Crowdsourced Indoor Wi-Fi REMs: Does the Spatial Interpolation Method Matter?

Zakarya El-friakh, Andra M. Voicu, Shaham Shabani, Ljiljana Simić, and Petri Mähönen
Institute for Networked Systems
RWTH Aachen University
Email: {zfr, avo, sha, lsi, pma}@inets.rwth-aachen.de

***Abstract*—Crowdsourcing is a promising approach to collect the received signal strength (RSS) measurements required to construct Wi-Fi radio environment maps (REMs) via spatial interpolation methods. However, crowdsourced RSS measurements are often unreliable and especially in indoor propagation environments they are strongly affected by shadowing and fading. It is thus important to understand which spatial interpolation method is better suited for such REMs. In this paper we present an extensive empirical performance evaluation of four major spatial interpolation methods for crowdsourced indoor Wi-Fi REMs: Inverse Distance Weighting (IDW), Gradient Plus Inverse Distance Squared (GIDS), Ordinary Kriging (OrK), and Universal Kriging (UnK). We evaluate them both quantitatively by the estimation error, and qualitatively by visually observing the resulting REMs from measured Wi-Fi beacon RSSs. In our experiments we used sixteen Raspberry Pi boards and we considered different measurement location densities and spatial distributions in an office environment with four rooms. Both the quantitative and qualitative analysis of our experimental results show that there is only a negligible difference in the accuracy of the REMs constructed using the four considered spatial interpolation methods, relative to the inherent variability in the RSS reported by different devices. This suggests that, in practice, for crowdsourced indoor Wi-Fi REMs, simpler methods like IDW may be preferred over more computationally complex methods like Kriging. Importantly, our analysis also shows that, for moderate measurement location densities, the global RSS estimation error over a large indoor area with multiple rooms does not comprehensively characterize the performance of spatial interpolation methods. In addition to considering this metric, a quantitative analysis of the local error specific to each room, and a qualitative evaluation of the observed topology of the resulting REM need to be performed. This is a novel observation and suggests that more sophisticated metrics should be investigated for different engineering applications of REMs.**

***Index Terms*—indoor REMs, Wi-Fi, crowdsourcing**

I. Introduction

Radio environment maps (REMs) are a powerful tool to spatially represent and predict the radio coverage, which can in turn be used for optimizing spectrum utilization. REMs were initially proposed to facilitate dynamic spectrum access (DSA) in bands where primary users must be protected from harmful interference from secondary users [1]. Recently, REMs have also been considered for cellular, Wi-Fi, and WiMAX networks [2]–[4]. Although different than for DSA, REMs could be equally useful for such broadband networks, where there is a large number of devices and deployments with high traffic load requirements, so careful network configuration is needed to manage interference and satisfy capacity demands.

Efficient interference management is especially challenging for the already ubiquitous Wi-Fi, which operates in the unlicensed bands, where any device is in principle allowed to transmit. Consequently, there may be many *distributed* Wi-Fi networks, which have to be configured individually, according to the changes that occur dynamically in other networks. As such, traditional walk or drive test approaches to construct REMs [5], [6] are not feasible for Wi-Fi networks in practice, due to the high cost and the infeasibility of continuously detecting dynamic changes in spectrum occupation.

Crowdsourcing is a cost-efficient, promising alternative to continuously collect measurements, due to the proliferation of low-cost end-user wireless devices [7]. Moreover, this is already a standardized feature for some modern wireless technologies, i.e. minimization of drive tests (MDT) for LTE [8], [9]. We note that continuous spectrum monitoring via crowdsourcing could be useful also for making regulatory decisions, especially for novel approaches such as risk-informed interference assessment [10], where a large number of measurements would facilitate identifying harmful interference and its likelihood.

Given the existing wide-spread Wi-Fi devices, crowdsourcing is thus also a natural candidate for collecting measurements for Wi-Fi REMs. However, raw crowdsourced measurements are typically unreliable, due to vendor-specific, low-precision device operation characteristics (e.g. for receiver sensitivity), potentially insufficient number of provided measurements, or even potential faulty device operation. Importantly, these aspects would impact the quality of the resulting REM. Furthermore, the REM quality is also affected by the different spatial interpolation techniques [11]–[14] that can be used for constructing the REM by estimating the received signal strength (RSS) at locations where no RSS measurements are available. For instance, more complex techniques like Kriging are expected to result in more accurate REMs, whereas less computationally demanding techniques like Inverse Distance Weighting (IDW) are expected to result in less accurate REMs. Moreover, especially in indoor environments with furniture, reflections, and object movement, the measured RSS values may vary significantly due to shadowing and fading. Given these multiple sources of measured RSS variation, it is critical to understand which spatial interpolation method generates better crowdsourced Wi-Fi REMs in indoor deployments and what the limits of different methods are in practice.

Existing work on building radio maps with crowdsourcing (or low-cost devices) considered Wi-Fi [12], [15]–[21] and cellular networks [22], [23]. We note that, although REMs for cellular networks are outside the scope of this paper, crowdsourcing and spatial interpolation methods investigated in this context may also be applicable for Wi-Fi. For Wi-Fi, most authors considered either initial measurement data processing without applying spatial interpolation methods [15], [16], or a single interpolation method, i.e. Kriging [17], [18], IDW [19], and bi-harmonic spline interpolation [20]. Consequently, it is difficult to directly compare the performance of such methods. Although several interpolation methods have been considered in [12], [21] for indoor Wi-Fi REMs, e.g. Kriging, IDW, Gradient Plus Inverse Distance Squared (GIDS), their performance was evaluated based on simulated rather than measured data. As such, the crowdsourcing-specific variability of the reported measurements from different sensing devices was not taken into account. For cellular networks, the authors in [22] presented an implementation of a framework for collecting measurements via crowdsourcing and investigated the limitations of the devices, but did not process the data further based on spatial interpolation. In [23] a performance comparison of different spatial interpolation techniques (Kriging, IDW, splines, etc.) with crowdsourced measurements for cellular networks was presented, but the authors focused on outdoor urban environments, which exhibit significantly different radio propagation characteristics compared to indoor environments. Overall, it is thus not yet clear which spatial interpolation method would be preferred for crowdsourced indoor Wi-Fi REMs and what are the practical limitations for such methods.

In this paper we present a comprehensive performance evaluation of four major spatial interpolation methods for indoor Wi-Fi REMs with crowdsourced measurements, i.e. IDW, GIDS, Ordinary Kriging (OrK), and Universal Kriging (UnK). We compare their performance both quantitatively and qualitatively, in terms of spatial interpolation error and the obtained RSS coverage map, respectively. To this end we conducted an extensive measurement campaign with Raspberry Pi (RPi) boards as sensor networks in a four-room indoor office environment, where we measured the Wi-Fi beacon RSS. Also, we varied the number of used measurement locations and their spatial distribution, in order to evaluate in detail how *real* crowdsourced measurements affect the quality of the constructed REM. Our results show that the difference in the performance of the considered interpolation methods is negligible, relative to the inherent variability of RSS reported by low-cost Wi-Fi devices. Consequently, simpler methods like IDW may be preferred over more computationally complex methods like Kriging for constructing indoor Wi-Fi REMs with crowdsourced measurements in practice. Furthermore, simply relying on the estimation error metric is not enough to comprehensively characterize the performance of different spatial interpolation methods for moderate measurement location densities. This should be complemented by qualitatively observing the REMs.

The remainder of this paper is organized as follows. Section II presents the considered spatial interpolation methods. In Section III our testbed and measurement methodology are described. Section IV presents and discusses our results, and Section V concludes the paper.

II. Interpolation Methods

In this section we present the four considered spatial interpolation methods: (i) IDW, (ii) GIDS, (iii) OrK, and (iv) UnK. We apply all considered spatial interpolation methods for finding the estimate RSS value $\hat{Z}(s_0)$ of the unknown real value $Z(s_0)$ at location $s_0 = (x_0, y_0)$, where (x_0, y_0) are the coordinates of location s_0. To achieve this, we use n measured RSS values $(Z(s_i))_{1\leqslant i\leqslant n}$ at locations $(s_i{=}(x_i, y_i))_{1\leqslant i\leqslant n}$ as input for the spatial interpolation methods.

A. IDW [24]

With IDW, the estimated value $\hat{Z}(s_0)$ is obtained as

$$\hat{Z}(s_0) = \sum_{i\in N(s_0)} \frac{Z(s_i) \times d_{i0}^{-\beta}}{\sum\limits_{i\in N(s_0)} d_{i0}^{-\beta}}, \tag{1}$$

where d_{i0} is the distance between locations s_0 and s_i, $N(s_0)$ is the set of measurement locations in the neighborhood of location s_0, and β is the *power parameter*. In our implementation we use a fixed neighborhood size $|N(s_0)|{=}N{=}13$,[1] which was found in the literature to be optimal for IDW [25]. We select $\beta{=}2$, which corresponds to the path loss exponent in free space. We note that we have also considered $\beta{=}2.8$ as for an office environment [26], but the resulting performance of IDW was similar, so these results are omitted in this paper for brevity.

B. GIDS [24]

GIDS improves IDW by using global knowledge about the measured values $(Z(s_i))_{1\leqslant i\leqslant n}$, instead of relying only on the immediate neighborhood of s_0. The estimate $\hat{Z}(s_0)$ is thus obtained as

$$\hat{Z}(s_0) = \sum_{i\in N(s_0)} \frac{(Z(s_i){+}(x_0{-}x_i){\times}\, C_x{+}(y_0{-}y_i){\times}C_y){\times}d_{i0}^{-\beta}}{\sum\limits_{i\in N(s_0)} d_{i0}^{-\beta}}, \tag{2}$$

where C_x and C_y are the gradient values at s_i, which are obtained through linear regression. In our implementation we used a fixed neighborhood value $N(s_0) = N = 13$ and fixed $\beta = 2$, consistent with the respective values selected for IDW. Parameters C_x and C_y are derived by fitting a first degree multi-variate polynomial to the measured RSS as a function of the coordinates.

C. Kriging [27]

We first present common aspects for both Kriging variants, i.e. OrK and UnK, and then we give the specifics of each variant. As is the case for IDW and GIDS, with Kriging the

[1] We note that, if fewer than thirteen measurement locations are available, all of them are considered to be in the neighborhood of s_0.

estimate $\hat{Z}(s_0)$ is given as a linear combination of the known measurement values $(Z(s_i))_{1\leqslant i\leqslant n}$, as follows

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i \times Z(s_i), \tag{3}$$

where $(\lambda_i)_{1\leqslant i\leqslant n}$ are the linear weights of the respective measured values $(Z(s_i))_{1\leqslant i\leqslant n}$. Unlike IDW and GIDS, Kriging aims to minimize the squared prediction error defined as

$$MSE[\hat{Z}(s_0)] = E\left[(\hat{Z}(s_0) - Z(s_0))^2\right]. \tag{4}$$

Kriging is based on the assumption that each measured value $(Z(s_i))_{1\leqslant i\leqslant n}$ is an instance of a random variable $(\mathbf{Z}(s_i))_{1\leqslant i\leqslant n}$ at location $(s_i)_{1\leqslant i\leqslant n}$. The co-variance among these random variables is used in the derivation of the linear coefficients $(\lambda_i)_{1\leqslant i\leqslant n}$. Since it is not possible to derive the co-variance between the different random variables $(\mathbf{Z}(s_i))_{1\leqslant i\leqslant n}$ using only their instantaneous measured values, stationarity models are adopted. In this paper we consider *intrinsic stationarity*, which is summarized in the two following assumptions [27]:

(i) $E[\mathbf{Z}(s)] - \mu = 0$;
(ii) $Var[\mathbf{Z}(s+h) - \mathbf{Z}(s)] =: 2\gamma(|h|)$ for all shifts h of the location vector,

where μ is a constant independent of the location s, $2\gamma(|h|)$ is referred to as the *variogram* function and $\gamma(|h|)$ as the *semi-variogram*, s is the location vector, and h is a shift vector. Assumption (i) implies that the mean of the random variable $\mathbf{Z}(s)$ is constant and independent of the location. Assumption (ii) implies that the variance of the difference $\mathbf{Z}(s+h) - \mathbf{Z}(s)$ depends only on the separation distance $|h|$ between the locations $s+h$ and s.

The solution to minimizing the prediction error results in values of $(\lambda_i)_{1\leqslant i\leqslant n}$ defined by

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ m/2 \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & 1 \\ \vdots & & & & \vdots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_{01} \\ \gamma_{02} \\ \vdots \\ \gamma_{0n} \\ 1 \end{pmatrix}, \tag{5}$$

where m is the Lagrangian multiplier used for the minimization resolution and $\gamma_{ij} = \gamma(d_{ij})$, with d_{ij} being the distance between locations s_i and s_j.

The values of the semi-variogram $\gamma(\bullet)$ are obtained by first estimating its values at the separation distances between the measurement locations and then fitting a parametric function to this estimator. In our study we used the first order estimator

$$\hat{\gamma}(h) = \frac{1}{2|M(h)|} \sum_{M(h)} [Z(s_i) - Z(s_j)]^2, \tag{6}$$

where $M(h)$ is the set of all pairs of measurements having separation h, and $|M(h)|$ is the number of such pairs. As a parametric function we adopted the exponential function family defined by

$$\zeta(h, \boldsymbol{\alpha}) = \alpha_0 + \alpha_1(1 - \exp(-h/\alpha_2)), \tag{7}$$

where $\alpha_0 \geqslant 0$ and $\alpha_2 > 0$. We note that $\zeta(h = 0, \boldsymbol{\alpha}) = \alpha_0$, where this value is known as the *nugget* and reflects the variance due to the measurement error. Additionally, $\lim_{h\to\infty} \zeta(h, \boldsymbol{\alpha}) = \alpha_0 + \alpha_1$, where this value is referred to as the *still*. Finally, α_2 reflects the range of the co-variance between the measurements and is simply referred to as the *range*. In our implementation, parameters α_0, α_1, and α_2 were obtained using Weighted Least Square fitting [27] of the semi-variogram estimator values $\hat{\gamma}(\bullet)$ to the function $\zeta(h, \alpha)$. The values obtained based on our measurements for these parameters are discussed in Section IV-B. We note that exponential semi-variograms have been successfully used in prior work for RSS value interpolation [28].

The Kriging variant described so far assumes that the mean μ of $\mathbf{Z}(s)$ is an unknown constant (*cf.* assumption (i) of intrinsic stationarity). This variant is known as *Ordinary* Kriging (OrK). *Universal* Kriging (UnK) is another variant, which assumes a non-constant mean. In UnK the random variable of interest $\mathbf{Z}(s)$ is expressed as

$$\mathbf{Z}(s) = \mu(s) + \mathbf{e}(s), \tag{8}$$

where $\mu(s)$ is a spatially-variant mean function and $\mathbf{e}(s)$ is a (intrinsic) stationary random variable. The first step in UnK is to estimate the mean function $\mu(s)$. In our implementation, this was achieved by fitting a second order multi-variate polynomial function to the measured RSS values. The expression of such a function is

$$\mu(s) = \sum_{0\leqslant k+l\leqslant 2} \delta_{kl} x^k y^l, \tag{9}$$

where $(\delta_{kl})_{0\leqslant k+l\leqslant 2}$ are parameters to be fitted to the measured values. In our implementation, the mean function was obtained by applying the Matlab function `fit` to the set of measured RSS values. After estimating the mean function $\mu(s)$, OrK is performed on the values $\mathbf{e}(s) = \mathbf{Z}(s) - \mu(s)$ and the estimate $\hat{e}(s_0)$ is obtained. Finally, the estimate $\hat{Z}(s_0)$ is determined using

$$\hat{Z}(s_0) = \mu(s_0) + \hat{e}(s_0). \tag{10}$$

III. Testbed and Measurements

For our experiments we deployed sixteen different RPi 3 Model B computer boards, in order to sense the RSS of Wi-Fi beacons from a given AP. These boards come equipped with a Broadcom BCM43438 Wi-Fi chipset operational in the 2.4 GHz band [29]. We used the `iwlist` GNU/Linux command to sample the RSS value at the locations of the RPi nodes. The collected RSS measurements were then regularly reported via Wi-Fi to an additional RPi, which acted as a central data collector. This RPi reported all collected measurements via Ethernet to a laptop, where further data processing and spatial interpolation was performed in Matlab.

Our measurements were conducted in an office environment, in an area spanning four rooms (i.e. *Library*, *Students room*, *Corridor 1*, and *Corridor 2*), where the sensing RPis were placed on the floor on a 20 cm-spaced regular grid, and shifted to cover a dense set of locations, as shown in Fig. 1. The total

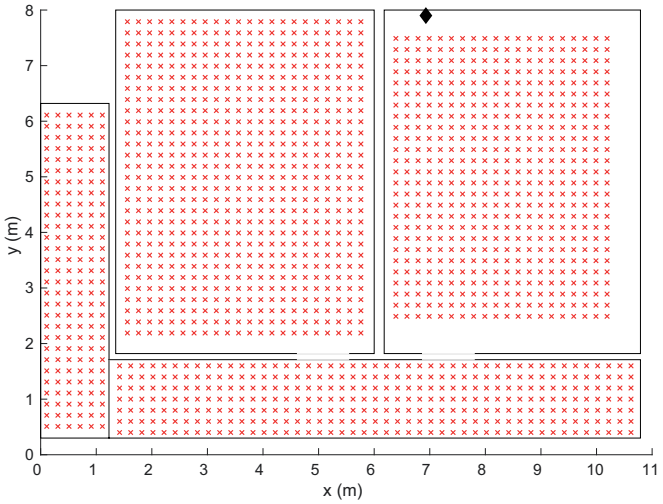


Fig. 1: Floor plan of the studied area showing the 1661 locations of the measurements (×) and the location of the transmitting AP (♦).

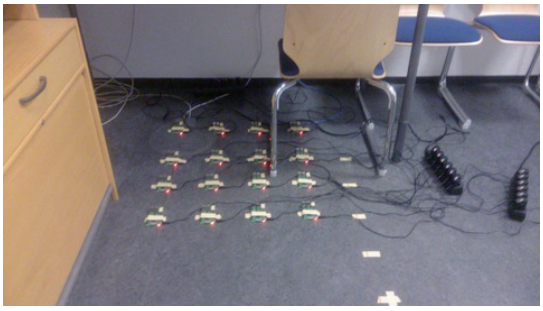


Fig. 2: Example RPi deployment for a set of the measurement locations in the *Library*.

covered area was 11×8 m^2 and comprised 1661 measurement locations, resulting in a density of 25 nodes/m^2. An example RPi deployment for a portion of the grid is shown in Fig. 2. *Library* contains a central wooden table, book shelves and multiple white boards, while *Students room* contains multiple desks and computers. Both rooms have large glass windows. All indoor separation walls are plaster-based, and all doors in *Corridor 2* are made of non-tinted glass, while the doors in *Corridor 1* are made of wood. This propagation environment results in non-negligible fading and shadowing effects, which are expected to increase the variation of the measured RSS values [26]. During our experiments, we measured the RSS of the beacons from one transmitting AP located in the *Library*.

Furthermore, multiple regular and random measurement location distributions were derived from the initial 20 cm-spaced grid. These distributions were obtained by keeping D = 50%, 25%, 11.11%, 6.25%, and 4% of the original measurement locations in each room. Two additional random distributions were obtained by keeping D = 1% and 0.5% of the original measurement locations. We note that in the case of regularly spaced locations, the distribution densities D are equivalent to grids with 30, 40, 60, 80 and 100 cm spacing. For each measurement location, we took 40 measurements, at 3-second time intervals. We note that in the case of the random distributions, for each density ten different realizations were selected from the initial 20 cm-spaced grid and used to generate the results shown in Section IV.

In order to quantitatively compare the performance of the different interpolation methods, we use the absolute estimation error

$$AEE(s_i) = |\hat{Z}(s_i) - Z(s_i)| \qquad \text{[dB]}, \tag{11}$$

which is calculated in all locations of the original 20 cm-spaced grid where an estimate RSS based on spatial interpolation is obtained, i.e. all locations in the 20 cm-spaced grid that are not selected to form the considered location distribution.

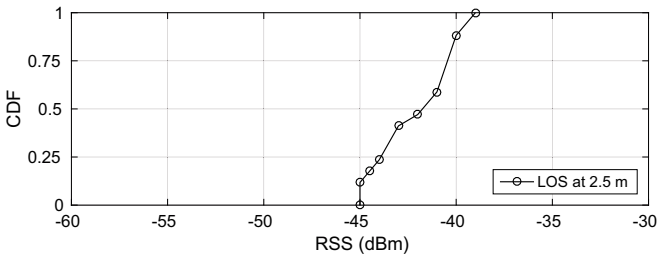
IV. Results

In this section, we first present in Section IV-A the results obtained from the calibration measurements of the RPi nodes and we discuss our method to establish the groundtruth for calculating AEE. We then present a comparative performance analysis of the four considered spatial interpolation methods via quantitative evaluation with AEE in Section IV-B and qualitative visual evaluation with REMs in Section IV-C.

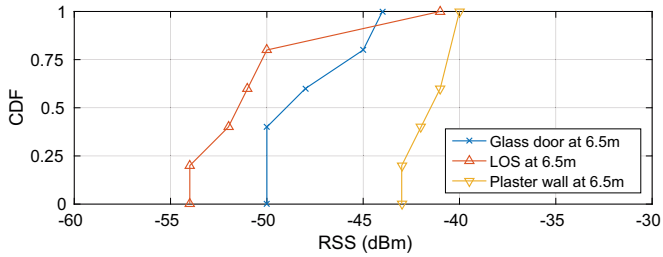
A. Calibration Measurements & Establishing Groundtruth

There are three types of possible sources of variation that affect the reported measured RSS values: (i) the low-precision receivers of different nodes, (ii) the dynamic propagation environment causing e.g. fast fading, and (iii) the antenna pattern and orientation of the sensing node relative to the transmitting AP. In order to quantify the impact of each of these aspects, 40 measurements where taken with each RPi with a 3 second interval between each two consecutive measurements, where the nodes were placed at the same given location, at a distance of 2.5 m from the AP with LOS conditions, i.e. coordinates (6.9 m, 5.4 m) in Fig. 1. Furthermore, for five selected RPis, additional locations at 6.5 m from the AP where considered: with LOS, i.e. coordinates (6.9 m, 1.4 m) in Fig. 1; with NLOS with an obstructing glass door, i.e. coordinates (6.9 m, 1.4 m); and with NLOS with an obstructing wall, i.e. coordinates (9 m, 1.4 m).

In order to isolate and characterize the effect of the hardware, only the median of the 40 measurements was considered for each node, so that the impact of fast fading is mitigated. Fig. 3(a) shows the distribution of the median RSS measured by each node at 2.5 m from the AP, where the variation across different nodes is up to 6 dB. Fig. 3(b) shows the median RSS per node, for 5 nodes, for a distance of 6.5 m from the AP. The RSS varies over 14, 6, and 3 dB for different devices in LOS conditions, NLOS with glass door, and NLOS with plaster wall, respectively. These results suggest that the effect of the hardware decreases as more obstacles obstruct the LOS from the transmitting AP to the receiving nodes. We also note that in our experiments, the variation across nodes in LOS was higher at 6.5 m from the AP than at 2.5 m from it. However,



(a)



(b)

Fig. 3: Distribution of the RSS from different RPis for the calibration measurements at (a) 2.5 m from the AP with LOS, for all 16 RPis; and (b) 6.5 m from the AP with LOS, NLOS with glass door, and NLOS with plaster wall, for 5 RPis.

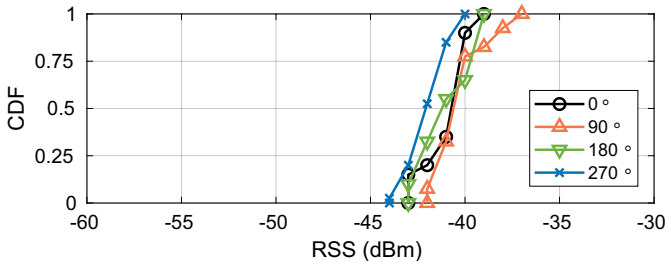
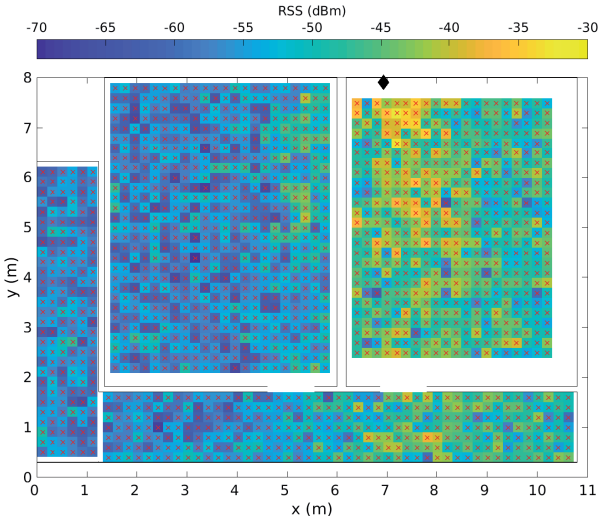
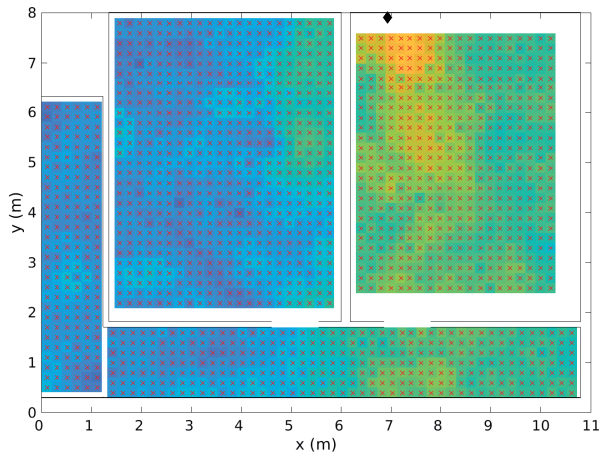


Fig. 4: Distribution of the the 40 RSS values measured over 120 s by a single selected RPi at 2.5 m from the AP, at the same location with coordinates (6.9 m, 5.4 m), across different orientations.



(a)



(b)

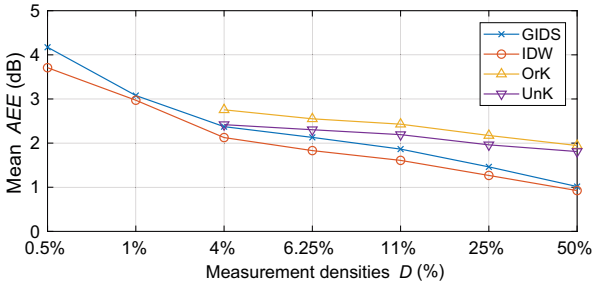
Fig. 5: Measured RSS values (a) before and (b) after spatial smoothing (a proxy for the *typical* crowdsourced RSS, averaging over the hardware variations of sensing devices).

we suspect that this is due to the presence of a door frame in the LOS path when the nodes were positioned at 6.5 m from the AP.

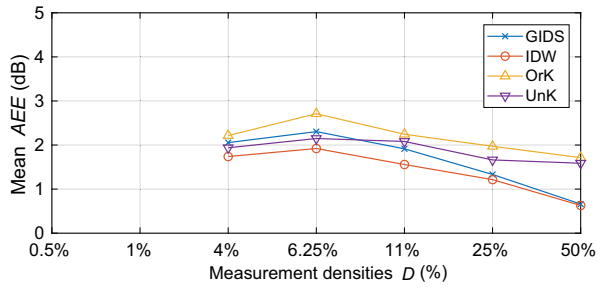
In order to evaluate the impact of the orientation of the receiving nodes relative to the transmitting AP, additional measurements were taken with each RPi, by also rotating them in a plane parallel to the floor. To also consider the variation due to fading, we keep the 40 measured RSS values instead of only considering their median. Fig. 4 shows the obtained results for a single selected RPi. We note that the variation of the RSS across different orientations is at most 2 dB. Furthermore, for all four orientations, the variation of the RSS across time is at most 5 dB.

Given this variation in the reported RSS, for the measurements considered as input for the interpolation methods we aim to use the typical value that would be reported in practice via crowdsourcing.[2] We obtain this as follows. In order to average the effect of fading, for each measurement location we consider only the median of the 40 measurements taken by each RPi at a given location. In order to average the effect of the sensing node orientation, we shift the RPis along the x and y axes (*cf.* Fig. 1), such that their relative orientation to the transmitting AP spans a wide range of angles. Finally, in order to average the variation due to the hardware differences between the sensing nodes, the measured RSS value at each location (i.e. the median of the 40 measurements at that location) was smoothed. This was obtained by estimating the RSS for each given location of the original 20 cm-spaced grid, while considering all other measurements on the grid as input for IDW. We note that IDW was selected as it is the least com-

[2]We note that in practice at each measured location there would be a distribution of RSSs from different devices.



(a)



(b)

Fig. 6: Mean estimation error for all interpolation methods, for all measurement location densities and over all considered rooms, for (a) random distributions and (b) regular distributions.

putationally demanding spatial interpolation method and thus straightforward to apply in practice. Fig. 5(a) shows the map of the raw measured RSS values, and Fig. 5(b) shows the map of the smoothed measurements. As expected, the smoothed map exhibits less abrupt RSS changes among neighboring locations, and represents a proxy for the *typical* crowdsourced RSS, which would have been obtained if averaging over the hardware variations from a large set of reporting devices. We consider these smoothed measurements as a baseline for evaluating different spatial interpolation methods via AEE in Section IV-B.

B. Quantitative Performance Analysis of Spatial Interpolation Methods

In this section we analyze quantitatively the performance of IDW, GIDS, OrK, and UnK, by means of the AEE for all interpolated values. Fig. 6 shows the mean AEE for both random and regular measurement location distributions, for different interpolation methods, and different values of D. We note that results for OrK and UnK are not shown for $D \leqslant 1\%$ (i.e. less than fifteen measurement locations in the total covered area), as Kriging can only be applied if a sufficient number of measurements are available to perform reliable curve fitting with (7), consistent with [11]. For the random distributions in Fig. 6(a), the mean AEE decreases as the density D increases, as more measurement locations are used as input for interpolation. We note that this trend is clearly evident for IDW and GIDS as the mean AEE drops from

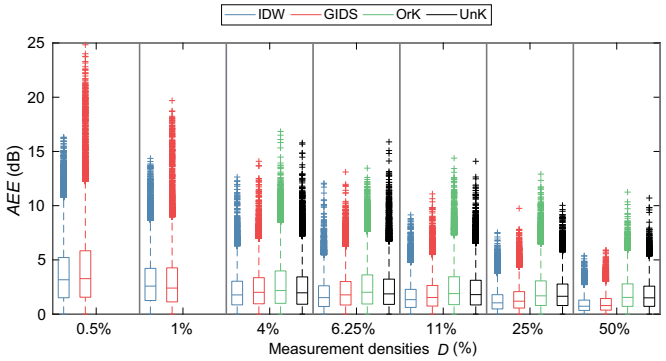


Fig. 7: Estimation error for IDW, GIDS, OrK and UnK for randomly distributed locations, for different location densities and over all rooms.

4 dB down to 1 dB as the density D is increased from 0.5% to 50%. This trend is less noticeable for OrK and UnK, for which the mean AEE variation is less than 1 dB as the density D is varied from 4% to 50%. We also note that IDW and GIDS consistently achieve lower mean AEE compared to OrK and UnK for all densities D. Fig. 6(b) shows similar trends for the regular distributions as for random distributions. However, the AEE is slightly lower for regular vs. random distributions, for an equal density D. We note that this is expected, since the regular distributions cover the space uniformly. Nonetheless, since random distributions are more representative of real measurement locations for crowdsourcing, in the following we focus on random distributions.

In order to analyze the performance of the considered interpolation methods in more depth, Fig. 7 shows the AEE distribution for different interpolation methods and measurement location densities, in the case of the randomly distributed locations, over all ten deployment realizations (*cf.* Section III). The median value of the AEE is at most 3 dB, regardless of the interpolation method and the measurement density. However, OrK and UnK result in a slightly higher maximum AEE compared to IDW and GIDS, especially for high D. Namely, for D=50%, the maximum is 3 dB for IDW and GIDS, and 6 dB for OrK and UnK. This suggests that Kriging does not lead to a higher interpolation accuracy, despite its higher computational complexity. Similar results were obtained for the regular location distributions, so these results are omitted here for brevity.

Since the propagation characteristics of the rooms are different, in Fig. 8 we also show the AEE for each individual room, in order to investigate whether the AEE variation is consistent across the rooms. We note that the range of the AEE is overall larger for rooms that are closer to the AP. For instance, the AEE in *Corridor 1* (Fig. 8(a)), which is the furthest from the AP, is at most 5 dB for D=50% for all interpolation methods, whereas in *Library*, where the AP is located (Fig. 8(d)), the maximum AEE for the same density D increases to 8 dB. This variation across rooms is consistent with the results of our calibration measurements in

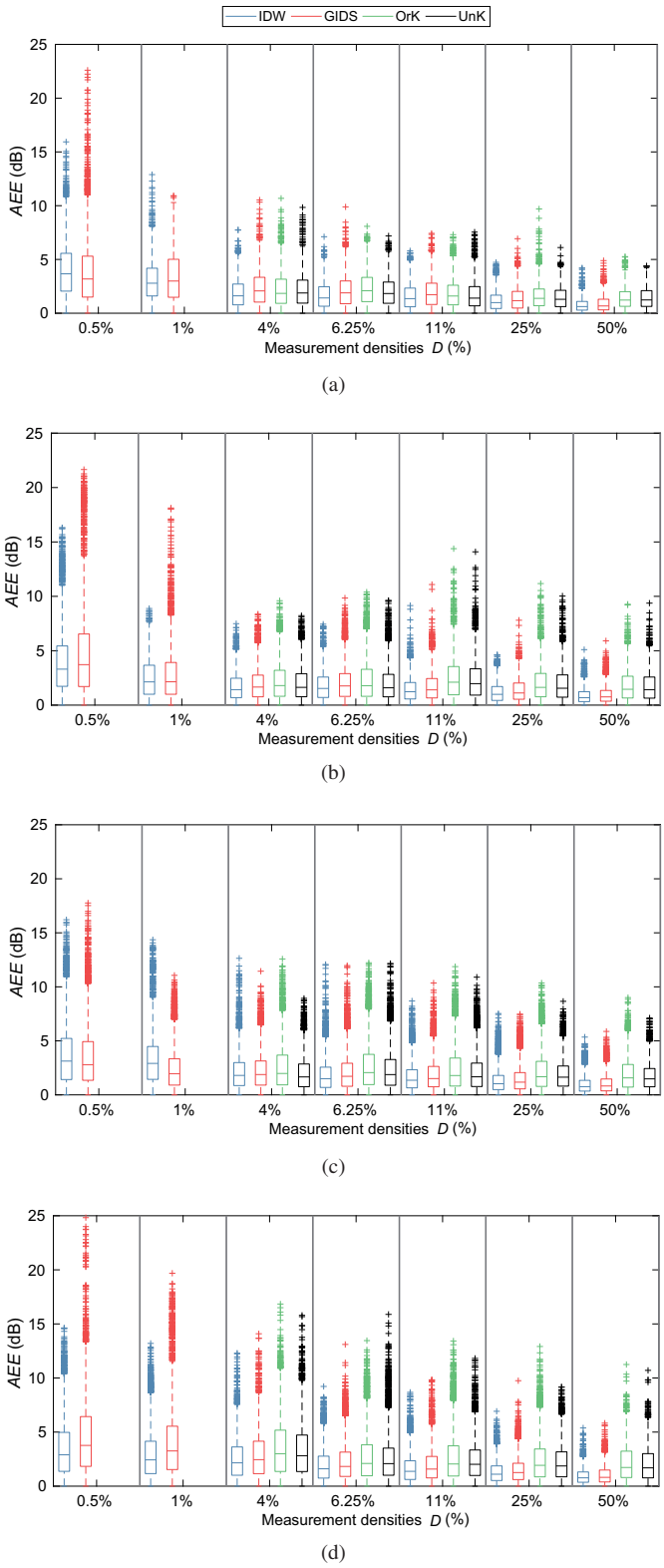


Fig. 8: Estimation error for IDW, GIDS, OrK and UnK in (a) *Corridor 1*, (b) *Corridor 2*, (c) *Students room*, and (d) *Library*.

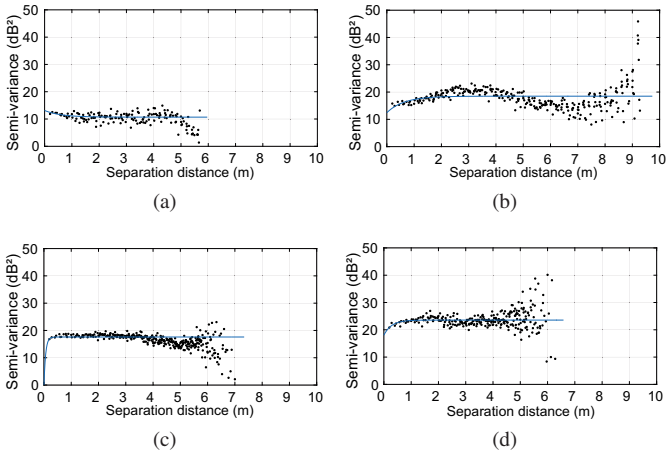


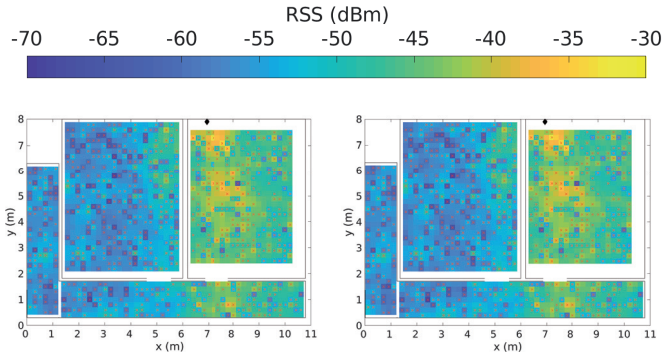
Fig. 9: Semi-variogram estimator (•) and exponential fit (—) for the different rooms: (a) *Corridor 1*, (b) *Corridor 2*, (c) *Students room*, (d) *Library*.

TABLE I: Variogram exponential fitting parameters for the different rooms.

Room	$\sqrt{\alpha_0}$ [dB]	$\sqrt{\alpha_0 + \alpha_1}$ [dB]	α_2 [cm]
Corridor 1	3.6	3.5	43.5
Corridor 2	2.69	8.3	565.4
Students room	3.81	4.9	136.7
Library	4.38	5.3	122.0

Section IV-A. Namely, the RSS values vary more for LOS between the considered transmitter and sensors, compared to RSS for NLOS with a wall between transmitter and sensors (*cf.* Fig. 3). This suggests that considering only the global AEE over all rooms is not enough to characterize the performance of different spatial interpolation techniques. The local AEE for individual rooms, or more sophisticated spatial statistics should also be analyzed.

Importantly, unlike IDW and GIDS, Kriging could not be performed using aggregated measurements from all rooms, since the semi-variogram estimator points were too scattered to fit to any function. Instead, we applied Kriging, i.e OrK and UnK, to each room separately. Fig. 9 shows the semi-variogram estimators and their fitting curves for the different rooms. We note that the obtained variance profiles are very distinct. Specifically, *Corridor 1* shows a very flat semi-variogram fit. We account for this by the non-negligible shadowing due to the number of obstacles separating the sensing nodes from the AP. For *Corridor 2*, the exponential fit does not match the semi-variogram well. This is likely due to the fact that depending on their location in *Corridor 2*, the RPi nodes are either separated from the AP by a wall, two walls, or a glass door. We note that in the case of *Library* and *Students room* the exponential fit matches the semi-variograms rather well. Table I shows the exponential fitting parameters α_0, α_1, and α_2 (*cf.* Section II-C) of the variogram, for each room using UnK. As a reminder, $\sqrt{\alpha_0}$ reflects the variation due to the



(b) IDW, D = 50% (c) GIDS, D = 50%

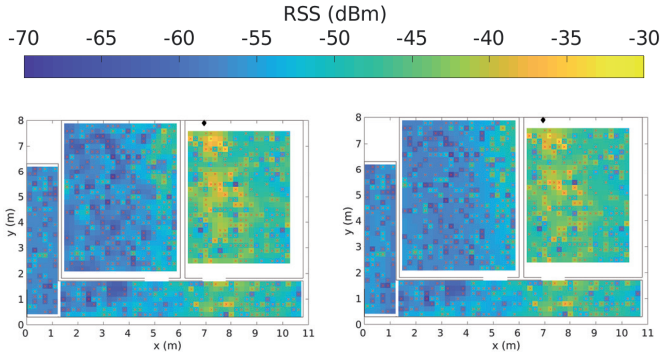
(d) IDW, D = 11% (e) GIDS, D = 11%

(f) IDW, D = 4% (g) GIDS, D = 4%

(h) IDW, D = 1% (i) GIDS D = 1%

(j) IDW, D = 0.5% (k) GIDS, D = 0.5%

Fig. 10: Example REMs generated using IDW (left) and GIDS (right) spatial interpolation, using measurements from a percentage D of original measurement locations, for a single random distribution of selected locations ($\times$).



(b) OrK, D = 50% (c) UnK, D = 50%

(d) OrK, D = 11% (e) UnK, D = 11%

(f) OrK, D = 4% (g) UnK, D = 4%

Fig. 11: Example REMs generated using OrK (left) and UnK (right) spatial interpolation, using measurements from a percentage D of original measurement locations, for a single random distribution of selected locations ($\times$).

measurement error. We note that the values obtained from the fitting are consistent with those obtained in the calibration step in Fig. 3. In particular, a variation of 3 dB is obtained in all rooms, in NLOS of the AP, with a slightly higher value of 4 dB in *Library*, where the AP is located.

Our AEE results quantifying the impact of the four considered spatial interpolation methods show overall that IDW, GIDS, OrK, and UnK have a similar performance for indoor Wi-Fi REMs, regardless of the measurement location density. This suggests that computationally demanding interpolation methods like Kriging may be less attractive for such deployments. Instead, less complex methods like IDW may be more feasible in practice, since they obtain a similar performance with less computational effort. Moreover, in order to apply Kriging, additional location information is needed, i.e. in which rooms the measurements were taken. As an insight, the AEE significantly increases only for $D \leqslant 1\%$. This

suggests that D=4%, i.e. 66 measurement locations for our total considered area, is sufficient to obtain accurate enough RSS estimations, with errors within the inherent variation of 6 dB across the RSS reported by different low-cost devices.

C. Qualitative Performance Analysis of Spatial Interpolation Methods

Our results so far suggest that all interpolation methods perform similarly. In this section we compare the performance of IDW, GIDS, OrK, and UnK, by observing the resulting coverage maps, for different measurement densities D. Fig. 10 shows the coverage maps obtained using IDW and GIDS for one realization of randomly distributed locations for each of the densities D=50%, 11%, 4%, 1%, and 0.5%. Fig. 11 shows the obtained maps using OrK, and UnK for the same realizations of location distributions having the densities D=50%, 11%, and 4%. For all interpolation methods, for each measurement location realization, interpolation was conducted only at locations where no measurement value was kept from the original 20 cm-spaced grid, i.e. for the other locations the actual measured RSS is shown on the map.

For D=50% all interpolation methods result in maps relatively similar to the original smoothed baseline map (*cf.* Fig. 5(b)). However, for D=11%, the resulting maps start to slightly diverge from the groundtruth baseline of Fig. 5(b). Finally, for values of $D\leqslant$4%, the obtained maps are significantly different than our groundtruth baseline, e.g. the blob of high RSS values surrounding the position of the transmitting AP is no longer present in the resulting coverage map.

According to the quantitative AEE results in Section IV-B, all four interpolation methods show a similar performance for all measurement densities from 50% down to 4%. By contrast, the results in this section suggest that the quality of the obtained coverage map for all interpolation methods starts degrading at densities as high as 11%. This suggests that evaluating the spatial interpolation methods based solely on the estimation error over all rooms is not sufficient to fully characterize crowdsourced indoor Wi-Fi REMs. This should be complemented by a qualitative evaluation of the REM topology and by considering more sophisticated statistics that quantify the impact of different methods for given engineering applications.

V. Conclusions

In this paper we presented an extensive empirical performance evaluation of four major spatial interpolation methods, for constructing crowdsourced indoor Wi-Fi REMs. We deployed low-cost spectrum sensors in an office indoor environment, for experimental data collection. We evaluated the considered spatial interpolation methods both quantitatively and qualitatively, for different measurement location distributions and a wide range of densities. Our results show that there is no significant difference in the accuracy of the REMs for the considered interpolation methods, relative to the inherent variability in the measurements reported by different low-cost devices. This suggests that, when building crowdsourced indoor Wi-Fi REMs in practice, less complex spatial interpolation methods like IDW may be preferred over more computationally demanding methods like Kriging variants. Furthermore, our analysis suggests that quantitative evaluation of REM accuracy via a globally-calculated first-order metric like average estimation error over the entire measurement area does not fully characterize REM accuracy. This is evident even by a qualitative inspection of our interpolated REMs as compared to the groundtruth coverage measurements. This is a novel observation for REM construction and evaluation, as it strongly suggests that the considered spatial interpolation methods and REM deployments should not be evaluated by simple metrics. As we have shown, the qualitative REM topology is also relevant for evaluation, and more sophisticated spatial statistics are actually needed for specific engineering applications of REMs, such as network coverage and QoS estimation. Thus, the engineering significance of interpolation and topology reproduction errors should be considered carefully in future work. A few possible quantitative metrics are, e.g., measuring the percentage of area covered by a certain threshold, extreme statistics to quantify the spatial granularity, and gradients of power slopes. We are currently developing and verifying such metrics as a part of our ongoing research work.

References

[1] H. B. Yilmaz, T. Tugcu, F. Alagöz, and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 162–169, Dec. 2013.

[2] J. van de Beek, T. Cai, S. Grimoud, B. Sayrac, P. Mähönen, J. Nasreddine, and J. Riihijärvi, "How a layered REM architecture brings cognition to today's mobile networks," *IEEE Wireless Commun.*, vol. 19, no. 4, pp. 17–24, Aug. 2012.

[3] J. Pérez-Romero, A. Zalonis, L. Boukhatem, A. Kliks, K. Koutlia, N. Dimitriou, and R. Kurda, "On the use of radio environment maps for interference management in heteregeneous networks," *IEEE Communications Magazine*, vol. 53, pp. 184–191, Aug. 2015.

[4] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *Proc. IEEE DySPAN*, Bellevue, WA, USA, Oct. 2012.

[5] H. Gacanin and A. Ligata, "Wi-Fi self-organizing networks: Challenges and use cases," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 158–164, 2017.

[6] A. Galindo-Serrano, B. Sayrac, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, "Automated coverage hole detection for cellular networks using radio environment maps," in *Proc. WiNMee*, Tsukuba Science City, Japan, May 2013.

[7] P. G. Lopez, A. Montresor, D. Epema, A. Datta, T. Higashino, A. Iamnitchi, M. Barcellos, P. Felber, and E. Riviere, "Edge-centric computing: Vision and challenges," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 5, pp. 37–42, Oct. 2015.

[8] *Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2*, 3GPP TS 37.320, V10.4.0, Release 10, Jan. 2012.

[9] *Harvesting MDT data: Radio environment maps for coverage analysis in cellular networks*, 2013.

[10] J. P. de Vries, "Risk-informed interference assessment: A quantitative basis for spectrum allocation decisions," *Telecommun. Policy*, vol. 41, no. 5-6, pp. 434–446, June 2017.

[11] M. Pesko, T. Javornik, A. Košir, M. Štular, and M. Mohorčič, "Radio environment maps: The survey of construction methods," *KSII Transactions on Internet and Information Systems*, vol. 8, no. 11, pp. 3789–3809, Nov. 2014.

[12] FARAMIR Consortium, "Neighbourhood mapping technology and algorithm report, D4.2," I. Dagres *et al.*, Eds., Oct. 2011. [Online]. Available: https://cordis.europa.eu/project/rcn/93764_en.html

[13] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, and B. Sayrac, "Fixed rank kriging for cellular coverage analysis," *IEEE Transactions on Vehicular Technology, vol. 66*, pp. 1–1, 2016.

[14] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, , and B. Sayrac, "Spatial prediction under location uncertainty in cellular networks," *IEEE Transactions on Wireless Communications, vol. 15*, vol. 15, pp. 7633–7643, 2016.

[15] W. Li, Y. Zhu, and T. He, "WiBee: Building WiFi radio map with ZigBee sensor networks," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012.

[16] S. Rosen, S.-J. Lee, J. Lee, P. Congdon, Z. M. Mao, and K. Burden, "MCNet: Crowdsourcing wireless performance measurements through the eyes of mobile devices," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 86–91, Oct. 2014.

[17] M. Portoles-Comeras, C. Ibars, J. Nunez-Martinez, and J. Mangues-Bafalluy, "Characterizing WLAN medium utilization for Radio Environment Maps," in *Proc. IEEE VTC Fall*, San Francisco, CA, USA, Sept. 2011.

[18] E. Meshkova, J. Riihijärvi, J. Ansari, and P. Mähönen, "Indoor coverage estimation from unreliable measurements using spatial statistics," in *Proc. ACM MSWiM*, Barcelona, Spain, Nov. 2013.

[19] R. P. Dionísio, T. F. Alves, and J. M. A. Ribeiro, "Experimentation with radio environment maps for resources optimisation in dense wireless scenarios," in *Proc. International Conference on Advances in Cognitive Radio (COCORA)*, Venice, Italy, Apr. 2017.

[20] J. Tervonen, M. Hartikainen, M. Heikkilä, and M. Koskela, "Applying and comparing two measurement approaches for the estimation of indoor WiFi coverage," in *Proc. IFIP International Conference on New Technologies, Mobility and Security*, Larnaca, Cyprus, Nov. 2016.

[21] M. Angjelicinoski, V. Atanasovski, and L. Gavrilovska, "Comparative analysis of spatial interpolation methods for creating Radio Environment Maps," in *Telecommunications Forum*, Belgrade, Serbia, Nov. 2011.

[22] A. Achtzehn, J. Riihijärvi, I. A. B. Castillo, M. Petrova, and P. Mähönen, "CrowdREM: Harnessing the power of the mobile crowd for flexible wireless network monitoring," in *Proc. International Workshop on Mobile Computing Systems and Applications*, Santa Fe, New Mexico, USA, Feb. 2015.

[23] M. Molinari, M.-R. Fida, M. K. Marina, and A. Pescape, "Spatial interpolation based cellular coverage prediction with crowdsourced measurements," in *Proc. ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*, London, United Kingdom, Aug. 2015.

[24] R. W. W. Ian A. Nalder, "Spatial interpolation of climatic normals: test of a new method in the canadian boreal forest," *Agricultural and Forest Meteorology*, vol. 92, no. 4, pp. 211–225, Dec. 1998.

[25] R. J. Renka, "Multivariate interpolation of large sets of scattered data," *ACM Transactions on Mathematical Software*, vol. 14, no. 2, pp. 139–148, Jun. 1988.

[26] T. S. Rappaport, *Wireless Communications: Principles and Practice, Second Edition*. Prentice Hall, 2002, vol. 27.

[27] M. Sherman, *Spatial statistics and spatio-temporal data: covariance functions and directional properties*. John Wiley & Sons, 2011.

[28] E. Meshkova, J. Riihijärvi, J. Ansari, and P. Mähönen, "Indoor coverage estimation from unreliable measurements using spatial statistics," in *Proceedings of the 16th ACM international conference on Modeling, analysis & simulation of wireless and mobile systems - MSWiM '13*. ACM Press, 2013.

[29] "Raspberry Pi 3 model b single-board computer with wireless lan and bluetooth connectivity." [Online]. Available: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/