applied sciences

Article

Automatic Clustering for Improved Radio Environment Maps in Distributed Applications

Haithem Ben Chikha [1,*,†] and Alaa Alaerjan [2,*,†]

[1] Department of Computer Engineering and Networks, College of Computer and Information Sciences, Jouf University, Sakaka 72341, Saudi Arabia
[2] Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka 72341, Saudi Arabia
* Correspondence: hbchikha@ju.edu.sa (H.B.C.); asalaerjan@ju.edu.sa (A.A.)
† These authors contributed equally to this work.

Abstract: Wireless communication greatly contributes to the evolution of new technologies, such as the Internet of Things (IoT) and edge computing. The new generation networks, including 5G and 6G, provide several connectivity advantages for multiple applications, such as smart health systems and smart cities. Adopting wireless communication technologies in these applications is still challenging due to factors such as mobility and heterogeneity. Predicting accurate radio environment maps (REMs) is essential to facilitate connectivity and improve resource utilization. The construction of accurate REMs through the prediction of reference signal received power (RSRP) can be useful in densely distributed applications, such as smart cities. However, predicting an accurate RSRP in the applications can be complex due to intervention and mobility aspects. Given the fact that the propagation environments can be different in a specific area of interest, the estimation of a common path loss exponent for the entire area produces errors in the constructed REM. Hence, it is necessary to use automatic clustering to distinguish between different environments by grouping locations that exhibit similar propagation characteristics. This leads to better prediction of the propagation characteristics of other locations within the same cluster. Therefore, in this work, we propose using the Kriging technique, in conjunction with the automatic clustering approach, in order to improve the accuracy of RSRP prediction. In fact, we adopt K-means clustering (KMC) to enhance the path loss exponent estimation. We use a dataset to test the proposed model using a set of comparative studies. The results showed that the proposed approach provides significant RSRP prediction capabilities for constructing REM, with a gain of about 3.3 dB in terms of root mean square error compared to the case without clustering.

Keywords: automatic clustering; edge computing; K-means clustering; Kriging technique; radio environment map; reference signal received power

check for updates

Citation: Ben Chikha, H.; Alaerjan, A. Automatic Clustering for Improved Radio Environment Maps in Distributed Applications. *Appl. Sci.* **2023**, *13*, 5902. https://doi.org/10.3390/app13105902

Academic Editor: Alessandro Lo Schiavo

Received: 7 April 2023
Revised: 4 May 2023
Accepted: 9 May 2023
Published: 10 May 2023

1. Introduction

In recent years, radio environment map (REM) technology has gained great attention in the field of wireless communication [1–7]. The generation and use of REM improve several aspects, such as wireless quality and efficiency. REM is considered to be an important tool, since it contributes to improving the spectral resources in wireless networks [8,9]. Hence, much effort has been expended in expanding the use of REM due to its potential in telecommunication for distributed applications.

REM provides key information about the behavior of wireless channels. It also has the capability of making decisions related to different wireless communication applications. For instance, REM improves the optimization of coverage tasks. It improves the predictability of wireless coverage and its associated analysis. Additionally, it contributes to the minimization of drive tests (MDT), which is critical in wireless networks [10–13].

Furthermore, it improves decision making and analysis in cognitive radio systems, especially in aspects related to resource allocation [14]. It also improves interference analysis and predictability of coverage opportunities in heterogeneous networks [5,15,16]. With regard to REM types, the proposed models in the literature are categorized into two types. The first type uses available prior knowledge of the environment, while the second type is based on measurement-based predictions [17]. As is evident, REM greatly contributes to the improvement of wireless communication.

Edge computing is one of the most promising fields in next-generation technologies. It is considered to be a new paradigm that aims at addressing several challenges currently encountered by the traditional cloud computing paradigm (e.g., latency, scalability) [18]. It addresses scalability and latency limitations [19,20]. It also addresses computation challenges, since edge computing aims at minimizing the process delay by bringing computation closer to data sources. Edge computing greatly depends on connectivity, especially wireless communication. Technologies such as 5G and 6G are considered essential to the continued progress of edge computing. This is due to the ability of wireless technologies to provide ubiquitous data access. Additionally, wireless technologies provide great support for heterogeneous devices and equipment, which is a key factor in enabling broader data exchange capabilities. Therefore, maximizing wireless capabilities and addressing coverage issues are much considered in research to improve the performance of distributed applications and edge computing [21].

Geolocalized radio measurements can be retrieved with the help of the MDT feature [10]. This has been possible since user equipment (UE) began including global positioning system information. Geolocalized measurements can be used with spatial interpolation techniques to construct an REM, which is measured by exploiting the collected readings gathered in a radio environment. The collected measurements are utilized to build an REM to characterize wireless channels. This is achieved through multiple propagation mechanisms [22–28]. In fact, these mechanisms affect the reference signal received power (RSRP), and may also affect the construction process for REMs.

Multiple research works have used the well-known Kriging technique [29] for geostatistical interpolation [16,30,31]. The Kriging technique depends on exploiting shadowing and median path loss. It is assumed that the fast-fading communication signal is averaged out at the receiver [32]. The REM can be employed to enhance the quality of service (QoS) for UEs in 5G, and beyond, networks. It builds the whole coverage map along the UE paths by interpolating measurements made in different places. A Bayesian Kriging spatial interpolation technique [33,34] was initially proposed for 3G. It targeted the prediction of the received signal code coverage [34]. It also initially worked for long-term evolution (LTE) RSRP coverage analysis [5].

Several research works have studied the main shortcomings in mobile and wireless networks. For example, the work in [35] showed that the relationship between logarithmic distance and RSRP values had an error of 8–9 dB in urban environments. It also showed that the error increased to around 15 dB in rural areas. This work assumed a single path loss trend for the whole area of interest, as is the case with the majority of the models available in the literature [36]. The assumptions may not provide accurate results, since the behavior of wireless communication is unpredictable.

The precise prediction of REM can be a challenging task, especially in large-scale and unstable environments [8]. This is due to several external factors, such as mobility and spectrum interference. Currently, many traditional learning-based models depend on a lengthy training process to predict an accurate REM. However, this can be challenging, since the learning algorithms are often complex, and require a significant amount of training sets [34,37]. Given these issues, this work proposes a dynamic approach that aims at improving REM construction in distributed applications. In the approach proposed, automatic clustering-based KMC is integrated with the Kriging technique. This integration process is proposed for several reasons. First, KMC is a flexible and efficient clustering algorithm, which enhances the efficiency of the clustering process. This is critical, due to the challenges that are being faced in wireless environments (e.g., mobility, instability) [38]. Second, KMC supports scalability as it reduces the calculation of the pairwise distance among different points in a given dataset. Additionally, KMC does not induce complexity, which is a key concern in most distributed applications, due to the large number of communicating entities. On the other hand, the Kriging technique provides several estimation advantages, including measuring errors and uncertainty in sample results. It also considers direct and dependent relationships. Furthermore, the Kriging technique is not restricted in terms of the range of observations. In fact, combining KMC with the Kriging technique not only improves prediction accuracy, but also promotes the efficiency of RSRP prediction.

The overall objective is to improve both the efficiency and accuracy of the construction process of REM. More specifically, we summarize the contributions of this work as follows:

- An automatic clustering approach using KMC in conjunction with the Kriging technique is proposed. In our proposal, two steps (i.e., clustering using KMC technique and the RSRP prediction using the Kriging technique) are needed to produce more accurate REMs. To the best of our knowledge, this is the first work that uses KMC to improve the accuracy of RSRP prediction. This justifies the novelty of the paper, since combining these two technologies proved to be significant, based on the results provided in this work. The proposed work in this paper also enhances the path loss exponent estimation through the distinction between different environments existing in the area of interest. This is considered to be a key contribution due to the importance of path loss exponent estimation in the field of wireless communication and REM construction.
- A comparative study with the proposed approach is performed, in which the Kriging technique, considered with only one cluster, is utilized as a benchmark. Here, the performance of the proposed REM construction technique using KMC is investigated and evaluated with the root mean square error (RMSE) metric. Multiple simulation tests were carried out to prove the superiority of our proposed work.
- The proposed approach is evaluated, in regard to its potential as a technique for constructing REMs, using KMC with R^2. This further proved the viability of the approach, because the method measures the variance of dependent variables in relation to independent variables in a dataset. In fact, using both RMSE and R^2 indicated the effectiveness of the proposed REM construction technique.

Given the above discussion, Table 1 provides the list of abbreviations to simplify the readability of the paper.

Table 1. List of abbreviations.

BS	base station
GoB	grid of beams
KMC	K-means clustering
LOS	line-of-sight
LTE	long-term evolution
MDT	minimization of drive tests
M2M	machine-to-machine
QoS	quality of service
REM	radio environment map
RMSE	root mean square error
RSRP	reference signal received power
SN	sensor node
UE	user equipment

To provide clarification in understanding the expressions in this paper, Table 2 defines the notations throughout the paper; lowercase letters in bold represent vectors and uppercase letter in bold indicate the matrices.

Table 2. List of notations.

$(\cdot)^T$	transpose operation
$(\cdot)^{-1}$	inverse operation
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathrm{card}(\mathcal{S})$	cardinality of the set $\mathcal{S}$

The rest of this paper is organized as follows: Section 2 describes the work on improving wireless coverage and wireless prediction models. It also describes the importance of wireless communication in multiple applications and domains, such as IoT and edge computing. Section 3 describes the related work and where this work fits in the literature. Section 4 describes the main approach for optimizing REM and RSRP predictions. Section 5 describes the utilized algorithm and how it is used in the clustering process to improve prediction models. Section 6 provides the research outcomes according to the proposed approach. Finally, Section 7 concludes the work with a brief discussion on future work.

2. Wireless Requirements and Applications

This section describes the requirements for improving new generation wireless technologies, such as those in 5G and 6G networks. It also describes the great benefits of multiple applications, such as IoT and edge computing, which have become possible due to new and improved wireless technologies.

2.1. Improving Wireless Coverage

One of the main challenges in the next generation of wireless technologies is the accuracy of the prediction models used in REM construction. It is essential to predict an accurate REM in order to determine several aspects that contribute to the quality of wireless communication. For example, predicting an accurate map contributes to determining the median of the relationship between RSRP and logarithmic distance in a wireless environment. Currently, most of the work done in improving RSRP has focused on single path loss in the considered range of communication. The single path approach is considered simple since it does not depend on complex mathematical operations. However, this approach does not provide accurate results, since it does not consider all communication factors in a wireless environment. For instance, the single path model does not capture the essential components in computing the path, such as the exponent of path loss [35].

Due to the challenges with a single path approach, multiple research works have considered other approaches to improve RSRP prediction. For example, the multiple-trend path loss model was explored to improve RSRP prediction capabilities. Some works targeted double-trend models where line-of-sight (LOS) possibilities were considered before and after the clearance distance. Other approaches looked at the median of RSRP values in the prediction process. This is because the median might be different, even with the same distance, mainly due to differences in environmental aspects, such as varying shadowing [39]. Clustering approaches have also been considered as a means to improve RSRP prediction. All these works, either on single or multiple approaches, emphasize the importance of finding an accurate prediction model and lowering the complexity of the prediction process at the same time.

2.2. Wireless Communication in New Applications

Several emerging technologies, such as IoT and edge computing, greatly depend on wireless communication [40,41]. Wireless communication systems are built with great consideration for aspects such as heterogeneity and mobility. For instance, consider the great benefits of machine-to-machine (M2M) communication in IoT applications where multiple devices (e.g., sensors and data aggregation units) need to directly communicate with each other. This advance has been possible due to the improved capabilities of new wireless technologies, such as 5G networks. Wireless connectivity improves other aspects of communicating applications. For example, due to wireless connectivity, M2M communication contributes not only to data connectivity but also to other quality factors, such as data availability, redundancy, and reliability. In IoT applications, wireless connectivity allows for bidirectional communications between field devices and equipment to control and data centers [42], which improves domain knowledge and promotes pervasive data aggregation.

Edge computing is a new computation paradigm, by means of which computation is brought closer to data sources to reduce latency and improve response time [19]. Wireless communication significantly contributes to the evolution of edge computing. This is because new wireless technologies provide multiple benefits, such as ultra-low latency and large throughput capabilities. Additionally, wireless communication supports mobility, which is a key factor in edge computing. Wireless communication also improves the management of applications that adopt the edge computing paradigm. For example, wireless network slicing allows an application to create different management layers based on exchanged data and communicating devices [43]. Figure 1 shows how wireless communication contributes to moving data among different domains in a smart city. The figure shows that different domains are managed through network slicing, where data is managed and controlled differently at each network tier. The slicing also enables edge computing to create sub-networks to manage the heterogeneity of devices and services.

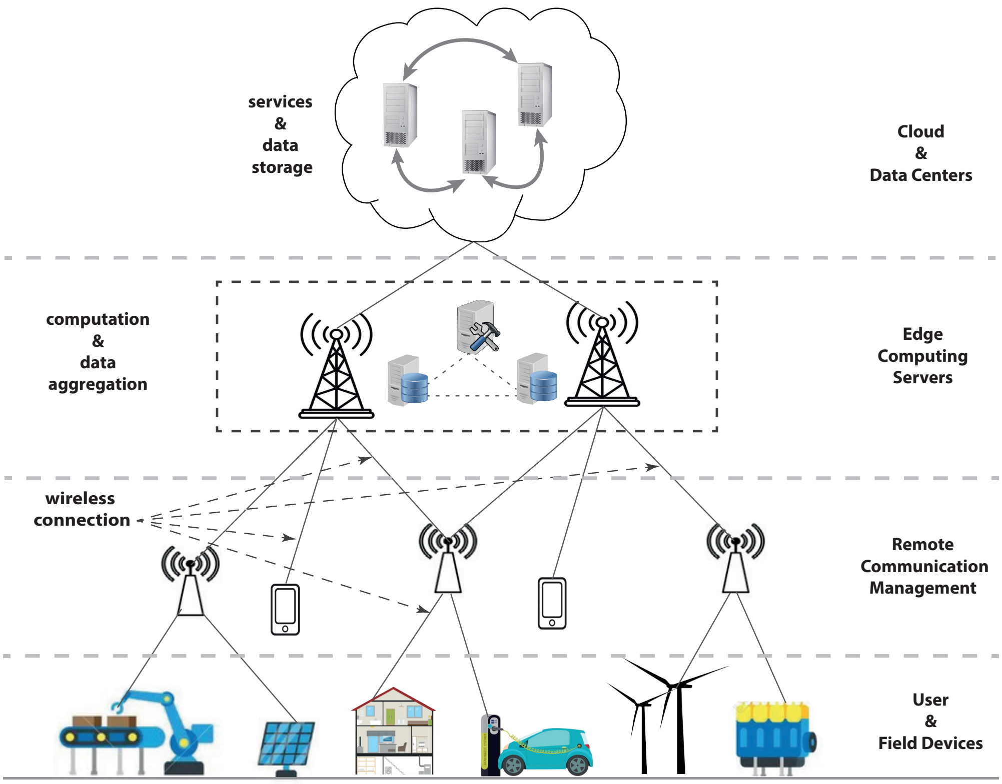

Figure 1. Multi-layer communication in edge computing-enabled applications.

3. Related Work

REMs play an important role in enhancing the QoS for UEs in 5G and IoT environments [44]. This was first introduced by [45], and aimed to spatially interpolate geolocation measurements to construct a whole coverage map. Interpolation techniques are mainly divided into deterministic and statistical approaches [46]. For deterministic interpolation, the geolocated measurements are combined to predict the corresponding value at a target location. In fact, larger weights are attributed to the measurements that are closest to the location of interest. However, this approach does not differentiate between the contributions of different measurements taken at the same distance from the target location.

Several research works have proposed improvements to the construction of REMs as they are key features for network planning and optimization of radio resources [32]. However, most of the proposed prediction techniques aim to construct only one model to construct an REM for the whole area of interest. Various works use the Kriging technique [47] as a prediction technique. This approach aims at observing the weights of the attributes in relation to neighboring observations. The overall objective of the observation is to enhance the prediction of the RSRP at an unobserved location. A Bayesian Kriging spatial interpolation technique [33] was first utilized in older generation wireless networks. It was used in 3G for predicting the received signal code power coverage [34]. It was also used in LTE for RSRP analysis of the wireless coverage [5]. The performance of the fixed rank Kriging technique has been studied for cellular networks. In fact, this later technology was also applied in these networks [3,48].

The use of Kriging and its derivatives showed good performance in terms of prediction accuracy. With the arrival of 5G, the use of narrow beams, steered simultaneously to different UEs, improves spectral efficiency through spatial multiplexing. To this end, a grid of beams (GoB) focused toward UEs is generated with the deployed antenna elements in both vertical and horizontal directions [49]. To construct an accurate REM in the context of 5G, and in different contexts and environments in the area of interest, being able to distinguish between heterogeneous environments is required. In fact, estimating a common path loss exponent for an entire area produces errors in the constructed REM. Hence, it is necessary to use automatic clustering to distinguish between different environments by grouping locations that exhibit similar propagation characteristics. This leads to better prediction of the propagation characteristics of other locations within the same cluster. In this work, we aimed to show the gains offered to RSRP prediction by the use of automatic clustering.

In [37], the authors propose a clustering mechanism to classify different wireless signal coverage based on a fixed threshold. This latter requires human intervention for each propagation environment. In practice, this involves a considerable amount of time, which results in additional deployment costs. To solve the problem, an automatic clustering technique can be deployed at the edge computing unit. A variety of clustering techniques are proposed in the literature. The KMC is a very popular method as it is simple to implement and has been applied in various applications [50].

Given the above reviews, it can be seen that the process of building REM can be challenging. It can also be seen that even if the above works address REM building issues, they involve a trade off between complexity and scalability. On the other hand, the proposed work in this paper provides a low-complexity approach to construct more accurate REM, based on an automatic clustering technique. The approach also considers scalability, as it reduces the REM computation process. In fact, one of the key differences is that our proposed approach has the capability to construct a more accurate REM for an area containing two heterogeneous environments.

4. Proposed Coverage Prediction Method using the Kriging Technique in Conjunction with Automatic Clustering

In this work, we apply the Kriging technique to predict the RSRP for heterogeneous environments differentiated by automatic clustering. In this section, we first present the Kriging technique. Then, we demonstrate how to perform automatic clustering to improve prediction accuracy.

Kriging Technique

To estimate the RSRP of UE located at $g \in \mathbb{R}^2$ and served by a base station (BS), we use the Kriging geostatistical approach. The purpose here is to build a map allowing the estimation of the RSRP given by

$$z(g) = \underbrace{p_{BS} + 10\mu \log_{10}(d(g)) + s(g)}_{q(g)} + \varepsilon(g), \tag{1}$$

where P_{BS} is the BS transmission power, $d(g)$ is the geometrical distance between BS and UE, and μ is the path loss exponent. The value $s(g)$ represents the effect of shadowing in a wireless environment. It is modeled through a log-normal distribution. It is also associated with the stationary covariance matrix $\mathbf{M}_{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha}$ is the set of parameters. Finally, $\varepsilon(g)$ represents the error of measurements and is modeled by means of a centered uncorrelated Gaussian process with variance σ^2. In this study, we realize a lognormal random process. This is performed using a standard deviation of 6 dB to model the shadowing. With this setting, we consider a Non-LOS Urban Macro-cell environment [51], where the shadowing decorrelation distance is equal to 20 m [52]. Furthermore, we assume that $s(g)$ and $\varepsilon(g)$ are independent processes. We also assume that the RSRPs are collected by K sensor nodes (SNs) located at $g_k, k = 1, 2, \ldots, K$. Here, the wireless sensor network is modeled using a connectivity graph $G(\mathbf{S}, \mathbf{L})$, in which the SNs are denoted as $\mathbf{S} = \{1, 2, \ldots, K\}$ and the links are defined by $\mathbf{L} \in \{L_n, m\}_{K \times K}(n, m \in \mathbf{V})$. The spatial location of SN$_k$ is denoted by $g_k = (x_k, y_k)$ and SN$_k$ measures the RSRP value at g_k. Thereafter, the collected RSRP values are transferred to the edge cloud. Next, the edge cloud builds an REM with the collected RSRPs.

The reported RSRP measurements by the K SNs can be defined as

$$\mathbf{r} = \mathbf{F}\mathbf{v} + \mathbf{s} + \mathbf{e}, \tag{2}$$

where $\mathbf{r} = (r(g_1), \ldots, r(g_K))^T$ represents a $K \times 1$ vector, $\mathbf{F} \in \mathbb{R}^{K \times 2}$ represents a deterministic matrix, in which the first column contains "1s", while the second is filled with $10\log_{10}(d(g_k))$. $\mathbf{v} = (p_{BS}, \mu)^T$ represents a 2×1 vector, and $\mathbf{s} = (s(g_1), \ldots, s(g_K))^T$ and $\mathbf{e} = (\varepsilon(g_1), \ldots, \varepsilon(g_K))^T$ are $K \times 1$ vectors. The RSRP vector, $\mathbf{r} \in \mathbb{R}^{K \times 1}$, is considered to be a multivariate Gaussian random variable, characterized by the mean $\mathbf{Fv}$ and covariance matrix $\mathbf{M}_{(\boldsymbol{\alpha},\sigma)} = \mathbf{M}_{\boldsymbol{\alpha}} + \sigma^2\mathbf{I}_K$. In this work, we use the Kriging technique to interpolate the RSRP at a location in the area of interest $g_0 \notin \{g_1, \ldots, g_K\}$. As $(q(g_0), \mathbf{r})^T$ is a Gaussian vector, the predicted RSRP value using the Kriging method at g_0 is the conditional expectation of $q(g_0)$, given the measurement vector $\mathbf{r}$. It can be expressed as [53]

$$\hat{q}(g_0) = \mathbf{f}(g_0)\hat{\mathbf{v}} + \mathbf{m}_{\hat{\boldsymbol{\alpha}}}\mathbf{M}^{-1}_{(\hat{\boldsymbol{\alpha}},\hat{\sigma})}(\mathbf{r} - \mathbf{F}\hat{\mathbf{v}}), \tag{3}$$

where $\mathbf{f}(g_0)$ is the drift function vector at the location g_0, $(\hat{\mathbf{v}}, \hat{\boldsymbol{\alpha}}, \hat{\sigma})$ represents the estimated value of $(\mathbf{v}, \boldsymbol{\alpha}, \sigma)$ and $\mathbf{m}_{\hat{\boldsymbol{\alpha}}}$ represents the vector of covariance between the target and the measurement values.

In this work, we use the method of moments of Matheron [24,54–57] to estimate the Kriging parameters. It can be described by

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N_h} \sum_{g_i - g_j = \mathbf{h}} \left[r(\mathbf{g}_j) - r(\mathbf{g}_i)\right]^2 \forall \mathbf{g}_i, i = 1, 2, \ldots, K, \tag{4}$$

where N_h is the number of pairs of points distant of $\mathbf{h}$.

In Figure 2, we present the three principal parameters of a variogram. The first parameter is the nugget. It represents the discontinuity at the origin caused by the measurement error. The second parameter is the sill. It represents the maximum semivariance value. It is equal to the process variance. Finally, the last parameter is the range, which represents the distance of decorrelation where the two samples become decorrelated and, consequently, the variogram reaches the sill. It is also shown that, at the same location, the semivariance value is equal to zero, i.e., $\gamma(0) = 0$.

In this work, we propose using automatic clustering to improve the RSRP prediction accuracy by enhancing the path loss exponent estimation used in (3). In the following, we describe our selected automatic clustering method.

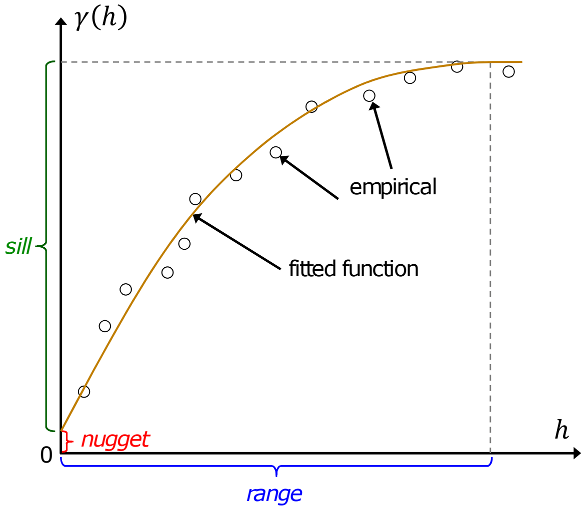

Figure 2. Variogram parameters.

Figure 3 illustrates the flow of the framework in this paper. Graph (a) describes the general workflow without clustering, while Graph (b) shows the flow of our approach using KMC clustering, described in Section 5. As seen here, in our approach we aggregate the predicted RSRP values for the two clusters built. This latter offers a gain in the RSRP prediction as the model parameters are better estimated for each cluster. In fact, this clustering approach greatly improves the efficiency of the training process, which contributes to accurate prediction results.

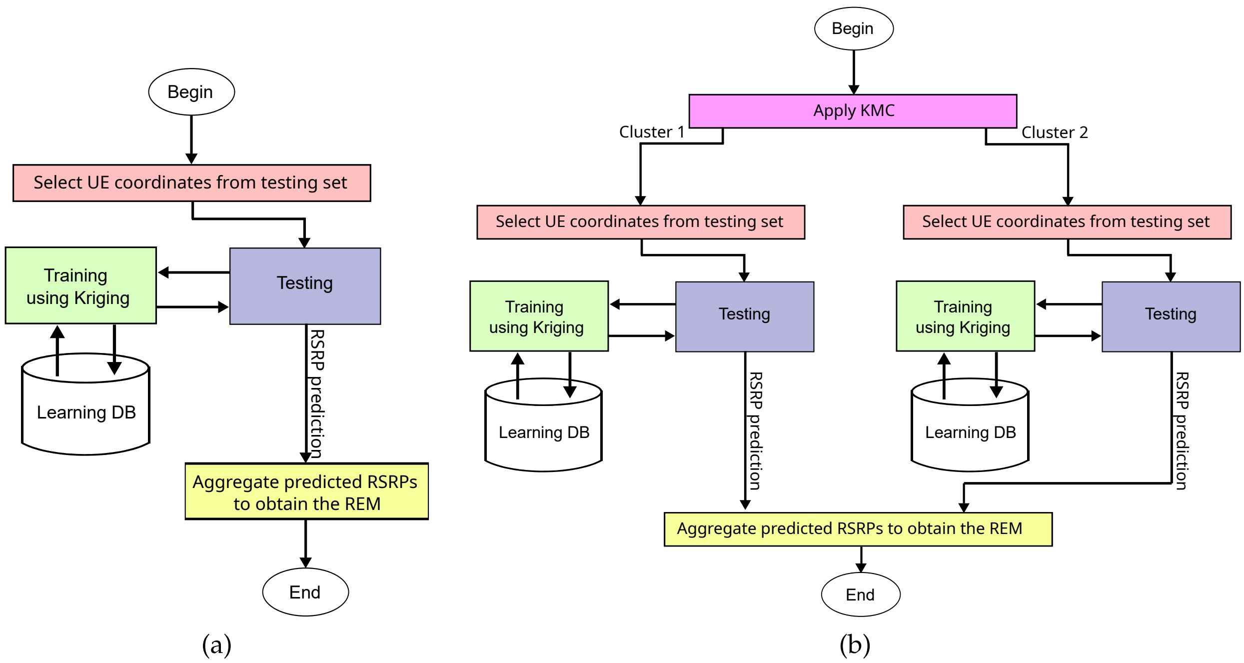

Figure 3. RSRP prediction comparison between the two approaches (**a**) without clustering, (**b**) with clustering.

5. K-Means Clustering Algorithm

In this work, we adopt the KMC algorithm to address our clustering prediction problem. Distributed applications, such as smart cities and smart transportation systems, require efficient and scalable technical solutions, which are low in complexity, to address communication challenges [58,59]. KMC is a clustering method that supports these requirements. As compared to other clustering approaches, such as Monte Carlo and Gillespie algorithms, KMC is an ideal candidate, specifically for unstable wireless environments. For instance, KMC is computationally efficient, as compared to the Monte Carlo and Gillespie algorithms [60]. This is a key concern in wireless environments, due to the large volume of data generated from various communicating entities. Additionally, the Monte Carlo algorithm greatly depends on the quality of the parameters and constraints. In contrast, KMC accounts for errors and uncertainty in clustering results. Based on the advantages of KMC, the model was adopted to perform clustering prediction in pour work.

The following describes the approach with KMC. Let $\mathscr{D}$ be the dataset that contains N quadri-dimensional normalized observations, i.e., $\mathscr{D} = \{\mathbf{o}_n \mid n = 1, \ldots, N\}$, where $\mathbf{o}_n$ is in quadri-dimensional space ($\mathbf{o}_n = \left(\underbrace{x_n^*}_{o_{n1}}, \underbrace{y_n^*}_{o_{n2}}, \underbrace{r_n^*}_{o_{n3}}, \underbrace{d_n^*}_{o_{n4}} \right) \in \mathbb{R}^4$). x_n^*, y_n^*, r_n^* and d_n^* are the normalized values of the x-coordinate, y-coordinate, RSRP and distance to BS, respectively. Here, we apply min–max normalization, and, hence, the normalized values of the x-coordinate, y-coordinate, RSRP, and distance to BS are respectively given by

$$x_n^* = \frac{x_n - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})}, \tag{5}$$

$$y_n^* = \frac{y_n - \min(\mathbf{y})}{\max(\mathbf{y}) - \min(\mathbf{y})}, \tag{6}$$

$$r_n^* = \frac{r_n - \min(\mathbf{r})}{\max(\mathbf{r}) - \min(\mathbf{r})}, \tag{7}$$

$$d_n^* = \frac{d_n - \min(\mathbf{d})}{\max(\mathbf{d}) - \min(\mathbf{d})}, \tag{8}$$

where $\mathbf{x} = [x_1, \ldots, x_n]$, $\mathbf{y} = [y_1, \ldots, y_n]$, $\mathbf{r} = [r_1, \ldots, r_n]$ and $\mathbf{d} = [d_1, \ldots, d_n]$ are the real x- and y-coordinates, RSRPs, and distances to BS, respectively (before the normalization). For a given $\mathscr{D}$ and a predefined number of clusters U, the clustering operation aims to divide the N observations into U clusters $C_u, u = 1, \ldots, U$, where the cluster centroid is considered the mean of the collected observations in the cluster. Note that the complexity of KMC is significantly lower in comparison with the complexity of an exhaustive search. The KMC process is summarized in Algorithm 1. The KMC procedure starts by randomly selecting a number of U observations from the dataset $\mathscr{D}$, and considering them to be the U initial set of centroids, denoted by $\mathbf{c}_u, u = 1, \ldots, U$. In step 2, the Euclidean distance is considered to determine the distance between each observation and the cluster centroids. It is expressed for two observations, $\mathbf{o}$ and $\mathbf{o}'$ as

$$d_{\mathbf{oo}'} = \sqrt{\sum_{i=1}^{4} \left(o_i - o_i'\right)^2}. \tag{9}$$

The next step consists of assigning each observation $\mathbf{o}_n$ to the cluster having a centroid closest to it. Then, it determines the new centroids of all the U clusters. In KMC, an iterative process is used to change the positions of the U cluster centroids until the positions remain unchanged.

Algorithm 1: KMC algorithm.

inputs : **Number of clusters:** U
Dataset: $\mathscr{D} = \{\mathbf{o}_n \mid n = 1, \ldots, N\}$
outputs : **Clusters:** $C_u, u = 1, \ldots, U$
1 **init:** Randomly select U observations from $\mathscr{D}$, and consider them to be the U initial set of centroids, denoted by $\mathbf{c}_u, u = 1, \ldots, U$.
2 Calculate the distance between each observation $\mathbf{o}_n$ and the cluster centroids using (9).
3 Assign each observation $\mathbf{o}_n$ to the cluster whose centroid is the closest to it.
4 Determine the new centroids $\mathbf{c}_u, k = 1, \ldots, U$ of all the U clusters.
5 Repeat steps 2–4 until the positions of the centroids remain unchanged.

Given the above description, we explain the complexity of the proposed framework in terms of the number of iterations and the number/size of clusters. The time complexity of the framework is given in Table 3. In the table, I represents the number of iterations. The number of clusters is represented by U. The size of the whole dataset is described by N. The training set for each observation is represented by K. The Kriging technique has a complexity of $\mathcal{O}(K^3)$ [6]. On the other hand, KMC has a complexity of $\mathcal{O}(UNI)$ [61]. Based on these definitions, the complexity is defined in terms of the numbers of clusters that are distinguished in this work. Since the Kriging technique is applied for each cluster, and since we have two clusters, then the framework's complexity can be defined as: $\mathcal{O}(UNI + 2K^3)$.

Table 3. Time complexity of our framework.

KMC	$\mathcal{O}(UNI)$
Kriging	$\mathcal{O}(K^3)$
Framework	$\mathcal{O}(UNI + 2K^3)$

6. Numerical Results

In this study, our dataset is generated through simulation and consists of simulated RSRPs, having the network parameters summarized in Table 4. The collected RSRPs are used as the ground-truth for the coverage in the area of interest since they can be considered realistic field measurements. In fact, they represent the received pilot power, as defined by the LTE standard, in the two different environments (i.e., urban and rural), as shown in Figure 4. These measurements are calculated over a uniform grid of size 1 km $\times$ 1 km.

Table 4. BS parameters and channel characteristics.

BS Parameters	
Number	1
Transmit power	46 dBm
Bandwidth	20 MHz
Channel Characteristics	
Thermal noise per Hertz	-174 dBm/Hz
Path Loss for urban area (d in km)	$128.1 + 37.6 \log_{10}(d)$dB
Path Loss for rural area (d in km)	$100.54 + 34.1 \log_{10}(d)$dB
Shadowing Log-normal	6 dB
Decorrelation distance	20 m

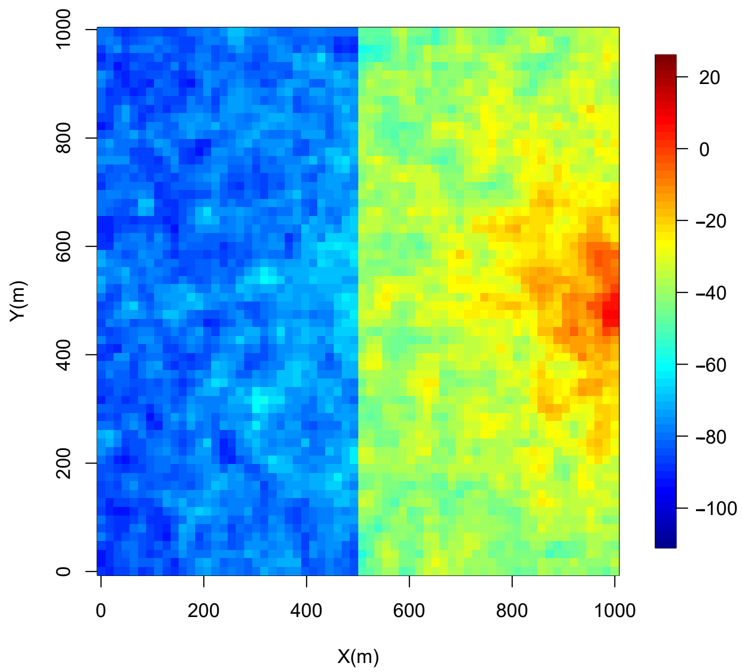

Figure 4. Real coverage map.

Here, we consider that the antenna is omnidirectional for this area. Given a resolution of 5 m × 5 m, we have a total of 40,000 RSRP samples. In Figure 4, we present the RSRP map corresponding to the values computed based on (3). As shown on the map, a remarkable difference in the RSRP values can be clearly observed in two different areas of the map. This can be explained by the presence of two different environments in the simulated area. To measure the accuracy of the proposed approach, we examined its ability to distinguish between different clusters based on their environments. The differences represent different attributes for each cluster. For example, clusters may be different in terms of geographical area or behavior of wireless channels. Regardless of the type of differences, a key testing element is to measure the ability of the model to differentiate designated clusters. Therefore, we assigned 'Cluster 1' to the area with buildings and 'Cluster 2' to the open area with gardens and parks. We applied the KMC method to split the dataset into two parts based on these defined clusters. At this stage of the test, the target was to clearly define the two different rural and urban environments. Figure 5 illustrates the performance of the model based on KMC. It is seen that the model correctly and clearly distinguished between the two different environments.

To assess the benefit of the clustering, we performed multiple steps. We first evaluated the performance of the Kriging prediction with the target dataset. Then, we modeled the measurement error using Gaussian noise with a zero mean and 3 dB variance. The generated noise was added to the RSRP measurements. Furthermore, we split the resulting RSRP dataset into learning and testing sets. Here, we used the learning set to estimate the predictor parameters, and the testing set to evaluate the performance of the predictor. It is worth mentioning that the learning set consisted of 10% of the whole dataset, wherein the points were uniformly selected over the area. To evaluate the performance of prediction, the RMSE was used as the performance metric. It is given by:

$$\text{RMSE} = \left(\frac{1}{\text{card}(\mathscr{T})}\sum_{g_0\in\mathscr{T}}\left(\hat{q}(g_0) - q(g_0)\right)^2\right)^{\frac{1}{2}}. \tag{10}$$

where $\mathscr{T}$ is the testing set. $\hat{q}(g_0)$ and $q(g_0)$ are the estimated and the real RSRP at g_0, respectively. We also used the R^2 metric expressed as

$$R^2 = 1 - \sum_{g_0 \in \mathscr{T}} \frac{\left(q(g_0) - \hat{q}(g_0)\right)^2}{\left(q(g_0) - \bar{q}\right)^2}, \tag{11}$$

where $\bar{q}$ is the mean of RSRP values. In the following, we evaluate the benefit of using clustering in RSRP map construction. To this end, we carried out a comparative study between the following two cases: (i) without clustering, and (ii) with two clusters. For the first case, we estimated the predictor parameters using a learning set that considered the entire area of interest, while for the second case, for each cluster, we estimated the predictor parameters using a learning set that considered the area covered by the cluster. The results of the prediction when considering the first case are presented in Figure 6. Here, the performance of the RSRP prediction when considering the entire area as one cluster is shown. As seen from this result, one can clearly see that there was a difference with the real coverage presented in Figure 4. In fact, an RMSE of 6.70 dB was observed in this case (i.e., without clustering).

For the second case, we estimated the predictor parameters of each cluster using its associated learning set. Then, the prediction for each testing set, associated with each cluster, was performed, based on the related estimated parameters. By combining the results for individual testing sets, we obtained the performance of the prediction for the entire testing set. The results of the prediction for each cluster are presented in Figure 7. In fact, we show, in Figure 7a,b, the constructed REMs for the first and second clusters, respectively. By substituting the two sub-REMs in Figure 7c, we obtain the REM for the whole area of interest. In Table 5, we present the comparison between the two cases, i.e., with and without clustering, in terms of RMSE. As seen from these results, the clustering offered a gain of about 3.3 dB and, consequently, led to a more accurately constructed map. Here, the standard deviation (std) of the RMSE was about 0.009 dB. To further examine the approach, we used R^2. Similarly, we compared the two cases with and without clustering. The R^2 of the generated model using the proposed approach achieved significant prediction results. Table 6 shows results of about 98%, which indicated high accuracy in terms of REM construction.

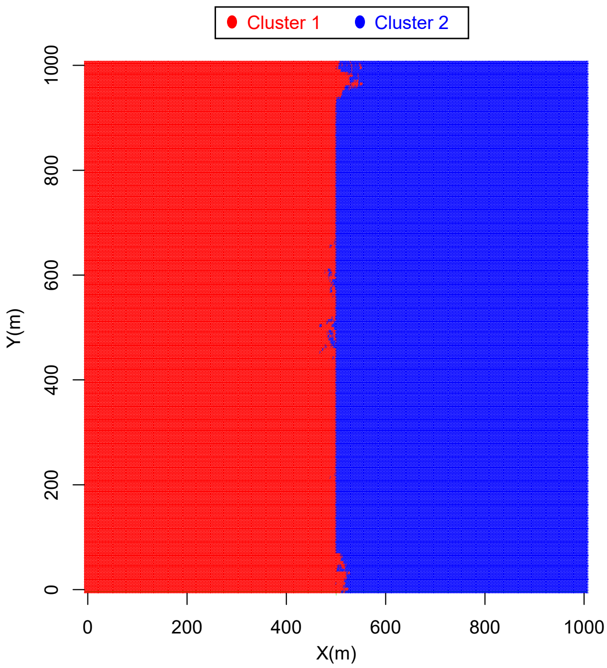

Figure 5. Clustering of the environment.

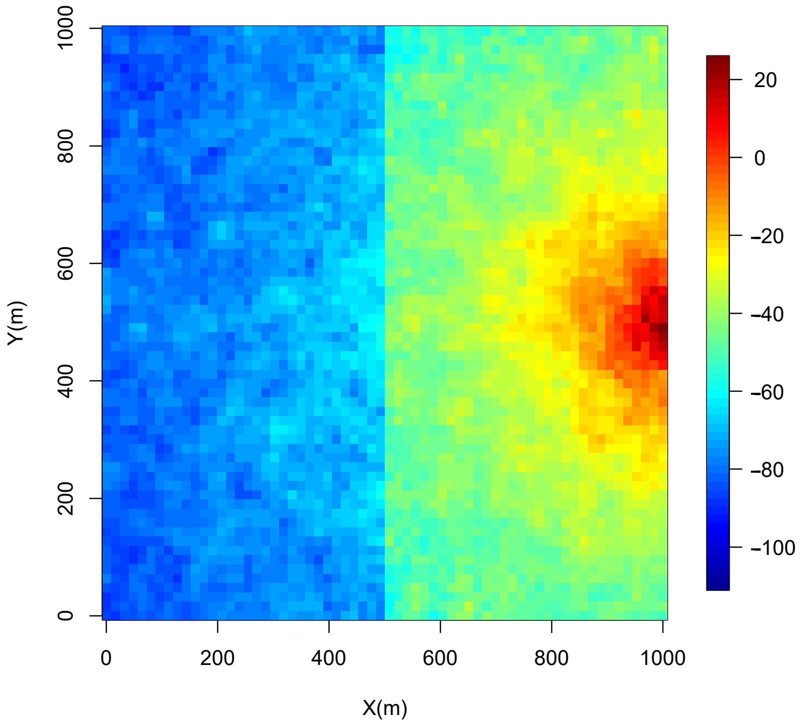

Figure 6. Predicted map without clustering.

Table 5. Comparison of the prediction error in terms of RMSE.

	Without Clustering	**With Clustering**	
RMSE in the whole area (dB)	6.703674	3.422388	
		1st cluster	2nd cluster
	RMSE per cluster (dB)	3.413063	3.431735
	std of the RMSE (dB)	0.009336	

Table 6. Comparison of the proposed approach with the case without clustering using R^2.

	Without Clustering	**With Clustering**	
R^2	0.921	0.979	
		1st cluster	2nd cluster
	R^2 per cluster	0.980	0.979

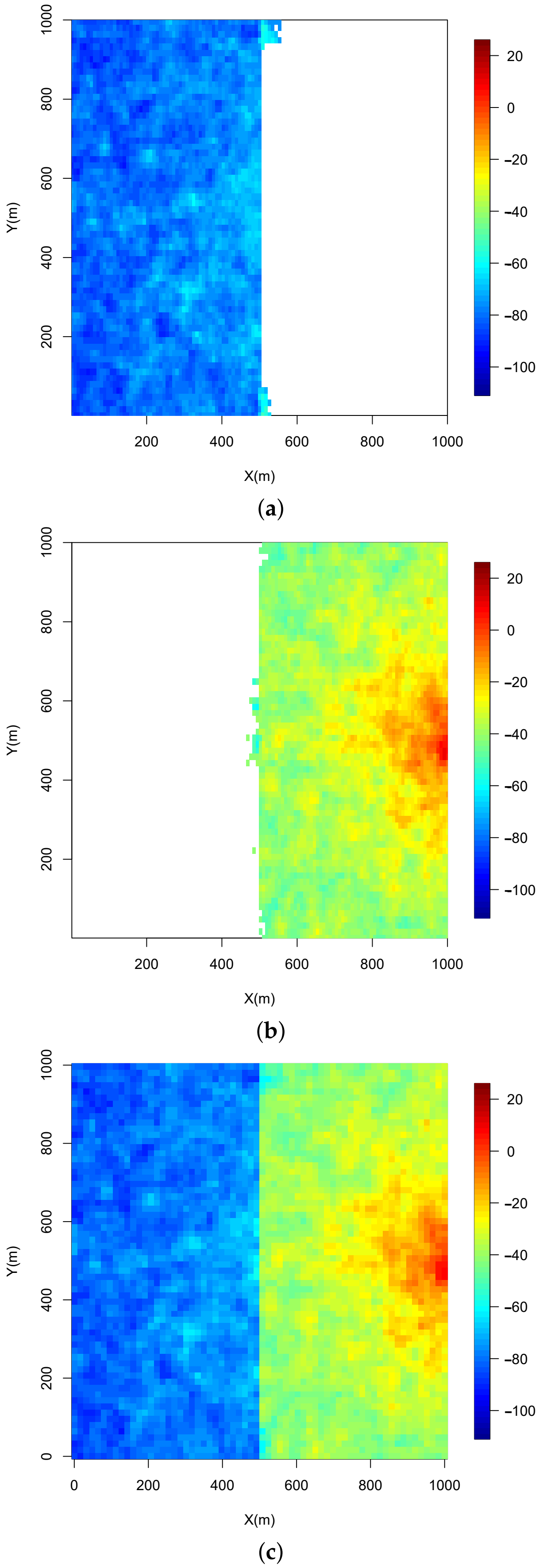

Figure 7. Prediction of the: (**a**) first cluster, (**b**) second cluster (**c**) both clusters.

7. Conclusions

In this work, we presented an approach to improve the prediction process for REMs. The objective of this work was to improve wireless connectivity in densely distributed applications, such as smart cities. In our approach, we first improve the prediction accuracy of RSRP through applying the K-means clustering technique. The aim of this clustering approach is to enhance the estimation of the path loss exponent. The proposed clustering technique contributes to distinguishing the different wireless environments, which enhances the prediction process. We then categorize each resulting cluster. For each cluster in our model, we use the Kriging technique to construct the accumulated results. This is performed in order to improve the overall prediction accuracy for the entire dataset. To test the viability of the approach, we carried out a set of comparative studies. In each test, we compared the prediction accuracy with clustering and then retested it without clustering. We observed the RMSE in each test to measure the progress of the prediction accuracy. The results showed that the proposed approach provided accurate predictions with a gain of about 3.3 dB in terms of RMSE compared to the case without clustering. In future work, we plan to test the approach with other trending models, such as the multiple-trend path loss model. We also plan to generalize the clustering approach using other techniques to test the performance of different algorithms. In fact, we target test the approach with a large number of heterogeneous clusters to cope with communication demands in several distributed applications.

Author Contributions: Conceptualization: H.B.C. and A.A.; methodology: H.B.C. and A.A.; validation: A.A. and H.B.C.; formal analysis: A.A. and H.B.C.; writing—original draft preparation: H.B.C. and A.A.; writing—review and editing: A.A. and H.B.C.; supervision: H.B.C. and A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Deanship of Scientific Research at Jouf University under grant No. (39/826).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Jouf University for funding this work through research grant No. (39/826).

Conflicts of Interest: The authors declare that there were no disclosed possible conflict of interest relevant to the research.

References

1. Yilmaz, H.B.; Tugcu, T.; Alagöz, F.; Bayhan, S. Radio environment map as enabler for practical cognitive radio networks. *IEEE Commun. Mag.* **2013**, *51*, 162–169. [CrossRef]
2. Perez-Romero, J.; Zalonis, A.; Boukhatem, L.; Kliks, A.; Koutlia, K.; Dimitriou, N.; Kurda, R. On the use of radio environment maps for interference management in heterogeneous networks. *IEEE Commun. Mag.* **2015**, *53*, 184–191. [CrossRef]
3. Braham, H.; Jemaa, S.B.; Fort, G.; Moulines, E.; Sayrac, B. Fixed rank kriging for cellular coverage analysis. *IEEE Trans. Veh. Technol.* **2016**, *66*, 4212–4222. [CrossRef]
4. Umbert, A.; Casadevall, F.; Rodriguez, E.G. An outdoor TV band Radio Environment Map for a Manhattan like layout. In Proceedings of the 2016 International Symposium on Wireless Communication Systems (ISWCS), Poznan, Poland, 20–23 September 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 399–403.
5. Galindo-Serrano, A.; Sayrac, B.; Jemaa, S.B.; Riihijärvi, J.; Mähönen, P. Automated coverage hole detection for cellular networks using radio environment maps. In Proceedings of the 2013 11th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), Tsukuba Science City, Japan, 13–17 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 35–40.
6. Braham, H.; Jemaa, S.B.; Sayrac, B.; Fort, G.; Moulines, E. Low complexity spatial interpolation for cellular coverage analysis. In Proceedings of the 2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Hammamet, Tunisia, 12–16 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 188–195.
7. Sodagari, S. A secure radio environment map database to share spectrum. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 1298–1305. [CrossRef]

8. Sayed, A.H.; Tarighat, A.; Khajehnouri, N. Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information. *IEEE Signal Process. Mag.* **2005**, *22*, 24–40. [CrossRef]
9. Celebi, H.; Arslan, H. Utilization of location information in cognitive wireless networks. *IEEE Wirel. Commun.* **2007**, *14*, 6–13. [CrossRef]
10. 3GPP. Study on minimization of drive-tests in next generation networks; (release 9). *3GPP TR 36. 805* .
11. Johansson, J.; Hapsari, W.A.; Kelley, S.; Bodog, G. Minimization of drive tests in 3GPP release 11. *IEEE Commun. Mag.* **2012**, *50*, 36–43. [CrossRef]
12. Hapsari, W.A.; Umesh, A.; Iwamura, M.; Tomala, M.; Gyula, B.; Sebire, B. Minimization of drive tests solution in 3GPP. *IEEE Commun. Mag.* **2012**, *50*, 28–36. [CrossRef]
13. Galindo-Serrano, A.; Sayrac, B.; Jemaa, S.B.; Riihijärvi, J.; Mähönen, P. Harvesting MDT data: Radio environment maps for coverage analysis in cellular networks. In Proceedings of the 8th International Conference on Cognitive Radio Oriented Wireless Networks, Washington, DC, USA, 8–10 July 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 37–42.
14. Dall'Anese, E.; Kim, S.J.; Giannakis, G.B. Channel gain map tracking via distributed kriging. *IEEE Trans. Veh. Technol.* **2011**, *60*, 1205–1211. [CrossRef]
15. Sato, K.; Fujii, T. Kriging-based interference power constraint: Integrated design of the radio environment map and transmission power. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 13–25. [CrossRef]
16. Sato, K.; Fujii, T. Kriging-based interference power constraint for spectrum sharing based on radio environment map. In Proceedings of the 2015 IEEE Globecom Workshops (GC Wkshps), San Diego, CA, USA, 6–10 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6.
17. Phillips, C.; Sicker, D.; Grunwald, D. A survey of wireless path loss prediction and coverage mapping methods. *IEEE Commun. Surv. Tutor.* **2012**, *15*, 255–270. [CrossRef]
18. Laghari, A.A.; He, H.; Khan, A.; Kumar, N.; Kharel, R. Quality of experience framework for cloud computing (QoC). *IEEE Access* **2018**, *6*, 64876–64890. [CrossRef]
19. Haibeh, L.A.; Yagoub, M.C.; Jarray, A. A survey on mobile edge computing infrastructure: Design, resource management, and optimization approaches. *IEEE Access* **2022**, *10*, 27591–27610. [CrossRef]
20. Laghari, A.A.; Jumani, A.K.; Laghari, R.A. Review and state of art of fog computing. *Arch. Comput. Methods Eng.* **2021**, *28*, 3631–3643. [CrossRef]
21. Wang, X.; Li, J.; Ning, Z.; Song, Q.; Guo, L.; Guo, S.; Obaidat, M.S. Wireless powered mobile edge computing networks: A survey. *ACM Comput. Surv.* **2023** . [CrossRef]
22. Malmirchegini, M.; Mostofi, Y. On the spatial predictability of communication channels. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 964–978. [CrossRef]
23. Szyszkowicz, S.S.; Yanikomeroglu, H.; Thompson, J.S. On the feasibility of wireless shadowing correlation models. *IEEE Trans. Veh. Technol.* **2010**, *59*, 4222–4236. [CrossRef]
24. Riihijärvi, J.; Mähönen, P. Estimating wireless network properties with spatial statistics and models. In Proceedings of the 2012 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), Paderborn, Germany, 14–18 May 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 331–336.
25. Riihijarvi, J.; Mahonen, P.; Wellens, M.; Gordziel, M. Characterization and modelling of spectrum for dynamic spectrum access with spatial statistics and random fields. In Proceedings of the 2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, Cannes, France, 15–18 September 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1–6.
26. Mirahsan, M.; Schoenen, R.; Szyszkowicz, S.S.; Yanikomeroglu, H. Measuring the spatial heterogeneity of outdoor users in wireless cellular networks based on open urban maps. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 2834–2838.
27. Perpinias, N.; Palaios, A.; Riihijärvi, J.; Mähönen, P. A measurement-based study on the use of spatial interpolation for propagation estimation. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 2715–2720.
28. Konak, A. Estimating path loss in wireless local area networks using ordinary kriging. In Proceedings of the 2010 Winter Simulation Conference, Baltimore, MD, USA, 5–8 December 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 2888–2896.
29. Gao, Y.; Fujii, T. A Kriging-based Radio Environment Map Construction and Channel Estimation System in Threatening Environments. *IEEE Access* **2023** , *11*, 38136–38148. [CrossRef]
30. Diago-Mosquera, M.; Aragón-Zavala, A.; Vargas-Rosales, C. The performance of in-building measurement-based path loss modelling using kriging. *IET Microwaves Antennas Propag.* **2021**, *15*, 1564–1576. [CrossRef]
31. Bi, J.; Wang, Y.; Li, Z.; Xu, S.; Zhou, J.; Sun, M.; Si, M. Fast radio map construction by using adaptive path loss model interpolation in large-scale building. *Sensors* **2019**, *19*, 712. [CrossRef]
32. Romero, D.; Kim, S.J. Radio map estimation: A data-driven approach to spectrum cartography. *IEEE Signal Process. Mag.* **2022**, *39*, 53–72. [CrossRef]
33. Omre, H. Bayesian Kriging—Merging observations and qualified guesses in Kriging. *Math. Geol.* **1987**, *19*, 25–39. [CrossRef]
34. Sayrac, B.; Riihijärvi, J.; Mähönen, P.; Ben Jemaa, S.; Moulines, E.; Grimoud, S. Improving coverage estimation for cellular networks with spatial bayesian prediction based on measurements. In Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design, Helsinki, Finland, 13 August 2012; pp. 43–48.

35. Zhang, X.; Andrews, J.G. Downlink cellular network analysis with multi-slope path loss models. *IEEE Trans. Commun.* **2015**, *63*, 1881–1894. [CrossRef]
36. Inaltekin, H.; Chiang, M.; Poor, H.V.; Wicker, S.B. On unbounded path loss models: Effects of singularity on wireless network performance. *IEEE J. Sel. Areas Commun.* **2009**, *27*, 1078–1092. [CrossRef]
37. Alam, A.M.; Benjemaa, S.; Romary, T. Clustering for high accuracy coverage mapping. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
38. Khan, J.Y. Basics of Communication Networks. *Internet Things* **2019** , 65–104.
39. Cheffena, M. Time-varying on-body wireless channel model during walking. *EURASIP J. Wirel. Commun. Netw.* **2014**, *2014*, 29. [CrossRef]
40. Hassan, R.; Qamar, F.; Hasan, M.K.; Aman, A.H.M.; Ahmed, A.S. Internet of Things and its applications: A comprehensive survey. *Symmetry* **2020**, *12*, 1674. [CrossRef]
41. Kazmi, S.H.A.; Qamar, F.; Hassan, R.; Nisar, K. Improved QoS in Internet of Things (IoTs) through Short Messages Encryption Scheme for Wireless Sensor Communication. In Proceedings of the 2022 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Penang, Malaysia, 22–25 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
42. Laghari, A.A.; Wu, K.; Laghari, R.A.; Ali, M.; Khan, A.A. A review and state of art of Internet of Things (IoT). *Arch. Comput. Methods Eng.* **2022**, *29*, 1395–1413. [CrossRef]
43. Alaerjan, A. Towards Sustainable Distributed Sensor Networks: An Approach for Addressing Power Limitation Issues in WSNs. *Sensors* **2023**, *23*, 975. [CrossRef]
44. Li, J.; Ding, G.; Zhang, X.; Wu, Q. Recent advances in radio environment map: A survey. In Proceedings of the Machine Learning and Intelligent Communications: Second International Conference, MLICOM 2017, Weihai, China, 5–6 August 2017; Proceedings, Part I 2; Springer: Berlin/Heidelberg, Germany, 2018; pp. 247–257.
45. Zhao, Y.; Reed, J.H.; Mao, S.; Bae, K.K. Overhead analysis for radio environment map enabled cognitive radio networks. In Proceedings of the 2006 1st IEEE Workshop on Networking Technologies for Software Defined Radio Networks, Reston, VA, USA, 25 September 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 18–25.
46. Bi, S.; Lyu, J.; Ding, Z.; Zhang, R. Engineering radio maps for wireless resource management. *IEEE Wirel. Commun.* **2019**, *26*, 133–141. [CrossRef]
47. Cressie, N. *Statistics for Spatial Data*; John Wiley & Sons: Hoboken, NJ, USA, 1993.
48. Braham, H.; Jemaa, S.B.; Fort, G.; Moulines, E.; Sayrac, B. Spatial prediction under location uncertainty in cellular networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 7633–7643. [CrossRef]
49. Ben Chikha, W.; Masson, M.; Altman, Z.; Jemaa, S.B. Radio Environment Map Based Inter-Cell Interference Coordination for Massive-MIMO Systems. *IEEE Trans. Mob. Comput.* **2022** .
50. Wu, X.; Kumar, V.; Ross Quinlan, J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]
51. 3GPP. *Study on Channel Model for Frequencies from 0.5 to 100 GHz*; Technical Report; 3rd Generation Partnership Project (3GPP): Biot, France, 2018; Volume 38.
52. Forkel, I.; Schinnenburg, M.; Ang, M. Generation of two-dimensional correlated shadowing for mobile radio network simulation. *WPMC Sep.* **2004**, *21*, 43.
53. Cressie, N.; Johannesson, G. Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B* **2008**, *70*, 209–226. [CrossRef]
54. Chiles, J.P.; Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 497.
55. Omre, H.; Halvorsen, K.B. The Bayesian bridge between simple and universal kriging. *Math. Geol.* **1989**, *21*, 767–786. [CrossRef]
56. Oliver, M.A.; Webster, R. *Basic Steps in Geostatistics: The Variogram and Kriging*; Springer: Berlin/Heidelberg, Germany, 2015.
57. Montero, J.M.; Fernández-Avilés, G.; Mateu, J. *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*; John Wiley & Sons: Hoboken, NJ, USA, 2015; Volume 998.
58. Syed, A.S.; Sierra-Sosa, D.; Kumar, A.; Elmaghraby, A. IoT in smart cities: A survey of technologies, practices and challenges. *Smart Cities* **2021**, *4*, 429–475. [CrossRef]
59. Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. *Future Internet* **2019**, *11*, 94. [CrossRef]
60. Horvat, M.; Jović, A.; Burnik, K. Assessing the robustness of cluster solutions in emotionally-annotated pictures using monte-carlo simulation stabilized K-means algorithm. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 435–452. [CrossRef]
61. Huang, Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min. Knowl. Discov.* **1998**, *2*, 283–304. [CrossRef]