calculated in ii) from n_{get} samples and calculates the average received signal power in each mesh. Then, we evaluate whether the calculated mean value is within I_{d}. In this dissertation, this evaluation index is defined as *degree of confidence*.

v). Procedures iii)-iv) are performed 1,000 times and the average received signal power samples are accumulated in each mesh. Subsequently, we calculate the standard deviation of the average received signal power in the meshes that not are inferred as normality to evaluate the usefulness of CLT.

vi). Procedure v) is performed 100 times and the average degree of confidence and average standard deviation are derived.

The common simulation parameters are shown in Table 4.1. In this simulation, $(1-\alpha)$ is the typical value in statistics.

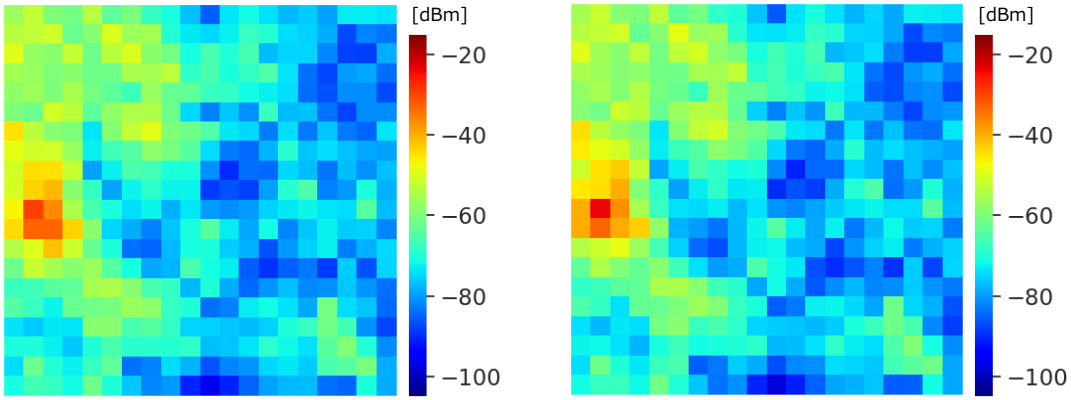
4.5 Simulation Results

This section presents the simulation results. First, we show the example of the radio map based on our formulations. Then, the average degree of confidence and average standard deviation of the average received signal power samples are described. Finally, this section illustrates the RMSE and accumulated data size of the instantaneous samples.

4.5.1 Example of Radio Maps

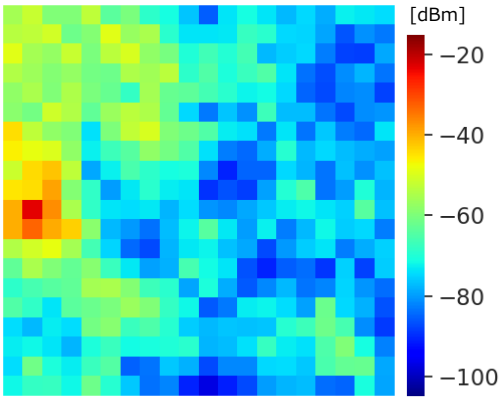
The example of the radio maps are presented in Fig. 4.1. Here, Fig. 4.1(a) is constructed by averaging 10,000 instantaneous samples for each mesh. Note that $(1-\alpha) = 0.99$, $\sigma_{\mathrm{d}} = 0.5$ [dB], and significance level of t-test is 0.01, respectively. Additionally, as simple comparisons, the two methods are evaluated; w/o deletion and forgetting factor. In the former method, the local database constructs the radio map without deleting the instantaneous samples. In Fig. 4.1(d), the average sample size in a mesh is 1,000. Meanwhile, in the forgetting factor-based method, the local database randomly extracts the instantaneous samples for 100η [%] from n_{get} samples. In the following simulation, the forgetting factor η is 0.1.

From these maps, it can be confirmed that the proposed methods represented in Figs. 4.1(b) and 4.1(c) can accurately estimate the radio environment in most meshes. However, in the forgetting factor-based method, the average received signal power is different from the True map in several meshes (e.g., the area within the red dotted frame). This is because the obtained samples are notably deleted owing to smaller η.

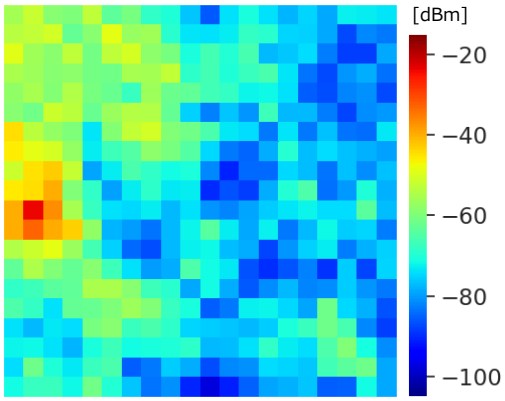


(a) True map.

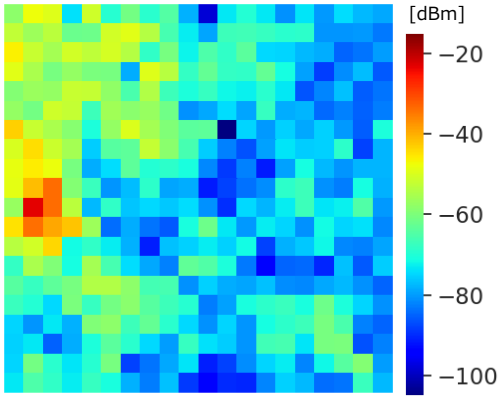
(b) Interval estimation and CLT.



(c) *t*-test.



(d) w/o deletion.



(e) Forgetting factor.

Fig. 4.1 Example of radio maps using minimum required sample sizes.

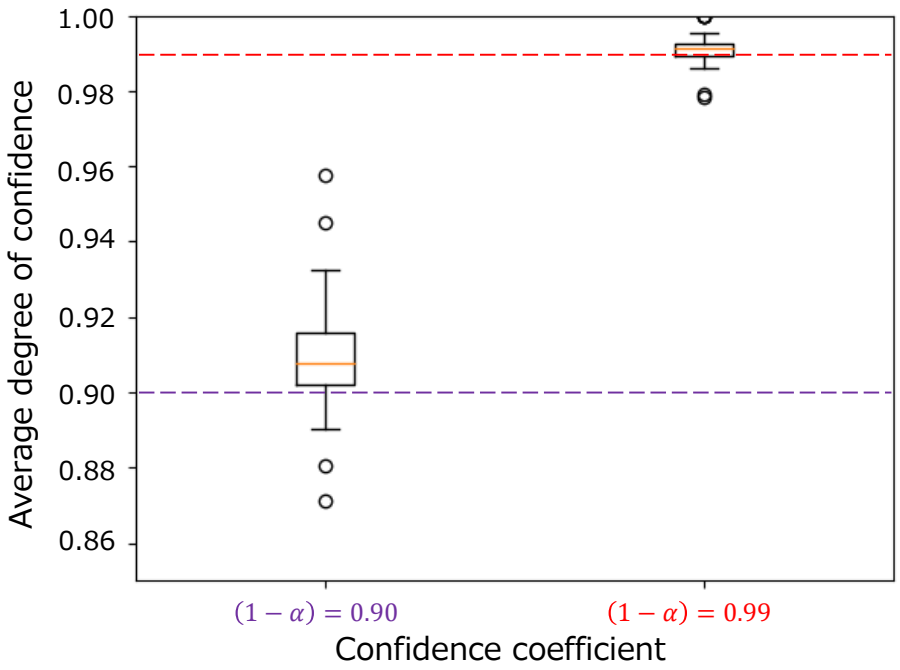


Fig. 4.2 Average degree of confidence © 2021 IEEE.

4.5.2 Degree of Confidence and Standard Deviation

In this section, the average degree of confidence and average standard deviation of sample means are described. The average degree of confidence is depicted in Fig. 4.2 using the box plot. These box plots are created in the meshes that are inferred as normality in Shapiro–Wilk test. In this figure, the horizontal axis is the confidence coefficient. The vertical axis means the probability that the sample mean using $n_{\mathrm{normal,min}}$ samples is within the desired confidence interval I_{d}. Here, the upper and lower confidence values are calculated by adding and subtracting $I_{\mathrm{d}}/2\,[\mathrm{dB}]$ from the initial average received signal power in each mesh. The initial average received signal power corresponds to the average value in the initial radio map. In the box plot, the yellow line and white circle represent the median value and outliers, respectively. Note that the outliers are calculated using IQR. Additionally, the red, black, and purple dotted lines denote each target confidence coefficient. Hence, each median value should be nearly equivalent to each target value. From these results, it can be found that the average received signal power can be included within I_{D} by increasing the confidence coefficient. This is because the required samples become large in an increase of confidence coefficient. Meanwhile, in the small confidence coefficient, the average degree of confidence is greatly different from the target value because of too small samples.

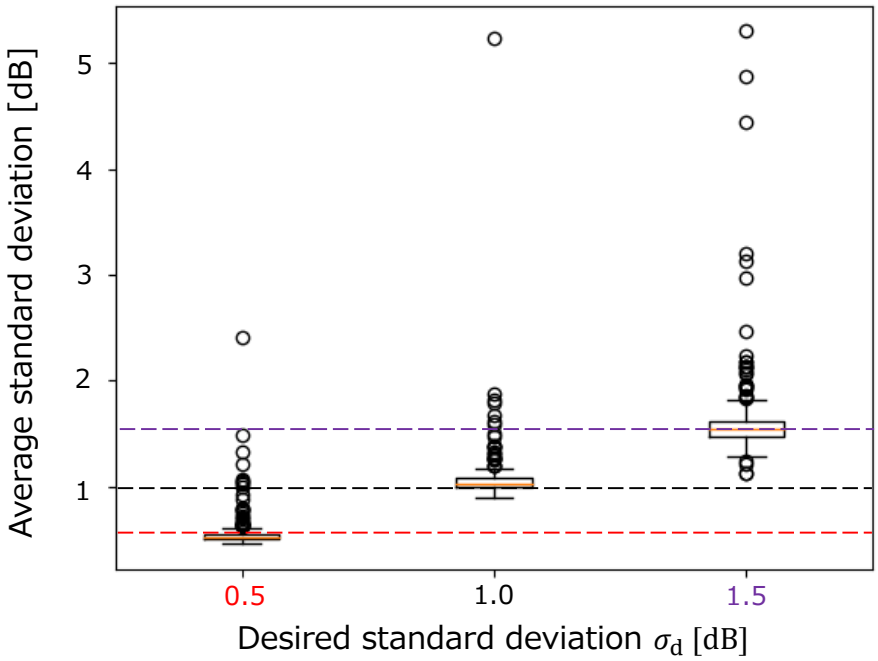


Fig. 4.3 Average standard deviation.

Next, the average standard deviation of sample means is plotted in Fig. 4.3. These box plots are generated in the meshes that are inferred as abnormality. In this figure, the horizontal and vertical axes are the σ_d and the average standard deviation of the sample means, respectively. Here, the red, black, and purple dotted lines denote the desired standard deviations, respectively. It can be seen that each median is almost equivalent to the desired value. Meanwhile, the outliers become large in increase of σ_d because n_{clt} is small owing to the large σ_d.

4.5.3 Estimation Accuracy and Accumulated Data Size

Finally, we evaluate the RMSE of the average received signal power and accumulated data size. The RMSE is defined as follows:

$$e_2 = \sqrt{\frac{1}{D}\sum_{m=0}^{D-1}(\bar{P}_{\text{true},m} - \bar{P}_{\text{req},m})^2} \quad [\text{dB}], \tag{4.13}$$

where e_2 is the RMSE, $\bar{P}_{\text{true},m}$ [dBm] is the true average received signal power in the m-th mesh that corresponds to Fig. 4.1(a). $\bar{P}_{\text{req},m}$ [dBm] is the estimated average received signal power using minimum required sample size in the m-th mesh.

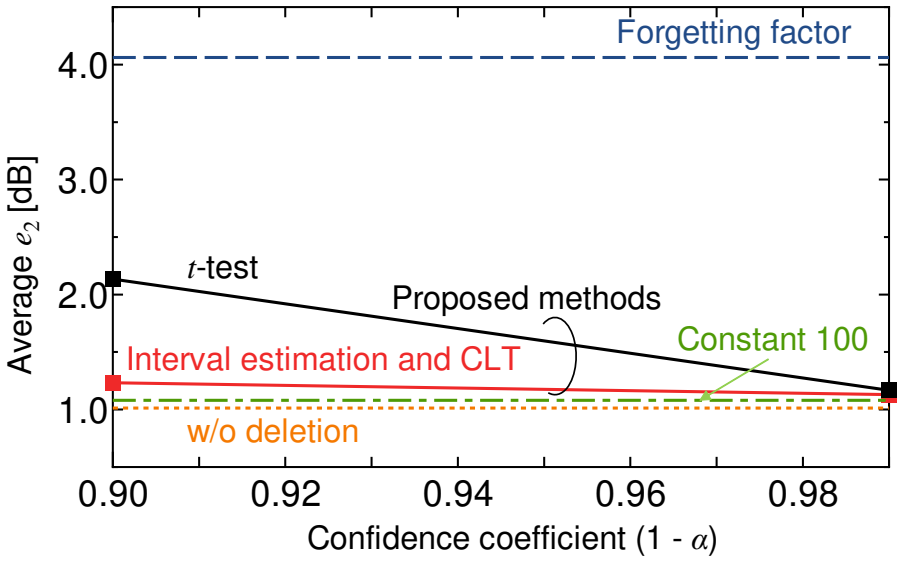


Fig. 4.4 Average RMSE.

The average RMSE is shown in Fig. 4.4. Here, as a simple comparison, the constant sample size-based method is described represented as 'Constant 100'. This method estimates the mean value of the received signal power using 100 samples in each mesh. Note that $\sigma_{\mathrm{d}} = 0.5\,[\mathrm{dB}]$. Here, the significance level of t-test corresponds to $1-(1-\alpha)$. From these results, it can be found that the interval estimation and CLT-based method is superior to t-test in smaller confidence coefficient because n_{t} tends to be calculated smaller compared to $n_{\mathrm{normal,min}}$. Meanwhile, the forgetting factor-based method cannot accurately estimate the radio environment due to the excessive deletion of samples. Although the average RMSE is small in Constant 100 comparing with the proposed methods, the accumulated data size described later is large.

The accumulated data size per mesh is defined considering the instantaneous samples and a 10m mesh code [54]. The 10m mesh code is represented as 16 [Byte] text type [54]. Additionally, the instantaneous received signal power can be accumulated as 8 [Byte] floating-point type [54]. Thus, the accumulated data size r_m [Byte] in the m-th mesh is calculated as follows:

$$r_m = 16 + 8N_m. \tag{4.14}$$

The average accumulated data size per mesh is shown in Fig. 4.5. Although the data size of w/o deletion is 8016 [Byte], it is not plotted in Fig. 4.5 by considering the visibility. It can be found that the accumulated data size of the proposed methods can be reduced by about $93\,[\%] - 99\,[\%]$ compared with the w/o deletion that is the conventional radio map

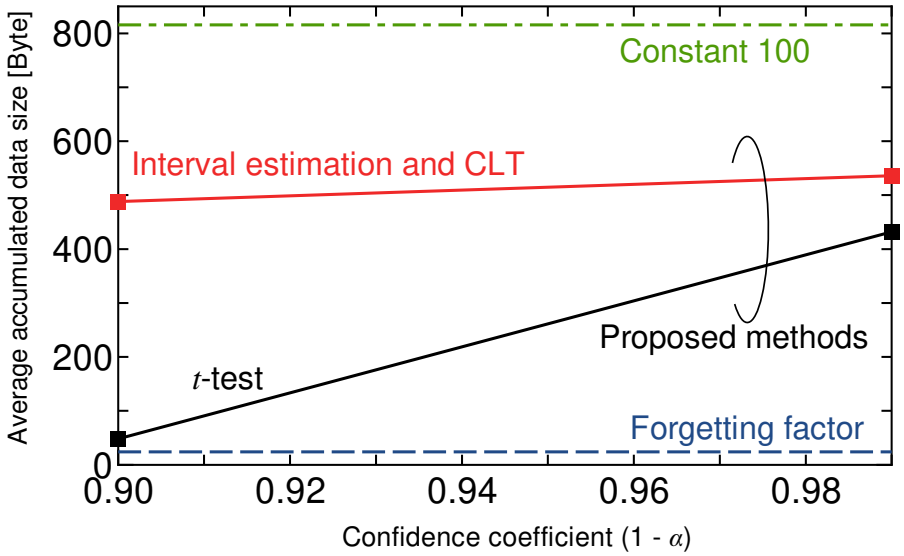


Fig. 4.5 Average accumulated data size of instantaneous samples.

construction. In the proposed methods, t-test can reduce the accumulated data size compared with the interval estimation and CLT. However, the average RMSE of t-test in $(1-\alpha) = 0.90$ is noticeably poor from Fig. 4.4. Thus, we argue that the t-test should be utilized in large $1-\alpha$. Meanwhile, the interval estimation and CLT can reduce the accumulated data size while the high estimation accuracy is kept in small $1-\alpha$.

4.6 Chapter Summary

This chapter has proposed the methods for determining the required sample size to reduce the registered data size. We have utilized three statistical methods: the confidence interval, CLT, and t-test to formulate sample size. The simulation results have revealed that the proposed methods can notably reduce the registered data size while the estimation accuracy keeps.

Chapter 5

Radio Map Updating using the Welch's *t*-test

Chapters. 2, 3, and 4 discuss the conventional radio map that is applied to television white space. In this system, a transmitter does not move from an initial location and the average received signal power is the constant value over the time domain. Meanwhile, the private 5G that provides the communication services by the non-public operator has been attracted attention to flexibly use finite spectrum resources. However, the location of the transmitter may vary in the private 5G; thus, the conventional radio map does not enable us to accurately estimate location-dependent radio propagation. In this chapter, we propose the dynamic radio map using SHT. The proposed method infers a significant difference between two average power values based on Welch's *t*-test. Then, the local database updates the information of the radio map according to the testing outcomes. The evaluation results clarify that the radio map can precisely be updated and estimate the average power via the proposed method.

5.1 Background

5G has been actively discussed worldwide. 5G is the new wireless communication system for achieving high transmission capacity, low delay, and efficient spectrum utilization. Focused on these previously capabilities, the private 5G has been discussed as a novel concept of wireless systems by many researchers in industry, academia, and government [75, 76]. For example, the local 5G and industrial 5G are considered in Japan and Europe, respectively. In this system, multiple transmitters are deployed by non-public operators in a local area, such as a factory to provide the wireless services. However, each non-public operator must share the same frequency band within the small coverage; thus, the inter-transmitter interference should

be appropriately coordinated to guarantee the communication quality of each transmitter. To meet this constraint, it is necessary to accurately understand surrounding radio propagation characteristics, such as path loss, shadowing, and multipath fading.

To estimate the radio propagation, we can utilize the radio map described Chapters. 2, 3, 4. As a utilization example of the radio map in the inter-transmitter interference management, an interferer adjusts the transmission power such that a desired SIR at the other cell edge is satisfied. When the interferer does not know any radio propagation characteristics, the interferer must excessively decrease its transmission power to avoid the inter-transmitter interference. By contrast, if the radio propagation characteristics, including the path loss and shadowing, are known by the radio map, it is not necessary to suppress the transmission power in the interferer. As the effectiveness of the radio map in the spectrum sharing, several researchers have reported that the accurate radio map has the potential to realize the efficient spectrum sharing in 5G systems [18–20]. Additionally, the achievable performance of the radio map using channel state information has been revealed [77, 8, 78]. In summary, we need to create a precise radio map in 5G systems, including private 5G systems, to efficiently utilize the finite spectrum resources. Most researchers have mainly utilized radio maps for television white space [79–81]. These systems assume that the transmitter does not move in an initial location and the average received signal power is constant over the time domain. Under the environment, we can precisely estimate the path loss and shadowing using the initially constructed radio map even if the average power is not updated.

However, the assumption may not be applied to a private 5G system. This chapter considers that a private 5G is deployed in a factory and a construction site to keep watch the construction progress. In this case, the worker may change the manufacturing configurations depending on the product and the working progress. This change may occur from week to week or month to month [82]; thus, it is necessary to move the transmitter of the private 5G system to keep watch the reconstructed production line. Under this environment, the received signal power values may dynamically change due to the movement of the transmitter. As a result, the initially constructed radio map may not accurately estimate radio propagation. In the case of private 5G, a spectrum should be spatially shared among multiple uncoordinated transmitters while inter-transmitter interference is avoided under constraints of changes in the transmitter positions.

Related works have been studied for Wi-Fi and millimeter-wave communications. For the radio map updating, [83, 84] utilize media access control (MAC) address of each access point (AP) in Wi-Fi positioning systems. However, these works cannot update the radio map if a single transmitter moves. This is because the radio map is updated based on whether new APs are installed or existing APs are removed. Another method is proposed for millimeter-wave

communication systems [85]. The updating rate is determined by referring to the coherence duration; thus, the radio map may be unnecessarily updated since the coherence duration is often given in the order of milliseconds. Additionally, [86] has proposed the real-time updating method of the radio map. This method enables us to detect instantaneous variations of received signal power by always collecting radio environment information. However, several costs increase significantly according to the range of the communication area. For instance, the observation time of the radio environment is considered. In summary, although several researchers have proposed radio map updating methods, there are some problems. In a private 5G system in the industrial scenario, the transmitter may move every a few days or weeks to monitor the production progress [87]. Thus, we need to update the radio map by considering this duration.

Several researchers have continuously discussed the utilization of ML to efficiently create a radio map [88]. However, it is necessary to prepare a large training dataset in the ML. Thus, we need to observe the radio environment repeatedly while changing the transmission location in cellular systems. As a result, we cannot use ML in the cellular systems. The main issue is to appropriately judge whether the radio map must be updated based on *only real-time datasets*. The other method of updating has been proposed by Mo *et al.* [89]. This work utilizes kernel principal component analysis (KPCA), that is, the features of the received signal power values are extracted. However, this method cannot be used if the inverse matrix cannot be calculated. This phenomenon occurs when the sample size of the measured data is smaller than the number of dimensions in KPCA.

The radio map can be simply updated by constantly calculating the average power using the measurement data reported from mobile terminals. The local database requires to provide the updated radio map to all mobile terminals. If the number of mobile terminals increases, the provision time of the radio map ma be long. Thus, the constant updating is inappropriate in crowdsourcing-assisted radio map construction.

We can use hypothesis testing [90, 91] to solve the main task. Hypothesis testing is divided into Bayesian hypothesis testing (BHT) [92–95] and SHT [71, 96–98]. The prior distribution is needed to use BHT; however, it may not be appropriately modeled since the radio propagation is complicated in a real environment owing to path loss, shadowing, and multipath fading.

SHT consists of parametric and non-parametric tests. The validity of the hypothesis cannot be properly inferred in non-parametric testing if heteroscedasticity exists between the two populations [99]. The variances between two datasets over the time domain may be unknown in a realistic environment. Here, the robustness of the t-test has been reported by [71]. The normality assumption for the population distribution is not required in the t-test

even if a small number of samples (e.g., 100) is used. However, homoscedasticity is assumed between the two populations in the general *t*-test.

Motivated by the tasks, this chapter proposes a radio map updating method based on Welch's *t*-test [100]. The local database infers a significant difference between the two average values using Welch's *t*-test and updates the statistical information according to the inferred outcomes. The major contributions of this chapter are presented as follows:

- The existing updating methods of the radio map, such as a BHT, a non-parametric method, and ML, were investigated. From the surveys, it can be found that the radio map should be updated only if the surrounding environment changes rather than the real-time updating. Additionally, we may not determine the prior distribution of the mean power due to the complicated radio propagation. Finally, robustness for heteroscedasticity is needed when hypothesis testing is utilized. Considering these factors, Welch's *t*-test is used to update the radio map in cellular systems. Through the simulation results, we can confirm that the radio map can be accurately updated via the proposed method even in a small number of samples.

- The usefulness of the proposal is verified using the observation datasets in the real environment. It reveals that the proposed method enables us to precisely estimate the radio propagation compared to the two updating methods: a unique averaging and forgetting factor.

5.2 System Model

The local database first creates an initial radio map based on crowdsourcing as shown in Sect. 1.4.2.

The radio map is assumed to be updated based on a simple algorithm. Here, we consider an outdoor smart factory as the private 5G. In this system, a transmitter and mobile terminals are deployed in the factory. Mobile terminals observe the radio environment, such as instantaneous received signal power, at each position and upload the measured data to the local database. Then, a radio map is created by deriving the mean power in the local database.

As described in Sect. **??**, an instantaneous power may fluctuate depending on the individual differences between mobile terminals [62] in the real environment. However, this chapter mainly evaluates the accuracy of the radio map updating; thus, the fluctuation of the received signal power is not considered in this chapter. To generate an accurate radio map further, the calibration technique will be utilized.

Here, in this chapter, the transmitter is assumed to be moved depending on the manufacturing process. Thus, the movement cycle of the transmitter is every a few days or weeks [87]; that is, the transmitter slowly varies in the communication area. Moreover, this chapter assumes that the location information of the transmitter is unknown because we aim to update the radio map based on the hypothesis testing without the location information of the transmitter.

For readability, we define that a *phase 0* and *phase 1* are the environments before and after the location of the transmitter is varied.

If the number of samples becomes n_{ini} in the m-th mesh, the mean power is calculated in phase 0 for the m-th mesh. After the local database generates the initial radio map, the radio environment is repeatedly observed by the mobile terminals in the same communication area. Then, mobile terminals upload the observed samples to the local database. If the average value is calculated of the m-th mesh in phase 0, the new uploaded samples are considered as time-series data $\boldsymbol{R}_m$ of the m-th mesh. Then, the local database divides $\boldsymbol{R}_m$ such that the sample size in each dataset is n_{new} if the number of new samples is larger than n_{new}. Thus, we define $\boldsymbol{R}_m$ and $\boldsymbol{R}_{m,T+\tau}$ as follows:

$$\boldsymbol{R}_m = (\boldsymbol{R}_{m,T}, \boldsymbol{R}_{m,T+1}, \cdots, \boldsymbol{R}_{m,T+N_s-1}), \tag{5.1}$$

$$\boldsymbol{R}_{m,T+\tau} = (P_{m,T+\tau,0}, \cdots, P_{m,T+\tau,n_{\mathrm{new}}-1}), \tag{5.2}$$

where T is an arbitrary time and τ ($\tau = 0, \cdots, N_s - 1$) is the time index in each dataset. N_s is the number of datasets and $P_{m,T+\tau,j}$ [dBm] is the j-th instantaneous received signal power of the τ-th time in the m-th mesh.

This chapter assumes that the transmitter moves over a period of a few weeks and the mobile terminals observe the radio environment every a few days. Then, the observed contents are uploaded to the local database after the observations because the radio map should be updated more frequently than the transmitter's movement.

5.3 Proposed Method

This section first summarizes statistical testing and Welch's t-test. After that, we propose the method for updating the radio map using hypothesis testing.

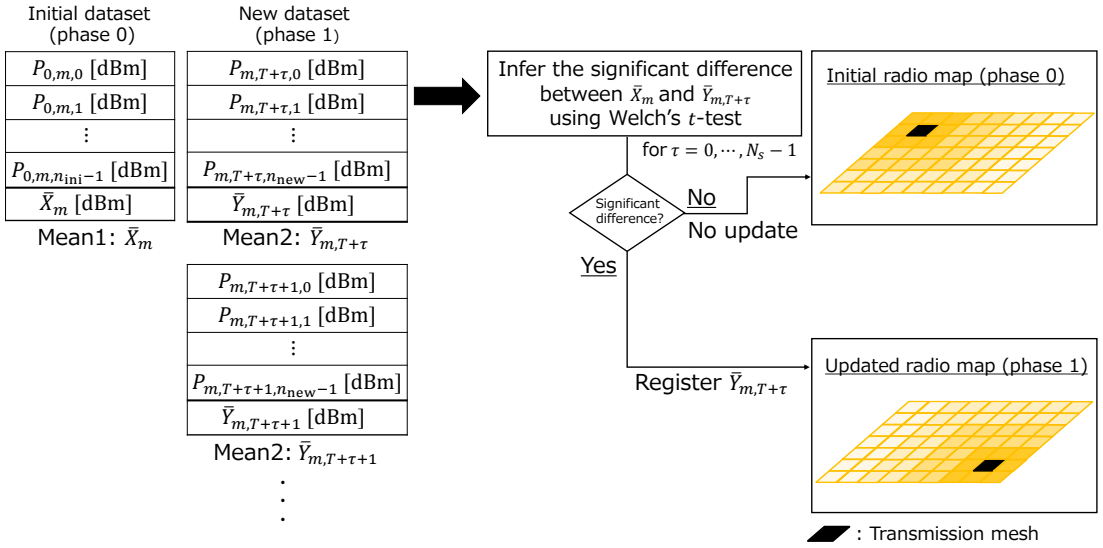


Fig. 5.1 The updating of the radio map.

5.3.1 Hypothesis Testing

In cellular systems, a large number of training datasets may not be obtained because it is difficult to repeatedly vary in the transmission location and observations. In addition, the prior distribution may not be accurately modeled because of the complicated radio propagation in the time domain. Thus, the radio map should be updated by determining whether the average received signal power significantly changes using only real-time datasets. To resolve this task, Welch's t-test is utilized. This method is established by improving the t-test. This testing method enables us to properly investigate the significant difference between two average values, regardless of the heteroscedasticity and homoscedasticity. Furthermore, Welch's t-test is superior to the general t-test and non-parametric testing [101]. Fig. 5.1 presents the concept of the proposed updating method. The local database infers the significant difference of the two average power values using Welch's t-test. A new mean value is stored in the mesh if the local database detects the significant difference.

5.3.2 Welch's t-Test

We define $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ as the datasets, each obtained from a log-normal distribution. For each dataset, the sample mean values are given as $\bar{X}_{\mathrm{w}}$ and $\bar{Y}_{\mathrm{w}}$. Moreover, the unbiased sample variances are represented as $S^2_{0,\mathrm{w}}$ and $S^2_{1,\mathrm{w}}$. The sample sizes of two datasets are n_{ini} and n_{new}, respectively. As a first step, we define two hypotheses: a null hypothesis H_0 and an

alternative hypothesis H_1, as follows:

$$H_0 : \mu_{0,\mathrm{w}} = \mu_{1,\mathrm{w}}, \tag{5.3}$$

$$H_1 : \mu_{0,\mathrm{w}} \neq \mu_{1,\mathrm{w}}, \tag{5.4}$$

where $\mu_{0,\mathrm{w}}$ [dBm] and $\mu_{1,\mathrm{w}}$ [dBm] denote population means of each log-normal distribution.

Next, the two variables: a t-value and a degree of freedom v, are represented as,

$$t = \frac{\bar{X}_{\mathrm{w}} - \bar{Y}_{\mathrm{w}}}{\sqrt{\frac{S_{0,\mathrm{w}}^2}{n_{\mathrm{ini}}} + \frac{S_{1,\mathrm{w}}^2}{n_{\mathrm{new}}}}}, \tag{5.5}$$

$$v = \frac{\left(\frac{S_{0,\mathrm{w}}^2}{n_{\mathrm{ini}}} + \frac{S_{1,\mathrm{w}}^2}{n_{\mathrm{new}}}\right)^2}{\frac{\left(\frac{S_{0,\mathrm{w}}^2}{n_{\mathrm{ini}}}\right)^2}{n_{\mathrm{ini}}-1} + \frac{\left(\frac{S_{1,\mathrm{w}}^2}{n_{\mathrm{new}}}\right)^2}{n_{\mathrm{new}}-1}}. \tag{5.6}$$

We can utilize the t-value to understand the difference between two average values. The t value is determined according to the difference between $\mu_{0,\mathrm{w}}$ and $\mu_{1,\mathrm{w}}$. The degree of freedom v means the variance of the Student's t distribution. Next, the probability distribution of the t-value named as the Student's t distribution is defined as follows:

$$f(t,v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}\left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}, \tag{5.7}$$

where $\Gamma(\cdot)$ denotes the gamma function. Here, it is necessary to calculate v as an integer; thus, we round off v if it is a decimal.

Next, a p-value is introduced to infer the significant difference. This value can be obtained by calculating the probability that the random variable $X \sim f(t,v)$ becomes higher or lower than the t-value. In other words, the p-value is equivalent to the tail region of the Student's t distribution. We calculate the p-value by considering a two-sided test. This method considers a rejection region in both tails of the Student's t distribution. In statistics, we can usually utilize a one-sided test only if the variation of the population mean can be calculated deterministically. However, the variation of the average received signal power value may not be known in advance because the location of the transmitter after the movement is not known. Additionally, in most cases, it is statistically recommended to use the two-sided test rather than the one-sided test. In the two-sided test, the alternative hypothesis H_1 is adopted if the

two-tailed p-value is below the significance level α. The conditional inequality is given by

$$\int_{-\infty}^{-t} f(\omega, v)\, d\omega + \int_{+t}^{+\infty} f(\omega, v)\, d\omega < \alpha, \tag{5.8}$$

where ω denotes the integral variable. The left hand side in the above equation is the two-tailed p-value. If there is a large difference between $\mu_{0,\mathrm{w}}$ and $\mu_{1,\mathrm{w}}$, the p-value becomes to 0. As a result, we tend to accept the alternative hypothesis.

5.3.3 Updating Procedures Using Welch's *t*-Test

We define that $\boldsymbol{W}_{0,m}$ denotes a dataset of an instantaneous power samples of the m-th mesh in phase 0. The function is defined as follows:

$$\boldsymbol{W}_{0,m} = (P_{0,m,0}, P_{0,m,1}, \cdots, P_{0,m,n_{\mathrm{ini}}-1}), \tag{5.9}$$

where $P_{0,m,j}$ [dBm] is defined as the j-th instantaneous power of the m-th mesh in phase 0.

The dataset for phase 1 is described. The mobile terminals upload the newly measured datasets to the local database after the measurement; thus, several datasets are obtained on each mesh. Here, the p-value of Welch's t-test becomes too small if the number of samples is too large. As a result, we may incorrectly adopt the alternative hypothesis H_1 due to the small p-value. To solve this issue, the hypothesis testing is conducted in each $\boldsymbol{R}_{m,T+\tau}$ expressed in Eq. (5.2) after the observation. Here, $\boldsymbol{W}_{1,m,T+\tau}$ is represented as $\boldsymbol{R}_{m,T+\tau}$ of the m-th mesh in phase 1. Moreover, we define $\bar{X}_m$ [dBm] and $\bar{Y}_{m,T+\tau}$ [dBm] as the mean values of $\boldsymbol{W}_{0,m}$ and $\boldsymbol{W}_{1,m,T+\tau}$, respectively. Each value can be defined as:

$$\bar{X}_m = \frac{1}{n_{\mathrm{ini}}} \sum_{j=0}^{n_{\mathrm{ini}}-1} P_{0,m,j}, \tag{5.10}$$

$$\bar{Y}_{m,T+\tau} = \frac{1}{n_{\mathrm{new}}} \sum_{j=0}^{n_{\mathrm{new}}-1} P_{m,T+\tau,j}. \tag{5.11}$$

The local database calculates these values in the decibel domain instead of the milliwatt domain since H_1 may not be properly accepted due to the outliers.

Next, the unbiased sample variances of $\boldsymbol{W}_{0,m}$ and $\boldsymbol{W}_{1,m,T+\tau}$ are introduced. These values are expressed as follows:

$$S_{0,m}^2 = \sum_{j=0}^{n_{\mathrm{ini}}-1} \frac{(P_{0,m,j} - \bar{X}_m)^2}{(n_{\mathrm{ini}} - 1)}, \tag{5.12}$$

$$S^2_{1,m,T+\tau} = \sum_{j=0}^{n_{\text{new}}-1} \frac{(P_{m,T+\tau,j} - \bar{Y}_{m,T+\tau})^2}{(n_{\text{new}} - 1)}. \tag{5.13}$$

Then, two important values: the t-value and degree of freedom are calculated as:

$$t_{m,T+\tau} = \frac{\bar{X}_m - \bar{Y}_{m,T+\tau}}{\sqrt{\frac{S^2_{0,m}}{n_{\text{ini}}} + \frac{S^2_{1,m,T+\tau}}{n_{\text{new}}}}}, \tag{5.14}$$

$$v_{m,T+\tau} = \frac{\left(\frac{S^2_{0,m}}{n_{\text{ini}}} + \frac{S^2_{1,m,T+\tau}}{n_{\text{new}}}\right)^2}{\frac{\left(\frac{S^2_{0,m}}{n_{\text{ini}}}\right)^2}{n_{\text{ini}}-1} + \frac{\left(\frac{S^2_{1,m,T+\tau}}{n_{\text{new}}}\right)^2}{n_{\text{new}}-1}}, \tag{5.15}$$

where $t_{m,T+\tau}$ is defined as the t-value of the m-th mesh at the k-th time. Additionally, $v_{m,T+\tau}$ denotes the degree of freedom of the m-th mesh at the k-th time. Here, $t_{m,T+\tau}$ follows the Student's t distribution. Finally, the local database derives the two-tailed p-value using Eq. (5.8).

Hypothesis testing is performed based on Eqs. (5.14), (5.15), and (5.8) in each mesh. Then, the local database judges the movement of the transmitter according to the tested results. If adopting H_1 in the m-th mesh, the local database accumulates the new mean value $\bar{Y}_{m,T+\tau}$ in the m-th mesh. Meanwhile, $\bar{Y}_{m,T+\tau}$ is not registered if H_1 is not adopted in the m-th mesh.

Although the radio map may be updated based on the threshold of the moving distance, this dissertation utilizes the significant difference between the two average values. This is because the fluctuation of the average value is notably different depending on the center frequency even if the moving distance is constant. Thus, it is difficult to determine the threshold of the moving distance for updating the radio map. As another criterion for updating the radio map, the spatial correlation of the shadowing may be used. Intuitively, the significant difference between two average power values may be small if the correlation coefficient is large. However, the shadowing value may be locally fluctuated according to the geographical conditions in a realistic environment even if the correlation coefficient is large. Thus, this dissertation infers the significant difference between two mean values in each mesh. Note that if the movement of the transmitter can be known by prior information, the radio map can be updated based on the prior information. Meanwhile, this chapter assumes that the location of the transmitter is unknown; thus, we must statistically update the radio map using a mathematical method, such as hypothesis testing.

If the number of $\boldsymbol{R}_{m,T+N_s-1}$ is not n_{new}, the local database does not perform hypothesis testing using $\boldsymbol{R}_{m,T+N_s-1}$. In statistics, hypothesis testing is sensitive to v that is determined according to the number of samples n_{new}.

Our scheme calculates Eqs. (5.10)–(5.13) to get $S^2_{0,m}$ and $S^2_{1,m,T+\tau}$. By substituting Eq. (5.10) into Eq. (5.12), it can be found that $S^2_{0,m}$ is calculated in the complexity of $O(n^2_{\text{ini}})$. Similarly, $S^2_{1,m,T+\tau}$ is derived in the complexity of $O(n^2_{\text{new}})$. Thus, our scheme has the complexity of $O(n^2_{\text{ini}} + n^2_{\text{new}})$.

Although we assume that two datasets $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ follow the log-normal distributions, Welch's t-test can be utilized even if each dataset is not obtained from the log-normal distribution because of the robustness of the t-test [71].

The proposed method updates the average received signal power in each mesh rather than the radio propagation model. If the local database updates the radio propagation model that was shown in Chapter 3, we need to consider a new method for inferring the significant variation of the radio propagation parameters, such as the path loss index. This task will be solved in our future work.

5.4 Statistical Power Analysis of Welch's t-Test

This chapter theoretically analyzes the statistical power of Welch's t-test for quantitative evaluation. In the following, we model the PDF of the t-value. Then, the statistical power is formulated. Finally, the simulation results of the statistical power are shown.

5.4.1 Theoretical Analysis of Statistical Power

It is well known that the statistical power [90] denotes the probability that the null hypothesis H_0 can be correctly rejected when it is false. Hence, the CDF of the test statistic can be utilized to derive the statistical power. The large statistical power means that significant difference tends to be inferred correctly.

This chapter assumes that two datasets, $\boldsymbol{A}_0$ and $\boldsymbol{A}_1$, are obtained. We define each dataset as:

$$\boldsymbol{A}_0 = (P_{0,0}, P_{1,0}, \cdots, P_{n_{\text{ini}}-1,0}), \tag{5.16}$$

$$\boldsymbol{A}_1 = (P_{0,1}, P_{1,1}, \cdots, P_{n_{\text{new}}-1,1}), \tag{5.17}$$

where $P_{i,0}$ [dBm] and $P_{i,1}$ [dBm] denote the i-th instantaneous received signal power in $\boldsymbol{A}_0$ and $\boldsymbol{A}_1$, respectively. As the simple radio propagation, we consider that each signal experiences a log-normal shadowing; thus, the probability distributions of $\boldsymbol{A}_0$ and $\boldsymbol{A}_1$ are

given as the the log-normal distributions $f_{\mathrm{LN}}\left(\mu_{0,\mathrm{w}}, \sigma^2_{0,\mathrm{w}}\right)$ and $f_{\mathrm{LN}}\left(\mu_{1,\mathrm{w}}, \sigma^2_{1,\mathrm{w}}\right)$, respectively. Here, the standard deviations are defined as $\sigma_{0,\mathrm{w}}$ [dB] and $\sigma_{1,\mathrm{w}}$ [dB], respectively.

If H_1 is correct, we first model t-value T_0 as follows

$$T_0 = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{S_0^2}{n_{\mathrm{ini}}} + \frac{S_1^2}{n_{\mathrm{new}}}}}, \tag{5.18}$$

where we define the sample means as $\bar{X}$ [dBm] and $\bar{Y}$ [dBm], respectively. $\delta = \mu_{0,\mathrm{w}} - \mu_{1,\mathrm{w}}$ [dB]. Additionally, the unbiased sample variances are expressed as S_0^2 and S_1^2, respectively. Each sample mean is calculated as,

$$\bar{X} = \frac{1}{n_{\mathrm{ini}}} \sum_{i=0}^{n_{\mathrm{ini}}-1} P_{i,0}, \tag{5.19}$$

$$\bar{Y} = \frac{1}{n_{\mathrm{new}}} \sum_{i=0}^{n_{\mathrm{new}}-1} P_{i,1}. \tag{5.20}$$

Additionally, S_0^2 and S_1^2 are given by

$$S_0^2 = \sum_{i=0}^{n_{\mathrm{ini}}-1} \frac{(P_{i,0} - \bar{X})^2}{n_{\mathrm{ini}} - 1}, \tag{5.21}$$

$$S_1^2 = \sum_{i=0}^{n_{\mathrm{new}}-1} \frac{(P_{i,1} - \bar{Y})^2}{n_{\mathrm{new}} - 1}. \tag{5.22}$$

Next, Eq. (5.18) is transformed as:

$$T_0 = \frac{\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}} + \frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}} + \frac{\delta}{\sqrt{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}} + \frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}}}{\sqrt{\frac{d_{\mathrm{w}} \frac{\frac{S_0^2}{n_{\mathrm{ini}}} + \frac{S_1^2}{n_{\mathrm{new}}}}{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}} + \frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}}{d_{\mathrm{w}}}}} = \frac{\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}} + \frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}} + \frac{\delta}{\sqrt{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}} + \frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}}}{\sqrt{\frac{\Omega}{d_{\mathrm{w}}}}}, \tag{5.23}$$

where we define d_{w} as a fixed number. This value is used in the non-central t distribution as the degree of freedom. Furthermore, Ω is defined as:

$$\Omega = d_{\mathrm{w}} \frac{\frac{S_0^2}{n_{\mathrm{ini}}} + \frac{S_1^2}{n_{\mathrm{new}}}}{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}} + \frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}. \tag{5.24}$$

It can be easily found that $\bar{X}$ and $\bar{Y}$ follow the log-normal distribution $f_{\mathrm{LN}}\left(\mu_{0,\mathrm{w}}, \sigma^2_{0,\mathrm{w}}/n_{\mathrm{ini}}\right)$, $f_{\mathrm{LN}}\left(\mu_{1,\mathrm{w}}, \sigma^2_{1,\mathrm{w}}/n_{\mathrm{new}}\right)$, respectively. Furthermore, $\delta = \mu_{0,\mathrm{w}} - \mu_{1,\mathrm{w}}$; hence, the first term of the numerator follows the log-normal distribution with the mean of 0 [dB] and the standard deviation of 1 [dB]. To formulate the statistical power, we need to show that the PDF of T_0 can be modeled as the non-central *t* distribution. In statistics, it is well known that if the PDF of Ω can be modeled as the Chi-squared distribution with d_{w} degree of freedom in Eq. (5.23), the non-central *t* distribution can be derived.

However, the probability distribution of Ω may not be theoretically formulated because this value is expressed by linearly combining the unbiased sample variances. Hence, we use the Welch–Satterthwaite equation [102] to approximately formulate the PDF of Ω by the Chi-squared distribution. The linear combination variable Φ is first defined as follows:

$$\Phi = \sum_{i=0}^{L-1} k_i s_i^2, \tag{5.25}$$

where s_i^2 is a sample variable. The degree of freedom of s_i^2 is expressed as v_i. k_i is a positive number. L denotes the number of sample variables. Next, we can approximately derive the effective degree of freedom $\hat{v}$ as,

$$\hat{v} \approx \frac{\left(\sum_{i=0}^{L-1} k_i s_i^2\right)^2}{\sum_{i=0}^{L-1} \frac{(k_i s_i^2)^2}{v_i}}. \tag{5.26}$$

The effective degree of freedom $\hat{d}_{\mathrm{w}}$ of Welch's *t*-test is derived by substituting $k_0 = \frac{c}{n_{\mathrm{ini}}}$, $k_1 = \frac{c}{n_{\mathrm{new}}}$, $v_0 = n_{\mathrm{ini}} - 1$, $v_1 = n_{\mathrm{new}} - 1$, $c = d_{\mathrm{w}}\left(\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}} + \frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}\right)$, $s_0^2 = S_0^2$, and $s_1^2 = S_1^2$ to Eq. (5.26). The value can be modeled as,

$$\hat{d}_{\mathrm{w}} \approx \frac{\left(\frac{S_0^2}{n_{\mathrm{ini}}} + \frac{S_1^2}{n_{\mathrm{new}}}\right)^2}{\frac{\left(\frac{S_0^2}{n_{\mathrm{ini}}}\right)^2}{n_{\mathrm{ini}}-1} + \frac{\left(\frac{S_1^2}{n_{\mathrm{new}}}\right)^2}{n_{\mathrm{new}}-1}}. \tag{5.27}$$

The above derivations are based on [102]. We may not utilize this method if k_i has a negative value; however, because c, n_{ini}, and n_{new} are positive values, we do not have to consider this problem.

Using Welch–Satterthwaite equation, we can model the probability distribution of T_0 as non-central *t* distribution. Because this chapter used a two-sided test, the CDF of the non-central *t* distribution is calculated by considering two tails. Thus, we can express the

statistical power $\pi_{\mathrm{w}}(n_{\mathrm{ini}}, n_{\mathrm{new}}, \delta)$ as,

$$\begin{aligned}\pi_{\mathrm{w}}(n_{\mathrm{ini}}, n_{\mathrm{new}}, \delta) &= 1 - \Pr\left(-t^{\frac{\alpha}{2}}_{\hat{d}_{\mathrm{w}}} \leq \frac{\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}}+\frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}} + \frac{\delta}{\sqrt{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}}+\frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}}}{\sqrt{\frac{d_{\mathrm{w}}\frac{s^2_0}{n_{\mathrm{ini}}}+\frac{s^2_1}{n_{\mathrm{new}}}}{\frac{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}}+\frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}{d_{\mathrm{w}}}}}} \leq +t^{\frac{\alpha}{2}}_{\hat{d}_{\mathrm{w}}} \right) \\ &\approx 1 - G_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}\left(+t^{\frac{\alpha}{2}}_{\hat{d}_{\mathrm{w}}}\right) + G_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}\left(-t^{\frac{\alpha}{2}}_{\hat{d}_{\mathrm{w}}}\right),\end{aligned} \tag{5.28}$$

where we express the CDF of the non-central t distribution as $G_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}(\cdot)$. $\hat{d}_{\mathrm{w}}$ denotes the effective degree of freedom. λ_{w} is the non-centrality parameter. $t^{\frac{\alpha}{2}}_{\hat{d}_{\mathrm{w}}}$ is t-value based on $\hat{d}_{\mathrm{w}}$. Here, $G_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}(\cdot)$ is represented as,

$$G_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}(t') = \begin{cases} F_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}(t') & (t' \geq 0) \\ 1 - F_{\hat{d}_{\mathrm{w}},-\lambda_{\mathrm{w}}}(t') & (t' < 0), \end{cases} \tag{5.29}$$

where we define the random variable that follows the non-central t distribution as t'. λ_{w} and $F_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}(t')$ are given by

$$\lambda_w = \frac{\delta}{\sqrt{\frac{\sigma^2_{0,\mathrm{w}}}{n_{\mathrm{ini}}}+\frac{\sigma^2_{1,\mathrm{w}}}{n_{\mathrm{new}}}}}, \tag{5.30}$$

$$F_{\hat{d}_{\mathrm{w}},\lambda_{\mathrm{w}}}(t') = \Psi(-\lambda_{\mathrm{w}}) + \frac{1}{2}\sum_{j=0}^{\infty}\left[\psi_j I\left(j+\frac{1}{2}, \frac{\hat{d}_{\mathrm{w}}}{2}\right) + \omega_j I\left(j+\frac{1}{2}, \frac{\hat{d}_{\mathrm{w}}}{2}\right)\right], \tag{5.31}$$

where the CDF of the standard normal distribution is expressed as$\Psi(\cdot)$. Additionally, we define the regularized incomplete beta function as I. Furthermore, ψ_j and ω_j are defined as

$$\psi_j = \frac{1}{j!}\exp\left(-\frac{\lambda^2_{\mathrm{w}}}{2}\right)\left(\frac{\lambda^2_{\mathrm{w}}}{2}\right)^j, \tag{5.32}$$

$$\omega_j = \frac{\lambda_{\mathrm{w}}}{\sqrt{2}\Gamma\left(j+\frac{3}{2}\right)}\exp\left(-\frac{\lambda^2_{\mathrm{w}}}{2}\right)\left(\frac{\lambda^2_{\mathrm{w}}}{2}\right)^j. \tag{5.33}$$

We can utilize the above distribution in Welch's t-test if the PDFs of $\boldsymbol{A}_0$ and $\boldsymbol{A}_1$ are represented as the log-normal distributions. The shadowing empirically follows the log-normal distribu-

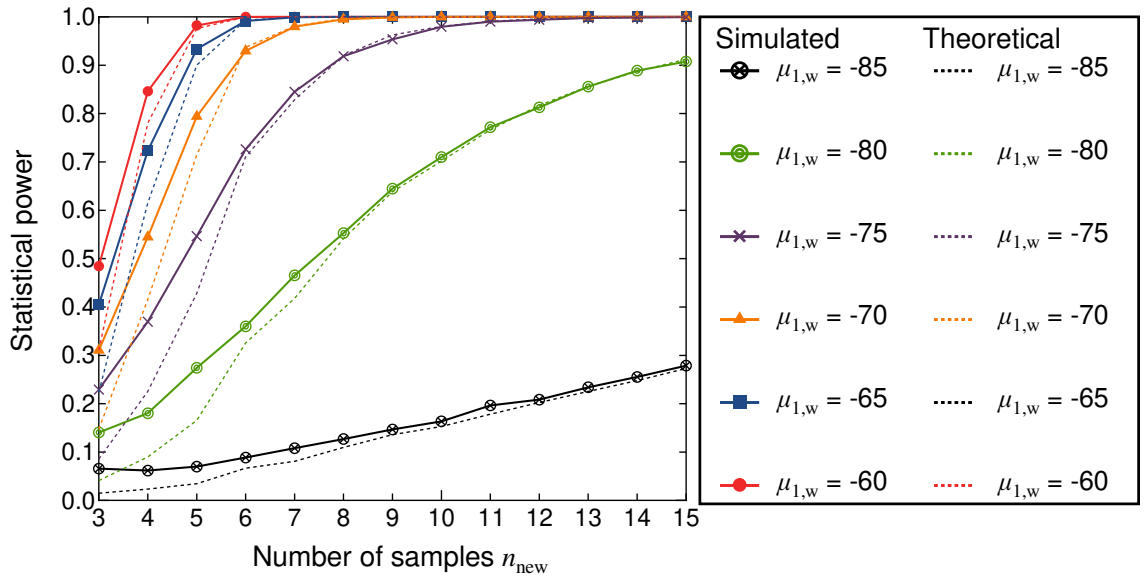


Fig. 5.2 Theoretical and simulated statistical power.

tion in the real environment. Considering this fact, the t distribution and the non-central t distribution can be used in the radio map construction.

5.4.2 Numerical Results of Statistical Power

This subsection presents the numerical results of the theoretical statistical power. In this simulation, n_{ini} random samples are created from $f_{\mathrm{LN}}(\mu_{0,\mathrm{w}}, \sigma^2_{0,\mathrm{w}})$. Next, n_{new} random samples are generated from $f_{\mathrm{LN}}(\mu_{1,\mathrm{w}}, \sigma^2_{1,\mathrm{w}})$. Then, Welch's t-test is used and the number of times that $H_1 : \mu_{0,\mathrm{w}} \neq \mu_{1,\mathrm{w}}$ is adopted is counted. This trial is repeated 10,000 times and the simulated statistical power is evaluated.

Fig. 5.2 presents numerical results of the statistical power where $\mu_{0,\mathrm{w}} = -80.0\,[\mathrm{dBm}]$, $\sigma_{0,\mathrm{w}} = \sigma_{1,\mathrm{w}} = 8\,[\mathrm{dB}]$, and $n_{\mathrm{ini}} = 50$. The theoretical value is calculated using $\pi_{\mathrm{w}}(n_{\mathrm{ini}}, n_{\mathrm{new}}, \delta)$. Note that α is 0.01. We can confirm that the simulation values are almost equivalent to the theoretical values. In the small sample size, the difference is slightly large owing to the deviation of sample means. Additionally, a large difference between $\mu_{0,\mathrm{w}}$ and $\mu_{1,\mathrm{w}}$ means that the significant difference can be correctly inferred even if the sample size is small.

5.5 Comparative Methods

This section presents several comparisons to verify the accuracy of updating the radio map. We first present a unique averaging-based method. Next, a forgetting factor is utilized to

update the radio map. Finally, as the non-parametric method, this chapter uses the Mann–Whitney U test.

5.5.1 Unique Averaging-Based Method

As the simple updating of the radio map, we can use a method that the average power is uniquely calculated in each mesh [17, 32]. The calculation equation of this method is defined as follows:

$$\bar{P}_{\mathrm{uni},m} = \frac{1}{n_{\mathrm{uni},m}} \left(\sum_{i=0}^{n_{\mathrm{ini}}-1} P_{0,m,i} + \sum_{\tau=0}^{N_s-1} \sum_{j=0}^{n_{\mathrm{new}}-1} P_{m,T+\tau,j} \right), \tag{5.34}$$

where the updated value is defined as $\bar{P}_{\mathrm{uni},m}$ [dBm]. This is derived based on the samples in phase 0 and phase 1 in the m-th mesh. $n_{\mathrm{uni},m}$ denotes the sample size of the received signal power accumulated in the m-th mesh in phases 0 and 1. After the measurements, the average value is derived in each mesh using Eq. (5.34); however, this method does not consider the movement of the transmitter. Hence, the mean value may not be accurately calculated due to the unique averaging; meanwhile, the radio map can be constructed with a small number of updates.

5.5.2 Forgetting Factor-Based Method

We utilize forgetting factor [103, 104] as the second comparison method. This method updates the average power based on the following equation:

$$\bar{P}_{\mathrm{forget},m,T+\tau} = \eta_{\mathrm{forget}}(\bar{P}_{\mathrm{forget},m,T+\tau-1}) + (1-\eta_{\mathrm{forget}})(\bar{Y}_{m,T+\tau}) \quad [\mathrm{dBm}], \tag{5.35}$$

where the updated value at time $T+\tau$ of the m-th mesh is defined as $\bar{P}_{\mathrm{forget},m,T+\tau}$. Moreover, $\bar{P}_{\mathrm{forget},m,T+\tau-1}$ [dBm] is the average power at time $T+\tau-1$. Additionally, the forgetting factor is given as η_{forget}. If using a smaller η_{forget}, the new value can be calculated using a larger weight. Meanwhile, the radio map may be created with large number of updates since it is necessary to continuously update the average value N_s times on the m-th mesh.

$\bar{P}_{\mathrm{forget},m,T+\tau-1}$ is defined as $\bar{X}_m$ to update the radio map in $\tau = 0$. Furthermore, the local database does not perform the aforementioned calculation if the sample size of $\boldsymbol{R}_{m,T+N_s-1}$ is not n_{new}. The mean value may not be accurately derived in such the case owing to the multipath fading; thus, we consider the sample size of $\boldsymbol{R}_{m,T+N_s-1}$.

5.5.3 Mann–Whitney U Test-Based Method

Finally, the Mann–Whitney U test [105], a non-parametric testing method, is explained. In a realistic environment, it is difficult to model the received signal power as a typical probabilistic distribution because of complicated radio propagation properties. Mann–Whitney U test enables us to test the difference between two average power values based on the median values. Additionally, this method does not have to assume a certain distribution; thus, we utilize this method as a comparison. Several researchers have used this test in recent years [105–107].

The statistical power of the Mann–Whitney U test is poorer than the one of Welch's t-test [108]. Thus, we may not appropriately infer the difference between two average power values even if the difference is very large.

5.6 Simulation Descriptions

This section describes computer simulations to evaluate the accuracy of the proposed method. In the following, the instantaneous propagation model is explained. Then, we present the simulation procedures.

5.6.1 Instantaneous Model

The instantaneous received signal power values in phases 0 and 1 are derived as follows:

$$P_0(\boldsymbol{x}_{\mathrm{Tx}}, \boldsymbol{x}_{\mathrm{Rx}}) = P_{\mathrm{Tx}} - L_{\mathrm{fspl}}(d_0) - 10C\log_{10}\left(\frac{\|\boldsymbol{x}_{\mathrm{Tx}} - \boldsymbol{x}_{\mathrm{Rx}}\|}{d_0}\right) + W_{0,\mathrm{s}} + F_0, \tag{5.36}$$

$$P_1(\boldsymbol{x}'_{\mathrm{Tx}}, \boldsymbol{x}_{\mathrm{Rx}}) = P_{\mathrm{Tx}} - L_{\mathrm{fspl}}(d_0) - 10C\log_{10}\left(\frac{\|\boldsymbol{x}'_{\mathrm{Tx}} - \boldsymbol{x}_{\mathrm{Rx}}\|}{d_0}\right) + W_{1,\mathrm{s}} + F_1, \tag{5.37}$$

where $P_0(\boldsymbol{x}_{\mathrm{Tx}}, \boldsymbol{x}_{\mathrm{Rx}})$ [dBm] is an instantaneous received signal power value in phase 0. Additionally, $P_1(\boldsymbol{x}'_{\mathrm{Tx}}, \boldsymbol{x}_{\mathrm{Rx}})$ [dBm] denotes an instantaneous received signal power value in phase 1. $\boldsymbol{x}_{\mathrm{Tx}}$ and $\boldsymbol{x}'_{\mathrm{Tx}}$ are the transmitter positions in phase 0 and phase 1, respectively, $\boldsymbol{x}_{\mathrm{Rx}}$ is the position of a terminal, $W_{0,\mathrm{s}}$ [dB] and $W_{1,\mathrm{s}}$ [dB] denote the shadowing values in phase 0 and phase 1, respectively. This dissertation assumes that these values follow the log-normal shadowing. Furthermore, we express the multipath fading in phase 0 and phase 1 as F_0 [dB] and F_1 [dB], respectively.

In this simulation, the shadowing has a spatial correlation. Each shadowing correlation of $W_{0,\mathrm{s}}$ and $W_{1,\mathrm{s}}$ is assumed to follow Eq. (4.10).

Table 5.1 The common simulation parameters

Communication area [m^2]	400×400
Mesh size [m^2]	10×10
The number of terminals H	20
The number of samples n_{ini}	10
Transmit power P_{Tx} [dBm]	29
Path loss coefficient C	4.5
Reference distance d_0 [m]	10
Shadowing $\sigma_{i,s}$ [dB]	8
F_0 and F_1	i.i.d. Rayleigh fading
Frequency [MHz]	3500
Average noise level [dBm]	-140
Forgetting factor η_{forget}	0.1
Significance level α	0.01

The fading values are modeled as independent and identically distributed (i.i.d.) Rayleigh distributions.

5.6.2 Simulation Procedures

The simulations are performed as follows:

i). The local database creates the initial radio map in phase 0 using instantaneous power sample defined by Eq. (5.36).

ii). The radio environment is observed by H mobile terminals in the communication area. Here, whether the transmitter moves or does not move is randomly selected.

iii). Hypothesis testing is performed based on Eqs. (5.14), (5.15), and (5.8) to infer whether or not the transmitter has moved.

iv). The local database updates the mean value in a mesh that Eq. (5.8) is met.

We repeated Step (i) 1,000 times, and Step (ii) to (iv) were performed 1,000 times each time Step (i) was performed. After that, the estimation accuracy of constructed radio maps was verified. Table 5.1 shows the common simulation parameters.

Here, the positional relation of the surrounding buildings is assumed not to be varied during the measurement and after the creation of a radio map. This chapter aims to verify the hypothesis testing-based updating method of the radio map when the transmitter moves. Thus, we consider such an assumption for the surrounding buildings.

The significance level α is 0.01. This is because it is a typical α in hypothesis testing.

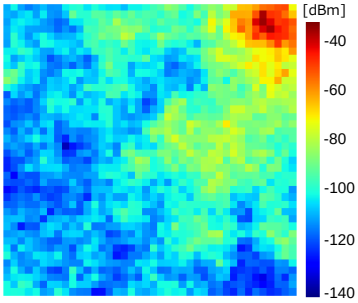
5.7 Simulation results

This section visualizes the updated radio maps. After that, the error characteristics of the radio maps are described. Finally, Type I error and Type II error are shown.

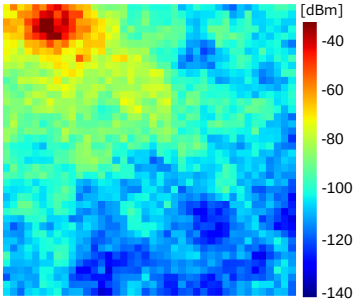
5.7.1 Example of Radio Maps

Fig. 5.3 shows an example of updated radio maps; Figs. 5.3(a) and 5.3(b) express the true radio maps. The sample size of each mesh is 10,000 in these maps. We use $n_{\mathrm{new}} = 50$ and $d_{\mathrm{cor}} = 20$ [m]. The transmitter moved 300 [m] to the left direction in the phase 1. By comparing with Figs. 5.3(b), 5.3(d), and 5.3(f), we can see that the radio map can be appropriately updated in each testing method. However, in Figs. 5.3(b) and 5.3(f), the large error can be confirmed for some meshes in the upper left of the area. This is because the Mann–Whitney U test may not accurately infer the significant difference compared to Welch's t-test even if the transmitter moves large distance. Hence, Welch's t-test is superior to the Mann–Whitney U test to update the radio map.

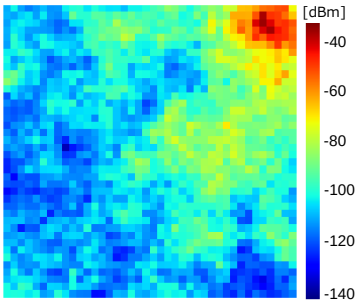
Meanwhile, it can be found that the accuracy of the updated radio maps is very degraded in the two comparison methods from Figs. 5.3(g) and 4.1(e) because the movement of the transmitter is not considered in these methods.



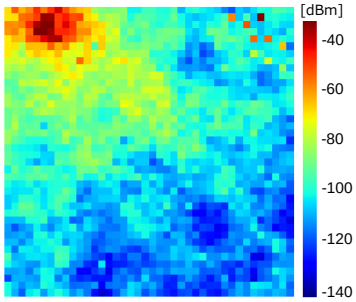
(a) True map (phase 0).



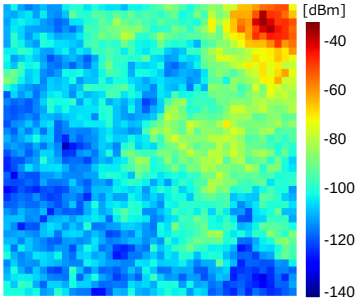
(b) True map (phase 1).



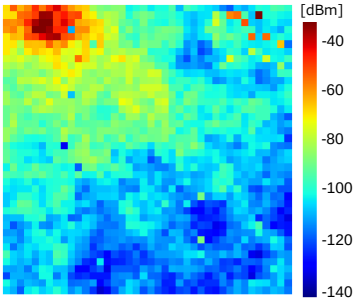
(c) Welch (phase 0).



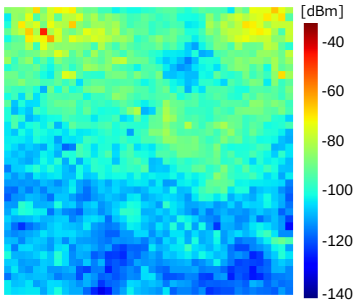
(d) Welch (phase 1).



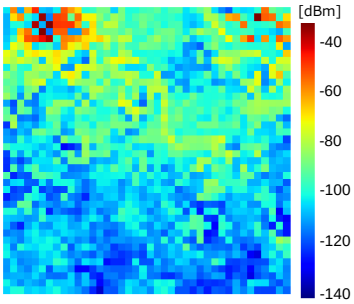
(e) Mann–Whitney (phase 0).



(f) Mann–Whitney (phase 1).



(g) Unique averaging.



(h) Forgetting factor.

Fig. 5.3 Example of updated radio maps.

5.7.2 Estimation Accuracy

This subsection evaluates the estimation accuracy of the radio maps based on mean absolute error (MAE). The MAEs in the phase 0 and phase 1 are given by

$$e_{0,\mathrm{MAE}} = \frac{1}{D_0} \sum_{m=0}^{D_0-1} |\bar{P}_{0,\mathrm{true},m} - \bar{P}_{0,m}| \quad [\mathrm{dB}], \tag{5.38}$$

$$e_{1,\mathrm{MAE}} = \frac{1}{D_1} \sum_{m=0}^{D_1-1} |\bar{P}_{1,\mathrm{true},m} - \bar{P}_{1,m}| \quad [\mathrm{dB}], \tag{5.39}$$

where we define the true mean value of the m-th mesh in phase 0 as $\bar{P}_{0,\mathrm{true},m}$ [dBm]. Similarly, $\bar{P}_{1,\mathrm{true},m}$ [dBm] is defined as the true mean value of the m-th mesh in phase 1, respectively. Moreover, an initial mean value of the m-th mesh is expressed as $\bar{P}_{0,m}$ [dBm] and an updated mean value of the m-th mesh is defined as $\bar{P}_{1,m}$ [dBm]. Additionally, D_0 is the number of meshes that H_0 is adopted on either Welch's t-test or the Mann–Whitney U test. D_1 is the number of meshes that H_1 is accepted on either Welch's t-test or the Mann–Whitney U test. The main purpose of this chapter is to verify the effectiveness of the hypothesis testing-based updating method; hence, we need to evaluate the accuracy in the meshes that each hypothesis is accepted on either Welch's t-test or the Mann–Whitney U test. Thus, D_0 and D_1 are defined separately.

If we use RMSE for the performance evaluation, the dynamic range wastefully increases because of the sum of squares of errors. Therefore, this chapter utilizes the MAE. The RMSE is nonlinearly calculated; meanwhile, the MAE is linearly derived using the absolute error. By this calculation, we can suppress the impacts of outliers.

The error characteristics are shown in Fig. 5.4 with $d_{\mathrm{cor}} = 20$ [m]. Here, the location of the transmitter is randomly changed in the area for each step (a). We can understand that Welch's t-test accurately estimates radio environment with the accuracy as the Mann–Whitney U test. This is because the transmitter may move too short distances. Here, the errors of these methods are 0 [dB] in phase 0 since those methods have the initial and updated radio maps separately. Obviously, the performances of the comparison methods are degraded compared to the hypothesis testing-based methods since the movement of the transmitter is not detected.

Fig. 5.5 shows the estimation accuracy versus the correlation distance d_{cor}. The transmitter randomly moved in each step (a). Note that n_{new} is 50. As can be seen, the error characteristics of the hypothesis testing-based methods are superior to the comparative methods regardless of d_{cor}.

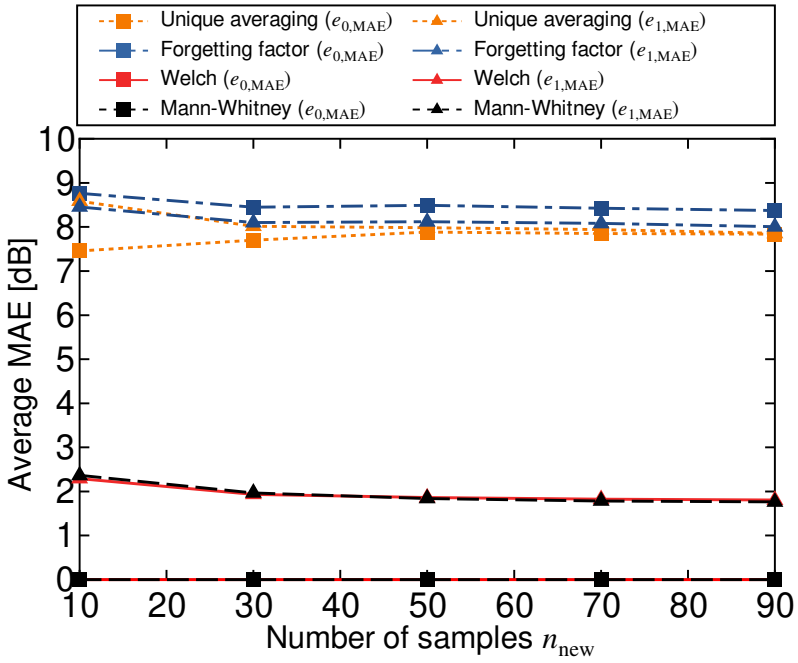


Fig. 5.4 Estimation accuracy versus the number of samples n_{new}.

Next, we show the estimation accuracy versus the movement distance of the transmitter in Fig. 5.6. The transmitter randomly moved according to the value of the horizontal axis. Note that $n_{\mathrm{new}} = 50$ and $d_{\mathrm{cor}} = 20$ [m] were utilized. We can argue that hypothesis testing-based methods more elaborately estimates the radio environment than the other methods. Moreover, Welch's t-test can improve the radio propagation estimation compared to the Mann–Whitney U test. This is because Mann–Whitney U test may not correctly infer the alternative hypothesis if the transmitter moves a large distance; therefore, the average MAE is slightly degraded.

Finally, Fig. 5.7 depicts the estimation accuracy versus the forgetting factor. The transmitter randomly moved 200 [m] in the communication area. Here, $n_{\mathrm{new}} = 50$ and $d_{\mathrm{cor}} = 20$ [m] were used. The results show that the forgetting factor-based method cannot skillfully update the radio map. Although $e_{0,\mathrm{MAE}}$ can be slightly small as the forgetting factor is 0.9, $e_{1,\mathrm{MAE}}$ becomes large owing to large factor value. Thus, there is a trade-off in the forgetting factor-based method. In the simulation, whether the transmitter moves or does not move is randomly selected. Thus, the estimation accuracy may not be improved even if the forgetting factor is changed. We conclude that if the estimation errors of the radio maps are guaranteed in both phases 0 and 1, the forgetting factor should be 0.5–0.7. Meanwhile, the factor value must be properly selected according to the importance of the new and old measurement data.

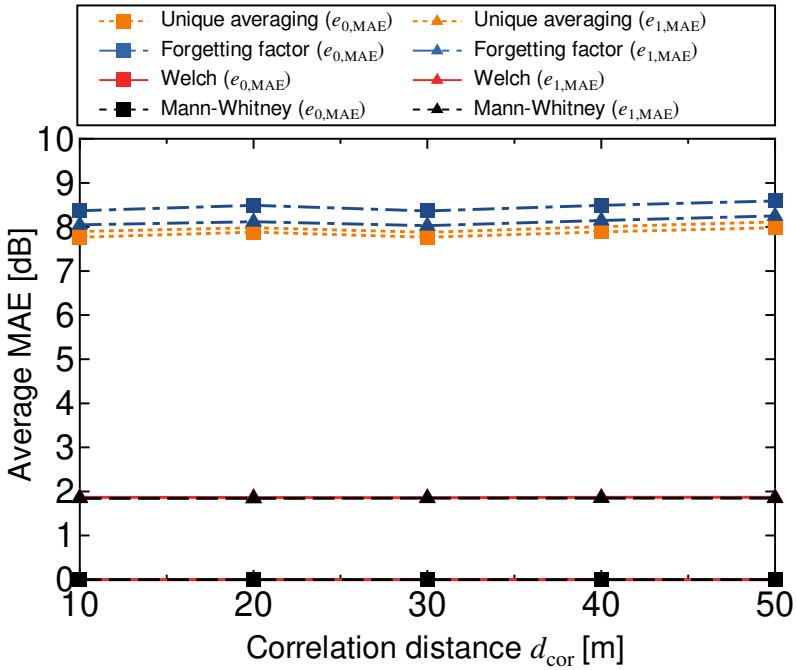


Fig. 5.5 Estimation accuracy versus the correlation distance d_{cor}.

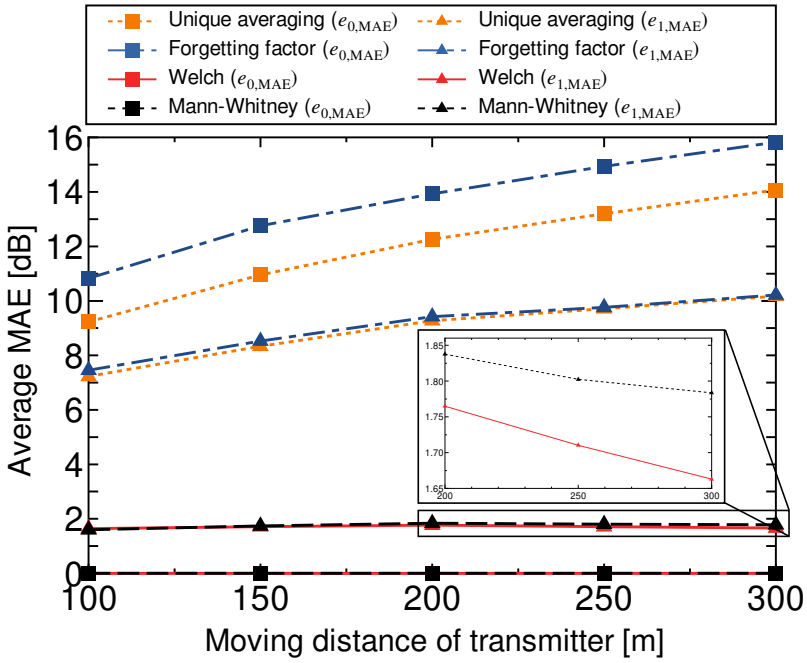


Fig. 5.6 Estimation accuracy versus moving distance of the transmitter.

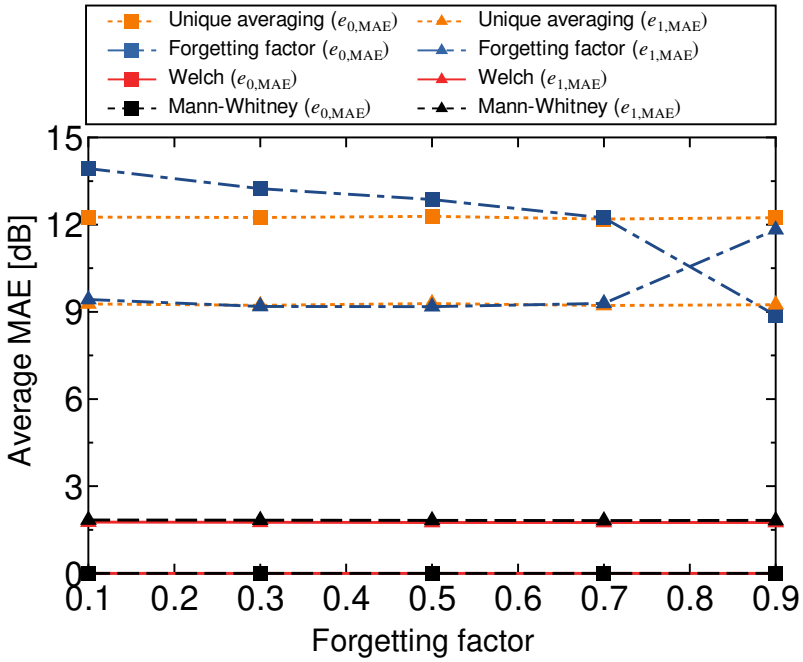


Fig. 5.7 Estimation accuracy versus forgetting factor.

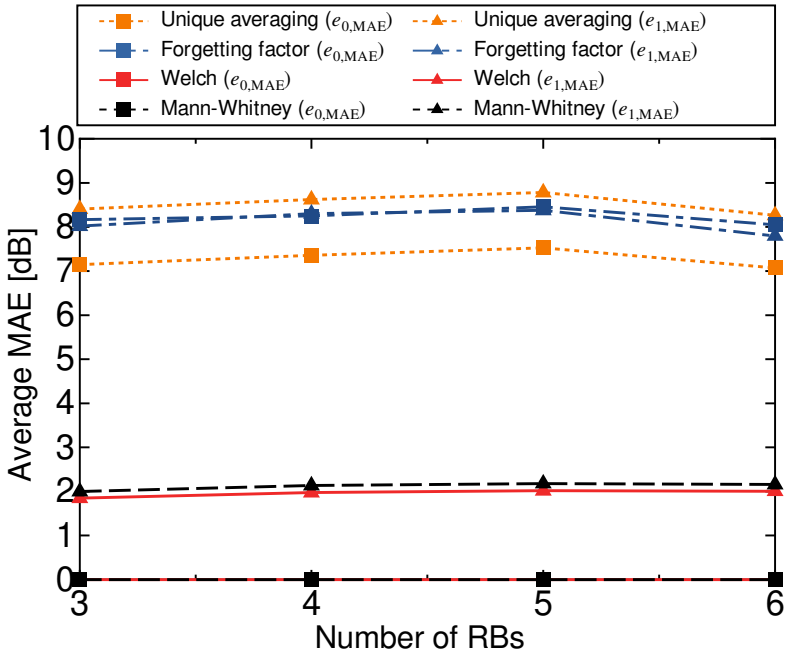


Fig. 5.8 Estimation accuracy over a frequency-selective fading channel.

5.7.3 Performance Evaluation over Frequency-Selective Fading Channel

The instantaneous power may be different according to the bandwidth. Here, a strong correlation exists between the mean power over the frequency domain [109, 110]. These facts may indicate that *the estimation accuracy of the mean power (and that of the proposed method) may not be affected by the bandwidth.*

Meanwhile, the instantaneous power fluctuates depending on the bandwidth; therefore, the frequency-selectivity should be considered in the performance evaluation. In the 3rd generation partnership project long-term evolution [111], the signal can be transmitted on multiple resource blocks (RBs) using orthogonal frequency division multiplexing (OFDM). The bandwidth of one RB is 1.4 [MHz]. This subsection assumes that each RB experiences i.i.d. Rayleigh fading.

The estimation accuracy is illustrated in Fig. 5.8. In this simulation, the frequency selectivity is strong in an increase of the number of RBs. We can argue that the performance of the proposed method is suitable even in the frequency-selective channel. The deviation of the multipath fading is smaller than the difference of the path loss and shadowing; hence, the high estimation accuracy may be realized in the proposed method even if the bandwidth changes.

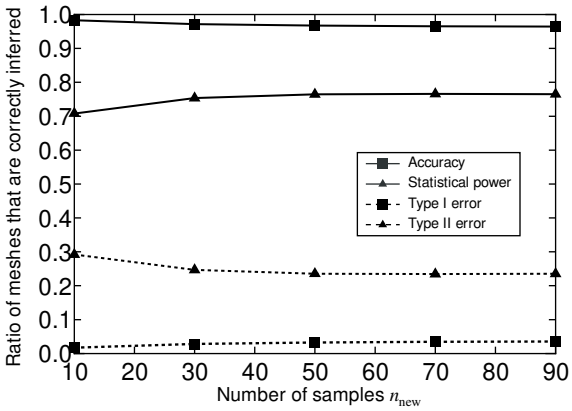


Fig. 5.9 Accuracy of testing versus n_1.

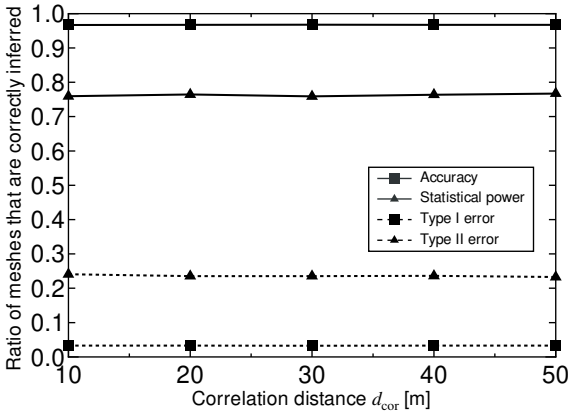


Fig. 5.10 Accuracy of testing versus d_{cor} © 2020 IEEE.

5.7.4 Accuracy of Testing

This subsection describes Type I and II errors to confirm whether the radio map can be correctly updated by the proposed method. In this dissertation, Type I error is defined as the ratio of meshes that the local database incorrectly updated the mean value, out of all 1,600 meshes when the transmitter does not move. Meanwhile, Type II error denotes the ratio of meshes that the local database does not correctly update the mean value, out of all 1,600 meshes when the location of the transmitter has been varied. Moreover, the accuracy and statistical power were calculated by subtracting the Type I and II errors from 1.

Fig. 5.9 is the accuracy of the testing in the proposed method. The location of the transmitter was randomly varied in each step (a), and d_{cor} is 20 [m]. As can be seen from the figure, the hypothesis can be correctly inferred in many meshes when the location of the transmitter is not varied. Meanwhile, the Type II error is degraded compared with the Type I error because the radio map may not properly be updated when the movement distance of

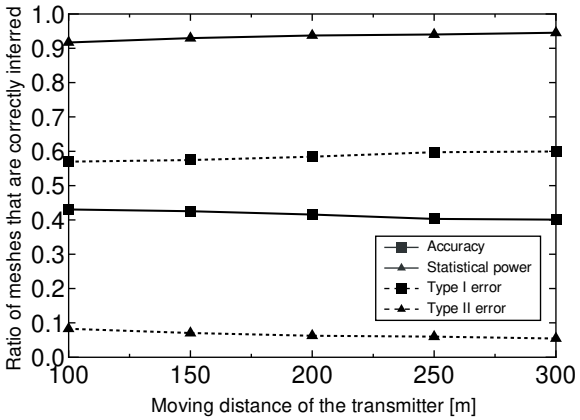


Fig. 5.11 Accuracy of testing versus moving distance of the transmitter © 2020 IEEE.

the transmitter is too small; that is, the significant difference is small between the two mean values. However, the estimation accuracy in such environments is significantly suitable as shown in Fig. 5.6.

Fig. 5.10 visualizes the accuracy of testing versus d_{cor} as n_{new} is 50. Here, the location of the transmitter was randomly changed in each step (a). This figure clarifies that the hypothesis can correctly be inferred in the proposed testing method regardless of the value of d_{cor}.

Finally, the accuracy of the testing versus the movement distance of the transmitter is presented in Fig. 5.11 as $n_{\mathrm{new}} = 50$ and $d_{\mathrm{cor}} = 20$ [m]. The results mean that the mean value can be updated with high accuracy in most meshes via the proposed method. Especially, the proposed method can improve the Type II error as the movement distance of the transmitter becomes large since the path loss greatly changes between phases 0 and 1. As a result, a difference of the mean values can be found; that is, the alternative hypothesis H_1 is accepted.

5.8 Performance Verification Using Measured Datasets

Finally, the performance verification was conducted of the radio map updating by utilizing measured samples that were obtained in our early work [112]. We had conducted the V2V communications to get the samples. These samples are not observed in cellular systems; however, we can utilize these samples to verify the effectiveness of our method since i) V2V communications uses OFDM and ii) we created several radio maps in each location.

The experiment was carried out in California Path, UC Berkeley, USA, and observations were conducted over two days in July 2018. As the experimental methods, after we prepared three vehicles implemented an onboard unit, each vehicle communicated with each other based on IEEE 802.11p standard. Fig. 5.12 is the observation route. The three vehicles

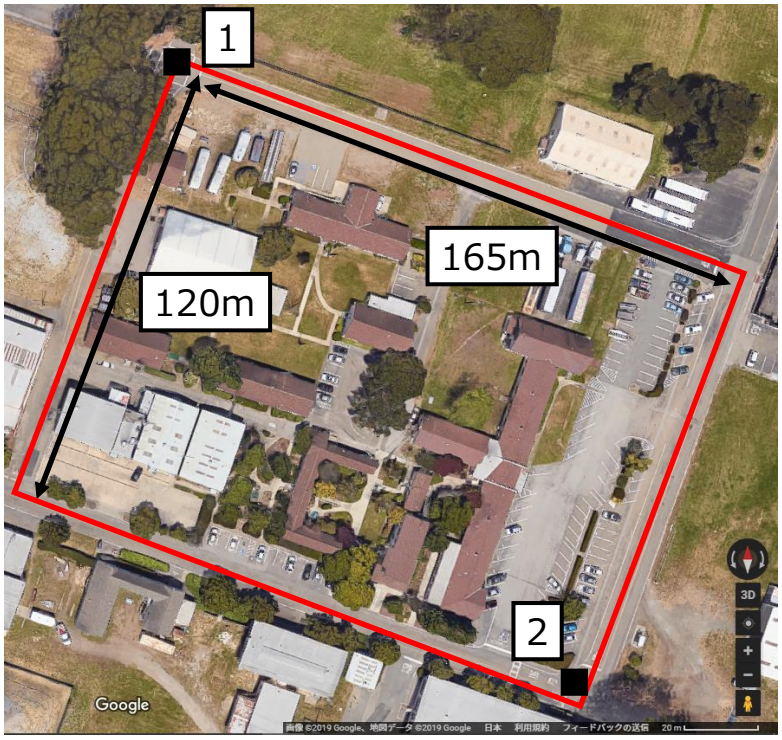


Fig. 5.12 Measurement area of California Path © 2020 IEEE.

recorded the received signal power in each position while moving on the red line for 6 hours. The location information of a transmission vehicle can be obtained from a transmission packet, and the reception position can be recorded using the GPS module. The experiment parameters are summarized in Table 5.2. 15,280,657 samples could be obtained in the campaign.

In V2V communications, since the position of each vehicle changes dynamically, many pairs of transmission and reception locations were created. If we had used all the samples in the emulation, the effectiveness of the proposed method may not be clearly shown because there are several datasets having a small difference in the path loss. Considering this fact, two locations (points 1 and 2) were selected to properly perform the emulation, as shown in Fig. 5.12. It is assumed that the transmitter moves from point 1 to point 2. The emulation procedures are presented as follows.

i). The local database calculated the actual average value in each 5 [m] mesh for each point.

Table 5.2 The experiment parameters

Onboard unit	MK5 OBU (Cohda Wireless)
Communication standard	IEEE 802.11p
Antenna	Omnidirectional
Frequency [MHz]	5890
Transmit power [dBm]	24
Modulation method	BPSK, QPSK, 16QAM
Coding rate	1/2
Transmission rate [packets/s]	200
Packet size [bytes]	100, 400

ii). We randomly selected n_{ini} instantaneous samples in each 5 [m] mesh when the transmitter locates in the point 1.

iii). Either point 1 or point 2 is randomly picked up as the position of the transmitter. Then, we virtually installed the receiver.

iv). The receiver collects n_{mew} samples when the transmitter locates in the position determined in (iii).

v). The radio map is updated based on Welch's *t*-test in each mesh. An average value calculated from n_{new} is registered if there is a significant difference between the two mean values.

After two datasets were created by dividing the measured samples, we utilized the first dataset as the true mean value samples. Meanwhile, the other was used to perform hypothesis testing. Procedures (iii) to (v) were repeated 1,000 times and these trials were conducted 100 times. After that, the performance of the proposed method was verified using the following MAEs:

$$e'_{1,\text{MAE}} = \frac{1}{D'} \sum_{m=0}^{D'-1} |\bar{P}'_{1,\text{true},m} - \bar{P}'_{1,m}| \quad [\text{dB}], \tag{5.40}$$

$$e'_{2,\text{MAE}} = \frac{1}{D'} \sum_{m=0}^{D'-1} |\bar{P}'_{2,\text{true},m} - \bar{P}'_{2,m}| \quad [\text{dB}], \tag{5.41}$$

where $\bar{P}'_{1,\text{true},m}$ [dBm] and $\bar{P}'_{2,\text{true},m}$ [dBm] denote the true mean values of the m-th mesh at positions 1 and 2, respectively. These values are calculated by averaging instantaneous samples in each 5 [m] mesh for each transmission point. $\bar{P}'_{1,m}$ [dBm] and $\bar{P}'_{2,m}$ [dBm] denote the initial and updated mean values of the m-th mesh, respectively. D' is defined as the number of meshes.

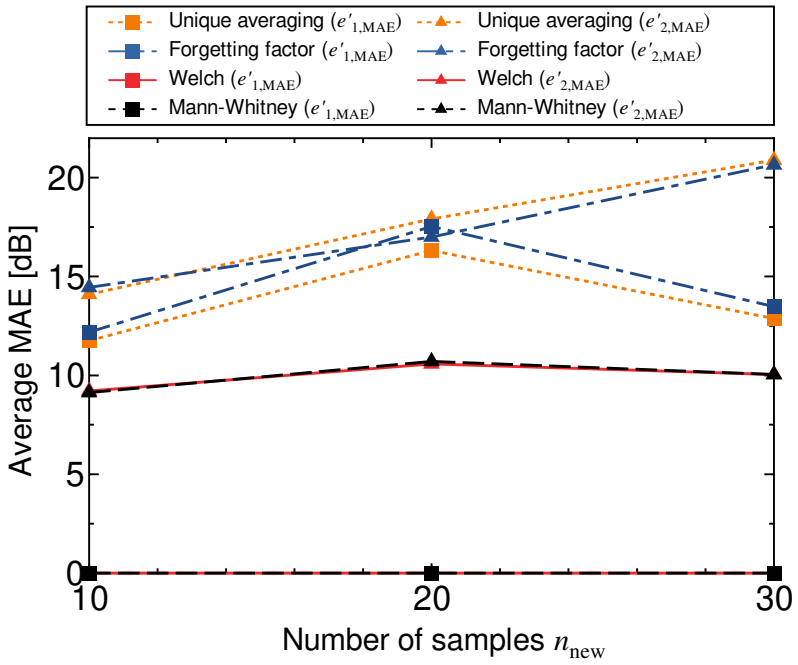


Fig. 5.13 The average MAE in real environment.

Fig. 5.13 depicts the emulation results. We can found that the movement of the transmitter can be properly detected via the hypothesis testing-based methods in a real environment. In this emulation, the communication distance is relatively shorter than in the cellular networks; thus, there may be no significant difference between the two mean values because of the small fluctuation in the path loss. As a result, the estimation accuracy of Welch's t-test is almost equivalent to those of the Mann–Whitney U test. Meanwhile, the comparison methods cannot precisely estimate the mean power comparing with the testing-based methods.

Furthermore, Fig. 5.14 shows the average MAE of hypothesis testing-based methods. These results mean that Welch's t-test can improve the estimation accuracy comparing with the Mann–Whitney U test in $n_{\mathrm{new}} = 20$. The statistical power is poor in the Mann–Whitney U test; thus, such the result may be obtained. On the other hand, Welch's t-test is slightly inaccurate in $n_{\mathrm{new}} = 10$ since the sample mean is notably fluctuated due to the small sample size. It can be seen that the accuracy is the same in each other at $n_{\mathrm{new}} = 30$ since the statistical power can be enhanced in the Mann–Whitney U test by increasing n_{new}.

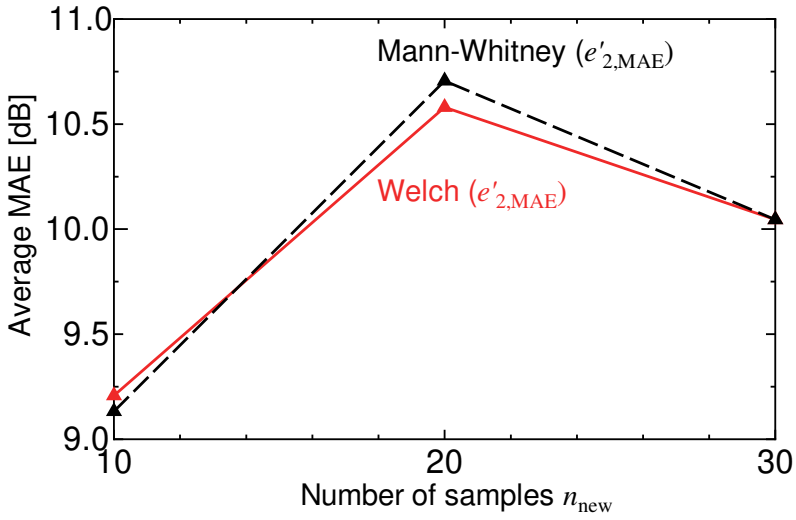


Fig. 5.14 The average MAE of hypothesis testing-based methods.

5.9 Chapter Summary

This chapter has discussed the updating procedures of the radio map by considering the movement of the transmitter. From the surveys, we have clarified the main technical challenge, that is, the radio map must be reconstructed using only real-time measurements in cellular systems based on whether the transmitter has moved. To resolve this task, we have proposed the radio map updating method based on Welch's t-test. This testing has three advantages: i) the updating in real-time, ii) no assumptions for modeling the prior distribution, and iii) robustness for heteroscedasticity. Our method has enabled us to test the difference between two average power values in phases 0 and 1. Through our verification, we have clarified that our method could correctly detect whether the transmitter has moved, and update the radio map.

Chapter 6

Radio Map Extrapolation under Interference-Limited Observations

The previous chapters considered the typical situation that a single transmitter locates in the communication area. Additionally, we assume that all target signals can be obtained in each mesh. However, if multiple transmitters operating at the same frequency exist in the area, several target signals may be missing due to interference from neighborhood transmitters. As a result, the average received signal power may be overestimated owing to the missing data having low signal-to-interference-plus-noise ratio (SINR).

This chapter discusses the extrapolation of the missing data in multiple-transmitter environments. In the proposed method, the empirical CDF of the received signal power is compensated by considering the number of missing data in each mesh. Then, we extrapolate the median received signal power from the compensated empirical CDF. Through the performance verification with the 3.5 GHz band datasets, we elucidate that the radio map can be accurately extrapolated via the proposed method than applying a non-compensation and with the multiple imputation (MI) and Kriging-based methods.

6.1 Background

Many researchers have discussed the effectiveness of the radio map in a wireless system where a target transmitter is located in a communication area. So far, the estimation accuracy of the radio map and performance of the spectrum sharing has been continuously evaluated. As a example of such an evaluation subject, television white spaces are well known [79–81]. In the fixed transmitter environment, the received signal power is determined according to the probabilistic radio propagation characteristics between the transmitter and a reception

position. Generally, by increasing the number of samples, the location-dependent radio propagation characteristics can be skillfully estimated thanks to the law of large numbers. Moreover, we can roughly know the multipath fading property by creating an empirical CDF. To reveal these facts, many researchers have implicitly assumed that the received signal power can be successfully obtained; that is, no missing data exists in the radio environment observations.

Meanwhile, the radio map enables us to enhance communication efficiency in multiple-transmitter environments. For instance, systems with multiple APs [21, 22, 113, 114] and in private networks [115, 116] are considered. To create radio maps in multiple-transmitter environments, the mobile terminal records the received signal power from each transmitter in association with the transmitter ID. Subsequently, the local database creates a radio map by averaging the received signal power samples for each transmitter ID. However, the radio map may not be precisely generated if there are several transmitters operating at the same frequency band. In such an environment, severe interference-limited observation may occur; thus, the received signal power from the target transmitter (hereafter, referred to as *target power*) may be missing. Especially, it may be difficult to obtain low SIR data even in areas where high SNR data can be measured. As a result, the average target power is overestimated compared to the actual value. [117] has reported such phenomenon as the *survivorship bias*.

For an example of the radio propagation estimation in the survivorship bias, Achtzehn *et al.* [118] has reported that the received signal power is overestimated 4 [dB] in a multiple-transmitter environment than in a single-transmitter environment in an urban area. If we simply utilize radio maps in such a situation, the coverage area is overestimated; that is, various performances, such as the transmission power efficiency and spectrum sharing, may be insufficient. As a promising technique for the compensation of the missing data, spatial interpolation is well known. However, spatial interpolation assumes that the absent data occurs inside of the known data range. In a survivorship bias environment, all data below a certain threshold are missing; that is, the absent data occurs outside of the known data range. This observation motivates us to extrapolate the missing target power in dense transmitter environments to create an accurate radio map.

As related studies, the generating methods of radio maps has been shown in [21, 22, 113, 114] by considering indoor multiple-transmitter environments. However, because the aggregate interference signal power is very small because of the walls and obstacles in indoor environments, there are no severe interference-limited observations. As the other method, a radio map is studied in multiple-transmitter environments where different frequency bands are used in each transmitter [119]. Although multiple transmitters are considered, the interference-limited observations do not occur in such a situation.

In this chapter, we propose an empirical CDF-based radio map extrapolation method. As the first step of the extrapolation, the target power samples measured by mobile terminals are uploaded to the local database. The local database compensates for the empirical CDF of the target power by considering the number of missing data. Finally, the median target power is extrapolated from the compensated empirical CDF. Through the performance verification based on the 3.5 GH datasets in a real environment, we can reveal the radio map can be skillfully extrapolated via the proposed method compared to a non-compensated radio map, MI and kriging-based methods without.

The following contents are the major contributions of this chapter.

- Conventional compensation methods of the missing data are comprehensively surveyed. In summary, parametric methods (e.g., truncated normal distribution, left-truncated exponential distribution, and Tobit model), an interpolation method (e.g. Kriging), and an extrapolation method (e.g., MI method) are investigated. By referring to these methods, we can first reveal that the parametric and typical extrapolation methods model the missing as certain distributions. Second, the interference-limited observations are not considered in the traditional interpolation method. Considering these motivations, we propose a novel empirical extrapolation method for interference-limited observations.
- It is confirmed that the proposed method elaborately extrapolates missing target power by a factor of approximately 1–8 [dB] compared with that obtained using the conventional methods.

6.2 Related works

6.2.1 Theoretical Analysis for Interference-Limited Observations

The theoretical analysis of the SIR in typical fading models (e.g., the Nakagami-*m* fading) has been conducted by modeling the probability distribution of the SIR [120–124]. As the major contributions, these works have evaluated the outage probability of the SIR. Although the outage probability can be theoretically modeled via these results, an accurate radio map is assumed to be constructed for each transmitter without missing data, even for interference-limited observations.

6.2.2 Compensation for Noise-Limited Observations

Several researchers have used the truncated normal distribution [63, 64], left-truncated exponential distribution [65], and Tobit model [125] to compensate for the missing data in

wireless systems. In these methods, the PDF of the missing data is modeled as an existing distribution, and the radio propagation parameters are estimated using the log-likelihood function. However, the target power is assumed to be missing if it is below the noise floor in noise-limited observations. Meanwhile, it is difficult to determine the threshold that the target power is missing owing to the stochastic fluctuation of the SIR in interference-limited observations. Consequently, the radio map may not be extrapolated by these methods in interference-limited observations.

6.3 System Model

This section explains the system model. The measured model in interference-limited observations is first shown. Then, we define the criterion for missing data and the mesh definitions.

6.3.1 Measured Model and Statistical Processing

Fig. 6.1 presents the assumed system in this chapter. There is a target transmitter in the communication area and sends the signal to a receiver. Here, the transmission power and center frequency are P_{Tx} [dBm], and f_0 [Hz], respectively. It is assumed the target transmitter communicates in certain coverage that is determined based on the permissible SNR. In the surrounding coverage, S_{I} interfering transmitters are installed, and each transmitter sends a signal to each position as the transmission power is $P_{\mathrm{Tx},s}$ [dBm] $(s = 0, 2, \cdots, S_{\mathrm{I}} - 1)$.

The target power samplers are measured by mobile terminals in each location and the observation data are uploaded to the local database. As the first statistical processing, M meshes are created to generate the radio map of the target transmitter; that is, the multipath fading effects are eliminated by averaging target power samples in each location. The local database stores this information as statistical data. Moreover, the measurement-based path loss model is estimated. The estimation procedures are based on sect. 3.3.2.

However, in dense transmitter environments, the performance of the statistical data may be insufficient because of survivorship bias; hence, the local database needs to extrapolate the missing target power in each mesh. To compensate for the empirical CDF of the target power, it is assumed that the location information of each transmitter and mobile terminal, the transmission cycle of the target transmitter, and the reception time of each mobile terminal are known.

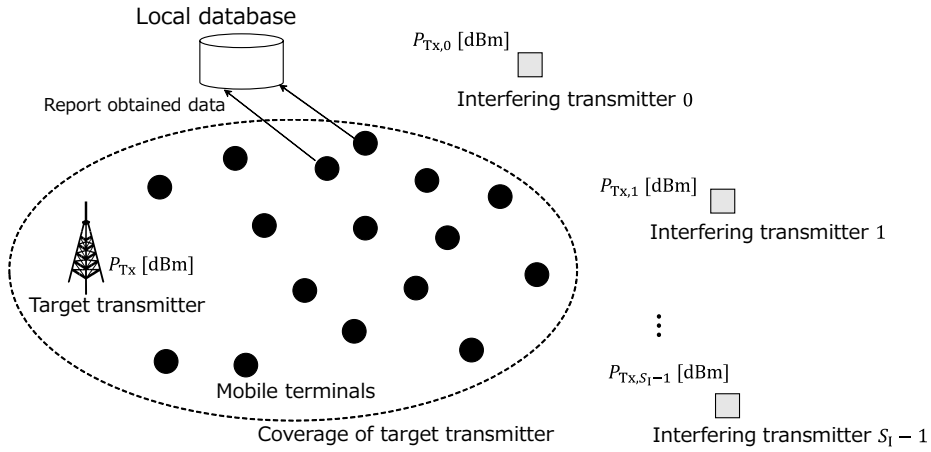


Fig. 6.1 The system model in interference-limited observations.

6.3.2 Criterion for Missing Data Based on Instantaneous SINR

In interference-limited observation, a target power may be probabilistically missing according to the ration of the target power and an aggregate interference power. In such a condition, the criterion for the missing data can be modeled using the SINR. Hence, it is assumed that the target power and ID of the target transmitter are obtained in each terminal if the following inequality is held:

$$\gamma_{g,m} \geq \gamma_{\mathrm{th}}, \tag{6.1}$$

where γ_{th} [dB] is the SINR threshold. Additionally, $\gamma_{g,m}$ [dB] is the instantaneous SINR for the g-th instantaneous target power in the m-th mesh, which is given by

$$\gamma_{g,m} = Q_{g,m} - \left[10\log_{10}\left(10^{\frac{I_{\mathrm{sum},m}}{10}} + 10^{\frac{N_0}{10}}\right)\right], \tag{6.2}$$

where $Q_{g,m}$ [dBm] $(g = 0, 1, \cdots, G-1)$ is the g-th instantaneous target power in the m-th mesh, G is the number of target power data before the interference occurs, and N_0 [dBm] is the noise floor of the mobile terminal. $I_{\mathrm{sum},m}$ [dBm] is the aggregate interference power in the m-th mesh, as explained later. In this chapter, we assume that the number of missing data in each mesh can be calculated by referring to the transmission interval of the target transmitter. Furthermore, the latitude and longitude in each reception location are assumed to be obtained from the GPS module implemented in the mobile terminal even if (6.1) is not held. After the measurement, the observed data are uploaded to the local database.

This chapter defines that target power below γ_{th} is a missing data. As a result, missing data occur outside of the known data range; thus, the spatial interpolations cannot be used in interference-limited observations.

It may be considered that $I_{\mathrm{sum},m}$ is large compared to N_0 because many interfering transmitters exist in a small area. Thus, the instantaneous SINR can be approximately modeled as follows:

$$\gamma'_{g,m} \approx Q_{g,m} - I_{\mathrm{sum},m}, \tag{6.3}$$

where $\gamma'_{g,m}$ [dB] is the approximate SINR for the g-th instantaneous target power in the m-th mesh. Although the approximate SINR can be used as the criterion for the missing data, we use the SINR criterion expressed in (6.2).

6.3.3 Mesh Definitions

As described in the previous section, the target power is probabilistically missing depending on the instantaneous SINR. Under this environment, the mesh can be categorized as follows:

a). Non-missing mesh: There are no missing data in the mesh. Extrapolation was not necessary for this mesh.

b). Partially-missing mesh: We define that $n_{\mathrm{loss},j}$ target power data in the j-th partially-missing mesh are missing owing to aggregate interference. Meanwhile, this dissertation defines that the j-th mesh has n_j instantaneous target power data $\boldsymbol{Q}_j = (Q_0, Q_1, \cdots, Q_{n_j-1})$, where Q_k [dBm]$(k = 0, 1, \cdots, n_j - 1)$ is the k-th instantaneous target power.

c). Complete-missing mesh: It is defined that all $n_{\mathrm{loss},u}$ target power data are missing in the u-th complete-missing. The extrapolation is required in this mesh.

6.4 Proposed method

This section explains two extrapolation methods based on the compensation of the empirical CDF. The following subsections present the extrapolation procedures for the partially-missing mesh and complete-missing mesh.

6.4.1 Extrapolation Method for Partially-Missing Mesh

If Eq. (6.3) is satisfied, the noise floor may be smaller than an aggregate interference power in each mesh. In other words, there may be no missing data below the noise floor. Thus, the range that target power is missing can be represented as $[N_0, q_{\mathrm{min},j}]$, where $q_{\mathrm{min},j}$ [dBm] is the minimum instantaneous target power in the j-th partially-missing mesh. The local

database compensates for the histogram of the target power in $[N_0, q_{\min,j}]$ by considering $n_{\text{loss},j}$. The detailed procedures are summarized as follows:

i). The local database divides $[N_0, q_{\min,j}]$ into B_j bins based on a class width w[dB], where $q_{\min,j}$ is divisible by w and less than $p_{\min,j}$.

ii). The histogram is compensated in each bin by adding the number of data by 1 from $q_{\min,j}$. When the number of added data becomes $n_{\text{loss},j}$, the local database finishes the compensation. Thus, $n_{\text{loss},j}$ must be satisfied by the following constraint:

$$n_{\text{loss},j} = \sum_{i=0}^{B_j-1} R_i, \tag{6.4}$$

where R_i is the number of added data points in the i-th bin. If the histogram in $[N_0, q_{\min,j}]$ has been compensated once, and the number of added data points does not become $n_{\text{loss},j}$, the histogram is compensated from $q_{\min,j}$ again until the above constraint holds. Fig. 6.2 shows the overview of the extrapolation as B_j is 4. The gray object in the histogram is the data points added by compensation. If the number of gray object becomes $n_{\text{loss},j}$, the compensation finishes. $\boldsymbol{Q}_{\text{part},j} = (Q_0, Q_1, \cdots, Q_{n_j-1+n_{\text{loss},j}})$ is defined as the compensated target power vector for the j-th partially-missing mesh.

iii). It can be finally obtained the empirical CDF of $\boldsymbol{Q}_{\text{part},j}$ as follows,

$$\hat{F}_j(Q_{\text{th}}) = \frac{1}{n_j + n_{\text{loss},j} - 1} \sum_{k=0}^{n_j+n_{\text{loss},j}-1} \chi(Q_k), \tag{6.5}$$

where $\hat{F}_j(Q_{\text{th}})$ is the empirical CDF of the j-th partially-missing mesh. Q_{th} [dBm] is an arbitrary instantaneous target power for calculating empirical CDF. Moreover, $\chi(Q_k)$ is an indicator function given by

$$\chi(Q_k) = \begin{cases} 1 & \text{if } (Q_k \leq Q_{\text{th}}) \\ 0 & \text{otherwise.} \end{cases} \tag{6.6}$$

iv). The median target power is extrapolated in the j-th partially-missing mesh using the median of $\hat{F}_j(Q_{\text{th}})$, where its value is defined as $Q_{\text{med},j}$ [dBm] and is registered in the local database.

The above calculation is conducted in all partially-missing meshes.

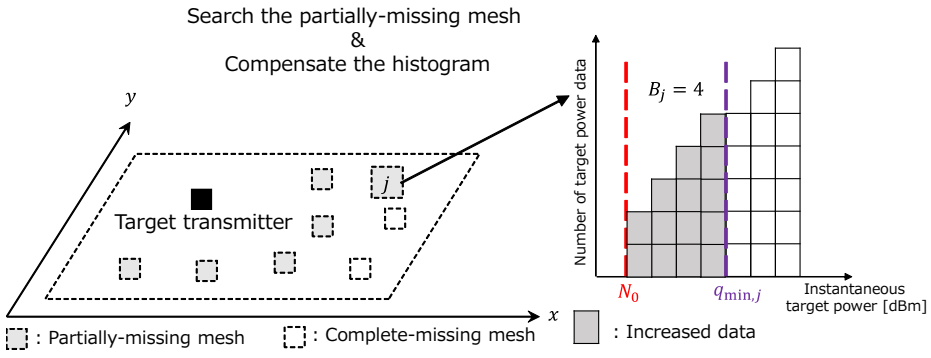


Fig. 6.2 Extrapolation for partially-missing mesh.

Conventional extrapolation methods, such as the truncated normal distribution and MI, assume that the missing target power follows existing probability distributions. Meanwhile, the proposed method can empirically extrapolate the missing target power by compensating for the histogram of the target power with consideration for the number of missing data. The empirical extrapolation is the most advantage of the proposed method.

6.4.2 Extrapolation for Complete-Missing Mesh

The extrapolation method for the u-th complete-missing mesh is explained. In the complete-missing mesh, all instantaneous target power data are missing; thus, we cannot find $q_{\min,j}$. Thus, the preprocessing is performed as follows:

i). The local database calculates the logarithmic distance $\log_{10}(d_u)$ between the target transmitter and the u-th complete-missing mesh.

ii). The non-missing meshes and partially-missing meshes that have the almost same distance as $\log_{10}(d_u)$ are picked up. If the D_{th}-th decimal place or higher is equivalent between $\log_{10}(d_u)$ and a mesh, the local database extracts the non-missing mesh and partially-missing mesh.

iii). Finally, the minimum target power is got from the picked non-missing and partially-missing meshes. We define the power as $q_{\min,u}$ [dBm] for the u-th complete-missing mesh.

Then, the extrapolation is performed as follows:

i). The same procedures shown in Sect. 6.4.1 i). and ii) are conducted with $q_{\min,u}$. $\boldsymbol{Q}_{\text{complete},u} = (Q_0, Q_1, \cdots, Q_{n_{\text{loss},u}-1})$ is defined as the compensated target power vector for the u-th complete missing mesh.

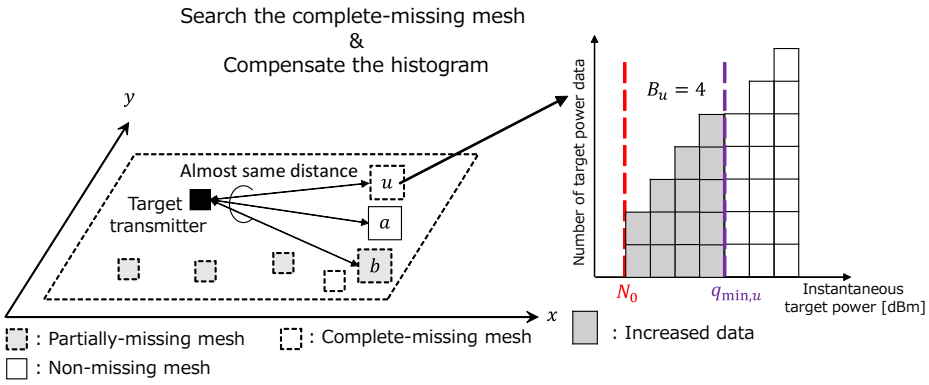


Fig. 6.3 Extrapolation for complete-missing mesh.

ii). The local database estimates the empirical CDF of $\boldsymbol{Q}_{\text{complete},u}$ as follows,

$$\hat{F}_u(Q_{\text{th}}) = \frac{1}{n_{\text{loss},u} - 1} \sum_{i=0}^{n_{\text{loss},u}-1} \chi(Q_i), \tag{6.7}$$

where $\hat{F}_u(Q_{\text{th}})$ is the empirical CDF of $\boldsymbol{Q}_{\text{complete},u}$.

iii). The database server extrapolates the median target power of the u-th complete-missing mesh using $\hat{F}_u(Q_{\text{th}})$. The value is represented as $Q_{\text{med},u}$ [dBm] and is registered on the local database.

The concept of the extrapolation for the complete-missing mesh is shown in Fig. 6.3 represents. In this figure, the u-th mesh is defined as a complete-missing mesh. Additionally, the a-th and b-th meshes denote non-missing and partially-missing, respectively. The three meshes are at approximately the same distance $\log_{10}(d_u)$. Thus, $q_{\min,u}$ is obtained from the a-th and the b-th meshes.

6.5 Conventional Compensation Methods

Two conventional compensation methods are presented in this section: an interpolation method and an extrapolation method.

6.5.1 Spatial Interpolation

Several researchers have utilized Kriging for radio map compensation [11, 58]. In this method, the unobserved received signal power for the u-th complete-missing mesh is interpolated as

follows:

$$\bar{Q}_{u,\mathrm{k}} = \sum_{o=1}^{O} \omega_o \bar{Q}_{o,\mathrm{dB}}, \tag{6.8}$$

where $\bar{Q}_{u,\mathrm{k}}$ [dBm] denotes the interpolated average received signal power in the u-th complete-missing mesh. $\bar{Q}_{o,\mathrm{dB}}$ [dBm] is the known average value in the o-th mesh, and O is the number of meshes with known average value. To perform Kriging, the local database needs to derive ω_o, which is the weight considering the spatial-covariance structure of the random field. Kriging methods consist of several types according to ω_o. This chapter uses the ordinary Kriging method because of its simplicity; that is, prior knowledge of the expectation of the received signal power is not necessary.

In the ordinary Kriging, the weights are derived so that the variance of the estimation error is minimized while satisfying the weight constraint as follows:

$$\begin{aligned} \min \quad & \sigma^2 = \mathbb{E}[(\bar{Q}_{u,\mathrm{k}} - Q_{u,\mathrm{true}})^2], \\ \text{s.t.} \quad & \sum_{o=1}^{O} \omega_o = 1, \end{aligned} \tag{6.9}$$

where σ^2 denotes the variance of the estimation error, and $Q_{u,\mathrm{true}}$ [dBm] is the true average received signal power in the u-th complete-missing mesh. Based on the Lagrange multiplier method, we can express the objective function $\psi_{\mathrm{Lag}}(\omega_o, \mu_{\mathrm{Lag}})$ as:

$$\psi_{\mathrm{Lag}}(\omega_o, \mu_{\mathrm{Lag}}) = \sigma^2 - 2\mu_{\mathrm{Lag}} \left(\sum_{o=1}^{O} \omega_o - 1 \right), \tag{6.10}$$

where μ_{Lag} is the Lagrange multiplier. Here, σ^2 can be represented using the semivariogram ξ as follows:

$$\sigma^2 = -\xi(d_{0,0}) - \sum_{l=1}^{O} \sum_{j=1}^{O} \omega_l \omega_j \xi(d_{l,j}) + 2 \sum_{l=1}^{O} \omega_l \xi(d_{l,0}), \tag{6.11}$$

where $d_{l,j}$ [m] denotes the distance between the l-th mesh and the j-th mesh. Because the empirical model of the shadowing correlation is well known as the exponential decay [31], we used the exponential semivariogram model for ξ. The definition is defined as follows:

$$\xi(d_{l,j}) = \theta_{\mathrm{n}}^2 + \theta_{\mathrm{sl}}^2 \left[1 - \exp\left(-\frac{d_{l,j}}{\theta_{\mathrm{r}}} \right) \right], \tag{6.12}$$

where θ_{n}^2, θ_{sl}^2, and θ_{r} are the nugget, sill, and range, respectively. The local database estimates these parameters by nonlinearly fitting the measured data into (6.12). Finally, we can express

the simultaneous equations by calculating the partial derivatives in (6.11) as follows:

$$\begin{pmatrix} \xi(d_{1,1}) & \cdots & \xi(d_{1,O}) & 1 \\ \xi(d_{2,1}) & \cdots & \xi(d_{2,O}) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \xi(d_{O,1}) & \cdots & \xi(d_{O,O}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_O \\ \mu_{\mathrm{Lag}} \end{pmatrix} = \begin{pmatrix} \xi(d_{1,0}) \\ \xi(d_{2,0}) \\ \vdots \\ \xi(d_{O,0}) \\ 1 \end{pmatrix}. \tag{6.13}$$

By solving the simultaneous equations, the weights that minimize σ^2 can be derived.

6.5.2 Spatial Extrapolation

As another compensation method, several researchers have utilized an extrapolation method [126, 127]. This method assumes that the missing data exists outside of the known data. In the interference-limited observations, no received signal power samples may be observed in the coverage area where the distance from the transmitter is above a certain distance. Thus, the phenomenon matches the assumption of extrapolation. Extrapolation mainly consists of two types: single imputation (SI) method [126] and MI method [127]. A representative value of the known data is imputed to the missing points in the SI method. For instance, a mean imputation [128], median imputation [129], and hot-deck imputation [130] have been proposed. However, the extrapolation accuracy may be degraded if the statistical properties are greatly different between the known and missing data.

Meanwhile, a posterior probability distribution of the missing data is modeled in the MI method. Then, V representative values are calculated by obtaining V datasets that follow the modeled distribution. Finally, these values are unified to a single value based on statistical processing, such as averaging and linear regression. The single value is utilized as the extrapolated value. If the posterior distribution can be appropriately modeled, the MI method enables us to roughly extrapolate the missing data. [127] has argued that the MI method is superior to the SI method. However, in an actual environment, the MI method may not accurately perform owing to the complicated radio propagation characteristics.

6.6 Emulation Setups

Emulation-based performance evaluation was conducted using a 3.5 GHz dataset measured over an actual cellular system. The detailed contents of this dataset are described in Sect. 3.5.1. In this emulation, this dataset is utilized as the target power. To create the interference-limited observation, we virtually deployed multiple interfering transmitters around the measurement

Table 6.1 The emulation parameters for the interfering transmitters

The number of interfering transmitters S_{I}	1–7
Transmission power $P_{\mathrm{Tx},s}$ [dBm]	40
Path loss index C_{I}	4.5
Reference distance d_0 [m]	10

area. Thus, the interference power was calculated over the computer simulation. Meanwhile, the average target power was calculated in each 10m mesh using 100423 observed samples. In the following, we describe the radio propagation model of the interfering transmitter and the installation location of those.

6.6.1 Radio Propagation Model for Interfering Transmitter

The simple radio propagation model for the interfering transmitter is modeled as follows:

$$I_{s,m} = P_{\mathrm{Tx},s} - L_{\mathrm{fspl}}(d_0) - 10C_{\mathrm{I}}\log_{10}\left(\frac{d_{s,m}}{d_0}\right), \tag{6.14}$$

where $I_{s,m}$ [dBm] is the instantaneous interference power in the m-th mesh from the s-th interfering transmitter. C_{I} is the path loss index for an interfering transmitter. $d_{s,m}$ [m] is the link distance between the s-th interfering transmitter and the m-th mesh. The aggregate interference power is expressed as,

$$I_{\mathrm{sum},m} = 10\log_{10}\left(\sum_{s=0}^{S_{\mathrm{I}}-1} 10^{\frac{I_{s,m}}{10}}\right). \tag{6.15}$$

Although the actual aggregate interference power may fluctuate owing to the phase rotation of each interference signal, this chapter defines $I_{s,m}$ by considering only the path loss effect. Because the interference power is considered depending on the Euclidean distance, the effect of the phase rotation is assumed to be eliminated.

Fig. 6.4 presents the locations of the interfering transmitters. The orange square and frame are the location and coverage of the target transmitter, respectively. The seven yellow squares are the locations of the interfering transmitters. The emulation parameters for the interfering transmitters are listed in Table 6.1.

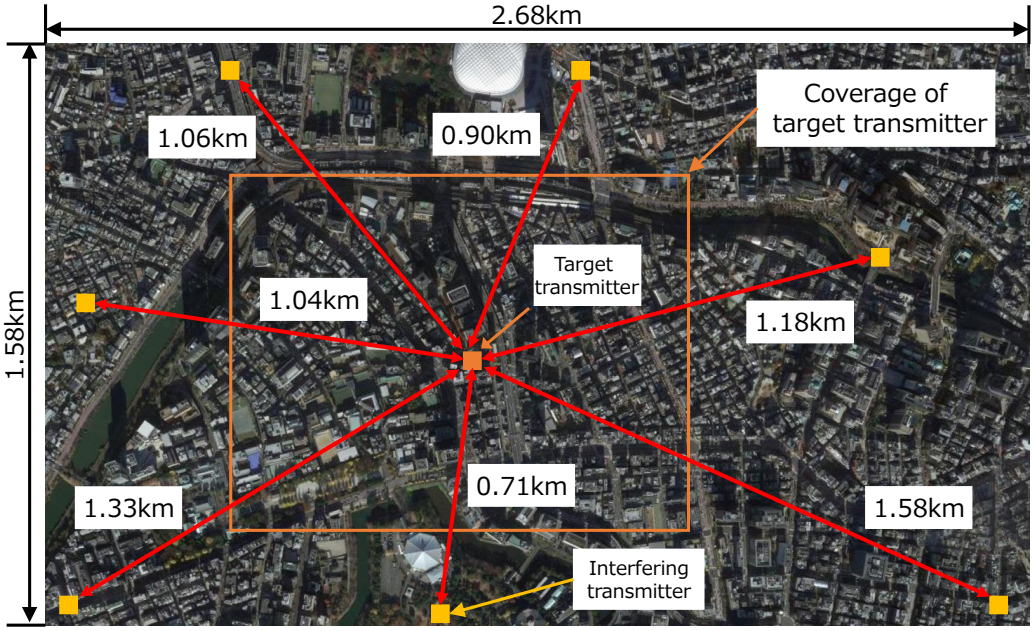


Fig. 6.4 The installation location of the interfering transmitters.

6.7 Emulation Results

Next, we explain the examples of radio maps, the measurement-based path loss model, and its accuracy. In the following, $D_{\text{th}} = 3$ is utilized.

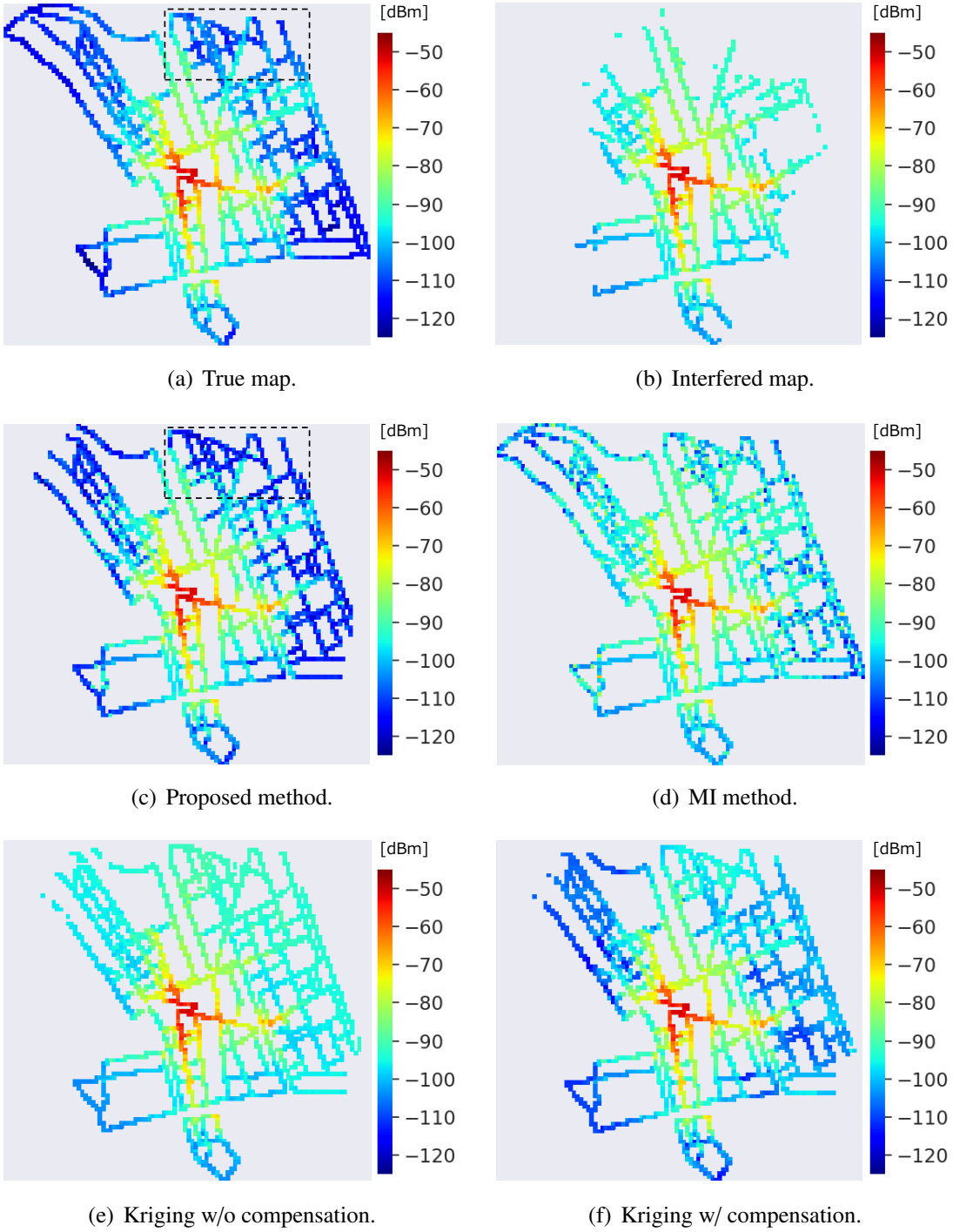
6.7.1 Example of Radio Maps

Fig. 6.5 illustrates an example of constructed radio maps as the mesh size is 10 [m], $\gamma_{\text{th}} = 7$ [dB], $S_{\text{I}} = 4$, and $w = 0.5$ [dB]. Fig. 6.5(a) denotes the average target power described in Sect. 6.6; that is, a true map without interference is created. Fig. 6.5(b) is the radio map with the interference-limited observations. As can be seen from it, there are many missing data. Figs. 6.5(c), 6.5(d), 6.5(e), and 6.5(f) present the extrapolated maps.

The missing data can be precisely extrapolated via the proposed method from Figs. 6.5(a), 6.5(b) and 6.5(c). However, in the severe interference area surrounded by the black dotted frame, the extrapolation performance is low. In this area, because $q_{\min,u}$ may be overestimated, the empirical CDF is inaccurately compensated. As a result, the extrapolation performance is insufficient. Moreover, the proposed method cannot extrapolate several target power in the area edge because there are few meshes that satisfy the conditions explained in Sect. 6.4.2-ii).

In the MI method, a target power is randomly obtained from the multivariate normal distribution in each complete-missing mesh. The results show that the target power cannot be accurately extrapolated in many meshes owing to the random sampling.

Finally, we use Kriging in the interference-limited observations as O is 8. Fig. 6.5(e) is created by applying the Kriging to the complete-missing meshes in Fig. 6.5(b). Meanwhile, Fig. 6.5(f) is constructed by compensating the empirical CDFs in the partially-missing meshes using the proposed method before the Kriging is applied. Note that the extrapolation cannot be performed in the non-colored meshes because there are no known target power data around the complete-missing mesh. These results clarify that the extrapolation performance is degraded in Kriging without median compensation. Additionally, Kriging with median compensation may inaccurately extrapolate the radio map in the coverage edge because of the severe missing in the target power data.



(a) True map.

(b) Interfered map.

(c) Proposed method.

(d) MI method.

(e) Kriging w/o compensation.

(f) Kriging w/ compensation.

Fig. 6.5 Example of extrapolated radio maps.

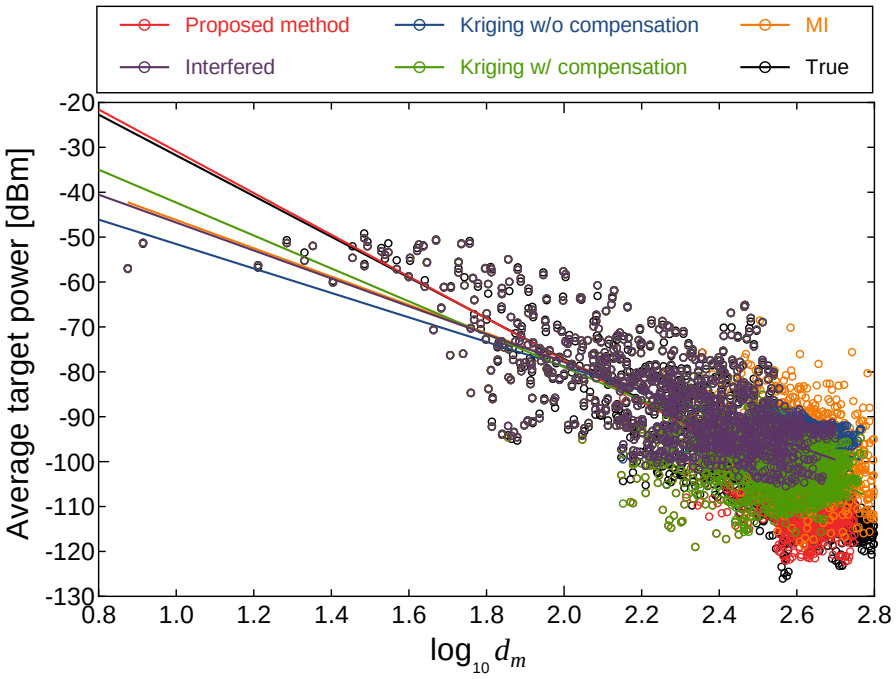


Fig. 6.6 An example of the path loss estimation ($\gamma_{\mathrm{th}} = 7\,[\mathrm{dB}]$).

6.7.2 Measurement-Based Path Loss Model

This subsection explains the measurement-based path loss model. We estimated the path loss model based on the procedures explained in Sect. 3.3.2.

Fig. 6.6 presents an example of the estimated path loss models. Each colored circle is the average target power in each mesh. An estimated path loss model is shown as a solid line. In this evaluation, $\gamma_{\mathrm{th}} = 7\,[\mathrm{dB}]$, $S_{\mathrm{I}} = 4$, and $w = 0.5\,[\mathrm{dB}]$. The black solid line denotes the true path loss model without interference-limited observations. Next. the estimated path loss model with interference-limited observations is represented as the purple solid line. It can be confirmed that the proposed method enables us to elaborately estimate the path loss model compared to the conventional methods, which cannot derive the path loss model compared with the true model owing to the missing of lower SINR data.

6.7.3 Estimation Accuracy

Finally, this section evaluates the estimation accuracy based on a signed mean error (SME) $e_{\mathrm{sme}}\,[\mathrm{dB}]$, which is defined as follows:

$$e_{\mathrm{sme}} = \frac{1}{N_{\mathrm{I}}} \sum_{m=0}^{N_{\mathrm{I}}-1} (\bar{P}_{\mathrm{true},m} - \bar{P}_{\mathrm{dB},m}), \tag{6.16}$$

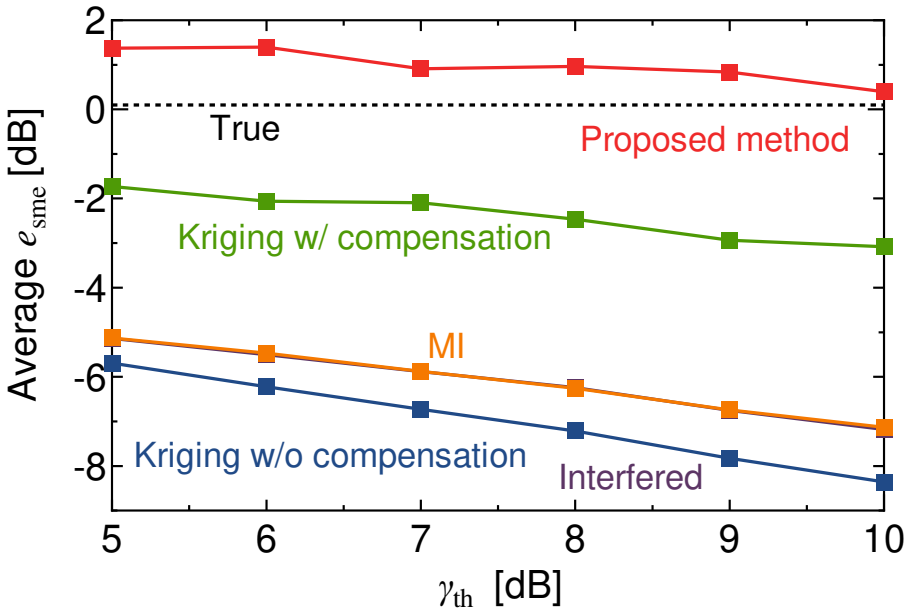


Fig. 6.7 The average SME.

where $\bar{P}_{\mathrm{dB},m}$ [dBm] denotes the median path loss value estimated using the measurement-based path loss model in the m-th mesh. N_{I} is the number of 10m meshes used in the evaluation. After dividing 100423 samples into three groups, we calculated the average e_{sme} based on the cross-validation. The SME is used instead of the RMSE because it is necessary to evaluate whether the missing data containing a lower SINR are accurately extrapolated.

Fig. 6.7 depicts the average SME as $S_{\mathrm{I}} = 4$ and $w = 0.5$ [dB]. In the MI method, a target power is randomly obtained in each complete-missing mesh and path loss parameters B and C. This operation is repeated by 20 times. Then, the parameters B and C are averaged and median path loss value is estimated. We can confirm that the median path loss can be precisely extrapolated in the proposed method compared to the other methods. Furthermore, the radio environment can be estimated with a certain degree of accuracy using Kriging with compensation; however, its performance is lower than the proposed method. The performances of the others are not skillful because the estimation accuracy of the path loss models is inaccurate.

Next, Fig. 6.8 shows the average SME versus the number of interfering transmitters as γ_{th} is 5 [dB] and w is 0.5 [dB]. The more accurate estimation can be realized via the proposed method; meanwhile, the others cannot improve the estimation error. If S_{I} is 5 or more, the estimation accuracy is degraded because of the severe aggregate interference power.

Finally, Fig. 6.9 depicts the error characteristics versus the class width w for $S_{\mathrm{I}} = 4$ and $\gamma_{\mathrm{th}} = 7$ [dB]. It can be seen that the performance of the proposed method degrades with

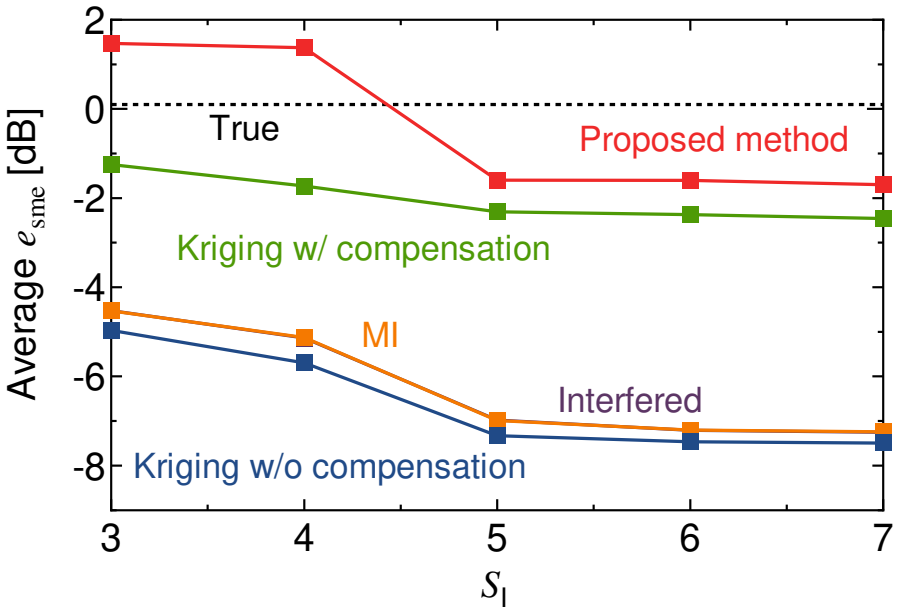


Fig. 6.8 The average SME versus S_{I}.

an increase in w. The proposed method repeatedly compensates for the empirical CDF by increasing the histogram. If w is large, the lower SINR data are excessively compensated; hence, the path loss index is larger than the true value. Furthermore, the estimation accuracy decreases in the smaller w because the lower SINR data cannot be accurately compensated for. Thus, we argue that w should be selected as approximately 0.5 [dB] to stably estimate the median path loss.

6.8 Chapter Summary

This chapter has proposed radio map extrapolation methods for interference-limited observation. In the conventional methods, there are several drawbacks, such as modeling of the specific distribution, where the compensation of missing data is not considered. To resolve these problems, a novel extrapolation method for the radio map has been presented. The proposed method compensates for the empirical CDF by adding the number of missing data in each bin. This idea is an advantage compared to the other methods. The emulation results have revealed that the more precise extrapolation can be realized via the proposed method compared to the MI method and Kriging-based methods.

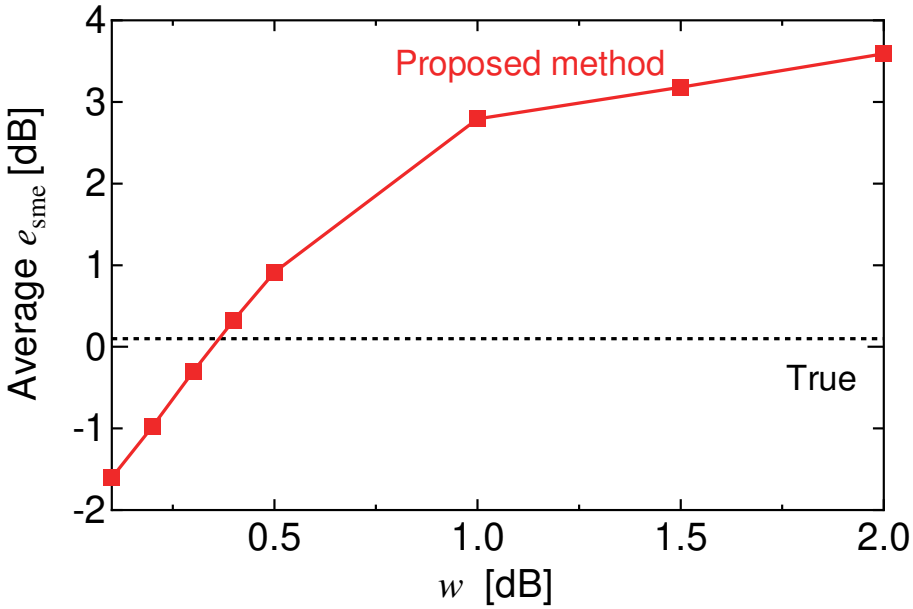


Fig. 6.9 The average SME versus w.

Chapter 7

Crowdsourcing-Assisted Radio Maps for MANETs

The previous chapters have focused on a wireless system where the location of a transmitter is fixed or is slowly varied. In such systems, we can estimate the average received signal power precisely by only processing the measured data in each received position.

With the development of wireless technology, MANETs, such as V2V communications and device-to-device (D2D) communications, are attracting attention. In these systems, both a transmitter and a receiver communicate with each other while dynamically moving within a communication area.

This chapter applies a radio map to MANETs to enhance the radio propagation estimation. We first explain the detailed motivation of the radio map construction for MANETs. Subsequently, the creation procedures of radio maps are presented. Finally, the performance evaluation of radio maps using measured datasets is described.

7.1 Background

The number of mobile terminals has drastically increased thanks to the rapid development of wireless technology, such as the Internet of things, and various wireless systems are connected to the Internet. Especially, as one of the wireless systems, V2V [131] and D2D communications [132] are well known and are called as MANETs in several researchers [133]. These applications enable us to exchange various data, such as the surrounding emergency information, in real-time manner. By sharing diverse data between terminals, each terminal can appropriately determine the modulation format and transmission timing

based on the surrounding environment. MANETs have the potential to enhance the reliability of each application (e.g., autonomous driving systems).

As one of the issues in MANETs, the degradation of communication reliability is considered. Although the location of a transmitter is fixed in the conventional downlink communications (e.g., the cellular networks), both a transmitter and a receiver dynamically move in MANETs. As a result, an *any-to-any* communication link is innumerably created in various environments. In other words, the radio propagation characteristics significantly change along with the movement of terminals. Additionally, due to the shortage of unused spectrum, it is difficult for each terminal to use a new spectrum. Under this environment, the packet collision and loss due to the hidden node problem and uncertainty of radio propagation become severe problems. To enhance the communication quality in overall wireless networks, each system must estimate the radio propagation characteristics with high accuracy and determine their parameters while considering the movement of terminals.

Many researchers have continuously used empirical propagation models to estimate radio propagation because various measured datasets had been used to construct these models. By appropriately determining parameters, such as the center frequency and antenna height, the median path loss value can be calculated with a certain degree of accuracy. However, because empirical models do not consider probabilistic radio propagation (i.e., the shadowing and multipath fading), the more accurate estimation may not be realized in these models. Especially, in the MANETs, since the surrounding environment dynamically changes, the estimation accuracy may be lower than in the conventional wireless systems.

To estimate such the probabilistic component, there are several probability distributions, for instance, the Gaussian distribution. However, these distributions are based on the rule of thumb and theoretical derivation; hence, the site-specific fluctuation of the received signal power may not be estimated. Furthermore, some composite models of shadowing and multipath fading have been established; however, the probability distributions are very complicated. Additionally, it is necessary to properly use the probability distributions according to the surrounding radio propagation characteristics [134–138].

Although a radio map can be utilized to estimate the path loss and shadowing in each location, it has been only applied to the systems that the location of the transmitter is not changed. Therefore, the conventional radio map cannot estimate the location-dependent radio propagation component in MANETs owing to the movement of both a transmitter and a receiver.

Here, when an arbitrary communication link is given, it can be considered that the path loss and shadowing components are static in the reception position. Focusing on this property, if a radio map is constructed in each transmission position, we can estimate the location-

dependent radio propagation even in MANETs. Similar work has been conducted in [139]. In this dissertation, mobile terminals share measured datasets using training symbols with each other. After that, the Kriged Kalman filtering is used to estimate the radio propagation in each environment. However, this dissertation has only clarified the simulation results. Additionally, multipath fading is assumed to be removed perfectly in each receiver.

This chapter proposes the construction of radio maps for MANETs. The proposed method gathers the received signal power samples by communicating between transmitters and receivers in a realistic environment; hence, the concept of crowdsourcing is extended to the MANETs to create radio maps. Using crowdsourcing, enormous received signal power samples can be collected in various locations. In IEEE 802.11p that is the traditional standard of the V2V, a beacon signal is sent with the transmitter position by a transmission vehicle to support traffic safety. The reception vehicle can obtain both transmission and reception position by utilizing the GPS in addition to the received signal power. After the observations, the local database processes the reported samples in each transmission / reception location and generates a radio map in each transmission position. By this processing, we can get radio maps even in MANETs.

To verify the usefulness of the proposed scheme, V2V communications had been conducted in two environments. Subsequently, a radio map was generated in each transmission position on the MySQL server. Emulation results make it clear that the proposed method can skillfully estimate the average received signal power in each position comparing with the empirical and theoretical propagation models. Furthermore, the power control in the unicast system is performed based on the proposed map and empirical CDF of the multipath fading to show the effectiveness of the proposed method more clearly. The verification results clarify that the transmission power can be reduced while the permissible communication quality can be kept.

Note that a spatial interpolation is utilized for the radio maps construction in this chapter because the main purpose of this chapter is to show the usefulness of the radio maps; that is, whether the location-dependent radio propagation can be estimated in each transmission / reception mesh. The measurement-based path loss model may interpolate the average received signal power. This is discussed in Sect.7.4.

7.2 Concept of Radio Maps for MANETs

This section explains the concept of radio maps for MANETs. In the following, the data gathering and statistical processing for constructing radio maps are shown. Subsequently, the remained tasks and utilization of radio maps are presented.

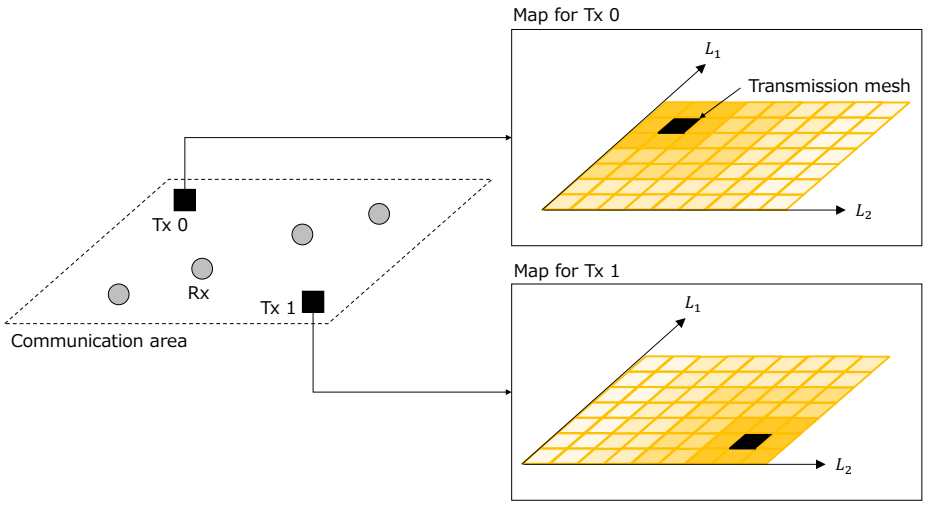


Fig. 7.1 Concept of radio maps for MANETs.

7.2.1 Data Gathering and Statistical Processing

The concept of radio maps for MANETs is presented in Fig. 7.1. The communication area is divided into two-dimensional meshes. By regarding each mesh as a transmission location, the radio maps are created for each transmission mesh: N_{Tx} maps are accumulated at the maximum when the number of meshes is N_{Tx} in the area. Although there are various methods to construct a radio map (e.g., empirical propagation models and ray tracing), this chapter creates a radio map using measured datasets. A radio map enables us to get statistical radio propagation characteristics thanks to its measured data; hence, the mean characteristics of radio propagation can be conjectured more clearly than the conventional path loss models.

Fig.7.2 shows the data gathering in MANETs. A packet is sent to the surrounding receivers by a transmitter. Here, it is assumed that the transmitter includes its location information and ID in a packet based on the IEEE 802.11p for supporting traffic safety. At each receiver, in addition to these data, the reception location, instantaneous power, datetime, center frequency and reception ID are recorded in their storage. Each receiver uploads the recorded data to the local database. In the actual case, we have to avoid upload congestion; thus, the recorded data are uploaded via a WLAN at time when the network is not congested. After massive datasets can be gathered in the local database, the radio maps are generated by processing the datasets.

A conventional radio map is created for a fixed transmitter; that is, most researchers have not attempted to generate multiple radio maps in MANETs. Therefore, the construction of a radio map in each transmitter location is the originality of our research.

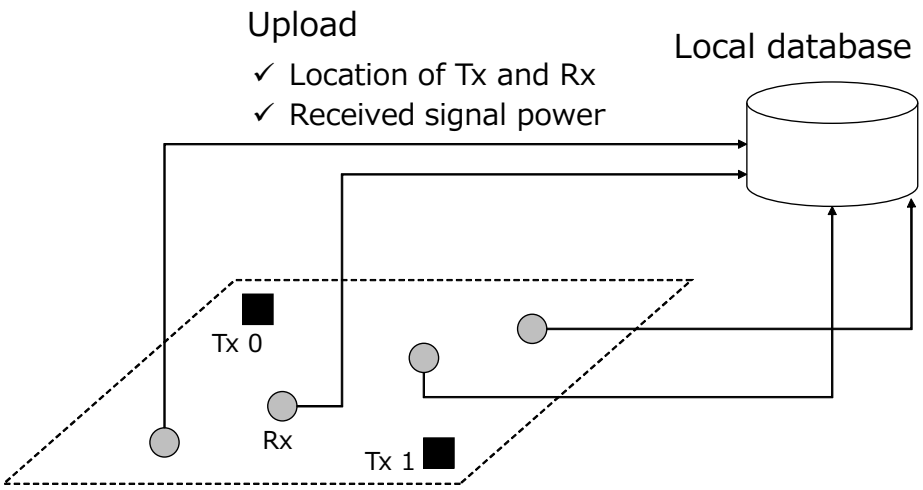


Fig. 7.2 Data gathering in MANETs.

7.2.2 Remained Tasks for Radio Maps

If the new datasets are reported from receivers after the creation of a radio map, those data are used again to calculate a mean value. This is because the location-dependent radio propagation, i.e., the average received signal power, is generally static over time domain when the surrounding structures do not vary. In such an environment, the effects of multipath fading may be mitigated, as the number of samples increases. Meanwhile, if surrounding structures vary in the realistic situation, the average received signal power significantly changes. As one of the solutions to this problem, we may use the updating method of the radio map based on Welch's *t*-test described in Chapter 5. As the testing procedures, distributed receivers observe received signal power samples in each location. After collecting these samples, the local database infers the significant difference between an initially estimated mean and newly calculated mean values in each transmission / reception location. If a significant difference is detected, the new value is registered in the mesh that corresponds to the reception location. Although the performance evaluation should be conducted using computer simulation and measured datasets, we have not verified the effectiveness yet. Thus, this task will be solved in our future work.

7.2.3 Utilization of Radio Maps

After the radio maps are created, the transmitter can adaptively determine the communication parameters based on the average received signal power in each mesh. When the transmitter sends a query with its position information, the local database gives a radio map that corresponds to the location of the transmitter. The average received signal power can be

Table 7.1 The measurement conditions in V2V communications

Measurement time	9:00–18:00
LOS environment	Straight road
Non-line-of-sight (NLOS) environment	Intersections

grasped precisely in each location using the radio map. For example, if the average received signal power in the radio map is low compared to the desired value, the transmitter can suppress the deterioration of the communication quality using high transmit power, low order modulation formats, and ad hoc communications. Meanwhile, if an estimated value is larger than the permissible value, the communication quality can be improved further by choosing the high order modulation formats and a high coding rate. The performance evaluation of the communication quality is discussed in Sect. 7.4.4.

7.3 Measurement Campaign in 700MHz Band

To obtain radio maps for MANETs, V2V communications have been carried out in Chofu City and Mitaka City, in Tokyo, Japan. These areas are typical suburbans. The measurement campaign had been conducted in January 2017 for 3 days. Table 7.1 shows the measurement conditions in V2V communications. The measurement setups are the same in [28].

7.3.1 Measurement Equipment

Fig. 7.3 shows an observation vehicle used in measurement campaign. To gather radio environment, we prepared three vehicles implemented in-vehicle devices. In the experiment, a transmission vehicle sent a packet to reception vehicles while moving the route presented by the red line in Fig.7.4. Each vehicle traveled the route at 40[km/h] by considering the average speed limit in the area. As the communication standard, ARIB STD-T109 [55] is utilized. The standard is established in Japan according to IEEE 802.11p. The center frequency is 760 [MHz] and the modulation format is OFDM.

A transmitter sent a beacon signal using the onboard device every 100 [ms]. The others vehicle observed the instantaneous power, datetime of the reception, transmission and reception IDs, and location information of both transmission and reception vehicles. The location information of a transmitter was included in the transmission packet. Additionally, the GPS module was implemented to each vehicle to obtain the location information of a receiver and the datetime. The GPS type was Garmin GPS 18x. The location information was obtained once per second. The positioning accuracy is 95 [%] within 15 [m].

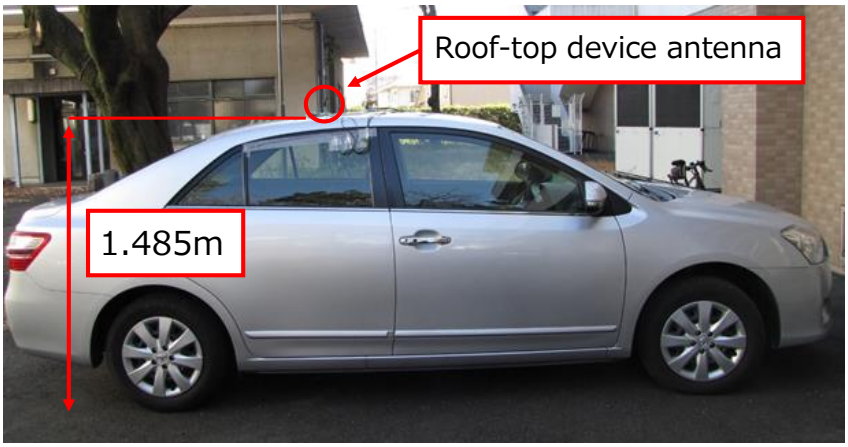


Fig. 7.3 Observation vehicle.

Table 7.2 On-vehicle device parameters

Standard	ARIB STD-T109
Modulation format	OFDM/QPSK(1/2)
Transmit power [dBm]	19.2
Frequency [MHz]	760
Header size [Byte]	61
Payload size [Byte]	77

Table 7.2 presents the specifications of the onborad device. 232 [μs] was necessary to send one packet. 200 [ms] was needed for updating the location information; thus, the information was updated every two messages in each vehicle. The antenna type was the monopole in all vehicles. Table 7.3 shows the antenna characteristics.

7.3.2 Statistical Processing on Local Database

Tables 7.4 and 7.5 present the registration elements in the database server. A mesh code is separately accumulated in each transmitting and receiving vehicle according to the mesh size. Transmitter and Receiver IDs are utilized to delete unnecessary datasets. A radio map was generated using the average received signal power in each transmission mesh. At that time, transmission and reception mesh codes were utilized to identify the location information.

We constructed radio maps on the local database. The construction environment of the database was Cent OS 7 and MySQL 5.7. Hypertext preprocessor (PHP) programs were used to process the measured datasets. The creation procedures of a radio map are defined as follows:

i). The measurement samples are uploaded to the local database.

Fig. 7.4 Measurement route.

ii). The MySQL server stores the uploaded samples in the table.

iii). The local database statistically processes the datasets. In this chapter, the average received signal power is calculated in each reception mesh by fixing a transmission mesh. Then, the local database stores the average value in the statistical table.

iv). When the terminal accesses a server, the local database searches the average value by referring to the MySQL server.

v). A radio map information can be used to estimate the radio propagation.

In the experiments, the observation dataset is recorded into comma-separated values (CSV) file in each receiving vehicle when a transmitted packet is successfully decoded. The

Table 7.3 Antenna characteristics

Type	Monopole antenna
Frequency range [MHz]	755-765
Absolute gain [dBi]	2.15
Element length [mm]	110

CSV is reported to MySQL server via each receiving vehicle. The local database accumulates the instantaneous received signal power samples with transmission and reception mesh codes, and statistically processes these samples.

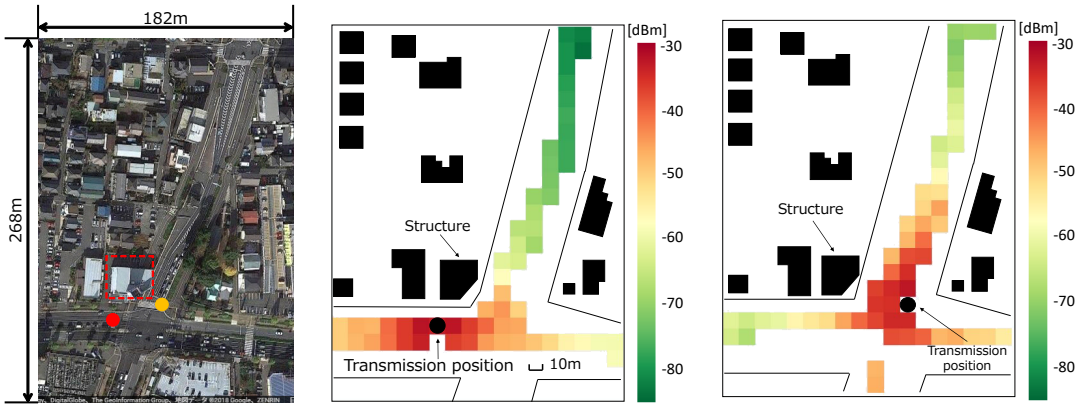
The local PC downloads a radio map in an arbitrary transmission mesh. If we observe new samples after the construction of a radio map, those samples are additionally registered to the MySQL server. Then, the local database calculates the average received signal power in each reception mesh for fixed transmission mesh.

Table 7.4 Registration contents in MANETs

Element	Type	Size [Byte]
Measurement datetime	datetime	8
Transmission latitude [deg]	double	8
Transmission longitude [deg]	double	8
Reception latitude [deg]	double	8
Reception longitude [deg]	double	8
Altitude [m]	double	8
Center frequency [Hz]	double	8
Instantaneous received signal power [mW]	double	8
Packet ID	integer	4
Transmitter ID	char	17
Receiver ID	char	17
Transmission mesh code (First)	char	4
Transmission mesh code (Second)	char	2
Transmission mesh code (Third)	char	2
Transmission mesh code (1/10)	char	2
Transmission mesh code (10m)	char	2
Transmission mesh code (5m)	char	3
Transmission mesh code (2m)	char	3
Transmission mesh code (1m)	char	3
Reception mesh code (First)	char	4
Reception mesh code (Second)	char	2
Reception mesh code (Third)	char	2
Reception mesh code (1/10)	char	2
Reception mesh code (10m)	char	2
Reception mesh code (5m)	char	3
Reception mesh code (2m)	char	3
Reception mesh code (1m)	char	3
Saved date	datetime	8

Table 7.5 Statistical contents in MANETs

Element	Type	Size [Byte]
Target start date	datetime	8
Target end date	datetime	8
Transmission mesh code	char	20
Reception mesh code	char	20
Transmission southwest latitude [deg]	double	8
Transmission southwest longitude [deg]	double	8
Transmission northeast latitude [deg]	double	8
Transmission northeast longitude [deg]	double	8
Reception southwest latitude [deg]	double	8
Reception southwest longitude [deg]	double	8
Reception northeast latitude [deg]	double	8
Reception northeast longitude [deg]	double	8
Averaged received signal power [dBm]	double	8
Saved date	datetime	8



(a) Structure. (b) NLOS case. (c) LOS case.

Fig. 7.5 Examples of radio maps in V2V communications.

7.4 Evaluation Results

This section presents the evaluation results. After an example of several radio maps is depicted, the RMSE is calculated to verify the accuracy of the radio maps.

7.4.1 Example of Radio Maps in LOS and NLOS

Figs 7.5(b) and 7.5(c) are examples of the constructed radio maps in Fig. 7.5(a). The local database processed the measured datasets with a 10 m-squared mesh. In Fig. 7.5(a), the red and yellow objects denote the transmission positions in NLOS and LOS environments, respectively. Each black circle in Figs. 7.5(b) and 7.5(c) corresponds to each transmission position. These maps clarify that the average received signal power values are different owing to the structure surrounded by the red dotted frame in Fig. 7.5(a). For instance, on the north side of the area, it can be found that there is a difference around 20 [dB] between average values. In Fig. 7.5(b), since the structure exists in front of the transmission position, the NLOS can be created in the environment. Meanwhile, there is LOS in Fig. 7.5(c); that is, the shadowing may not be dominant in the environment.

These results reveal that the radio maps can skillfully estimate the path loss and shadowing in each location. Here, because the receiving vehicles could not observe the received signal power in several positions, there are missing data in some meshes. In realistic communication, we may obtain enormous received signal power samples via crowdsourcing; hence, the number of missing data will be decreased.

Additionally, this subsection presents the load for creating radio maps. The evaluation environment is PHP 7.0.12 in Cent OS 7 with Intel(R) Xeon(R) CPU E5-2407 @ 2.20 GHz.

Table 7.6 Processing load in the local database

Day	Number of samples	Number of 10m meshes	Processing time [s]	Average time per mesh [s]
All days	2839101	69158	24.0	0.00036
Days 1 and 2	1940364	53096	17.0	0.00033
Day 3	898691	37786	10.0	0.00028

Table 7.6 shows an example of the processing load on the local database. The first row of the Table 7.6 is all datasets in this measurement campaign.

7.4.2 Propagation Analysis

This subsection estimates the statistical characteristics of the shadowing in the measurement area to confirm that whether the estimated properties match with the typical examples. The local database first averages all datasets in each 10m mesh. Next, the communication distance between a transmitter and a receiver was calculated to estimate the path loss. Finally, the scatter diagram was created using the average received signal power and the communication distance.

In the experiment, the noise floor in a receiver was about -96.0 [dBm]. To suppress the underestimation of the average value owing to the noise floor, the maximum distance was limited to 100 [m] in the evaluation. If the datasets over a long communication distance are processed, the average received signal power is inaccurately estimated because of missing data below the noise floor.

This subsection defines the path loss model L_{loss} as follows:

$$L_{\mathrm{loss}}(d_{l,j}) = \begin{cases} \alpha_1 \log_{10}(d_{l,j}) + \beta_1 & (d_{l,j} \leq R_{\mathrm{b}}) \\ \alpha_2 \log_{10}(d_{l,j}) + \beta_2 & (d_{l,j} > R_{\mathrm{b}}) \end{cases} \quad [\mathrm{dB}], \tag{7.1}$$

where α_1, α_2 are the path loss coefficients. β_1 [dB] and β_2 [dB] denote the effects of the transmit power and antenna. R_{b} [m] is the distance from a transmitter to the break point. It is defined in [13] and its equation is given by

$$R_{\mathrm{b}} = \frac{4 h_{\mathrm{Tx}} h_{\mathrm{Rx}}}{\lambda} \quad [\mathrm{m}], \tag{7.2}$$

where h_{Tx} [m] and h_{Rx} [m] are heights of a transmitter and a receiver. These values are 1.485 [m], respectively. The local database estimated α_1, α_2 and β_1, β_2 by applying the least-squares method to the scatter diagram. As a result, the following characteristics were

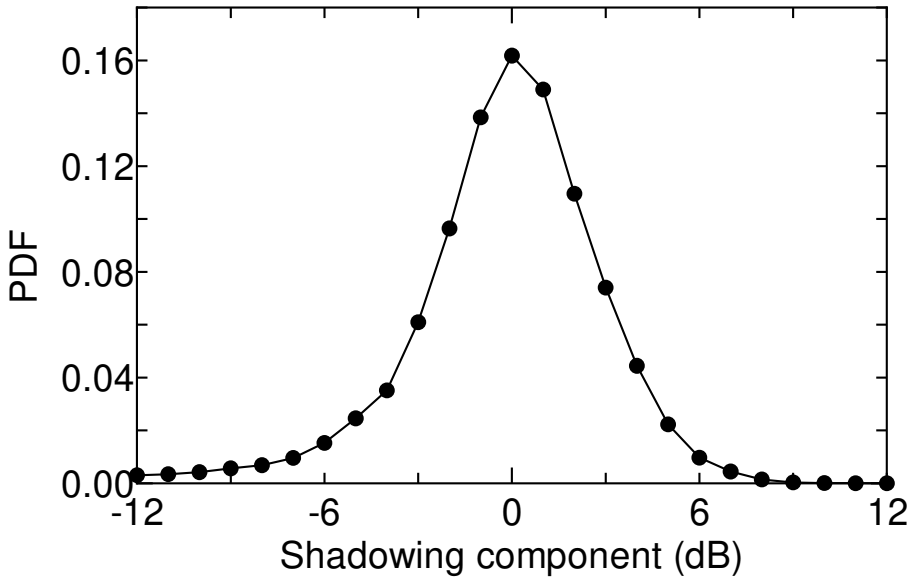


Fig. 7.6 PDF of shadowing component.

obtained:

$$L_{\mathrm{loss}}(d_{l,j}) = \begin{cases} -5.4919\log_{10}(d_{l,j}) - 30.1360 & (d_{l,j} \leq R_{\mathrm{b}}) \\ -31.925\log_{10}(d_{l,j}) + 7.9852 & (d_{l,j} > R_{\mathrm{b}}) \end{cases} \quad [\mathrm{dB}], \tag{7.3}$$

From these value, we can find that the path loss coefficient is 3.1925 for $d_{l,j} > R_{\mathrm{b}}$. The local database estimated the shadowing components by calculating the difference between the path loss from the average received power in each mesh. The estimated PDF of the shadowing is depicted in Fig. 7.6. This figure clarifies that the shape of the PDF may be equivalent to the log-normal distribution. Additionally, we confirmed that the mean is 0.0005198 [dB] and the standard deviation is 3.776 [dB]. A good matching can be found because these are the typical values in suburban areas [15, 140, 31].

7.4.3 Estimation Accuracy

The estimation accuracy of radio maps is verified. The datasets of day 1 and day 2 were used to create radio maps. Then, the local database statistically processed these datasets using several mesh scales, and estimated the average power in each reception mesh. Finally, we utilized the datasets of day 3 as the actual received signal power value. The RMSE was derived by calculating the difference between the estimated and actual values.

As the comparative method, this subsection uses various path loss models. First, the measurement-based path loss model was estimated using the datasets of day 1 and day 2. Several complicated methods have been established; however, the accuracy may be poor [9]. Hence, we use the measurement-based method in path loss estimation. The path loss model was estimated based on the same procedures described in Sect. 3.3.2. To improve the estimation accuracy of the path loss, the break point was considered. This method estimated the path loss model before and after the break point separately. We could obtain the path loss models as follows:

$$L_{\mathrm{loss}}(d_{l,j}) = \begin{cases} -8.5478\log_{10}(d_{l,j}) - 25.795 & (d_{l,j} \leq R_{\mathrm{b}}) \\ -43.364\log_{10}(d_{l,j}) + 27.312 & (d_{l,j} > R_{\mathrm{b}}) \end{cases} \quad [\mathrm{dB}], \tag{7.4}$$

The above parameters are different from Eq. (7.3) since Eq. (7.3) is calculated by utilizing all measured datasets. Meanwhile, Eq.(7.4) is estimated using datasets of the days 1 and 2.

The measurement-based path loss method estimated the average received signal power by subtracting the path loss from the transmit power. Then, the RMSE was derived by calculating the difference between the estimated and actual powers. The datasets of day 3 were treated as actual powers.

Moreover, the ITU-R P.1411 [13] was used to evaluate the estimation accuracy when a theoretical model is considered. The propagation model of the ITU-R P.1411 is given by

$$L_{\mathrm{ITU}}(d_{l,j}) = L_{\mathrm{b}} + 6 + \begin{cases} 20\log_{10}\left(\frac{d_{l,j}}{R_{\mathrm{b}}}\right) & (d_{l,j} \leq R_{\mathrm{b}}) \\ 40\log_{10}\left(\frac{d_{l,j}}{R_{\mathrm{b}}}\right) & (d_{l,j} > R_{\mathrm{b}}) \end{cases} \quad [\mathrm{dB}], \tag{7.5}$$

where L_{b} [dB] is defined as the propagation loss from a transmitter to the break point. The equation is defined as,

$$L_{\mathrm{b}} = \left| 20\log_{10}\left(\frac{\lambda^2}{8\pi h_{\mathrm{Tx}} h_{\mathrm{Rx}}}\right)\right|, \tag{7.6}$$

The local database calculated the average received signal power by subtracting the derived propagation loss from the transmit power. Then, the estimation accuracy was evaluated.

Furthermore, the two-ray path loss model [14] and Okumura–Hata model [12] were utilized. The former considers the direct and reflection waves. This assumption matches with the V2V communications in the LOS environment. Meanwhile, the latter is the famous empirical model.

Fig. 7.7 depicts the estimation accuracy of each method. From this figure, it can be confirmed that the fitted path loss model could estimate the radio propagation with the error of 5.546 [dB]. Meanwhile, the proposed maps were superior to the path loss model. The

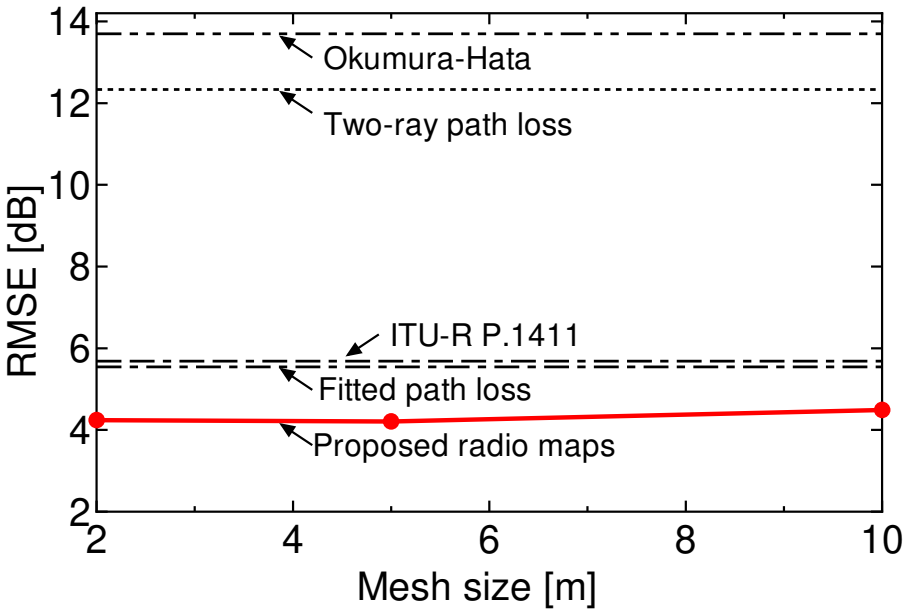


Fig. 7.7 Mesh size versus RMSE.

RMSEs are around 4.2–4.5 [dB] in the proposed method. When we consider the standard deviation of the shadowing and the effect of the small-scale fading, the proposed method can achieve the small RMSEs. Additionally, the proposed radio maps can perfectly estimate the path loss and shadowing in each mesh because the smallest RMSE of the radio map is empirically around 4 [dB] [57]. On the other hand, in the comparatives, the estimation accuracy is poor in each method. These results reveal that the proposed maps enable us to skillfully estimate the radio propagation even in MANETs. The communication efficiency will be improved by appropriately designing parameters based on the radio maps.

Note that the missing data can be estimated using the fitted path loss model. Hence, the error characteristics mean the performance of the fitted path loss-aided interpolation.

7.4.4 Evaluation of Communication Efficiency

Finally, the communication efficiency is evaluated when the radio maps are used. In the unicast V2V communication, the interference may occur at surrounding vehicles except for the destination if a packet is sent with high transmission power. Thus, it is necessary to properly determine the transmission power to avoid interference. Thus, simple power control is conducted in this subsection. In the proposed power control, the transmission power is determined so that the permissible outage probability is guaranteed. Here, the

outage probability is defined that the instantaneous received signal power is less than the desired power. We show the procedures of the power control as follows:

i). The local database creates the radio maps as mesh size is 10m using the datasets of day 1 and day 2.

ii). We calculate the empirical CDF of the multipath fading by calculating the difference between instantaneous received signal power and the average received signal power that contains the path loss and shadowing. The shape of the CDF is assumed to be static in the evaluation area.

iii). The datasets of day 3 are utilized as actual received signal power values. The instantaneous received signal power is calculated in each mesh using the estimated statistical information.

iv). The power control is performed using the estimated average values and the empirical CDF of the multipath fading so that the permissible outage probability is guaranteed.

v). We apply this power control to measured instantaneous datasets of day 3.

The power control is evaluated in the area shown in Fig. 7.5. As a comparative, the empirical path loss model is used. This method calculates the average received signal power using the measurement-based path loss model fitted to Eq.(7.1). We obtained the following path loss models via the linear regression:

$$L_{\mathrm{loss}}(d_{l,j}) = \begin{cases} -9.6183\log_{10}(d_{l,j}) - 24.335 & (d_{l,j} \leq R_{\mathrm{b}}) \\ -43.470\log_{10}(d_{l,j}) + 26.699 & (d_{l,j} > R_{\mathrm{b}}) \end{cases} \quad [\mathrm{dB}]. \tag{7.7}$$

The above parameters are different from both Eqs. (7.3) and (7.4) since Eq. (7.7) is calculated using datasets in Fig. 7.5. Meanwhile, we calculate both Eqs. (7.3) and (7.4) based on the datasets in all areas.

The power control was conducted as the permissible outage probability is from 0.1 to 0.2 and the desired received signal power is -82.0 [dBm]. Fig. 7.8 presents the outage characteristics. The estimation accuracy of the radio propagation may be poor when the path loss-based method is applied. Thus, the outage probability is smaller than the permissible value owing to the inaccurate determination of the transmission power. Meanwhile, the proposed method can skillfully guarantee the permissible outage value compared to the path loss-based method. However, it can be found that the slight increase of the outage probability occurs even in the proposed method. We consider the reasons as follows:

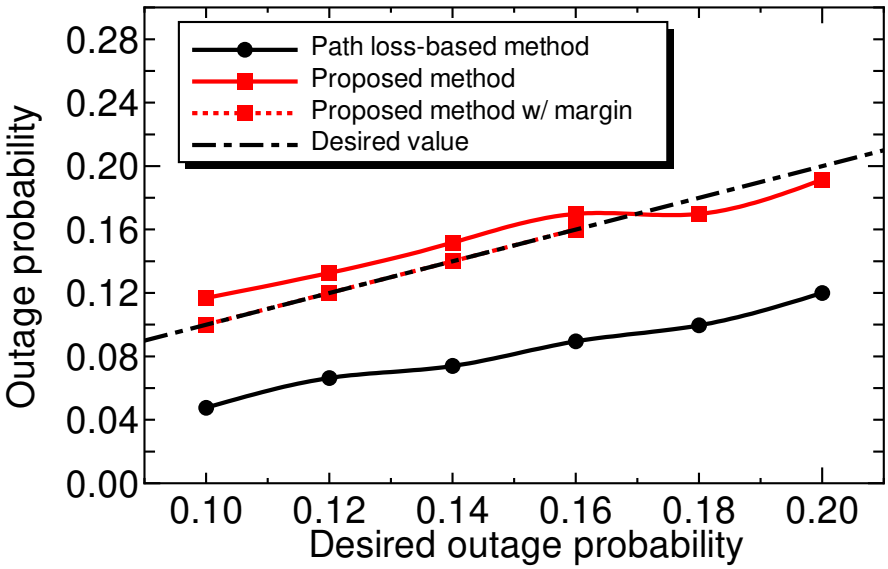


Fig. 7.8 Outage probability characteristics after power control.

- The estimation accuracy of the empirical CDF may be slightly degraded in the tail area. Although the CDF in the tail area should be estimated with high accuracy to improve the packet delivery ratio, the acquisition of such information may be very difficult because of the limited datasets.

- The maximum transmission power is limited around 19.2 [dBm]; that is, several samples having low SNR may be missing.

To accurately guarantee permissible outage probability in such the constraints, the margin should be added to the transmission power. We illustrate the outage probability and the average transmission power with margins as the red dotted lines in Figs. 7.8 and 7.8, respectively.

These results reveal that the permissible outage probability can be guaranteed with the small margin in the proposed method. Additionally, the average transmission power is low compared to the path loss-based method even if the margin is added. In summary, the proposed method can improve communication efficiency in V2V communications. This fact will enhance spectral efficiency.

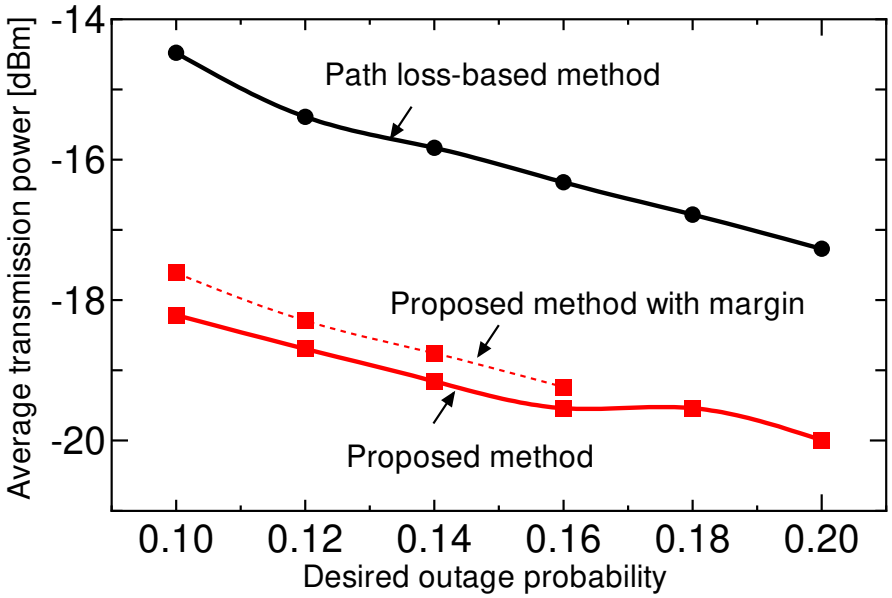


Fig. 7.9 Average transmission power.

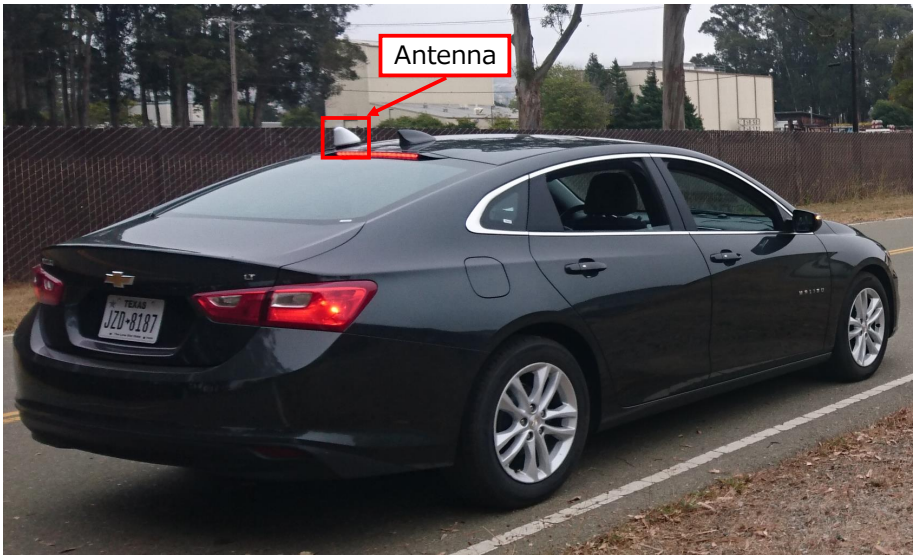


Fig. 7.10 An observation vehicle.

7.5 Measurement Campaign in 5.8GHz Band

In the previous campaign, since the number of datasets was small in several meshes, the packet delivery rate (PDR) could not be accurately derived. To solve this task, we had conducted the measurement campaign of V2V communications as described in Sect. 5.8.

We prepared an observation vehicle as shown in Fig. 7.10 and mounted an on-board unit on the vehicle. In the experiment, three vehicles communicated with each other and observed the radio environment on the route shown in Fig. 7.11. The observation routes consisted of two test courses, and the measurement campaign was performed for 6 hours in each route. There are many buildings on the southern route and only a few on the northern route. Therefore, we can compare the radio environment between NLOS and LOS. To obtain enough datasets, two vehicles traveled one lap on the measurement course at 20 [km/h] and 0.2 [km/h], respectively. The other vehicle kept moving on the course at 20 [km/h].

The communication method is based on dedicated short range communication and the center frequency is 5890 [MHz]. An in-vehicle device is connected to the laptop PC via an Ethernet cable and the beacon signal is transmitted and received by executing shell scripts in Ubuntu 16.04. Each vehicle observes the transmitted and received time, the received signal power, packet ID, the transmitted and received position and so on. After the execution of the shell scripts, these measurement datasets are recorded to the transmission and reception log-file. Note that the MAC address is set to each in-vehicle device and recorded to the

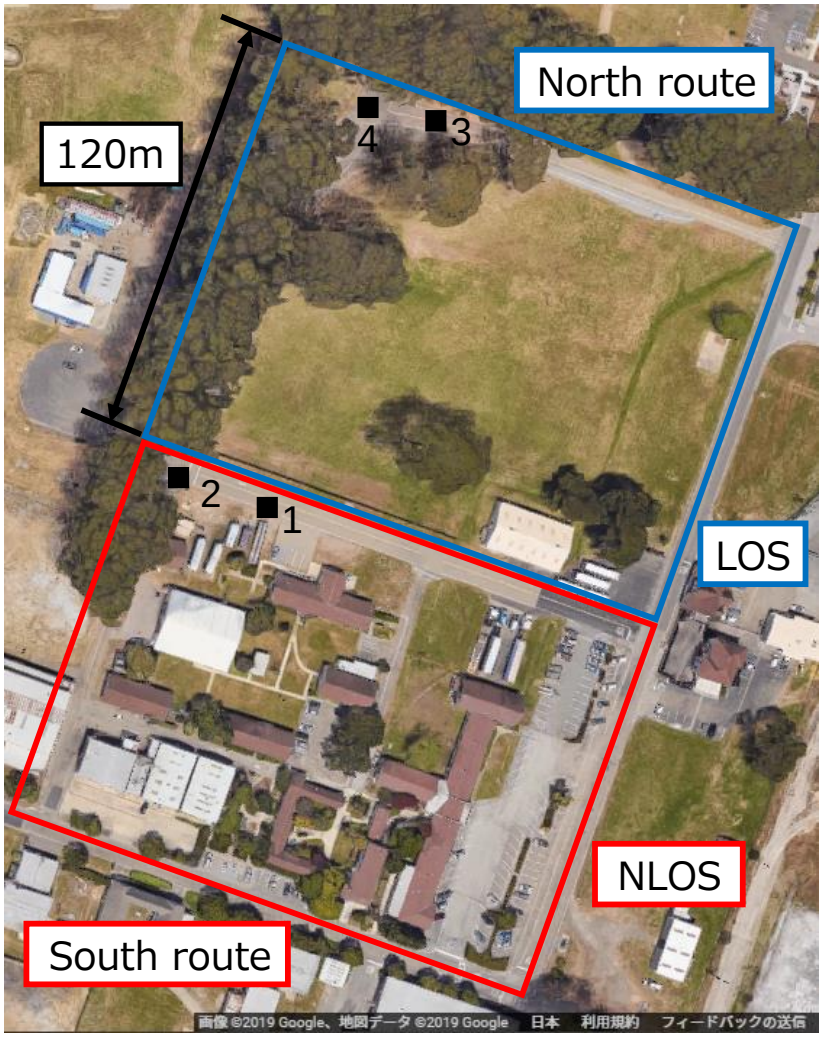
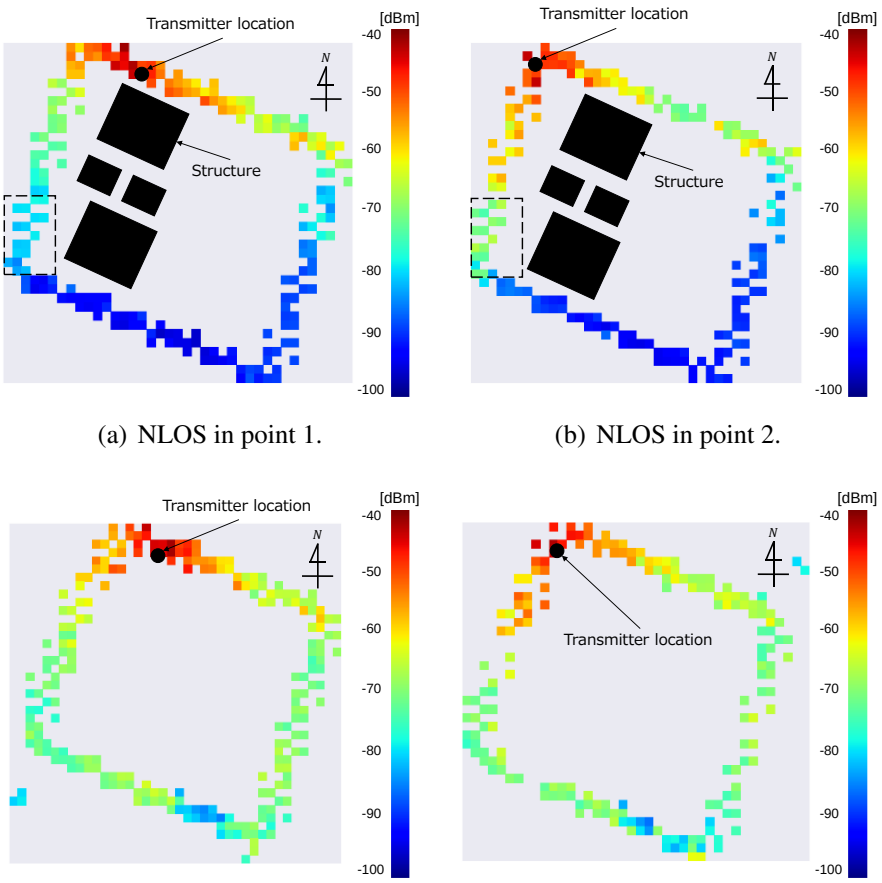


Fig. 7.11 The observation routes in California Path.

reception log file. The number of measurement datasets are 15,280,657 and 14,760,806 in NLOS and LOS, respectively and these datasets are statistically processed in MySQL server.

7.6 Measurement Results in California Path

This section presents the measurement results in the California path. Examples of radio maps are first explained. Then, the scatter plots of the average received signal power are shown. Finally, we describe the PDR characteristics.



(a) NLOS in point 1.

(b) NLOS in point 2.

(c) LOS in point 3.

(d) LOS in point 4.

Fig. 7.12 Example of radio maps in California path.

7.6.1 Example of Radio Maps in NLOS and LOS

Fig. 7.12 shows the example of radio maps in NLOS and LOS, respectively. In these maps, a black circle denotes the transmitter location corresponding to the black square in Fig. 7.11. Here, the mesh size is 5 [m]. We show the four radio maps to explain difference of the structure-dependent radio propagation. In Figs. 7.12(a) and 7.12(b), the received signal power is fluctuated by changing the transmitter location. For example, in the west area represented the dotted line, around 20 [dB] difference can be confirmed. The main factor of the difference is shadowing. In Fig. 7.12(a), since the structures are presented between the transmitter mesh and the dotted line area, the received signal power is relatively lower by the shadowing. On the other hand, the structures are not presented in Fig. 7.12(b): the shadowing does not occur in this area. Next, we explain the radio maps in LOS. As can be seen from Figs. 7.12(c) and 7.12(d), the difference of the received signal power in each mesh is small.

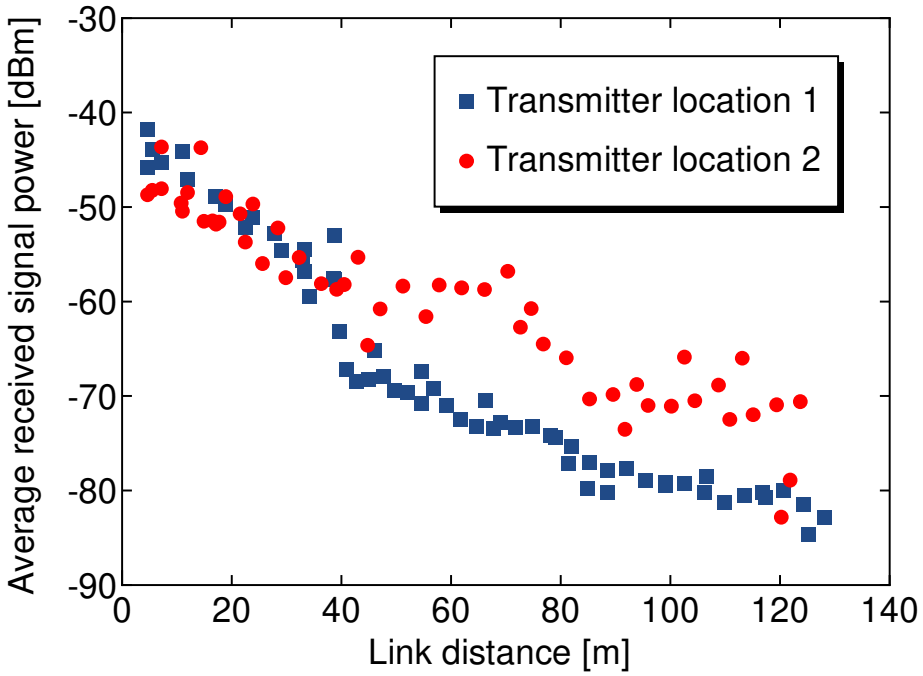


Fig. 7.13 Scatter plot in NLOS.

7.6.2 Scatter Plots of Average Received Signal Power

Next, the detailed fluctuation of the average received signal power is described. Figs. 7.13 and 7.14 show the scatter plots in NLOS and LOS environments, respectively. In these figures, the average received signal power values are plotted in each transmitter location shown in Fig. 7.11. From Fig. 7.13, we can confirm that the average received signal power is different around 20–25 [dB] in a received position on NLOS. This phenomenon is shadowing in the measurement route. Meanwhile, the significant difference is not observed in the LOS environment from Fig. 7.14.

7.6.3 RMSE Characteristics

This subsection shows the RMSE of the proposed radio maps to confirm the estimation accuracy of the average received signal power. Each measured data in NLOS and LOS environments were divided into three groups and two datasets are used to create radio maps. The other dataset is used as the test data. The RMSE was derived by calculating the difference between the estimated and actual values. The worst RMSE is derived based on the cross-validation.

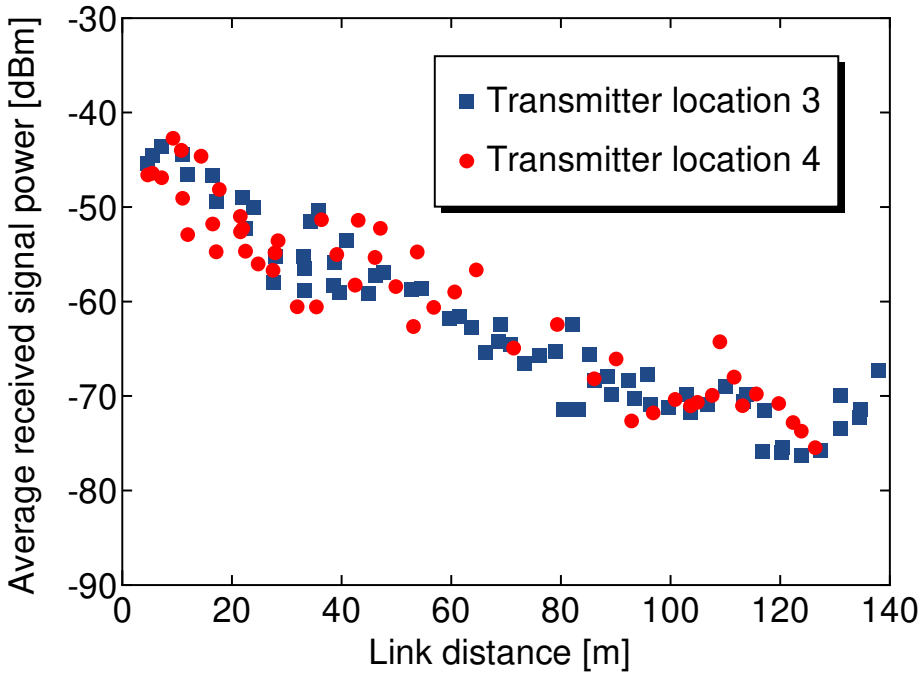


Fig. 7.14 Scatter plot in LOS.

Fig. 7.15 shows the RMSE in the 5.8GHz band datasets. The solid and the dotted lines denote the RMSEs in NLOS and LOS environments, respectively. The red line prensents the RMSE of the proposed radio maps. Meanwhile, the black line is the RMSE of the fitted path loss model. In the path loss estimation, the logarithmic distance between a transmission position and a reception position was first calculated. Next, the scatter diagram was created as the horizontal axis is the logarithmic distance and the vertical axis is the average power in each mesh. Finally, the path loss parameters B and C were estimated using the least squared method.

The results mean that the proposed radio maps can accurately estimate the received signal power compared to the fitted path loss model. Since the fitted path loss model cannot consider the shadowing deviation in each mesh, the RMSE is larger than the proposed radio maps. We can confirm that the RMSEs become large by increasing the mesh size in both methods. This is because the estimation accuracy of the average power in each mesh may be degraded owing to the fluctuation of an instantaneous power.

7.6.4 PDR Characteristics

Finally, this section describes the PDR characteristics in NLOS and LOS environments. For calculating PDR in each 5m-mesh, we count the number of transmitted and received

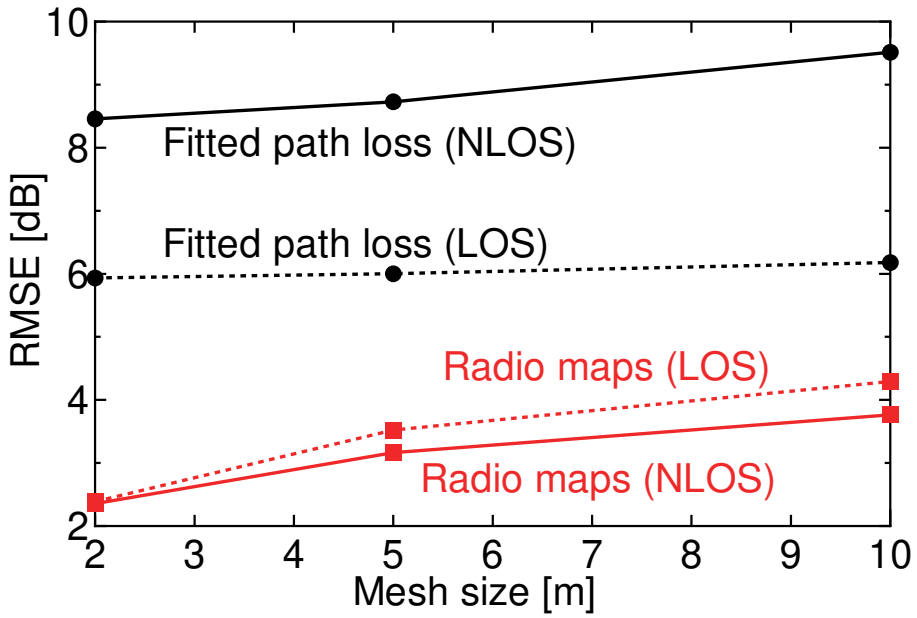
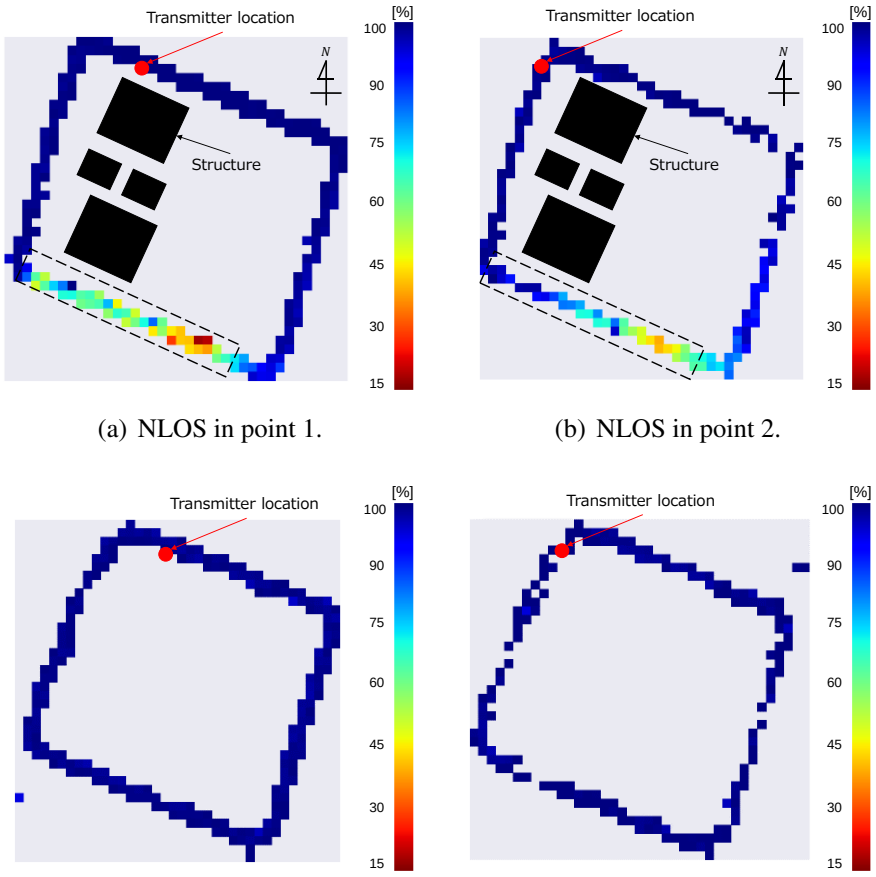


Fig. 7.15 RMSE characteristics in 5.8GHz.

packets in each transmitter / receiver mesh from the log-files. Note that the corresponding transmission and reception log-files are linked by using packet ID, MAC address, and the transmission / reception time.

Fig. 7.16 shows the PDR in each 5m-mesh. A red circle denotes the transmitter location corresponding to the black circle in Fig. 7.12. In NLOS, it can be found that the PDR is relatively lower in the south area representing the dotted line. This is because the shadowing occurs due to the structures. On the other hand, in LOS, the PDR is higher in all meshes, and the minimum PDR is around 90.0 [%]. From these results, we can confirm that radio maps and PDR maps can be used for accurately estimating the structure-dependent radio environment. It is expected that the proposed database can drastically improve the communication efficiency of V2V communications. Note that in V2V communications, since both of a transmitter and a receiver dynamically move, it is difficult to construct the radio maps and PDR maps in each transmission / reception position. However, from the PDR maps in LOS, it can be found that the radio environment is stable in all meshes. Therefore, in future V2V communications, we generate the radio maps and PDR maps in some meshes such as NLOS environments.



(a) NLOS in point 1.

(b) NLOS in point 2.

(c) LOS in point 3.

(d) LOS in point 4.

Fig. 7.16 Example of PDR in California path.

7.7 Chapter Summary

This chapter has proposed radio maps for MANETs. The evaluation results showed that the radio maps can accurately estimate the radio propagation fluctuations compared to some path loss models including the measurement-based path loss model and ITU-R P.1411. Additionally, the transmission power can be significantly decreased via the proposed power control method. Furthermore, the PDR characteristics have been clarified in NLOS and LOS environments. The measurement results have shown that the PDR is notably degraded in NLOS where there are structures near the transmitter and the receiver. We conclude that the measurement-based radio maps and PDR maps can be used for estimating radio propagation characteristics in not only conventional applications where the transmitter is fixed, but also MANETs.

Chapter 8

Conclusions and Future Works

This chapter concludes our dissertation. After summarizing the research contents of each chapter, we explain our future works.

8.1 Conclusions

This dissertation dealt with several issues of the radio map construction based on statistical inference. To expand the utilization of the radio map, we considered novel methods for constructing the radio map. The research contents are summarized as follows.

Chapter 1 explained the background of our research. The relation between the communication quality and radio propagation estimation was first shown. Then, the radio map and its remaining tasks are summarized.

Chapter 2 introduced the concept of the spectrum database and the radio map. The hierarchical architecture of the spectrum database was first presented. Then, crowdsourcing-assisted radio map construction and its problem were explained. Finally, the mesh definitions and utilization procedures of the radio map were introduced.

In chapter 3, the shadowing classifier was proposed to reduce the registered data size of the radio map. This classifier was created by quantizing the measured shadowing values by a certain step width. After that, the quantized shadowing value was assigned to multiple meshes that have similar shadowing components based on the proposed objective function. We conducted the emulation using the two measured datasets that were obtained in the actual cellular systems and V2V communications. The evaluation results clarified that the proposed classifier could reduce the registered data size while keeping the high estimation accuracy even if there were outliers of the average received signal power. Meanwhile, two conventional clustering methods: k-means++ and GMM could not skillfully classify the shadowing values owing to the outliers.

Additionally, motivated by the reduction of instantaneous sample size, Chapter 4 formulated the minimum required sample size for estimating the average received signal power. This chapter used three statistical methods: the confidence interval, the CLT, and t-test to calculate the sample size. The simulation results showed that the radio map could be accurately constructed with a small number of samples compared to the conventional method that was without consideration of the required sample size.

Next, Chapter 5 proposed the hypothesis testing-based updating method of the radio map. Based on our investigations for the radio map updating, we used Welch's t-test that is an improved method of the t-test. The proposed method inferred the significant difference between two mean values before and after the movement of the transmitter. Then, the radio map was updated if a significant difference was detected. The simulation results revealed that the proposed method enabled us to correctly update the radio map compared to conventional updating methods that were the unique averaging and forgetting factor-based methods. Additionally, we clarified that Welch's t-test was superior to the Mann–Whitney U-test in terms of the statistical power.

Considering interference-limited observations in multiple transmitters environment, Chapter 6 proposed the novel extrapolation method based on the compensation of the empirical CDF. The empirical CDF was first estimated using the measured target power samples. Then, the proposed method compensated for the empirical CDF by taking into account the number of missing data. Finally, the median target power was extrapolated using the compensated empirical CDF. The emulation results revealed that the radio map could be extrapolated with high accuracy compared with conventional interpolation and extrapolation methods.

Finally, Chapter 7 applied the radio map to MANETs. This scheme first collected received signal power samples in each pair of a transmitter and a receiver. By averaging these samples, a radio map could be created in each transmission position. To verify the usefulness of the application, we had conducted two measurement campaigns in Japan and the United States of America. The experimental results clarified that the proposed application could accurately estimate the location-dependent radio propagation even in MANETs. Additionally, the PDR characteristics were studied in the LOS and NLOS environments. The results showed that the PDR was notably reduced in the NLOS environment.

This dissertation constructs the radio maps over 3.5 [GHz] and 760 [MHz]. For using radio maps in various systems, we need to evaluate the accuracy of radio maps over several frequencies. Here, [109, 110, 45] have described that the average received signal power indicates a strong correlation over the frequency domain even if the two measurement signals are separated over hundred MHz in cellular systems. These results imply that the radio map may be accurately created over various frequencies.

8.2 Status of Our Research and Future Outlook

In wireless communications, radio propagation estimation has been eagerly studied for a long time by many researchers. As one of the methods for estimating radio propagation, a radio map is well known by several researchers. So far, the effectiveness of the radio map has been evaluated in a typical wireless system that a transmitter is singly located in a fixed position. The estimation of television white spaces is a typical utilization example of the radio map. As the construction method of the radio map, most conventional works have only considered the spatial interpolation in the fixed transmitter environment.

However, the situation where a transmitter is fixed is very limited in realistic wireless communications. With the rapid development of wireless technology in past decades, various systems have appeared, such as the MANETs, the private 5G, and WLANs. Unfortunately, the radio map cannot be applied or the accuracy of the map may be degraded in such systems because of the following reasons: i) the incompatibility for the dynamic movement of both transmitter and receiver, ii) no updating methods for the map in the private 5G with the movement of a transmitter, and iii) no extrapolation schemes in interference-limited observations (i.e., dense APs environment). Additionally, the increase of the registered data size has not been considered in conventional radio maps. As a result, the continuous development of wireless communications may be disturbed owing to the inaccuracy of the radio propagation estimation. For future wireless systems, it is necessary to estimate radio propagation more accurately and to properly determine communication parameters according to the estimation results. Motivated by the facts, the purpose of our research has been determined to expand the use of the radio map for drastically improving the communication efficiency in any system.

Thus, we have comprehensively attempted to solve these problems from both theoretical and experimental aspects. In our dissertation, all proposed methods are based on various statistical inferences rather than deterministic methods. Such statistical methods enable us to probabilistically consider the uncertainty of radio propagation. Especially, in the radio map construction, by the proposed statistical inference methods, the updating and extrapolation can be realized in addition to the spatial interpolation. *We consider that the updating and extrapolation of the radio map are the main contributions of our dissertation in the academic viewpoints.* The accurate radio maps can be created in various wireless systems other than the cellular networks by the proposed statistical inference-based methods. As a result, more wireless systems can communicate with each other in various environments while efficiently utilizing finite spectrum resources. In the future, we believe that the radio map may significantly change the design of wireless communications, such as interference management in spectrum sharing.

8.3 Future Works

Statistical inference methods could enhance the accuracy of the radio map. However, there are several remaining tasks to more accurately generate the radio map. The remaining tasks are summarized as follows.

The Calibration of Received Signal Power in Crowdsourcing

There are several problems to realize crowdsourcing-assisted radio map construction. First, the measurement accuracy of smartphones is very low compared to the spectrum analyzer. Especially, the instantaneous received signal power may sometimes fluctuate around 10 [dB] depending on the types of smartphones [23]. This is the most difficult task in crowdsourcing. Hence, many researchers have considered the solution of this task and one of the famous solutions is the *calibration* [23, 62]. As a simple calibration, the offset between smartphones and the spectrum analyzer is estimated using linear regression. If there is a constant offset, the measurement error of smartphones can be easily calibrated. However, the offset strongly depends on the types of smartphones. Thus, to construct an accurate radio map, it is necessary to estimate various offsets by considering many types of smartphones in real environments. Additionally, the offset significantly fluctuates according to how each observer holds their smartphone. For example, the smartphone is held in hand or in the bag. Although we must know these measurement situations using external information, the detailed method has not been established in the radio map construction. Since these tasks are very challenging, we consider the problems as future works.

Radio Map Extrapolation Under Noise-Limited Observations

We proposed the extrapolation method for the radio map in Chapter 6. However, this method could be accurately performed if the noise floor of a mobile terminal is smaller enough than the aggregate interference power in each position. If not, the number of missing data may increase; thus, such a situation can be considered as noise-limited observations. Under the environment, the radio map may be inaccurately extrapolated by the proposed method. In the spectrum sharing, a secondary user must determine their transmission power so that the interference power is lower than the noise floor of a primary receiver. For instance, the threshold is 10 [dB]. To this end, the radio propagation below the noise floor should be extrapolated.

Optimization of Mesh Size

As described in Sect. 1.4.3, the typical mesh size is 10 [m] in an urban area considering the correlation distance of the shadowing. However, the correlation distance may be different according to the geographic conditions, such as the number of structures. There is a trade-off between the mesh size and the estimation accuracy of the radio propagation. Hence, it is necessary to optimize the mesh size by taking into account the estimation accuracy.

8.3.1 Privacy Protection in Crowdsourcing

The privacy protection of ordinary people has been considered in several works [141, 142]. Ordinary people participating in crowdsourcing send their location information and terminal identification to the local database. However, several researchers have been concerned that persons do not want to be known about these individual data considering privacy protection. This problem is also discussed in the connected car-based V2V communications [143]. If private information is abused, who is responsible for it? To solve this problem, we consider that legal development is required for crowdsourcing. Maybe, it will take a long time. All researchers should pay attention to this field.

8.3.2 Investigation of Correlation Characteristics over Frequency Domain in MANETs

Although the shadowing correlation over the frequency domain has been confirmed in cellular systems, the correlation characteristics in MANETs have not been clarified yet. In future, we need to investigate the shadowing correlation over various frequencies via experiments and related works.

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2017–2022,," [Online]. available: https://s3.amazonaws.com/media.mediapost.com/uploads/CiscoForecast.pdf.

[2] S. B. of Japan, "Method of demarcation for grid square," [Online]. available: https://www.stat.go.jp/english/data/mesh/05.html.

[3] MIC, "The radio use web site," [Online]. available: https://www.tele.soumu.go.jp/e/adm/freq/search/myuse/0002/index.htm.

[4] NTIA, "United states frequency allocation chart," [Online]. available: https://www.ntia.doc.gov/files/ntia/publications/january_2016_spectrum_wall_chart.pdf.

[5] L. Beltramelli, A. Mahmood, P. Ferrari, P. Österberg, M. Gidlund, and E. Sisinni, "Synchronous lora communication by exploiting large-area out-of-band synchronization," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 7912–7924, December 2020.

[6] H. S. Oh, D. G. Jeong, and W. S. Jeon, "Joint radio resource management of channel-assignment and user-association for load balancing in dense wlan environment," *IEEE Access*, vol. 8, pp. 69 615–69 628, April 2020.

[7] K. H. Anabi, R. Nordin, and N. F. Abdullah, "Database-assisted television white space technology: challenges, trends and future research directions," *IEEE Access*, vol. 4, pp. 8162–8183, November 2016.

[8] A. Ghasemi and E. S. Sousa, "Fundamental limits of spectrum-sharing in fading environments," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 649–658, February 2007.

[9] C. Phillips, D. Sicker, and D. Grunwald, "Bounding the error of path loss models," in *Proc. DySPAN 2011*, May 2011.

[10] A. F. Molisch, *Wireless Communications*. Wiley Publishing, 2005.

[11] K. Sato and T. Fujii, "Kriging-based interference power constraint: Integrated design of the radio environment map and transmission power," *IEEE Trans. Cong. Commun. Netw.*, vol. 3, no. 1, pp. 13–25, March 2017.

[12] Y. Okumura, T. Ohmori, E. Kawano, and K. Fukuda, "Field strength and its variability in vhf and uhf land mobile radio service," *Rev. Electr. Commun. Lab.*, vol. 16, no. 9-10, pp. 825–828, October 1968.

[13] I. Recommendation, "Propagation data and prediction methods for the planning of shortrange outdoor radiocommunication systems and radio local area networks in the frequency range 300 mhz to 100 ghz," in *Rec. ITUR P.1411-8*, July 2015.

[14] A. J. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[15] J. E. Berg, R. Bownds, and F. Lotse, "Path loss and fading models for microcells at 900mhz," in *Proc. IEEE VTC*, May 1992.

[16] Y. Ma, X. Zhang, and Y. Gao, "Joint sub-nyquist spectrum sensing scheme with geolocation database over tv white space," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3998–4007, December 2018.

[17] A. Achtzehn, J. Riihijärvi, and P. Mähönen, "Improving accuracy for tvws geolocation databases: Results from measurement-driven estimation approaches," in *Proc. IEEE DySPAN 2014*, April 2014.

[18] E. Dutkiewicz, B. A. Jayawickrama, and Y. He, "Radio spectrum maps for emerging iot and 5g networks: Applications to smart buildings," in *Proc. 2017 ICECOS*, August 2017.

[19] I. Kakalou, K. Psannis, S. K. Goudos, T. V. Yioultsis, N. V. Kantartzis, and Y. Ishibashi, "Radio environment maps for 5g cognitive radio network," in *Proc. 2019 MOCAST*, May 2019.

[20] E. Dutkiewicz, Y. He, B. A. Jayawickrama, and H. V. Abeywickrama, "Radio environment maps generation and spectrum sensing testbed for spectrum sharing in 5g networks," in *Proc. 2017 IEEE-APS APWC*, September 2017.

[21] X. Wang, X. Wang, S. Mao, S. Zhang, S. C. G. Periaswamy, and J. Patton, "Indoor radio map construction and localization with deep gaussian processes," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11 238–11 249, May 2020.

[22] S. H. Jung and D. Han, "Automated construction and maintenance of wi-fi radio maps for crowdsourcing-based indoor positioning systems," *IEEE Access*, vol. 6, pp. 1764–1777, December 2017.

[23] 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. 25, no. 11, pp. 2669–2681, December 2015.

[24] X. Wang, M. Umehira, B. Han, P. Li, Y. Gu, and C. Wu, "Online incentive mechanism for crowdsourced radio environment map construction," in *Proc. IEEE ICC 2019*, May 2019.

[25] S. H. Jung and D. Han, "Automated construction and maintenance of wi-fi radio maps for crowdsourcing-based indoor positioning systems," *IEEE Access*, vol. 6, pp. 1764–1777, December 2017.

[26] Y. Ye and B. Wang, "Rmapcs: Radio map construction from crowdsourced samples for indoor localization," *IEEE Access*, vol. 6, pp. 24 224–24 238, April 2018.

[27] X. Ying, S. Roy, and R. Poovendran, "Pricing mechanism for quality-based radio mapping via crowdsourcing," in *Proc. 2016 IEEE GLOBECOM*, December 2016.

[28] K. Katagiri, K. Sato, and T. Fujii, "Crowdsourcing-assisted radio environment database for v2v communications," *Sensors*, vol. 18, no. 4, pp. 1–17, April 2018.

[29] F. Wu, N. Kubo, and A. Yasuda, "Performance evaluation of gps augmentation using quasi-zenith satellite system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 4, pp. 1249–1260, October 2004.

[30] K. Sato, M. Kitamura, K. Inage, and T. Fujii, "Measurement-based spectrum database for flexible spectrum management," *IEICE Trans. Commun.*, vol. E98-B, no. 10, pp. 2004–2013, October 2015.

[31] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron Lett.*, vol. 27, no. 23, pp. 2145–2146, November 1991.

[32] M. Barrie, S. Delaere, G. Sukarevicienė, J. Gesquiere, and I. Moerman, "Geolocation database beyond tv white spaces? matching applications with database requirements," in *Proc. IEEE DySPAN 2012*, October 2014.

[33] M. Denkovska, P. Latkoski, and L. Gavrilovska, "Geolocation database approach for secondary spectrum usage of tvws," in *2011 TELFOR Proc. Papers*, November 2011.

[34] Y. Chen, H. Zhang, H. Hu, and Q. Wang, "A new cooperative spectrum sensing with radio environment map in cognitive radio networks," in *Proc. 2015 Int. Conf. Intell. Comut. Internet Things*, January 2015.

[35] B. Huang, Z. Xu, B. Jia, and G. Mao, "An online radio map update scheme for wifi fingerprint-based localization," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6909–6918, April 2019.

[36] X. Mo, Y. Huang, and J. Xu, "Radio-map-based robust positioning optimization for uav-enabled wireless power transfer," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 179–183, October 2020.

[37] M. Suchanski, P. Kaniewski, J. Romanik, and E. Golan, "Radio environment map to support frequency allocation in military communications systems," in *2018 URSI*, May 2018.

[38] K. Sato, K. Inage, and T. Fujii, "On the performance of neural network residual kriging in radio environment mapping," *IEEE Access*, vol. 7, pp. 94 557–94 568, July 2019.

[39] Y. Q. Xu, B. Zhang, G. Ding, B. Zhao, S. Li, and D. Guo, "Radio environment map construction based on spatial statistics and bayesian hierarchical model," *IEEE Trans. Cong. Commun. Netw.*, vol. 7, no. 3, pp. 767–779, September 2021.

[40] T. Fujii, "Smart spectrum management for v2x," in *Proc. IEEE DySPAN 2018*, October 2018.

[41] M. Lee and D. Han, "Voronoi tessellation based interpolation method for wi-fi radio map construction," *IEEE Commun. Lett.*, vol. 16, no. 3, pp. 404–407, February 2012.

[42] J. Ojaniemi, J. Kalliovaara, J. Poikonen, and R. Wichman, "A practical method for combining multivariate data in radio environment mapping," in *Proc. 2013 IEEE PIMRC*, September 2013.

[43] S. Debroy, S. Bhattacharjee, and M. Chatterjee, "Spectrum map and its application in resource management in cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 4, pp. 406–419, December 2015.

[44] T. Farnham, "Radio environment map techniques and performance in the presence of errors," in *Proc. 2016 IEEE PIMRC*, September 2016.

[45] K. Sato, K. Suto, K. Inage, K. Adachi, and T. Fujii, "Space-frequency-interpolated radio map," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 714–725, January 2021.

[46] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Structured group sparsity: A novel indoor wlan localization, outlier detection, and radio map interpolation scheme," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6498–6510, November 2017.

[47] A. E. C. Redondi, "Radio map interpolation using graph signal processing," *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 153–156, October 2018.

[48] M. Zhou, Y. Tang, Z. Tian, and X. Geng, "Semi-supervised learning for indoor hybrid fingerprint database calibration with low effort," *IEEE Access*, vol. 5, pp. 4388–4400, March 2017.

[49] G. Boccolini, G. Hernandez-Penaloza, and B. Beferull-Lozano, "Wireless sensor network for spectrum cartography based on kriging interpolation," in *Proc. 2012 IEEE PIMRC*, September 2012.

[50] K. Katagiri, K. Onose, K. Sato, K. Inage, and T. Fujii, "Highly accurate prediction of radio propagation using model classifier," in *Proc. IEEE VTC-Spring 2019*, April 2019.

[51] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. 29, no. 3, pp. 317–325, August 1980.

[52] K. Katagiri, K. Sato, and T. Fijii, "Crowdsourcing-assisted radio environment database for v2v communication," *Sensors*, vol. 18, no. 4, pp. 1–17, April 2018.

[53] R. M. Esteves, T. Hacker, and C. Rong, "Cluster analysis for the cloud: Parallel competitive fitness and parallel k-means++ for large dataset analysis," in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci.*, December 2012.

[54] R. Hasegawa, K. Katagiri, K. Sato, and T. Fijii, "Low storage, but highly accurate measurement-based spectrum database via mesh clustering," *IEICE Trans. Commun.*, vol. E101.B, no. 10, pp. 2152–2161, October 2018.

[55] *ARIB STD-T109: 2013. 700MHz band intelligent transport systems version 1.2.* Association of Radio Industries and Businesses, 2013.

[56] J. Kim, "Using median as a threshold in determining anomaly in back-end authentication," in *Proc. ICCIA*, July 2018.

[57] K. Sato, "Measurement-based spectrum database for spatial spectrum sharing," *Doctor of Philosophy, The University of Electro-Communications*, March 2018.

[58] Y. Deng, L. Zhou, L. Wang, M. Su, J. Zhang, J. Lian, and J. Wei, "Radio environment map construction using super-resolution imaging for intelligent transportation systems," *IEEE Access*, vol. 8, pp. 47 272–47 281, March 2020.

[59] C. Zhang and D. Tao, "Structure of indicator function classes with finite vapnik–chervonenkis dimensions," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 24, no. 7, pp. 1156–1160, March 2013.

[60] N. Morgan and H. Bourlard, "Generalization and parameter estimation in feedforward nets: Some experiments," *Advances Neural Inf. Process. Syst.*, vol. II, no. 3852, pp. 630–637, June 1990.

[61] J. Robinson, R. Swaminathan, and E. W. Knightly, "Assessment of urban-scale wireless networks with a small number of measurements," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw.*, September 2008.

[62] Z. Fang, Z. Zhao, D. Geng, Y. Xuan, L. Du, and X. Cui, "Rssi variability characterization and calibration method in wireless sensor network," in *Proc. ICINFA 2010*, June 2010.

[63] G. Callebaut and L. V. Perre, "Characterization of lora point-to-point path loss: Measurement campaigns and modeling considering censored data," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1910–1918, March 2020.

[64] C. Gustafson, T. Abbas, D. Bolin, and F. Tufvesson, "Statistical modeling and estimation of censored pathloss data," *IEEE Wireless Commun. Lett.*, vol. 4, no. 5, pp. 1–7, October 2015.

[65] K. Sato, K. Inage, and T. Fujii, "Compensation of survivorship bias in path loss modeling," in *Proc. IEEE PIMRC 2017*, October 2017.

[66] J. Pek, A. C. M. Wong, and O. C. Y. Wong, "Confidence intervals for the mean of non-normal distribution: Transform or not to transform," *Open J. Statist.*, vol. 7, no. 3, pp. 405–421, June 2017.

[67] Y. Wang, Z. Lu, and S. Xian, "Parametric bootstrap confidence interval method for the power law process with applications to multiple repairable systems," *IEEE Access*, vol. 6, pp. 49 157–49 169, August 2018.

[68] N. M. Razali and B. W. Yap, "Power comparisons of shapiro–wilk, kolmogorov–smirnov, lilliefors and anderson–darling tests," *J. Statist. Model. Analytics*, vol. 2, no. 1, pp. 21–33, January 2011.

[69] S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, December 1965.

[70] G. L. Hickey, S. W. Grant, J. Dunning, and M. Siepe, "Statistical primer: Sample size and power calculations–why, when and how?" *Eur. J. Cardio-Thoracic Surv*, vol. 54, no. 1, pp. 4–9, July 2018.

[71] T. Lumley, P. Diehr, S. Emerson, and L. Chen, "The importance of the normality assumption large public health data sets," *Annu. Rev. Public Health*, vol. 23, pp. 151–169, February 2002.

[72] R. Zhang, J. Wei, D. G. Michelson, and V. C. M. Leung, "Outage probability of mrc diversity over correlated shadowed fading channels," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 516–519, July 2012.

[73] K. Sato and T. Fujii, "Average interference power constraint with measurement-based spectrum database," in *Proc. IEEE WCNC 2015*, March 2015.

[74] V. Kristem, C. U. Bas, R. Wang, and A. F. Molisch, "Outdoor wideband channel measurements and modeling in the 3-–18 ghz band," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4620–4633, April 2018.

[75] A. Rostami, "Private 5g networks for vertical industries: Deployment and operation models," in *Proc. 2019 IEEE 2nd 5GWF*, September 2019.

[76] M. M. Blue, S. Yrjola, V. Deppanen, P. Ahokangas, H. Hammainen, and M. L. Aho, "Analysis of spectrum valuation elements for local 5g networks: Case study of 3.5-ghz band," *IEEE Trans. Cong. Commun. Netw.*, vol. 5, no. 3, pp. 741–753, May 2019.

[77] A. Laourine, A. Stephenne, and S. Affes, "On the capacity of log-normal fading channels," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1603–1607, June 2009.

[78] K. Sato, K. Inage, and T. Fujii, "Modeling the kriging-aided spatial spectrum sharing over log-normal channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 749–752, June 2019.

[79] A. Kliks, P. Kryszkiewicz, A. Umbert, J. P. Romero, F. Casadevall, and L. Kulacz, "Application of radio environment maps for dynamic broadband access in tv bands in urban areas," *IEEE Access*, vol. 5, pp. 19 842–19 863, September 2017.

[80] A. Umbert, F. Casadevall, and E. G. Rodriguez, "An outdoor tv band radio environment map for a manhattan like layout," in *Proc. 2016 ISWCS*, September 2016.

[81] Y. Huang, X. Yan, K. Chen, H. Zhou, Y. Zhang, and Z. Feng, "Tv band radio environment mapping in beijing," in *Proc. IEEE Mil. Commu. Conf.*, October 2014.

[82] Z. Yuxian and L. Qingxia, "The construction and reconstruction of production line based on modularize," in *Proc. 2009 ICIII*, December 2009.

[83] X. C. Liu, S. Zhang, H. H. Lu, and X. K. Lin, "Method for efficiently constructing and updating radio map of fingerprint positioning," in *Proc. 2010 IEEE GLOBECOM Workshops*, December 2010.

[84] J. S. Lim, W. H. Jang, G. W. Yoon, and D. S. Han, "Radio map update automation for wifi positioning systems," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 639–696, March 2013.

[85] S. W. Ellingson, "Continuous updating of radio environment maps for millimeter-wave networks with beamforming," in *Proc. 2017 IEEE Int. Symp. Antennas Propag. USNC/URSI Nat. Radio Sci. Meeting*, July 2017.

[86] H. Wang, D. N. Nguyen, D. T. Hoang, E. Dutkiewics, and Q. Cheng, "Real-time crowdsourcing incentive for radio environment maps: A dynamic pricing approach," in *Proc. 2018 IEEE GLOBECOM*, December 2018.

[87] D. Zurita, E. Sala, J. A. Carino, M. Delgado, and J. A. Ortega, "Industrial process monitoring by means of recurrent neural networks and self organizing maps," in *Proc. 2018 IEEE GLOBECOM*, September 2016.

[88] S. He and S. H. G. Chan, "Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Commun. Surv. & Tut.*, vol. 18, no. 1, pp. 466–490, August 2015.

[89] Y. Mo, Z. Zhang, W. Meng, and G. Agha, "Space division and dimensional reduction methods for indoor positioning system," in *Proc. 2015 IEEE ICC*, September 2015.

[90] D. P. Turner, H. Deng, and T. T. Houle, "Statistical hypothesis testing: Overview and application," *Headache*, vol. 60, no. 2, pp. 302–308, February 2020.

[91] A. Chehade and Z. Shi, "Sensor fusion via statistical hypothesis testing for prognosis and degradation analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1774–1787, February 2019.

[92] M. Korki, H. Zayyani, and J. Zhang, "Bayesian hypothesis testing for block sparse signal recovery," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 494–497, March 2016.

[93] H. Zayyani, M. B. Zadeh, and C. Jutten, "Bayesian pursuit algorithm for sparse representation," in *Proc. 2009 IEEE ICASSP*, April 2009.

[94] T. Halme, M. Golz, and V. Koivunen, "Bayesian multiple hypothesis testing for distributed detection in sensor networks," in *Proc. 2019 IEEE Data Sci. Workshop*, June 2019.

[95] J. Wahlstrom, I. Skog, F. Gustafsson, A. Markham, and N. Trigoni, "Zero-velocity detection-a bayesian approach to adaptive thresholding," *IEEE Sens. Lett.*, vol. 3, no. 6, pp. 1–4, June 2019.

[96] U. Palukha and Y. Kharin, "Performance analysis for statistical testing of random and pseudorandom generators by entropy statistics," in *Proc. 2019 IDT*, June 2019.

[97] H. P. Nguyen, A. Delahaies, F. Retraint, D. H. Nguyen, M. Pic, and F. M. Nicolier, "A watermarking technique to secure printed qr codes using a statistical test," in *Proc. 2017 IEEE GlobalSIP*, November 2017.

[98] M. R. Leonard, M. Stiefel, M. FunB, and A. M. Xoubir, "Robust sequential testing of multiple hypotheses in distributed sensor networks," in *Proc. 2018 IEEE ICASSP*, April 2018.

[99] D. W. Zimmerman, "Inflation of type i error rates by unequal variances associated with parametric, nonparametric, and rank-transformation tests," *Psicologica*, vol. 25, no. 1, pp. 103–133, January 2004.

[100] V. Immler, M. Hiller, J. Obermaier, and G. Sigl, "Take a moment and have some t: Hypothesis testing on raw puf data," in *Proc. 2017 IEEE HOST*, May 2017.

[101] G. D. Ruxton, "The unequal variance t-test is an underused alternative to student's t-test and the mann–whitney u test," *Behavioral Ecology*, vol. 17, no. 4, pp. 688–690, May 2006.

[102] F. E. Satterthwaite, "An approximate distribution of estimates of variance components," *Biometrics Bulletin*, vol. 2, no. 6, pp. 110–114, December 1946.

[103] J. Dai, A. Xu, X. Liu, C. Yu, and Y. Wu, "Online sequential model for multivariate time series prediction with adaptive forgetting factor," *IEEE Access*, vol. 8, pp. 175 958–175 971, September 2020.

[104] W. Cao, Z. Ming, Z. Xu, J. Zhang, and Q. Wang, "Online sequential extreme learning machine with dynamic forgetting factor," *IEEE Access*, vol. 7, pp. 179 746–179 757, December 2020.

[105] C. Yang and H. Fang, "A new nonlinear model-based fault detection method using mann–whitney test," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10 856–10 864, December 2019.

[106] Y. Wang and L. Tian, "The equivalence between mann–whitney wilcoxon test and score test based on the proportional odds model for ordinal responses," in *Proc. 2017 4th IEIS*, July 2017.

[107] M. Zhou, Y. Wang, Z. Tian, and Q. Zhang, "Marvel: Mann–whitney rank-sum testing via segments labeling for indoor pedestrian localization," in *Proc. 2018 IEEE ICC*, May 2018.

[108] D. Rasch, K. D. Kubinger, and K. Moder, "The two-sample t test: pre-testing its assumptions does not pay off," *Statistical Papers, Springer*, vol. 52, no. 1, pp. 219–231, February 2011.

[109] P. E. Mogensen, P. Eggers, C. Jensen, and J. B. Andersen, "Urban area radio propagation measurements at 955 and 1845 mhz for small and micro cells," in *Proc. 1991 IEEE GLOBECOM*, December 1991.

[110] B. V. Laethem, F. Quitin, F. Bellens, C. Oestges, and P. D. Doncker, "Correlation for multi-frequency propagation in urban environments," *Prog. Electromagn. Res. Lett.*, vol. 29, no. 2, pp. 151–156, January 2012.

[111] K. P. Snehal and M. G. Shajan, "Efficient resource allocation in downlink lte networks," in *Proc. 2015 Int. Conf. Commun. Syst. Netw. Technol.*, April 2015.

[112] K. Katagiri and T. Fujii, "Demo: Highly accurate prediction of radio environment for v2v communications," in *Proc. IEEE DySPAN 2019*, November 2019.

[113] H. Zou, M. Jin, H. Jiang, L. Xie, and C. J. Spanos, "Winips: Wifi-based non-intrusive indoor positioning system with online radio map construction and adaptation," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 8118–8130, October 2017.

[114] D. Han, A. Sahar, J. Berkner, and G. Lee, "City radio map construction for wi-fi-based citywide indoor positioning," *IEEE Access*, vol. 7, pp. 99 867–99 877, July 2019.

[115] M. Girmay, V. Maglogiannis, D. Naudts, J. Fontaine, A. Shahid, E. D. Poorter, and I. Moerman, "Adaptive cnn-based private lte solution for fair coexistence with wi-fi in unlicensed spectrum," in *Proc. IEEE Conf. Comput. Commun. Workshops*, August 2020.

[116] T. Matsumura, K. Ibuka, K. Ishizu, H. Murakami, and F. Kojima, "A prototype of multi-ran enb with dynamic baseband resource allocation for heterogeneous private lte networks," in *Proc. 21st WPMC*, May 2019.

[117] S. J. Brown, W. N. Goetzmann, R. G. Ibbotson, and S. A. Ross, "Survivorship bias in performance studies," *Rev. Financ. Stud.*, vol. 5, no. 4, pp. 553–580, October 1992.

[118] A. Achtzehn, J. Riihijärvi, G. M. Vargas, M. Petrova, and P. Mähönen, "Improving coverage prediction for primary multi-transmitter networks operating in the tv whitespaces," in *Proc. 9th Annual IEEE SECON*, June 2012.

[119] F. Bouali, O. Sallent, J. P. Romero, and R. Agusti, "A novel spectrum selection strategy for matching multi-service secondary traffic to heterogeneous primary spectrum opportunities," in *Proc. IEEE 22nd PIMRC*, September 2011.

[120] J. Anastasov, G. T. Djordjevic, and M. C. Stefanovic, "Analytical model for outage probability of interference-limited systems over extended generalized-k fading channels," *IEEE Commun. Lett.*, vol. 16, no. 4, pp. 473–475, February 2012.

[121] T. Soithong, V. A. Aalo, G. P. Efthymoglou, and C. Chayawan, "Outage analysis of multihop relay systems in interference-limited nakagami-*m* fading channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1451–1457, March 2012.

[122] A. Chopra and B. L. Evans, "Outage probability for diversity combining in interference-limited channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 550–560, February 2013.

[123] J. A. Anastasov, G. T. Djordjevic, and M. C. Stefanovic, "Outage probability of interference-limited system over weibull-gamma fading channel," *Electronics Lett.*, vol. 48, no. 7, pp. 408–410, March 2012.

[124] F. Calmon and M. Yacoub, "A general exact formulation for the outage probability in interference-limited systems," in *Proc. 2008 IEEE GLOBECOM*, November 2008.

[125] K. G. Lee and S. J. Oh, "Comparison of the millimeter wave path loss estimation model with outages," in *Proc. IEEE ICCEM*, April 2017.

[126] S. Rana, A. H. John, and H. Midi, "Robust regression imputation for analyzing missing data," in *Proc. ICSSBE*, September 2012.

[127] N. Anindita, H. A. Nugroho, and T. B. Adji, "A combination of multiple imputation and principal component analysis to handle missing value with arbitrary pattern," in *Proc. 7th InAES*, August 2017.

[128] W. Kim, W. Cho, J. Choi, J. Kim, C. Park, and J. Choo, "A comparison of the effects of data imputation methods on model performance," in *Proc. 21st ICACT*, February 2019.

[129] H. L. Shashirekha and A. H. Wani, "Analysis of imputation algorithms for microarray gene expression data," in *Proc. iCATccT*, October 2015.

[130] W. Z. hong, "Numeric missing value's hot deck imputation based on cloud model and association rules," in *Proc. 2nd Int. Workshop Educ. Technol. Comput. Sci.*, March 2010.

[131] M. Sepulcre and J. Gozalvez, "Heterogeneous v2v communications in multi-link and multi-rat vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 20, no. 1, pp. 162–173, September 2019.

[132] X. Yuan, H. Tian, and B. Fan, "Mobility-aware joint resource allocation and power allocation for d2d communication," in *Proc. 2019 IEEE WCNC*, April 2019.

[133] S. Wang, A. Huang, and T. Zhang, "Performance evaluation of ieee 802.15.4 for v2v communication in vanet," in *Proc. 2013 Int. Conf. Comput. Inf. Sci.*, June 2013.

[134] J. Reig and L. Rubio, "Estimation of the composite fast fading and shadowing distribution using the log-moments in wireless communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 3672–3681, May 2013.

[135] T. T. Tjhung and C. C. Chai, "Fade statistics in nakagami-lognormal channels," *IEEE Trans. Commun.*, vol. 47, no. 12, pp. 1769–1772, December 1999.

[136] G. Rafiq and M. Patzold, "The influence of the severity of fading and shadowing on the statistical properties of the capacity of nakagami-lognormal channels," in *Proc. 2008 IEEE GLOBECOM*, December 2008.

[137] P. M. Shankar, "Outage probabilities in shadowed fading channels using a compound statistical model," *IEE Proc. Commun.*, vol. 152, no. 6, pp. 828–832, December 2005.

[138] A. J. Coulson, A. G. Williamson, and R. G. Vaughan, "Improved fading distribution for mobile radio," *IEE Proc. Commun.*, vol. 145, no. 3, pp. 197–202, June 1998.

[139] E. Dall'Anese, S. J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, March 2011.

[140] A. J. Goldsmith and L. J. Greenstein, "A measurement-based model for predicting coverage areas of urban microcells," *IEEE J. Select. Areas Commun.*, vol. 11, no. 7, pp. 1013–1023, September 1993.

[141] X. Xin, W. Da, Z. Xiaoxu, and L. Xuejie, "Traffic information collection system based on crowdsourcing and related privacy issues," in *Proc. WiSPNET 2017*, March 2017.

[142] Z. Chi, Y. Wang, Y. Huang, and X. Tong, "The novel location privacy-preserving ckd for mobile crowdsourcing systems," *IEEE Access*, vol. 6, pp. 5678–5687, December 2017.

[143] R. Hussain and S. Zeadally, "Autonomous cars: Research results, issues, and future challenges," *IEEE Commun. Surv. Tut.*, vol. 21, no. 2, pp. 1275–1313, September 2019.

Publications

List of Publications Directly Related to The Dissertation

Journal Papers

i). Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "Dynamic radio map using statistical hypothesis testing, " *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 3, pp. 752-766, December 2020.

ii). Keita Katagiri, Koya Sato, and Takeo Fujii, "Crowdsourcing-assisted radio environment database for V2V communication, " *Sensors*, vol. 18, no. 4, pp. 1-17, April 2018.

Refereed International Conference Papers

i). Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "Experimental verification of shadowing classification for radio map, " in *Proc. IEEE VTC2020-Fall*, November 2020, pp. 1-7.

ii). Keita Katagiri and Takeo Fujii, "Demo: Highly accurate prediction of radio environment for V2V communications, " in *Proc. IEEE DySPAN 2019*, November 2019, pp. 1-2.

iii). Keita Katagiri, Keita Onose, Koya Sato, Kei Inage, and Takeo Fujii, "Highly accurate prediction of radio propagation using model classifier, " in *Proc. IEEE VTC2019-Spring*, April 2019, pp. 1-5.

List of Referenced Publications

Journal Papers

i). Yoji Uesugi, Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "Clustering for signal power distribution toward low storage crowdsourced spectrum database, " *IEICE Trans. Commun.*, vol. E104-B, no. 10, pp. 1237-1248, October 2021.

ii). Rei Hasegawa, Keita Katagiri, Koya Sato, and Takeo Fujii, "Low storage, but highly accurate measurement-based spectrum database via mesh clustering, " *IEICE Trans. Commun.*, vol. E101-B, no. 10, pp. 2152-2161, October 2018.

Refereed International Conference Papers

i). Keita Katagiri and Takeo Fujii, "Classification and discretization of shadowing toward low storage radio map, " in *Proc. ICAIIC 2022*, online, February 2022, pp. 1-6.

ii). Keita Katagiri and Takeo Fujii, "Partitioned path loss models based on coefficient of determination, " in *Proc. ICOIN 2022*, Online, January 2022, pp. 198-203.

iii). Shougo Matsuo, Sunao Miyamoto, Keita Katagiri and Takeo Fujii,"Radio propagation extrapolation by using multiple separated propagation paths estimated by radio map, " in *Proc. ICOIN 2022*, Online, January 2022, pp. 192-197.

iv). Keita Katagiri and Takeo Fujii, "Low storage radio map using statistical inference, " in *Proc. IEEE DySPAN2021*, Online, December 2021, pp. 123-130.

v). Keita Katagiri, Koya Sato, and Takeo Fujii, "Empirical CDF-based power control method for obstructed V2V communications, " in *Proc. ICETC2021*, Online, December 2021, pp. 1-4.

vi). Keita Katagiri, Koya Sato, and Takeo Fujii, "Estimation of the number of obstacles based on *p*-value for V2V communications, " in *Proc. ICETC2021*, Online, December 2021, pp. 1-4.

vii). Shinichiro Kakuda, Yudai Yamazaki, Keita Katagiri, Takeo Fujii, Koichi Adachi, Osamu Takyu, and Mai Ohta, "Channel allocation for LoRaWAN considering intra-system and inter-system interferences, " in *Proc. IEEE VTC2021-Fall*, online, September 2021, pp. 1-7.

viii). Keita Katagiri and Takeo Fujii, "Mesh-clustering-based radio maps construction for autonomous distributed networks, " in *Proc. ICUFN 2021*, online, August 2021, pp. 345-349.

ix). Sunao Miyamoto, Keita Katagiri, Koya Sato, Koichi Adachi, and Takeo Fujii, "Highly accurate prediction of radio propagation based on compensation of clutter loss for spectrum sharing, " in *Proc. ICETC 2020*, online, December 2020, pp. 1-2.

x). Keita Katagiri and Takeo Fujii, "Radio environment map updating procedure considering change of surrounding environment, " in *Proc. IEEE WCNC Workshop on Smart Spectrum (IWSS 2020)*, online, May 2020, pp. 1-6.

xi). Keita Katagiri and Takeo Fujii, "Measurement-based spectrum database for wireless distributed networks using shadowing classifier, " in *Proc. ICAIIC 2020*, Fukuoka, Japan, February 2020, pp. 64-69.

xii). Keita Katagiri and Takeo Fujii, "Radio environment map updating procedure based on hypothesis testing, " in *Proc. IEEE DySPAN 2019 DD-DSS*, Newark, NJ, USA, November 2019, pp. 1-6.

xiii). Yoji Uesugi, Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "Clustering of signal power distribution toward low storage crowdsourced spectrum database, " in *Proc. VTC-Fall2019 Workshop on 5G and Beyond Technologies for Ultra-Dense Environments*, Honolulu, Hawaii, USA, September 2019, pp. 1-6.

xiv). Hirofumi Nakajo, Yuya Aoki, Keita Katagiri, and Takeo Fujii, "High density spectrum sharing method among micro operators considering spectrum database, " in *Proc. WCNC Workshop on Smart Spectrum (IWSS 2019)*, Marrakech, Morocco, April 2019, pp. 1-5.

xv). Keita Katagiri, Koya Sato, and Takeo Fujii, "Crowdsourcing-assisted radio environment maps for V2V communication systems, " in *Proc. IEEE VTC 2017-Fall*, Toronto, Canada, September 2017, pp. 1-5.

Domestic Conference Papers (in Japanese)

(1) Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "[Invited Lecture] Radio map extrapolation using compensated empirical CDF under interference-limited observations, " in *IEICE Tech. Rep.*, SR2021-92, March 2022.

(2) Atsuki Mukoda, Ayumu Ueda, Keita Katagiri, and Takeo Fujii, “Constructing packet delivery rate-based map for V2X using confidence interval, ” in *IEICE Tech. Rep.*, ITS2021-36, February 2022.

(3) Keita Katagiri and Takeo Fujii, “Calibration of measurement error for radio map, ” in *IEICE Tech. Rep.*, SR2021-64, January 2022.

(4) Keita Katagiri and Takeo Fujii, “Low storage radio map based on classification and quantization of shadowing, ” in *IEICE Tech. Rep.*, November 2021.

(5) Keita Katagiri and Takeo Fujii, “Partitioned path loss models based on coefficient of determination, ” in *IEICE Tech. Rep.*, SR2021-54, November 2021.

(6) Yuto Uchida, Keita Katagiri, and Takeo Fujii, “Spectrum sharing method utilizing radio environment information of primary and secondary users, ” in *Proc. IEICE Society Conf. ’21*, B-17-5, September 2021.

(7) Keita Katagiri and Takeo Fujii, “Mesh-clustering-based radio map construction for autonomous wireless distributed environments, ” in *Proc. IEICE Society Conf. ’21*, B-17-7, September 2021.

(8) Shougo Matsuo, Sunao Miyamoto, Keita Katagiri, and Takeo Fujii, “Radio propagation extrapolation based on propagation path division using surrounding transmitter information, ” in *655th URSI-F Meeting*, June 2021.

(9) Shinichiro Kakuda, Yudai Yamazaki, Keita Katagiri, Takeo Fujii, Osamu Takyu, Mai Ohta, and Koichi Adachi, “Channel allocation for LoRaWAN considering intra-system and inter-system interferences, ” in *IEICE Tech. Rep.*, vol. 121, no. 30, SR2021-12, pp. 79-85, May 2021.

(10) Keita Katagiri and Takeo Fujii, “Statistic sample size determination for average received signal power using statistical inference, ” in *IEICE Tech. Rep.*, vol. 121, no. 30, SR2021-3, pp. 16-23, May 2021.

(11) Shougo Matsuo, Sunao Miyamoto, Keita Katagiri, and Takeo Fujii, “Radio propagation extrapolation based on separated estimation of path loss for spectrum sharing, ” in *Proc. IEICE Gen. Conf. ’21*, B-17-1, March 2021.

(12) Shinichiro Kakuda, Yudai Yamazaki, Keita Katagiri, Takeo Fujii, Osamu Takyu, Mai Ohta, and Koichi Adachi, “Channel selection for LPWA using interference power

distribution in multiple channel environment, ” in *Proc. IEICE Gen. Conf. ’21*, B-17-26, March 2021.

(13) Sunao Miyamoto, Keita Katagiri, Koya Sato, Koichi Adachi, and Takeo Fujii, “Compensation of clutter loss considering antenna height difference and dominant path for spectrum sharing, ” in *IEICE Tech. Rep.*, vol. 120, no. 238, SR2020-39, pp. 108-113, November 2020.

(14) Keita Katagiri, Koya Sato, and Takeo Fujii, “Estimation of the number of obstacles based on p-value for V2V communications, ” in *Proc. IEICE Society Conf. ’20*, B-17-6, September 2020.

(15) Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, “Experimental verification of shadowing classification for measurement-based spectrum database, ” in *IEICE Tech. Rep.*, vol. 120, no. 90, SR2020-20, pp. 63-70, July 2020.

(16) Keita Katagiri and Takeo Fujii, “Radio propagation extrapolation based on correction of empirical CDF, ” in *IEICE Tech. Rep.*, vol. 120, no. 53, SR2020-2, pp. 9-16, June 2020.

(17) Keita Katagiri and Takeo Fujii, “Measurement-based spectrum database for wireless distributed networks using shadowing classifier, ” in *Proc. IEICE Gen. Conf. ’20*, B-17-35, March 2020.

(18) Keita Katagiri and Takeo Fujii, “Radio environment map updating procedure considering changes in surrounding environments, ” in *IEICE Tech. Rep.*, vol. 119, no. 449, SR2019-132, pp. 105-112, March 2020.

(19) Keita Katagiri and Takeo Fujii, “Radio environment map updating procedure based on hypothesis testing, ” in *IEICE Tech. Rep.*, vol. 119, no. 325, SR2019-104, pp. 87-94, December 2019.

(20) Yoji Uesugi, Keita Katagiri, and Takeo Fujii, “[Poster Presentation] Signal power map prediction method considering structure information, ” in *IEICE Tech. Rep.*, vol. 119, no. 325, SR2019-90, pp. 25-26, December 2019.

(21) Sunao Miyamoto, Keita Katagiri, and Takeo Fujii, “[Poster Presentation] Radio propagation estimation considering altitude for spectrum sharing between LTE/5G and different systems, ” in *IEICE Tech. Rep.*, vol. 119, no. 325, SR2019-89, pp. 23-24, December 2019.

(22) Yoji Uesugi, Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "[Poster Presentation] Clustering of signal power distribution toward low storage crowdsourced spectrum database, " in *IEICE Tech. Rep.*, November 2019.

(23) Keita Katagiri and Takeo Fujii, "[Technology Exhibit] Highly accurate prediction of radio environment for V2V communication systems, " in *IEICE Tech. Rep.*, vol. 119, no. 62, SR2019-18, pp. 107-112, May 2019.

(24) Keita Katagiri, Keita Onose, Koya Sato, Kei Inage, and Takeo Fujii, "Highly accurate estimation of radio propagation using model classifier, " in *IEICE Tech. Rep.*, vol. 118, no. 421, SR2018-108, pp. 79-84, January 2019.

(25) Keita Onose, Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "A study on frequency-domain interpolation of a radio environment map for bands of cellular system, " in *IEICE Tech. Rep.*, vol. 118, no. 421, SR2018-116, pp. 123-128, January 2019.

(26) Hirofumi Nakajo, Yuya Aoki, Keita Katagiri, and Takeo Fujii, "High density spectrum sharing method among micro operators considering mesurement-based spectrum database, " in *IEICE Tech. Rep.*, vol. 118, no. 421, SR2018-115, pp. 117-122, January 2019.

(27) Yoji Uesugi, Keita Katagiri, Koya Sato, Kei Inage, and Takeo Fujii, "Spectrum database of signal power distribution using clustering, " in *IEICE Tech. Rep.*, vol. 118, no. 421, SR2018-118, pp. 135-140, January 2019.

(28) Hirofumi Nakajo, Yuya Aoki, Keita Katagiri, and Takeo Fujii, "[Poster Presentation] Spatial spectrum sharing among micro operators considering distribution on measurement-based spectrum database, " in *IEICE Tech. Rep.*, vol. 118, no. 274, SR2018-78, pp. 45-46, October 2018.

(29) Keita Katagiri, Koya Sato, and Takeo Fujii, "[Poster Presentation] Power control exploiting spectrum database for V2V communication systems, " in *IEICE Tech. Rep.*, vol. 118, no. 274, SR2018-76, pp. 39-40, October 2018.

(30) Keita Katagiri, Koya Sato, and Takeo Fujii, "Highly accurate prediction of signal variation distribution using surrounding vehicles information, " in *IEICE Tech. Rep.*, vol. 117, no. 410, SR2017-104, pp. 63-68, January 2018.

(31) Keita Katagiri, Koya Sato, and Takeo Fujii, "[Technology Exhibit] Construction of radio environment maps for improving communication reliability of cooperative

automated vehicles, ” in *IEICE Tech. Rep.*, vol. 117, no. 56, SR2017-13, pp. 75-80, May 2017.

(32) Keita Katagiri and Takeo Fujii, “Reception success rate calculation algorithm based on received signals in V2V communications, ” in *Proc. IEICE Gen. Conf. ’17*, B-17-12, March 2017.